Study of an Exported Protein of *Mycobacterium avium subspecies paratuberculosis*

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science
in
Biochemistry

at Massey University Palmerston North,
New Zealand.

Susan Maree Copland

2004

Acknowledgements

Firstly, I would like to thank Dr Alan Murray for providing the project, laboratory space and materials, and a special thanks to Dr Kathryn Stowell who was a constant source of enthusiam and supported me with intellectual and technical expertise throughout my postgraduate studies.

Secondly I would like to thank my lab colleges Jeremy, Sandy, Kalyani, Miho and Becky who allowed me the privilege of their friendship, advice and technical knowledge in and out of the lab. Thanks to Dr Des Collins from Wallaceville Wellington, for supplying genomic DNA from *M. ptb* isolates and to Professor Brigitte Gicquel from the Pasteur Institute Paris, for supplying the vectors pJEM11 and pMIP12. Lorraine Berry undertook a large amount of DNA sequencing on my behalf, an exercise for which I am grateful and thanks to Chris Dupont for her technical assistance.

Thanks to those people in IVABS and IMBS who chose to help me in technical areas and gave me their friendship in the process, you were the icing on the cake and made the whole experience more enjoyable.

I would like to take this opportunity to acknowledge the Institute of Veterinary, Animal and Biomedical Sciences, the Phyllis Irene Grey Fellowship and the C. Alma Baker Trust of Massey University, Meat New Zealand and Wool Pro for their kind financial assistance.

Last but not least, thanks to my partner Luc. No only were you my personal computer technician and principle car mechanic, you were also the one who showered me with support and encouragement no matter what challenge I faced.

Abstract

Johne's disease is a chronic, progressive enteric disease of ruminants caused by infection with *Mycobacterium avium subspecies paratuberculosis* (*M. ptb*) from the MAIS complex (*M. avium*, *M. ptb*, *M. intracellulare* and *M. scrofulaceum*). The lack of specific and sensitive diagnostic tests often leads to *M. ptb* infected animals being diagnosed with bovine tuberculosis, a member of the MTB complex (*M. tb*, *M. bovis*, *M. bovis* BCG, *M africanum*, *M. microti* and *M. canetti*). Secreted proteins from pathogenic mycobacteria have been found to be important for the development of protective immunity, namely a cell mediated immune response (CMI). The development of reliable differential diagnostic tests will require the use of species-specific secreted protein antigens and the CMI response.

Due to the taxonomic distance between the MAIS and MTB complexes our hypothesis was that the *M. ptb* genome may encode for secreted proteins that are absent from members of the MTB complex. If such proteins can stimulate an immune response they may be suitable for use as antigens in a differential diagnostic test for Johne's disease. To this end, the secreted protein library clone pJEM11-*M. ptb*281 was examined and its insert found to contain the 5' region of the hypothetical *M. ptb*281 ORF fused in frame with *phoA*. The entire ORF was determined using *M. avium* and *M. ptb* database sequences then cloned into *E. coli* and mycobacterial expression systems. These systems incorporate 6x histidine (His_6) affinity tags into recombinant proteins allowing them to be semi-purified by Ni-NTA affinity chromatography. Semi-purified recombinant proteins tested positive by western blot analysis to highly specific anti-His_6-tag antibodies. Amino acid sequencing to confirm the identity of these recombinant proteins and screening for their ability to stimulate an immune response were prevented by time constraints.

Homologs to *M. ptb*281 were absent from *M. tb*, *M. bovis* and *M. bovis* BCG but present in the MAIS complex, making this protein unsuitable for use as an antigen to differentiate between MAIS complex species in a diagnostic test. *M. ptb*281 homologs found in the genomes of two members of the Acetomycetes order corresponded to hypothetical proteins predicted by computer software programs trained to identify genes, which may indicate that the hypothetical *M. ptb*281 ORF may encode a functional protein.

Abbreviations and Definitions

ADC	albumin-dextrose-catalase
Amp^r	gene conferring ampicillin resistance
AP	alkaline phosphatase
ATCC	American type culture collection
BCG	*M. bovis* Bacilli Calmette-Guerin
BCIP	5-bromo-4-chloro-3-indoyl phosphate disodium
bp	base pair
CDP^{star}	Disodium 2-chloro-5-(4-methoxyspiro {1,2-dioxetane-3,2´(5´-chloro) tricyclo ($3.3.1.1^{3.7}$) decan}-4-yl)-1-phenyl phosphate
CSPD	Disodium 3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'-chloro) tricyclo [$3.3.1.1^{3.7}$] decan}-4-yl) phenyl phosphate
C-terminal	carboxyl terminal
DIG	Digoxigenin
DIG-11-dUTP	Digoxigenin-11-2'-deoxy-uridine-5'-triphosphate
DMSO	dimethyl-sulphoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotide triphosphates
E. coli	*Escherichia coli*
hr	hour
IPTG	isopropyl β-D-thiogalactoside
kan	kanamycin sulphate
Kan^r	gene conferring kanamycin resistance
kb	kilo base
kDa	kilodalton
LB	Luria-Bertani
MAIS complex	*M. avium, M. ptb, M. intracellulare, M. scrofulaceum* complex

continued over page

Abbreviations and Definitions continued

M. ptb	*M. paratuberculosis*
M. tb	*M. tuberculosis*
M-MPTB281	pMIP MPTB281 recombinant protein
MPTB281	hypothetical MPTB281 protein
mRNA	messenger ribonucleic acid
MTB	*Mycobacterium tuberculosis* complex (*M. tb*, *M. bovis*, *M. bovis* BCG, *M africanum, M. microti* and *M. canetti*).
NBT	Nitro blue tetrazolium chloride
Ni-NTA	nickel nitrilo-tri-acetic acid
N-terminal	amino terminal
OADC	oleic acid-albumin-dextrose-catalase
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
pers. comm.	personal communication
phoA	Truncated alkaline phosphatase gene
PhoA	Truncated alkaline phosphatase protein
$PhoA^+$	Alkaline phosphatase phenotype
phoA-281-DIG	*phoA-M. ptb*281
pI	iso-electric point
SDS	sodiumdodecyl sulphate
Sec	secretion
soln.	solution
sp.	species
TAE	tris acetate EDTA
Taq	*Thermus aquaticus* DNA polymerase
TEMED	N, N', N, N-tetramethyl ethylene diamine
UV	ultra violet light
xg	multiplied by gravity
X-MPTB281	pPROEX MPTB281 recombinant protein

List of Figures

Page

Figure 1: The pJEM11 library and PhoA$^+$ selection....28

Figure 2: NCBI BLASTX search results....30

Figure 3: Location of *katE* homolog within the pJEM11-*M. ptb*281 insert....31

Figure 4: Alignment of the 107 bp *M. ptb* sequence to λGAM22....32

Figure 5: Agarose gel analysis of *phoA*-281 and MPB70 PCR products....35

Figure 6: Analysis of *Eco* R1 digested Neoparasec and BCG....36

Figure 7: Reference photo of Neoparasec and BCG *Eco* R1 Southern blot set 1....37

Figure 8: Hybridization of *phoA*-281-DIG using low stringency conditions....38

Figure 9: Neoparasec and BCG *Eco* R1 Southern blot, set 1 standard curve....39

Figure 10: Reference photos of Neoparasec and BCG *Eco* R1 Southern blots, set 2....40

Figure 11: Hybridization of *phoA*-281-DIG using high stringency conditions....41

Figure 12: Hybridization of MPB70-DIG using high stringency conditions....42

Figure 13: BLASTN result from the *M. avium* database....45

Figure 14: The most significant *M. ptb* database result....46

Figure 15: The *M. ptb*281 stop codon....47

Figure 16: Inspection of *M. ptb* and....48

Figure 17: Potential start codon for MPTB281....50

Figure 18: The full hypothetical *M. ptb*281 ORF and translated sequence....51

Figure 19: MPTB281F2 & MPTB281R1 primer check....55

Figure 20: Preparation of vectors for cloning....57

Figure 21: PCR screen to identify *M. ptb*281 positive clones....59

Figure 22: PMIP12-*M. ptb*281 and pPROEX-*M. ptb*281....60

Figure 23: Nucleotide and translated sequence of pPROEX-*M. ptb*281....61

Figure 24: DNA and translated sequence of pMIP-*M. ptb*281....62

Figure 25: Enriched recombinant X-MPTB281 66

Figure 26: Enriched recombinant M-MPTB281 68

Figure 27: Western transfer of X-MPTB281 recombinant protein 69

Figure 28: Immunodection of X-MPTB281 and M-MPTB281 western blots 70

Figure 29: CCGB database alignment to *M. ptb.* 73

Figure 30: Sanger *M. marinum* database alignment 74

Figure 31: TIGR database alignment to *M. avium* 75

Figure 32: TIGR database alignment to *M. smegmatis* 76

Figure 33: NCBI database alignments 77

Figure 34: Reference photo of *Bam* H1 digested mycobacterial DNA 79

Figure 35: High stringency hybridization of *M. ptb*281-DIG 80

Figure 36: Low stringency hybridization of *M. ptb*281-DIG 82

Figure 37: 16S control probe 84

Figure 38: PCR screen of selected mycobacterial species 85

Figure 39: The 62° C PCR screen of selected mycobacterial species 86

Figure 40: PLATINUM *Pfx* PCR screens. PCRs were performed with 87

Figure 41: PCR screen of multiple *M. ptb* isolates 88

List of Tables.

Page

Table 1: Bacterial strains used in this study 16

Table 2: Summary of *M. ptb*281 species distribution 90

Contents

Page

Abstract i

Abbreviations and Definitions ii

List of Figures iv

List of Tables. vi

Contents vii

Chapter 1: Introduction 1

1.1 Johne's Disease 1

1.1.1 The Organism 1

1.1.2 Historical Background 1

1.1.3 Symptoms and Stages of Johne's Disease 2

1.1.4 Control of Johne's Disease 2

1.1.5 Crohn's Disease 3

1.2 Economic impact to New Zealand 3

1.3 Host-pathogen Interaction 3

1.4 Detection of Johne's Disease 4

1.4.1 Direct Detection 4

Culture 4

Microscopy 5

Polymerase Chain Reaction (PCR) 5

1.4.2 Immunological Detection 6

Humoral immune response. 6

Cell mediated immune response (CMI). 6

1.5 Bovine Tuberculosis 7

1.6 Hypothesis and Aims 8

Chapter 2: Materials and Methods 9

2.1 Materials 9

2.2 pJEM11-*M. ptb* Secreted Protein Library 10

2.3 Automated DNA Sequencing 11

2.4 Computer Analysis 11

2.4.1 BLAST Search Databases 11

2.4.2 Translation, Molecular Weight and pI Prediction 12

2.4.3 Signal Peptide Search 12

2.5 Mycobacterial Genomic DNA 13
2.5.1 Growth of Mycobacteria on 7H11 Agar Slopes 13
2.5.2 Recovery of Mycobacterial Genomic DNA 13
2.6 Analysis of DNA by Agarose Gel Electrophoresis 14
2.7 Purification of DNA Fragments 14
2.8 Luria-Bertani (LB) Agar and Broth 14
2.9 Sauton's Media 15
2.10 Recovery of Plasmid Vectors 15
2.11 Storage of Bacterial Cultures Grown in Liquid Medium 15
2.12 Bacterial Strains 16
2.13 Preparation of Southern Blots 17
2.13.1 Separation of Genomic DNA 17
2.13.2 Southern Transfer 17
2.14 Digoxigenin (DIG) Labelling. 18
2.14.1 Estimating the Yield of a DIG-labelled Probe 19
2.15 Southern Blot Analysis Using DIG-labelled Probes 19
Removal of probe. 20
2.15.1 Stringency Washes 20
Low Stringency Washes 20
High Stringency Washes 20
2.15.2 Chemiluminescent Detection of a Hybridized Southern Blot 21
2.16 Standard Polymerase Chain Reaction (PCR) 21
2.17 Restriction Endonuclease Digests 22
2.17.1 Restriction Endonuclease Digests of Genomic DNA 22
2.17.2 *Kpn* I/*Bam* H1 Restriction Endonuclease Digests 22
2.18 Ligation 22
2.19 Transformation 23
2.20 Electrocompetent *M. smegmatis* mc^2155 23
2.21 Preparation of *M. smegmatis* mc^2155-pMIP-*M. ptb*281 template for PCR 23
2.22 Protein Expression 23
2.22.1 Small Scale Induction of pPROEX Clones 23
2.22.2 Large Scale Induction of pPROEX Clones 24
2.22.3 Protein Expression of pMIP12 Clones 24
2.23 Sonication 24
2.24 15% SDS PAGE Gel Electrophoresis 24

2.25 Immobilized Metal Affinity Chromatography (IMAC) 25
2.26 Centricon Centrifugal Filter Device 25
2.27 Western Blots 25
2.27.1 Immunodetection 26
Chapter 3: Results 27
3.1 Introduction 27
3.1.1 pJEM11-*M. ptb* Secreted Protein Library 27
3.1.2 Sequencing of the pJEM11-*M. ptb*281 Insert 29
Part A: Preliminary Analysis of pJEM11-M. ptb281 29
3.2 Introduction 29
3.3 Preliminary Database Search 29
3.4 Preliminary Southern Blot Analysis 33
3.4.1 Preparation of Probes 33
phoA-281 33
MPB70 34
3.4.2 Neoparasec and BCG *Eco* R1 Southern Blot Set 1 35
Low Stringency Hybridization 37
3.4.3 Neoparasec, BCG *Eco* R1 Southern Blot Set 2 39
High Stringency Hybridization 41
3.4.4 Hybridization of the BCG Positive Control Probe MPB-DIG 42
3.4.5 Conclusion to Preliminary Analysis of pJEM11-*M. ptb*281 43
Part B: Identification of the Full Length Hypothetical M. ptb281 ORF 44
3.5 Introduction 44
3.6 MPTB281 C-terminal Search 44
3.6.1 *M. ptb* Lambda Library 44
3.6.2 TIGR Database 44
M. avium Database Search 45
M. ptb Database Search 46
Search for the first downstream Stop Codon 47
3.7 MPTB281 N-terminus Search 49
3.7.1 Start of Translation 49
3.7.2 Hypothetical Shine-Dalgarno Sequence 50
3.7.3 Hypothetical Promoter Element 50
3.7.4 Signal Peptide 51
3.8 The Hypothetical *M. ptb*281 ORF 51

3.9 Conclusion to Identification of Hypothetical *M. ptb*281 ORF 52
Part C: Cloning the Hypothetical M. ptb281 ORF *53*
3.10 Introduction to Directional Cloning 53
3.10.1 Prokaryote Expression Vectors 53
pMIP12 53
pPROEX-Htb 54
3.11 *M. ptb*281 Insert Preparation 54
3.12 pMIP12 and pProEX Preparation 56
3.13 Ligation and Transformation 57
3.14 Screening of Recombinant Clones 58
3.15 Transformation of *M. smegmatis* mc^2155 with pMIP-*M. ptb*281 61
3.16 Conclusion to Cloning the Hypothetical *M. ptb*281 Gene 63
Part D: Protein Expression and Western Analysis *64*
3.17 Introduction 64
3.18 Protein Expression from *E. coli* DH10B-pPROEX-*M. ptb*281 65
3.19 Protein Expression from *M. smegmatis* mc^2155-pMIP-*M. ptb*281 67
3.20 Preparation of Western Blots 68
3.21 Immunodetection 69
3.22 Conclusion to Protein Expression and western Analysis 71
Part E: Species Distribution of the Hypothetical M. ptb281 ORF *72*
3.23 Introduction 72
3.24 Database Search 72
3.25 Southern Blot Analysis of Selected Mycobacterial Species 78
3.25.1 Preparation of Probes 78
*M. ptb*281-DIG 78
16S-DIG 78
3.25.2 *Bam* H1 Southern Blot #3 78
3.25.3 High Stringency Conditions 79
3.25.4 Low Stringency Conditions 81
3.25.5 16S Positive Control 83
3.26 PCR Analysis of Selected Mycobacterial Species 84
3.27 PCR Analysis of Selected *M. ptb* Isolates 87
3.28 Conclusion to Species Distribution 89
Chapter 4: Discussion and Conclusions 91
4.1 Hypothetical *M. ptb*281 ORF 91

Inverted repeats 91
Promoter regions 92
Signal sequence 94
Recombinant MPTB281 95
4.2 Species distribution 95
4.3 Potential use for MPTB281 96
4.4 Future Work 98
References 100
Appendix A: Nucleotide sequence. 109
Appendix B: Plasmid Maps. 113
Appendix C: Primers. 116

Chapter 1: Introduction

1.1 Johne's Disease

Johne's disease, otherwise known as paratuberculosis infects wild and domesticated ruminants such as cattle, sheep and goats, as well as deer, antelope, lamas and bison on a worldwide scale. The causative agent of Johne's disease is the obligate parasite and pathogen *Mycobacterium avium subspecies paratuberculosis* (*M. ptb*) (1). Johne's disease is an infectious chronic enteritis characterized by an incubation period that is usually measured in years which is untreatable and invariably fatal. The disease has been recognized as a significant condition of ruminants for over a century but knowledge of the biology of the causative organism and host response to infection has remained limited (2).

1.1.1 The Organism

M. ptb belongs to the taxonomic group from the Actinomycetales order that includes the pathogenic genera of *Corynebacterium, Mycobacterium, Streptomyces* and *Nocardia.* Within the mycobacteria taxon is the MAIS complex which includes *Mycobacterium avium* (*M. avium*), its derivative *M. ptb*, *M. intracellulare* and *M. scrofulaceum. M. ptb* is genotypically distinguished from *M. avium* its closest relative by the presence of 15-20 copies of an insertion element (IS) designated IS900 (3). As a weakly gram-positive rod shaped bacterium 0.5 μm wide and 1.5 μm in length, *M. ptb* is resistant to drying, acid conditions, certain disinfectants and can survive for months in water, soil and faeces (4).

1.1.2 Historical Background

In 1895, Johne and Frothingham were the first to clearly describe paratuberculosis in a cow with chronic enteritis (4). The disease was recognised as being non-tuberculosis in 1906 by Band who called the condition Johne's disease or pseudotuberculosis (4). In 1912, Twort was the first person to successfully grow *M. ptb* in culture and in 1914 showed the organism produced experimental enteritis (5).

1.1.3 Symptoms and Stages of Johne's Disease

Clinical symptoms of Johne's disease in cattle are chronic progressive weight loss with chronic or intermittent diarrhoea with a reduction in milk production and fertility. Ovine and caprine clinical cases resemble those of cattle with the exception that diarrhoea may not be present (4). Ruminants with Johne's disease do not have a fever, continue to eat and generally appear to feel well despite their obvious illness (1).

Johne's disease progresses through three distinct phases. Stage one animals are infected but do not shed the organism in their faeces and give the appearance of good health. Stage two animals intermittantly shed the organism in their faeces but are clinically normal. Stage three animals shed the organism in their faeces and have clinical symptoms of the disease. Livestock in stages one and two are classified as preclinical and shed minimal amounts of *M. ptb* in their faeces however over time this shedding results in the significant contamination of the environment (6). During the clinical terminal stages of the disease, livestock can shed up to 10^{10} organisms per gram of faeces leading to substantial environmental contamination.

1.1.4 Control of Johne's Disease

There are no requirements in New Zealand for producers to implement Johne's management programs in herds; therefore control of the disease is optional. Herd management programs usually centre on detection and removal of infected animals by test and cull measures and are aimed at reducing transmission to susceptible stock and reducing environmental contamination. Vaccines currently available in New Zealand for the protection of cattle, sheep and goats against Johne's disease are considered likely to provide an important tool for the strategic control of disease. When considering their use producers must be aware that they do not give 100% protection and do not prevent faecal shedding (7-9). In addition the live vaccine commonly causes lesions at the injection site and the vaccine stain can persist in the draining lymph nodes. As a result vaccinated animals can be mistakenly diagnosed with bovine tuberculosis at slaughter, leading to downgrading of carcasses and financial loss to farmers. Vaccinated animals may also be excluded from export markets or semen collection.

In animals already infected, vaccination may actually exacerbate development of clinical disease including shedding and losses. Vaccination also reduces the sensitivity

of certain tests used to monitor the prevalence of Johne's disease, encumbering herd management strategies. Approval from the Ministry of Agriculture and Fisheries (MAF) must first be gained before cattle can be vaccinated due to the presence of bovine tuberculosis and deer are currently unable to be vaccinated because it sensitises them to bovine tuberculosis testing.

1.1.5 Crohn's Disease

M. ptb has been detected in the intestinal tissues of human patients with Crohn's disease; a chronic enteritis of unknown etiology with pathological and clinical similarities to Johne's disease. However, there remains uncertainty over the role and relationship that *M. ptb* may play in Crohn's disease (10-13).

1.2 Economic impact to New Zealand

Johne's disease is known to affect around 12% of dairy herds in New Zealand but is unoffically thought to affect upwards of 60% (14). The prevalence of Johne's disease in sheep herds is difficult to determine because *M. ptb* sheep strains have proved to be notoriously difficult to culture (15). Less is known of its prevalence in deer and goat herds. Evaluation of economic losses due to Johne's disease in herds tends to focus on the reduced productivity of animals in the clinical stages of the disease. Little is known about the economic losses incured during the preclinical stages of Johne's disease but field experience suggests losses during the preclinical stages are insignificant compared to the clinical stages. The economic cost of Johne's disease on the livestock industry in New Zealand has been officially estimated to be just under $5 million/per year whereas unofficial estimates put the cost at $29.2 million/per year (14).

1.3 Host-pathogen Interaction

Pathogenic mycobacteria such as *M. ptb* target phagocytic sub-epithelial and intraepithelial mononuclear cells (monocyte/macrophage) (16). The method by which *M. ptb* enters host cells is not known. Once engulfed by macrophages *M. ptb* avoids cell lysis by unknown mechanisms to survive and replicate within the host macrophage. Other well-known bacterial intracellular parasites have evolved several mechanisms to adapt to or modify the intracellular environment. Well-documented strategies employed

by these bacteria are 1) modification of the phagosome that inhibits acidification, phagosome-lysosomal fusion and lysosomal enzyme activities, 2) resistance to or neutralization of the damaging effects caused by reactive intermediates, and 3) suppression of macrophage responsiveness to activating cytokines such as gamma interferon (IFN-γ) (16).

Details of the host's immune response to infection by *M. ptb* are not well known, however, they can be divided into two main stages; the preclinical stage characterised by a strong cell-mediated immune response (CMI) and the clinical stages where a strong humoral immune response dominates (16). The CMI response, encompassing the complex relationship between T-cells and macrophages infected with mycobacteria, are believed to be particularly important for the development of protective immunity to *M. ptb* (9). Thus the humoral immune response is not central to the development of protective immunity to intracellular parasites (16).

The greater degrees of protection seen with live tuberculosis vaccines than with dead vaccines suggests that proteins secreted by growing *M. tuberculosis* play an important role in eliciting protective immunity (17). Mycobacterial secreted protein antigens are recognised by T cells during the initial immune response to mycobacterial infection in human and in animal models. Proliferation of T cells and production of gamma interferon (INF-γ) in response to mycobacterial antigens are generally considered to be indicators of an antigen-specific CMI response. Therefore secreted *M. ptb* proteins that stimulate a CMI response may be useful as antigens in a diagnostic test that measures γINF.

1.4 Detection of Johne's Disease

Diagnosis of Johne's disease can be achieved by either the direct detection of *M. ptb* in faecal and tissue samples or indirectly by use of immunological assays.

1.4.1 Direct Detection

Culture

Faecal cultures directly identify *M. ptb* in samples and are considered the gold standard by which all other tests are measured (18). *M. ptb* is differentiated from other

mycobacteria by its characteristic slow growth, mycobactin dependence, colony morphology, conventional biochemical tests and antimicrobial susceptibility tests (19).

Cultures can be performed on pooled faecal samples, to ascertain the presence of Johne's disease in a herd, or on individual samples to identify infected animals. Of the tests presently available for the detection of *M. ptb,* culturing of individual faecal samples is considered to be the most accurate with a specificity of 100% and sensitivity of about 50% (20). Diagnosis by conventional culture can take as long as 12-16 weeks before a positive result can be achieved and up to 6 months have been recorded for a definitive negative. Radiometric culture allows more rapid detection of *M. ptb* in culture (as early as 9 days), however, it is relatively expensive (21).

Diagnosis by culture is time consuming, labour intensive and contamination is often a problem when *M. ptb* is cultured from faecal specimens, even when two-step decontamination procedures are performed (22). Faecal culture tests also lack sensitivity due to the intermittent and inconsistent shedding of *M. ptb* in the faeces of animals in the preclinical stages of the disease (6,18).

Microscopy

Microscopic examination of faecal or culture samples using Ziehl-Neelsen or Kinyon staining methods that rely on the ability of the bacilli to resist destaining when treated with acid or alcohol are of limited value as they can not distinguish between *M. ptb* and other mycobacteria. Fluorescence staining that uses fluorescein-labeled immunoglobulins is a more sensitive method however it has a tendency to give false positive results (23). The presence of mycobacterial clumps in granulomatous tissue from the terminal ileum and mesenteric lymph nodes during histological analysis is considered confirmatory for Johne's disease (5). For practical reasons histological examination of tissue samples is not a viable method for whole herd testing although it is useful at slaughter for monitoring the prevalence of Johne's disease in the herd.

Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR), involving the amplification of specific nucleotide sequences from genomic DNA, combined with restriction endonuclease digestion can differentiate between *M. ptb* and closely related mycobacteria (24).

However, contaminants in faecal and tissue samples can inhibit PCR reactions. As with faecal culture tests PCR analyses are unable to detect all animals in the preclinical stages of the disease due to intermitted shedding.

1.4.2 Immunological Detection

A number of tests are available that utilize humoral or CMI responses to detect Johne's disease.

Humoral immune response.

Serological tests detect antibodies directed against *M. ptb* antigens as a means to diagnose Johne's disease. The agarose gel immunodiffusion test (AGID) and the complement fixation test (CFT) are serological assays that are easy and rapid to perform. These tests use purified protein derivative (PPD) prepared from *M. ptb* (Johnin) which contains antigens common to other mycobacteria resulting in low specificity. The Parachek ™ kit, a solid phase enzyme linked immunosorbant assay (ELISA) currently available in New Zealand for the detection of Johne's disease in cattle, uses a complex mixture of *M. ptb* antigens adsorbed to a solid support. To improve the specificity, test serum is pre-incubated with *M. phlei* antigens to remove cross-reacting antibodies prior to performing the assay.

While these tests are suitable for detecting clinical cases of Johne's disease they do not detect all preclinical animals due to the absence of a strong humoral immune response during the early stages of the disease (9,25).

Cell mediated immune response (CMI).

The preclinical stages of Johne's disease are characterised by a strong CMI response and a weak humoral immune response. Diagnostic tests based on the CMI response make use of the complicated interplay between phagocytic cells and their interaction with T-cells. T-cells that recognise *M. ptb* antigens presented to them by phagocytes in association with major histocompatibility class II molecules undergo cellular proliferation and/or secrete cytokines such as γ-interferon (γINF). The response of circulating T-cells to *M. ptb* antigens can be measured by the incorporation of radioactive thymine into actively replicating cells, or by using the commercial

Bovigam® kit that measures γINF in an ELISA assay (26,27). The γINF ELISA assay is more sensitive than the lymphocyte proliferation assay and does not require the expensive and time consuming isolation of circulating T-cells, or the use of hazardous radioisotopes. Additionally, proliferation assays can take up to 7 days to generate results whereas the γINF assay can generate results within 24 hours (28).

A T-cell response to antigens can be detected using the delayed-type hypersensitivity reaction (DTH). This involves the indra-dermal injection of PPD (0.1 mL) into the caudal fold of cattle or the cervical region of the neck in deer (29). Recognition of antigens by T-cells is measured 72 hours later by the swelling at the site of the injection. This method is used extensively for TB herd testing using *M. bovis* PPD. When there is reason to doubt the results, perhaps due to the presence of *M. avium* in the area that can transiently infect cattle, a comparative cutaneous test can be performed. The comparative test uses *M. bovis* PPD and *M. avium* PPD injected indra-dermally at different sites. The injection sites are inspected 72 hours later with a larger swelling at the *M. bovis* PPD site than at the *M. avium* PPD site being interpreted as a positive result for bovine tuberculosis (5). At the present time, using a DHT test to detect Johne's disease in herds is not feasible due to the lack of suitable antigens for distinguishing between *M. ptb* and *M. avium* infections (30).

These CMI based methods are relatively fast compared to faecal cultures, however, they have serious specificity problems due to the lack of species-specific antigens. Tests that utilise antibodies to detect *M. ptb* infections are not useful for early diagnosis because seroconversion does not occur until the late clinical stages of Johne's disease. Since infected animals in the preclinical stages outnumber those in the clinical stages of Johne's disease, early diagnosis is crucial to identify potential bacterial shedders to avoid spreading the infection. In order to diagnose animals in the preclinical stages of Johne's disease, *M. ptb* specific diagnostic antigens need to be identified for use in CMI based tests.

1.5 Bovine Tuberculosis

Bovine tuberculosis is a chronic, infectious disease caused by *M. bovis*, a member of the *M. tb* (MTB) complex. Primarily infecting cattle, *M. bovis* also infects a wide range of host species including humans. Introduced to New Zealand with the first cattle and settlers in the 1840s, bovine tuberculosis has become a serious public health and

economic problem for New Zealand. Test and cull programs during the 1940s were successful in reducing the incidence of bovine tuberculosis so that by the 1960s it was believed it would be eradicated from the nation's herds. The introduction to New Zealand of the Australian Brush Tail Possum (*Trichosurus vulpecular*), a species highly susceptible to *M. bovis* infection, has hampered measures to eradicate bovine tuberculosis and is recognized as the primary feral reservoir re-infecting herds in New Zealand. Of major concern is the inability of current diagnostic tests to reliably and consistently differentiate between *M. ptb* and *M. bovis* infections.

1.6 Hypothesis and Aims

Due to the taxonomic distance between the MAIS (*M. avium*, *M. ptb*, *M. intracellulare* and *M. scrofulaceum*) and MTB (*M. tb*, *M. bovis*, *M. bovis* BCG, *M africanum*, *M. microti* (31), and *M. canetti* (32)) complexes, the hypothesis to be tested was that a number of secreted proteins from *M. ptb* may be absent from members or a member of the MTB complex. *M. ptb* secreted proteins that are absent from *M. bovis* may stimulate a CMI and/or a humoral immune response in *M. ptb* vaccinated sheep and cattle. Antigens such as these could improve the specificity of immunological based diagnostic tests for identifying animals in the preclinical stages of Johne's disease.

Using a PhoA$^+$ *M. ptb* secreted protein library clone whose sequence appeared to be absent from the MTB complex, the aims were to:

- Identify the *M. ptb* ORF responsible for the PhoA$^+$ phenotype
- Determine the species distribution of the ORF
- Clone and express the ORF
- Purify the recombinant protein for immunological analyses

Chapter 2: Materials and Methods

2.1 Materials

Antibiotics, DNA ladders, *E. coli* DH10B ElectroMAX, pPROEX-Htb vector, restriction endonucleases and DNA polymerases were obtained from Invitrogen Inc, USA whom also manufactured the oligonucleotide primers.

Proteinase K, T4 ligase, pre-stained Precision Protein Standard, mouse monoclonal anti-his_6-peroxidase, Southern blot blocking reagent, DIG-dUTP, anti-digoxygenin-Ab fab fragments, Dig Easy Hyb Granules, DIG-labelled control DNA, CDP-star, NBP/BCIP stock solution and all deoxynucleotide triphosphates were purchased from Roche Molecular Biochemicals, Germany.

The vectors pJEM11 and pMIP12 were gifts of Professor Brigitte Gicquel, Pasteur Institute, Paris, France.

QIAquick Gel Extraction kit, Concert Rapid Gel Extraction System, High Pure Plasmid Isolation Kit and Phase Lock Gel Light were obtained from the following sources: Qiagen, Germany; Invitrogen, Inc., USA; Roche Molecular Biochemicals; Germany; Eppendorf, USA.

Nylon membrane Biodyne B and Nitrocellulose membrane (0.45 μm) were obtained from Biodyne®, Pall Corporation, USA and BioRad Laboratories, USA respectively. Premixed preweighed acrylamide/bis and 0.025 μm filter discs were supplied by Millipore Corp., USA.

Electroporation chambers and GenePulser were supplied by Bio-Rad Laboratories Inc., USA, while BCIP was provided by Sigma Aldrich Ltd, USA.

Middlebrook 7H11 Agar medium and Middlebrook OADC enrichment were supplied by Difco Laboratories, USA and Becton Dickenson, New York, USA, while Mycobactin J was obtained from Allied Monitor, Inc., USA.

HiTrap 1 mL chelating columns were purchased from Amersham Pharmacia Biotech, Stockholm, Sweden.

Digoxigenin-labeled control DNA and CSPD was provided by Roche, Germany and BioMax MR film was obtained by Kodak, USA.

Centricon YM-3, 3,000 MW cut-off centrifugal filters were obtained from Millipore Corp. USA and SuperSignal® West Femto Maximum Sensitivity Substrate solutions were provided by Pierce, USA.

Automated DNA sequencing was performed using an ABI prism 377-64 DNA Sequencer from Applied Biosystems Inc, USA.

Spectrophotometer Helios λ was purchased from Unicam, UK., while the PCR Gene Amp 9600 cycler was obtained from Perkin Elmer, USA.

2.2 pJEM11-*M. ptb* Secreted Protein Library

A secreted protein library of the New Zealand *Mycobacterium avium* subspecies *paratuberculosis* (*M. ptb*) field isolate (ATCC 53950) had been made previously by Chris Dupont (33) using the vector pJEM11 (34) (Appendix A). In brief, the library was prepared as follows: *M. ptb* genomic DNA was partially digested with *Sau* 3A and size fractionated fragments of 300-3500 bp ligated into pJEM11. The vector DNA was digested with *Bam* H1 and treated with alkaline phosphatase (AP) to remove the 5' phosphate group. Ligated pJEM11 was dialysed through 0.025 μm filter discs and used to transform *Escherichia coli* (*E. coli*) DH10B by electroporation in a 0.2 cm gap chamber. The transformation mixture was spread onto Luria-Bertani (LB) plates containing 30 μg/mL kanamycin sulphate (kan) and incubated at 37° C overnight. The recombinant plasmids were recovered from *E. coli* using standard techniques and then electrocompetent *M. smegmatis* mc^2155 were transformed with the plasmid to facilitate expression of *M. ptb* genes in a homologous host. The transformation mixture was spread onto LB/kan plates supplemented with 40 μg/mL 5-bromo-4-chloro-3-indoyl phosphate disodium (BCIP) for selection of colonies with the alkaline phosphatase phenotype ($PhoA^+$). The plates were incubated at 37° C for five days then transferred to

4° C for a further 40 days to detect secreted phoA recombinant proteins expressed at low levels. During the 45 day period $PhoA^+$ colonies were collected and stored as glycerol stocks. Selected $PhoA^+$ recombinant plasmids were individually transferred back into *E. coli* DH10B for sequencing purposes.

2.3 Automated DNA Sequencing

Automated DNA sequence analysis of recombinant plasmids and PCR fragments was performed using an ABI prism 377-64 DNA Sequencer using Big Dye version 3.0 and Ampli*Taq*® DNA polymerase and the appropriate primer.

2.4 Computer Analysis

2.4.1 BLAST Search Databases

Nucleotide and translated sequences were submitted to National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/blast/blast.cgi), The Institute for Genomic Research (TIGR, http://tigrblast.tigr.org/cmr-blast), The Sanger Centre for Computational Genomics (http://www.sanger.ac.uk), and Bioinformatics (CCBG, http://www.ccgb.umn.edu/cgi-bin/common/web_blast.cgi) Protein Extraction, Description, and ANalysis Tool (PEDANT; http://pedant.gsf.de) databases and analysed by basic local alignment search tool (BLAST), (35) and FASTA (36) computer programs. Expect (E) Values $\leq$ 0.0 were considered significant. The E value is a statistically significant threshold for reporting matches against database sequences; the default value is 10, meaning that 10 matches are expected to be found merely by chance (37). Lower E thresholds are more stringent; hence they are more significant (www.ncbi.nlm.nih.gov/blast/html/blastcgihelp.html#expect). Identical nucleotide or amino acid matches were defined as identities while amino acids and their conservative substitutions were defined as similarities.

BLAST tools

BLASTN	Nucleotide vs nucleotide comparisons
BLASTP	Protein vs protein comparisons
BLASTX	Compares a nucleotide query sequence translated in all reading frames against a protein sequence database
TBLASTN	Takes a protein query sequence and compares it against an NCBI nucleotide database which has been translated in all six reading frames
TBLASTX	Converts a nucleotide query sequence into protein sequences in all 6 reading frames and then compares this to an NCBI nucleotide database which has been translated in all six reading frames

2.4.2 Translation, Molecular Weight and pI Prediction

Expert **P**rotein **A**nalysis **Sy**stem translation tool (ExPASy, http://us.expasy.org/) was used to translate nucleotide sequences into all six reading frames while Bionavigator (http://www.bionavigator.com) was used for molecular weight and pI prediction.

2.4.3 Signal Peptide Search

Signal peptide searches were performed with the first fifty N-terminal amino acid residues against PSORT (prediction of protein sorting signals and localization sites in amino acid sequences, http://psort.nibb.ac.jp/) and SignalP (Signal peptide, http://www.cbs.dtu.dk/services/SignalP/) servers using both gram negative and gram positive parameters. SignalP V 1.1 and the beta release V 2.0 were used. PSORT predictions were based on the modified McGeoch's method (38) and the Heijne's method (39). SignalP V 1.1 based its predictions on SignalP-NN (neural networks trained on SWISS-PROT rel. 29) (40) and SignalP V 2.0 based its predictions on SignalP-NN (neural networks trained on SWISS-PROT rel. 35) and SignalP-HMM (based on hidden Markov models) (41,42).

2.5 Mycobacterial Genomic DNA

2.5.1 Growth of Mycobacteria on 7H11 Agar Slopes

Mycobacterial species and strains used in this study were routinely grown at 37° C on Middlebrook 7H11 Agar medium slopes supplemented with Middlebrook oleic acid-albumin-dextrose-catalase (OADC) enrichment for 1 month or until a lawn of bacteria covered the agar surface. Slopes growing *M. ptb* contained an additional supplement of 1 mg/mL Mycobactin J enrichment, an iron-chelating agent required for the growth of *M. ptb*.

2.5.2 Recovery of Mycobacterial Genomic DNA

Mycobacterial genomic DNA was prepared by a modification of the method described in the thesis of Dr Susanne Borich (43). Using sterile technique, a lawn of cells was scraped from the 7H11 agar slopes and re-suspended in 1.5 mL pre-lysis solution (25% sucrose, 50 mM Tris pH 8.0, 25 mM EDTA) and heat treated at 70° C for 2 hr. Cells were then incubated at 37°C over night in 1.5 mL lysis solution (500 μg/mL lysozyme in pre-lysis solution). A 4 mL Proteinase K solution (100 mM Tris pH 8.0, 1.0% SDS and 400 μg/mL Proteinase K) was added and incubated at 55° C for 2 hr. After addition of 0.1 mL of 5 M NaCl, the DNA was extracted with phenol-chloroform, and ethanol precipitated at –20° C for 1 hr. The DNA was centrifuged at 12,000 rpm for 30 minutes, air-dried and re-suspended in 1 mL dH_2O. The aqueous DNA was added to a 15 mL tube containing approximately 2 mL of Phase Lock Gel Light (PLG) then subjected to 3 more phenol-chloroform extractions within the same 15 mL tube. PLG is an inert material that migrates to form a tight seal between the phases of an aqueous/organic extraction during centrifugation. The organic phase and the interphase materials are effectively trapped in or below the barrier, thus enabling complete and easy decanting or pipetting of the entire aqueous phase without interphase contamination. The aqueous phase was transferred to a fresh tube and a final chloroform wash and ethanol precipitation performed as previously described. DNA was re-suspended in 1–3 mL dH_2O and its concentration estimated on a 1% agarose gel or by absorbance at 260 nm in a 0.5 cm path length quartz cuvette using a double-beam spectrophotometer.

2.6 Analysis of DNA by Agarose Gel Electrophoresis

Loading buffer (30% glycerol, 0.1 g xylene cyanol 0.05 g bromo-phenol blue) was mixed at a ratio of approximately 1:6 with a sample of DNA and loaded onto a 1% agarose gel made with 1x Tris-acetate-EDTA (TAE) buffer (40 mM Tris, 1 mM EDTA, 20 mM glacial acetic acid) and electrophoresis in 1x TAE buffer at 100 volts for ~40-45 min. Following electrophoresis the agarose gel was soaked in solution of ethidium bromide (~2 mg/mL) for ~15-20 min then visualized by UV light on a transilluminator and a digital image was collected on Biorad GelDoc 2000 with software Quantity One, version 4.

2.7 Purification of DNA Fragments

Samples containing DNA were subjected to agarose gel electrophoresis (section 2.6). The gel was soaked in ethidium bromide (~2 mg/mL) and visualized by UV light on a transilluminator. The desired band(s) was excised and purified using either the QIAquick Gel Extraction kit or Concert Rapid Gel Extraction System as recommended by the manufacturers and stored at -25° C. DNA purification was based on the principle that in the presence of chaotropic salt, DNA will adsorb to a silica membrane (44). The silica membrane, immobilized in a plastic filter tube allows smaller contaminating molecules to be washed away from the adsorbed DNA. The purified DNA can then be eluted from the silica membrane using water or a low salt buffer. An aliquot of the purified sample was analysed by agarose gel electrophoresis to verify the integrity of the sample before being stored at -20° C.

2.8 Luria-Bertani (LB) Agar and Broth

LB broth (1% Bacto tryptone, 0.5% yeast extract, 1% NaCl,) was made in accordance with manufacturer's instructions, autoclaved and the appropriate antibiotic (50 μg/mL kan or 100 μg/mL ampicillin (amp)) added before use. LB agar plates were made by adding 1.5% Bacto agar to the LB broth prior to autoclaving. The agar was cooled to 50° C before addition of the appropriate antibiotic.

2.9 Sauton's Media

Sauton's media (2.87 mM K_2HPO_4, 4.15 mM $MgSO_4$, 30.28 mM asparagine, 0.159 mM ferric ammonium citrate, 10.4 mM citric acid, 6% glycerol, 0.1% Tween® 80, 0.1 mL/Liter 1% $ZnSO_4$ solution, pH adjusted to 6.8-7.2 with NaOH and autoclaved) was modified from Sauton, B (45). Antibiotic (50 μg/mL kan) was added before use.

2.10 Recovery of Plasmid Vectors

A single colony of the desired recombinant *E. coli* clone was transferred from an LB agar plate (1% Bacto tryptone, 0.5% yeast extract, 1% NaCl, 1.5% Bacto agar) containing the appropriate antibiotic (50 μg/mL kan or 100 μg/mL amp) to 5 mL LB (1% Bacto tryptone, 0.5% yeast extract, 1% NaCl) containing antibiotic and grown overnight with agitation at 37° C. The recombinant plasmids were recovered using the High Pure Plasmid Isolation Kit as recommended by the manufacturers. Lysis of *E. coli* bacterial cells was based on the alkaline lysis method that uses a 0.2 M NaOH, 1% SDS solution (46). Simultaneously cellular RNA was removed by RNase A in this step. The lysis solution disrupts the cell and denatures high molecular weight chromosomal DNA while covalently closed circular plasmid DNA remains intact. Upon neutralization, chromosomal DNA renatures to form an insoluble clot, that is precipitated along with cellular debris leaving plasmid DNA in the supernatant. Purification and recovery of plasmid DNA was based on the principle that in the presence of chaotropic salt, DNA will adsorb to a silica membrane (44). The silica membrane, immobilized in a plastic filter tube allows smaller contaminating molecules to be washed away from the adsorbed DNA. The purified plasmid DNA can then be eluted from the silica membrane using water or a low salt buffer. A sample of purified plasmid was analysed by agarose gel electrophoresis (section 2.6) prior to storage at -20° C.

2.11 Storage of Bacterial Cultures Grown in Liquid Medium

Samples of bacteria grown to stationary phase in liquid media were stored in 15% sterile glycerol at -70° C in a sterile airtight tube.

2.12 Bacterial Strains

Table 1: Bacterial strains used in this study

Strain	Source/Reference
M. ptb (*M. paratuberculosis*)	New Zealand field isolate ATCC 53950
Neoparasec®	*M. ptb* vaccine strain 316F (Merial)
M. bovis	ATCC 35726
M. bovis Bacilli Calmette-Guerin (BCG)	Pasteur
M. fortuitum	ATCC 6841
M. gordonae	ATCC 14470
M. intracellulare	ATCC 35848
M. kansasii	ATCC 12478
M. marinum	ATCC 927
M. phlei	Wallaceville, field isolate
M. scrofulacium	ATCC 19981
M. smegmatis mc^2155	(47)
M. terrrae	ATCC 15755
M. tb (*M. tuberculosis*) H37Ra	ATCC 25177
E. coli DH10B™ Genotype: F⁻*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80d*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ⁻ *rps*L *nup*G	ELECTROMAX™, Invitrogen Inc., USA.

2.13 Preparation of Southern Blots.

2.13.1 Separation of Genomic DNA

Aliquots of restriction endonuclease digested genomic DNA were loaded onto 0.7% agarose gels made in 1x TAE buffer; each lane containing approximately 1 μg of DNA. Electrophoresis was carried out at 96 volts for ~1 1/2 hr in fresh 1x TAE buffer. The upper left hand corner of the gel was cut and the orientation recorded to enable the correct identification of the genomic samples after Southern transfer. The gel was soaked in ethidium bromide (~2 mg/mL) diluted in 1x TAE for ~10-15 min before being aligned next to a fluorescent ruler on the surface of a clean transilluminator. The gel was visualized by UV light and the image recorded using Biorad GelDoc 2000 Software: Quantity One version 4.

2.13.2 Southern Transfer

The following method was modified from Current Protocols in Molecular Biology (48). The procedure was carried out while wearing fresh gloves with clean apparatus and bench tops. Membranes were handled with clean tweezers and gloves.

After electrophoresis, the gel was submerged in excess denaturation solution (0.5 M NaOH, 1.5 M NaCl) for 30 min with shaking to separate DNA strands then briefly washed 3x in dH_2O. The agarose gel was then submerged in excess neutralization solution (1 M Tris, 1.5 M NaCl, pH 7.5) with shaking for 2x 15 min intervals.

Four Whatman filters were cut ~4 cm larger than the 0.7% agarose gel. An excess amount of 20x SSC (3 M NaCl, 0.3 M sodium citrate, pH 7.0) transfer solution was poured into a reservoir tray. The first filter was laid on top of the solution and then more 20x SSC was poured on top. The second filter was laid on top of the first and air bubbles removed by rolling a glass rod over the filters. More 20x SSC transfer solution was poured onto the stack of filters and the process repeated until four filters had been added. A small amount of transfer solution was poured on top of the stack of filters and then the gel was laid on top so that the bottom of the wells was upper most. Air bubbles were removed then a small amount of transfer solution was poured on top of the gel. Four strips of plastic wrap were then placed over the edges of the gel. The positively-charge nylon membrane Biodyne B was cut slightly larger than the gel and with one corner cut. To enable the correct identification of the genomic samples after Southern

transfer the membrane was laid on top of the gel so their cut corners aligned. Air bubbles were removed and two dry filters cut to the same size as the membrane were laid on top followed by a stack of paper towel ~4 cm high cut to the same size as the membrane. A glass plate was placed on top of the paper towel stack and a 500 g weight placed on top.

Following overnight transfer the paper towels and filter papers were removed and the position of the wells marked on the membrane in pencil to ensure that the correct orientations could be established. The membrane was then rinsed in excess 2x SSC to remove any agarose and laid on the surface of a clean transilluminator and then exposed for 4 min with UV light to permanently cross link the DNA to the membrane. Southern blots were stored dry or in 2x SSC in a sealed bag. To assess efficiency of the transfer, the agarose gel was soaked in dilute ethidium bromide then visualized by UV light.

2.14 Digoxigenin (DIG) Labelling.

Nucleic acid probes were labelled using the DIG Nonradioactive Nucleic Acid Labelling system. This system used digoxigenin (DIG), a steroid hapten, coupled to dUTP via an alkali-labile ester-bond forming digoxigenin-11-2'-deoxy-uridine-5'triphosphate alkali-labile (DIG-11-dUTP). DIG-11-dUTP is a substrate for *Taq* DNA polymerase and replaces dTTP during PCR labelling of the desired DNA fragments. DIG-labelling of DNA fragments was routinely carried out by PCR using the thermal cycle program as follows: pre-incubation at 95° C for 10 min followed by 35 cycles of 94° C melting temperature for 45 sec, 60° C anneal for 1 min, and a 72° C extension for 2 min with a final 72° C extension for 10 min. The 50 µL PCR mixture contained 1.5 units of *Taq* DNA polymerase with 1.5 mM $MgCl_2$ and 1x *Taq* polymerase buffer, 20 pmol of each primer, in the presence of 10% dimethyl-sulphoxide (DMSO), 20 µM of DIG-dUTP, and 0.5 mM of dATP, dCTP and dGTP with ~ 10 ng of template. After labelling, the PCR mixture was analysed by agarose gel electrophoresis (section: 2.6) and DNA fragments of the expected size excised and purified using either the QIAquick Gel Extraction kit or Concert Rapid Gel Extraction Kit as previously described (section: 2.7) and stored at -25° C.

2.14.1 Estimating the Yield of a DIG-labelled Probe

To estimate the yield of a DIG-labelled probe a 1 μL sample of the newly labelled experimental DNA probe was taken and serially diluted. A dot blot was prepared by taking 1 μL aliquots from each dilution and spotting them in a row on a positively-charge nylon membrane. In a second row 1 μL spots of DIG-labelled control DNA ranging in concentration from 0.01 pg to 10 pg were spotted adjacent to the experimental probe spots. After the spots had dried the DNA was cross linked to the membrane by exposure to UV light for 4 min before being rinsed briefly in excess wash buffer (maleic acid buffer with 0.3% Tween® 20). Block solution (1x) was prepared in maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl, pH 7.5; autoclaved) from a 10x stock solution (10% w/v Block reagent dissolved in maleic acid buffer). The dot blot was incubated in excess 1x Block solution at room temperature with gentle agitation for 15 min. An anti-digoxigenin alkaline phosphatase sheep Fab fragment (anti-DIG-AP) solution was prepared by diluting anti-DIG-AP 1:5000 (150 mU/mL) in 1x Block solution. The dot blot and 2 mL of anti-DIG-AP solution were transferred to a sealed container and incubated with gentle agitation at room temperature for 15 min. The dot blot was removed from the container and washed 2x 5 min in excess wash buffer before being equilibrated for 2 min in excess detection buffer (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5: autoclaved). From a 50x NBT (nitro blue tetrazolium chloride)/BCIP (5-bromo-4-chloro-3-indolyl phosphate disodium) stock, (18.75 mg/mL NBT, 9.4 mg/mL BCIP in 67% DMSO (v/v)), a 1x NBT/BCIP AP substrate solution was made in detection buffer to a final concentration 375 μg/mL NBT and 188 μg/mL BCIP. The dot blot was laid in a shallow dish and covered with 1x NBT/BCIP AP substrate solution. Excess solution was poured off and the membrane incubated in the dark until the desired degree of development had occurred. This was a subjective measure gauged visually by the formation of a blue precipitate in the vicinity of the control and experimental probe spots. Direct comparison between the intensities of the diluted probe and the control DNA spots after colorimetric detection enabled the DIG-labelling yields to be estimated.

2.15 Southern Blot Analysis Using DIG-labelled Probes

The following stringency conditions were adopted based on previous experience in the laboratory (pers. comm.). A Southern blot was placed in a plastic bag containing

2.5 mL DIG Easy Hyb solution per 100 cm^2 of membrane and sealed to exclude air. The bag was place in a gently agitating water bath set to the appropriate hybridization temperature and incubated for 30 min. Under high stringency conditions the hybridization temperature was 42° C (high stringency hybridization) and for low stringency conditions the hybridization temperature was 40° C (low stringency hybridization). The probe, used at 25 ng/mL hybridization solution was prepared by adding the probe to 50 µl of water in a 1.5 mL tube. The tube was placed in a boiling water bath for 10 min to render the probe single stranded. The tube was briefly centrifuged and its contents quickly added to the hybridization solution in the plastic bag avoiding direct contact with the membrane. The hybridization bag was resealed to exclude air and returned to the water bath to incubate overnight. A glass plate was placed over the bag to keep it flat. On completion of the incubation period the hybridization solution was discarded and the blot washed at the appropriate stringency with post- hybridization solutions.

Removal of probe.

Membrane was rinsed in dH_2O for 1 min then incubated with agitation 2x 10 mins in alkaline probe stripping solution (0.2% NaOH, 0.1% SDS) at 37° C. Membrane was rinsed 2x 5 min in 2x SSC and stored a 4° C in fresh 2x SSC.

2.15.1 Stringency Washes

Low Stringency Washes

Unless stated otherwise all washes were performed using gentle agitation in wash solutions made from 20x SSC and 10% sodium dodecylsulphate (SDS) solutions. Following low stringency hybridization the blot was subjected to low stringency washes as follows; 2x 5 min in excess 2x SSC, 0.1% SDS at room temperature, 2x 15 min in excess 1x SSC, 0.1% SDS at 68° C and 2x 15 min in excess 0.5x SSC, 0.1% SDS at 68° C.

High Stringency Washes

Following high stringency hybridization the Southern blot was subjected to high stringency washes as follows; 2x 5 min in excess 2x SSC, 0.1% SDS at room

temperature, followed by 2x 15 min in excess 1x SSC, 0.1% SDS, 2x 15 min in excess 0.5x SSC, 0.1% SDS and a final 2x 15 min wash in excess 0.1x SSC, 0.1% SDS at 68° C

2.15.2 Chemiluminescent Detection of a Hybridized Southern Blot

The following incubations were performed at room temperature with gentle agitation. After hybridization and stringency washes the membrane was briefly rinsed for 1 min in excess wash buffer. The membrane was incubated in 1x Block solution at a ratio of 100 mL 1x Block soln./100 cm^2 of membrane for 30 min. For chemiluminescent detection the anti-DIG-AP solution was prepared by diluting anti-DIG-AP 1:10000 (75 mU/mL) in 1x Block solution. The membrane was transferred to the anti-DIG-AP solution and incubated for 30 min. The anti-DIG-AP solution was discarded and the membrane washed for 2x 15 min in excess wash buffer before being equilibrated for 2 min in 20 mL/100 cm^2 detection buffer. Excess detection buffer was drained from the Southern blot and then laid onto a plastic sheet. The two AP substrates used for chemiluminescent detection were CSPD® and CDP^{star}. Approximately 0.5 mL of CDP^{star} for every 100 cm^2 of membrane or 1 mL of CSPD® diluted 1/100 in detection buffer was evenly distributed around the perimeter of the blot. The substrate was evenly distributed over the membrane by repeatedly folding and lifting the other half of the plastic sheet over the membrane. The blot was sealed inside the folded plastic sheet and placed in an X-ray cassette with a reflective screen. The chemiluminescent signal was detected by BioMax MR film which was subsequently developed.

2.16 Standard Polymerase Chain Reaction (PCR)

Amplification was achieved using a Gene Amp 9600 cycler and a thermal cycle program of 94° C for 10 min followed by 35 cycles of 94° C melting temperature for 30 sec, 55° C annealing for 30 sec, and a 72° C extension for 1 min with a final 72° C 10 min extension unless stated otherwise. The standard PCR used 2.5 units of *Taq* DNA polymerase per 20 µL reaction with 1.5 mM $MgCl_2$, 1x *Taq* polymerase buffer (20 mM Tris-HCL (pH 8.2), 50 mM KCl), 10 pmol of each primer and 0.1 mM of each dNTP. PCRs were performed in the presence of 10% DMSO except when 16S primers (246

and 264) and MPB70 primers (MPB71 and MPT72) were used. Approximately 5 ng of template was used per 20 µL reaction.

Where stated 1 unit of PLATINUM *Pfx* DNA Polymerase with 2 mM $MgSO_4$ and 1x *Pfx* Amplification Buffer (Invitrogen Inc, USA) were used instead of *Taq* DNA polymerase, $MgCl_2$ and 1x *Taq* polymerase buffer in a standard (20 µL) PCR.

2.17 Restriction Endonuclease Digests

2.17.1 Restriction Endonuclease Digests of Genomic DNA

Genomic DNA was digested at 37° C for ~16 hr with restriction endonucleases (~1 µg DNA/5 units restriction enzyme) *Eco* R1 and *Bam* H1, used in accordance with manufacturers instructions. Samples were analysed on 1% agarose gels by electrophoresis in 1x TAE buffer to check the DNA was fully digested.

2.17.2 *Kpn* 1/*Bam* H1 Restriction Endonuclease Digests

Restriction endonucleases, *Kpn* 1 and *Bam* H1 were used in accordance with manufacturers' instructions. Aliquots of purified *M. ptb*281 PCR product containing approximately 150 ng of DNA or vector containing approximately 500 ng of DNA were separately incubated with 20 units of *Kpn* 1 in 1x buffer (20 mM Tris-HCL (pH 7.4), 5 mM $MgCl_2$, 50 mM KCL) in a volume of 20 µL and incubated overnight at 37° C. After this period 4 µL of vector was removed and stored for analysis by agarose gel electrophoresis. The remaining *Kpn* 1 digests were subjected to *Bam* H1 digestion overnight at 37° C by adding 20 units *Bam* H1 in 1x buffer (10 mM Tris-HCL, 10 mM NaCl, 5 mM $MgCl_2$, 1 mM 2-mercaptoethanol, pH 8.0) to the *Kpn* 1 digest and the volume made up to 40 µL. The *Kpn* 1/*Bam* H1 digests were then purified by agarose gel electrophoresis as described in section 2.7.

2.18 Ligation

Ligations were carried out overnight at room temperature with T4 DNA Ligase in accordance with manufacturer's recommendations. A ratio of approximately 1:3 (vector:insert) was mixed with 1 unit of T4 Ligase in a total volume of 40 µL.

2.19 Transformation

A 20 µL sample of ligation mixture or 20 µL of purified plasmid vector diluted in a ratio of 1:10 was dialysed on a 0.025 µm filter disc for 20 mins. The dialysed sample was mixed with 20 µL of chilled *E. coli* DH10B cells or electrocompetent *M. smegmatis* mc^2155 (43) and transferred to a chilled 0.1 cm gap electroporation chamber and electroporated as per the manufacturer's recommendations. LB broth (200 µL) was added to the chamber and the mixture incubated for 2-2½ hours at 37° C. The transformation mixture was spread on an LB agar plate containing 50 µg/mL of kan or 100 µg/mL amp and incubated at 37° C overnight for *E. coli* DH10B or 72 hours for *M. smegmatis* mc^2155.

2.20 Electrocompetent *M. smegmatis* mc^2155

Electrocompetent *M. smegmatis* mc^2155 cells were previously made using a method modified from Ranes *et al.* J. Bact. 1990: 199 p2195-2197 that substituted glycerol for sucrose (43).

2.21 Preparation of *M. smegmatis* mc^2155-pMIP-*M. ptb*281 template for PCR

A single colony was transferred to 5 mL Sauton's media containing 50 µg/mL kan and incubated at 37° C with agitation for 72 hours. A 1 mL sample from the Sauton's culture was centrifuged at 14000 g and the supernatant discarded. The pellet was washed twice by re-suspending in sterile water and repeating the centrifugation step. The washed pellet was re-suspended in 400 µL of sterile water and boiled for 10 mins before being frozen at -70° C. The sample was thawed, mixed and centrifuged at 14000 g to pellet cellular debris before a 5 µL aliquot was used as template in a standard PCR.

2.22 Protein Expression

2.22.1 Small Scale Induction of pPROEX Clones

A starter culture was prepared by inoculating 5 mL LB broth containing 100 µg/mL amp with a single colony of the desired *E. coli* DH10B pPROEX clone and incubating overnight at 37° C with agitation. The starter culture was used to inoculate 10 mL LB/amp broth to obtain a dilution with an OD_{600nm} reading of 0.1 and the culture incubated at 37° C with agitation. When the culture had reached an OD_{600nm} reading

between 0.5 and 1.0 a 1 mL pre-induction sample was collected, centrifuged at 14000 g, the supernatant discarded and the pellet stored at -70° C. The remaining culture was induced with IPTG to a final concentration of 0.2 mM or 1 mM and incubated at 37° C with agitation. After 4 hours a 1 mL induced sample was collected, centrifuged, the supernatant discarded and the pellet stored at -70° C. The remaining culture was incubated overnight before being centrifuged, the supernatant discarded and the wet weight of the pellet determined prior to storage at -70° C.

2.22.2 Large Scale Induction of pPROEX Clones

Induction of a 150 mL LB/amp culture with the desired pPROEX *E. coli* DH10B clone was performed as stated above (section 2.22.1) using a 10 mL starter culture and inducing to a final concentration of 0.2 mM IPTG overnight.

2.22.3 Protein Expression of pMIP12 Clones

Sauton's starter cultures (5 mL) containing 50 μg/mL kan were inoculated with a single colony of the appropriate *M. smegmatis* mc^2155-pMIP clone and incubated with agitation at 37° C for 4 days. Fresh 50 mL Sauton's/kan cultures were prepared and inoculated with a 1 mL aliquot of the appropriate starter culture and incubated as above for 7 days. The 50 mL cultures were centrifuged at 14000 g and the pellets stored at -70° C until needed.

2.23 Sonication

Pellets were re-suspended in phosphate buffered saline (PBS; 137 mM NaCl, 2.68 mM KCl, 4.02 mM Na_2HPO_4, 1.76 mM KH_2PO_4, pH 7.4) or start buffer (20 mM Na_2HPO_4, 0.5 M NaCl, 15 mM imidzole, pH 7.4) as indicated and sonicated on ice 8x 30 sec (*E. coli*) or 1x 40 sec (*M. smegmatis* mc^2155).

2.24 15% SDS PAGE Gel Electrophoresis

Protein samples were boiled for 10 mins in 1x loading buffer (2x stock; 0.125 M Tris-Cl pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.2% w/v bromophenol blue), applied to a 15% polyacrylamide gel (PAGE) (Stacking gel; 13%

acrylamide, 0.125 M Tris-HCl, pH 6.8, 0.1% SDS, 0.05% w/v ammonium persulphate, 0.05% v/v TEMED, Resolving gel; 15% acrylamide, 0.375 M Tris-HCL, pH 8.8, 0.1% SDS, 0.1% w/v ammonium persulphate, 0.07% v/v TEMED) and electrophoresed in tank buffer (0.02 M Tris base, 0.192 M glycine, pH 8.3, 0.1% SDS) until the dye front reached the end of the gel. Gels were soaked in stain (0.125% Coomassie blue, 50% methanol, 10% acetic acid) with gentle agitation for ~15 mins and de-stained (50% methanol, 10% acetic acid) with gentle agitation until protein bands were clearly visible.

2.25 Immobilized Metal Affinity Chromatography (IMAC)

HiTrap 1 mL Chelating columns were used in accordance with manufacturer's recommendations. A column was charged with 1 mL 0.1 M $NiCl_2$ and equilibrated with 5 mL start buffer (20 mM Na_2HPO_4, 0.5 M NaCl, 15 mM imidzole, pH 7.4). Samples were loaded onto the column in 1 mL aliquots, each of which was left to adsorb for ~2 mins before the next aliquot was pushed through and a fraction collected. Unbound proteins were washed from the column using 1 mL start buffer and the fraction collected. Bound proteins were eluted from the column using 1 mL aliquots of elution buffer containing increasing amounts of imidazole (start buffer containing 50 mM, 100 mM or 250 mM imidazole) and 1 mL fractions collected.

2.26 Centricon Centrifugal Filter Device

A Centricon YM-3, 3,000 MW cut-off centrifugal filter was used in accordance with the manufacturer's instructions.

2.27 Western Blots

Protein samples were subjected to 15% SDS PAGE gel electrophoresis (section 2.24) with a sample of Pre-stained Precision Protein Standard and the gel soaked in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol, pH 9.5) 2x 10 min to equilibrate. Nitrocellulose membrane (0.45 μm) was cut to slightly larger than the gel, pre-wetted in methanol and soaked in transfer buffer for 2x 10 min. Two sheets of filter paper were cut to a size larger than the membrane and soaked in transfer buffer with two pads. The polyacrylamide gel was laid on a sheet of filter paper, overlaid with the membrane and the second sheet of filter paper was laid on top. The filter paper

containing the gel and membrane was sandwiched between the pads and placed inside a porous plastic transfer cassette. The cassette was placed in a tank containing transfer buffer so that the membrane was between the gel and the anode. Electrophoresis was performed with cooling at 100 volts for ~45 min. The sandwich was then disassembled and the position of the lanes transferred to the membrane before the gel was removed and stained with Coomassie blue. The membrane was rinsed in dH_2O and soaked in 1x Ponceau S (10x stock; 2 g Ponceau S dissolved in 30 mL glacial acetic acid and brought up to 100 mL with dH_2O) for 5 mins and the excess washed off with dH_2O until the bands were resolved. The membrane was photocopied, the lanes cut into individual strips, rinsed in dH_2O for 10 mins to remove the Ponceau S and stored at 4° C in a sealed bag until needed.

2.27.1 Immunodetection

All incubations were performed at room temperature. Western blots were equilibrated in western wash (20 mM Tris pH 7.4, 100 mM NaCl, 0.1% Tween 20) for ~2 min then incubated with gentle agitation for 1 hr in excess western block solution (20 mM Tris pH 7.4, 100 mM NaCl, 0.1% Tween 20, 5% Skim milk). Mouse monoclonal anti-his_6-peroxidase was used in accordance with manufacturer's instructions and diluted 1/500 in western block solution (100 mU/mL). The membrane was incubated in anti-his_6-peroxidase/block solution with gentle agitation for 1 hr, washed for 6x 5 min in 1x western wash and placed in an open plastic bag. Chemiluminescent SuperSignal® West Femto Maximum Sensitivity Substrate solutions were mixed 1:1 and 5 μL per strip pipetted onto the membrane. The other half of the plastic bag was used to draw the substrate across the length of the membrane and the bag sealed. The membrane was then exposed to BioMax MR film for 1 sec, 10 sec, 1 min, 30 min, overnight and the film developed.

Chapter 3: Results

3.1 Introduction

During the preclinical stage of Johne's disease the host's response to the organism is in the form of a strong cell-mediated immune response (CMI) (16). Work with pathogenic mycobacteria has demonstrated that part of the CMI response is directed against secreted proteins (17,49,50). At present CMI based detection methods are unable to reliably differentiate between mycobacterial infections such as bovine tuberculosis and Johne's disease caused by *M. bovis* and *M. ptb* respectively due to the lack of species specific antigens. Thus the identification of species specific secreted proteins from *M. ptb* that are recognised by the CMI response could potentially improve the specificity of current CMI based diagnostic tests.

Reporter gene technology has been widely used to identify and isolate secreted proteins from a number of different micro-organisms. Previously the *Escherichia coli* (*E. coli*) alkaline phosphatase (*phoA*) reporter gene was used to identify secreted proteins from *M. ptb* expressed in a mycobacterial model (33).

3.1.1 pJEM11-*M. ptb* Secreted Protein Library

The secreted protein library of the New Zealand *Mycobacterium avium* subspecies *paratuberculosis* (*M. ptb*) field isolate (ATCC 53950) made previously (section: 2.2) by Dupont et. al. was constructed using the vector pJEM11 (Appendix B) supplied by Professor Brigitte Gicquel (Pasteur Institute, Paris). The pJEM11 vector contains a truncated *Escherichia coli* (*E. co*li) alkaline phosphatase (*phoA*) reporter gene (Acc#01659) devoid of its own promoter, ribosome binding site, N-terminal signal sequence and start codon and is unable to drive its own expression and export. Adjacent to *phoA* is a multiple cloning site that allows the fusion of genomic material directly to *phoA*. PhoA activity, otherwise known as the PhoA phenotype ($PhoA^+$) will only occur if a genomic fragment carrying the N-terminal encoding region of a secreted protein with its promoter, ribosome binding site, start codon and signal sequence is fused in frame with *phoA*. Hence, information within the *M. ptb* genomic fragment is ultimately responsible for the expression and export of the PhoA-hybrid protein. Export of the PhoA-hybrid protein out of the reducing environment of the cytosol across the plasma

membrane causes the formation of disulfide bonds essential for the correct formation and activation of the PhoA (51). Interaction of active PhoA with its chromogenic substrate 5-bromo-4-chloro-3-indolyl phosphate (BCIP) causes a blue precipitate to form turning the colony blue, enabling un-aided visual detection of $PhoA^+$ clones (34). Thus only blue colonies will contain recombinant plasmids carrying the N-terminal encoding region of secreted *M. ptb* protein conveniently fused in frame with and adjacent to *phoA* (Figure 1).

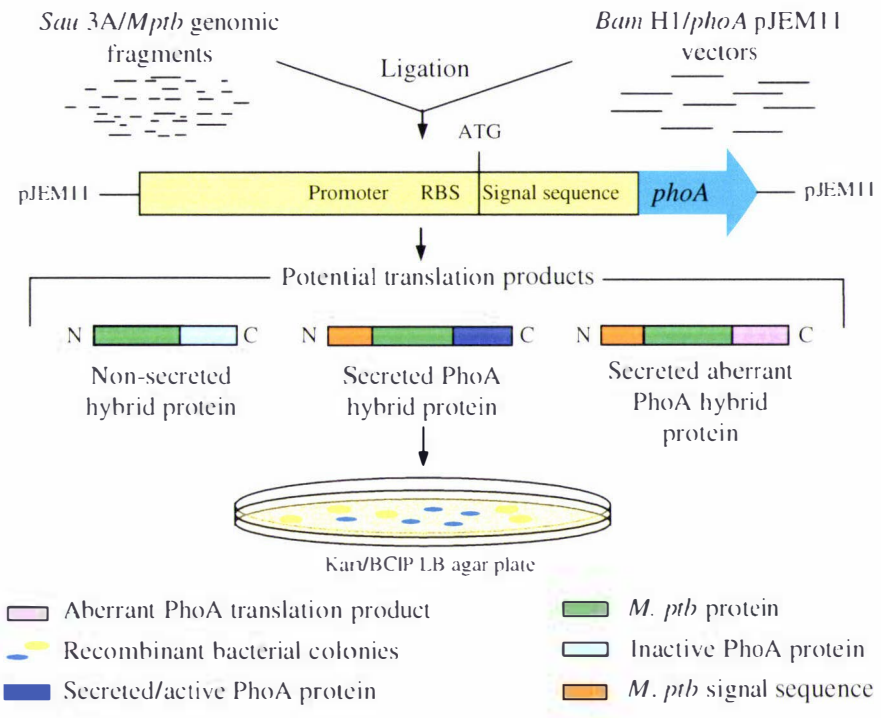


Figure 1: The pJEM11 library and $PhoA^+$ selection. Genomic fragments, randomly ligated into pJEM11 vectors whose transcriptional products run into and in frame with *phoA* will produce a PhoA-hybrid protein with detectable $PhoA^+$ activity and turn the colony blue. N = amino termini; C = carboxyl termini; ATG, start codon; RBS, ribosome binding site.

Designed to bind to sequences flanking the *M. ptb* genomic insert, the pJEM11 sequencing primers JEM1 and JEM2 (Appendix C) allow the intervening sequences to be determined by DNA sequencing (33). The JEM2 primer binds to the reading strand within *phoA* approximately 100 bp from the insert enabling the nucleotide sequence

adjacent to *phoA* to be determined thus enabling the N-terminal encoding region of the gene responsible for the $PhoA^+$ activity to be identified.

3.1.2 Sequencing of the pJEM11-*M. ptb*281 Insert

The pJEM11-*M. ptb*281 insert was estimated to be ~3100 bp (52), a length exceeding the ability of DNA sequencing to cover with the two flanking primers JEM1 and JEM2. To sequence the entire insert the internal primers KatE-i1, KatE-i2, KatE-i3 and KatE-i4 (Appendix C) were used with JEM1 and JEM2 which revealed a 2043 bp insert (Appendix A).

Part A: Preliminary Analysis of pJEM11-*M. ptb*281

3.2 Introduction

The starting material for this work was the nucleotide sequence of the 2043 bp insert from the pJEM11-*M. ptb* plasmid library clone, pJEM11-*M. ptb*281 (Appendix A) (52). In this section the genomic insert from pJEM11-*M. ptb*281 was analysed using internet based bioinformatics tools to determine the identity of the gene(s) responsible for the $PhoA^+$ activity and to ascertain its distribution across bacterial species.

3.3 Preliminary Database Search

The National Centre for Biotechnology Information (NCBI) database was searched using basic local alignment search tools (BLAST) with the *M. ptb* insert sequence from pJEM11-*M. ptb*281. The most significant BLASTN alignment was to *katE* which encodes for the catalase HPII of *Mycobacterium avium* (*M. avium*), a close relative of *M. ptb* (Appendix A). The BLASTN alignment covered the entire 2043 bp *M. ptb* query sequence and shared 98% identity with the *M. avium* subject sequence with an expected (E) value of 0.0.

An artefact of this type of library is the generation of inserts made up from multiple instead of single *Sau* 3A digested genomic fragments. This can occur when *Sau* 3A digested genomic fragments are randomly ligated together prior to their ligation into the *Bam* H1 digested vector. Thus the order of the genomic fragments within the insert will not reflect the order found in the native genomic material. If this event is not identified in a clone, erroneous conclusions may be drawn from database search results.

The un-interrupted nucleotide alignment of the entire *M. ptb* query sequence to the *M. avium* sequence is evidence that the insert was derived from a single *Sau* 3A digested fragment.

A BLASTX search also reported a significant alignment to *M. avium katE* (Figure 2). The alignment had a significant E value of 0.0, shared 96% identities, 97% positives and was in the +2 reading frame, not the *phoA* +1 reading frame.

```
gi|762829|gb|AAC18407.1|   catalase HPII [Mycobacterium avium]
Expect = 0.0, Identities = 621/644 (96%), Positives = 628/644 (97%), Frame+2

Query: 2    ITHFDHERIPERVVHARGAGAYGYFEPYDDRLAQYTAAKFLTSPGTRTPVFVRFSTVAGS 181
            ITHFDHERIPERVVHARGAGAYGYFEPYDDRLAQYTAAKFLTSPGTRTPVFVRFSTVAGS
Sbjct: 63   ITHFDHERIPERVVHARGAGAYGYFEPYDDRLAQYTAAKFLTSPGTRTPVFVRFSTVAGS 122

Query: 182  RGSADTVRDVRGFATKFYTEQGNYDLVGNNFPVFFIQDGIKFPDFVHAVKPEPHNEIPQA 361
            RGSADTVRDVRGFATKFYTEQGNYDLVGNNFPVFFIQDGIKFP  VHAVKPEPHNEIPQA
Sbjct: 123  RGSADTVRDVRGFATKFYTEQGNYDLVGNNFPVFFIQDGIKFPVLVHAVKPEPHNEIPQA 182

Query: 362  QSAHDTLWDFVSLQPETLHAIMWLMSDRALPRSYRMMQGFGVHTFRLVNARGRGTFVKFH 541
            QSAHDTLWDFVSLQPETLHAIMWLMSDRALPRSYRMMQGFGVHTFRLVNARGRGTFVKFH
Sbjct: 183  QSAHDTLWDFVSLQPETLHAIMWLMSDRALPRSYRMMQGFGVHTFRLVNARGRGTFVKFH 242

Query: 542  WKPRLGVHSLIWDECQKIAGKDPDYNRRDLWEAIESRQYPEWELGVQLVAEDDEFSFDFD 721
            WKPRLGVHSLIWDECQKIAGKDPDYNRRDLWEAIES QYPEWELGVQLVAEDDEFSFDFD
Sbjct: 243  WKPRLGVHSLIWDECQKIAGKDPDYNRRDLWEAIESGQYPEWELGVQLVAEDDEFSFDFD 302

Query: 722  LLDATKIIPEEQVPVLPVGKMVLNRNPDNFFAETEQVAFHTANVVPGIDFTNDPLLQFRN 901
            LLDATKIIPEEQVPVLPVGKMVLNRNPDNFFAETEQVAFHTANVVPGIDFTNDPLLQFRN
Sbjct: 303  LLDATKIIPEEQVPVLPVGKMVLNRNPDNFFAETEQVAFHTANVVPGIDFTNDPLLQFRN 362

Query: 902  FSYLDTQLIRLGGPNFAQLPVNRPVAQVRTNQHDGYGQHTIPQGRSSYFKNSIGGGCPAL 1081
            FSYLDTQLIRLGGPNFAQLPVNRPVAQVRTNQHDGY QH IPQGRSSYFKNSIGGGCPAL
Sbjct: 363  FSYLDTQLIRLGGPNFAQLPVNRPVAQVRTNQHDGYAQHAIPQGRSSYFKNSIGGGCPAL 422

Query: 1082 ADENVFRHYTQRVDGQTIGKRAEAFQNHYGQARMFFKSMSPVEAEHIVAAFAFELGKVEM 1261
            ADE+VFRHYTQRVDGQTIGKRAEAFQNHYGQARMFFKSMSPVEAEHIVAAFAFELGKVEM
Sbjct: 423  ADEDVFRHYTQRVDGQTIGKRAEAFQNHYGQARMFFKSMSPVEAEHIVAAFAFELGKVEM 482

Query: 1262 PEIRSAVVAQLARVDDQLAAQVAAKLGLPEPPEEQVDESAPVSPAVSQVTDGGDTIASRR 1441
            PEIRSAVVAQLARVDDQLAAQVAAKLGLPEPPEEQVDESAPVSPA+SQVTDGGDTIASRR
Sbjct: 483  PEIRSAVVAQLARVDDQLAAQVAAKLGLPEPPEEQVDESAPVSPALSQVTDGGDTIASRR 542

Query: 1442 IAVLAPDGVNVVGTHRFTELMEHPGAVVKVLAPVPGRTLAGGSARKLRVDRSFTTMASVL 1621
            IAVLA DGV+VVGT RFTELME  GAVV+VLAPV G TLAGGS  +LRVDRSFTTMASVL
Sbjct: 543  IAVLAADGVDVVGTQRFTELMEQRGAVVEVLAPVAGGTLAGGSGGELRVDRSFTTMASVL 602

Query: 1622 YNAVVVACEPRSVSTLSEQRYAVHFVTEAYKHLKPIGAYGAGVDLLRKAGIDNRLAEDTD 1801
            Y+AVVVAC PRSVSTLS+  YAVHFVTEAYKHLKPIGAYGAGVDLLRKAGIDNRLAEDTD
Sbjct: 603  YDAVVVACGPRSVSTLSDDGYAVHFVTEAYKHLKPIGAYGAGVDLLRKAGIDNRLAEDTD 662

Query: 1802 VLNDQAVVTTKAAADELPERFAEEFAAALAQHRCWQRRTDAVPA 1933
            VLNDQAVVTTKAAADELPERFAEEFAAALAQHRCWQRRTDAVPA
Sbjct: 663  VLNDQAVVTTKAAADELPERFAEEFAAALAQHRCWQRRTDAVPA 706
```

Figure 2: NCBI BLASTX search results. Alignment between the translated pJEM11-*M. ptb*281 insert query sequence and *M. avium katE*. Query = translated nucleotide sequence from pJEM11-*M. ptb*281 insert; Sbjct = amino acid sequence of KatE; + = conserved residues. Query numbers refer to the nucleotide sequence residue whereas the subject numbers refer to the amino acid sequence residues.

The alignment included the first 1936 bp of the 2043 bp insert sequence to the last 643 amino acid residues of the 706 residue KatE protein indicating the 5' region of *katE* was absent from the insert while the 3' region was located upstream of *phoA*.

The combined BLAST database results indicate the presence of a *M. ptb* homolog of *M. avium katE* in the pJEM11-*M. ptb*281 insert. The published *M. avium katE* sequence (53) and database results were used to locate the *M. ptb katE* gene within the pJEM11-*M. ptb*281 insert (Figure 3).

```
GATCACCCACTTCGACCACGAGCGCATCCCGGAGCGTGTGGTGCATGCGCGCGGGGCCGG   60
CGCCTACGGCTATTTCGAACCGTACGACGACCGGTTGGCGCAGTACACGGCGGCGAAATT  120
TCTGACCTCGCCGGGCACCAGGACGCCGGTGTTCGTGCGTTTTTCGACGGTCGCCGGATC  180
GCGCGGTTCGGCCGACACCGTCCGCGACGTGCGCGGGTTCGCCACCAAGTTCTATACCGA  240
ACAAGGCAATTACGACTTGGTGGGCAACAACTTCCCGGTGTTCTTCATCCAGGACGGCAT  300
CAAGTTCCCCGACTTCGTGCACGCGGTGAAACCCGAGCCGCACAACGAGATTCCGCAGGC  360
GCAGTCGGCGCACGACACGCTGTGGGACTTCGTGTCGCTGCAGCCCGAGACGTTGCACGC  420
CATCATGTGGCTGATGTCGGACCGGGCGCTGCCGCGCAGCTACCGCATGATGCAGGGGTT  480
CGGGGTGCACACCTTCCGGCTGGTGAACGCCCGCGGCCGAGGGACTTTCGTGAAGTTCCA  540
CTGGAAGCCCCGACTCGGCGTGCACTCGCTGATCTGGGACGAATGCCAGAAGATCGCCGG  600
CAAAGACCCCGATTACAACCGCCGCGACCTGTGGGAGGCCATCGAATCCCGCCAGTACCC  660
GGAGTGGGAGCTGGGCGTGCAGCTGGTCGCCGAGGACGACGAGTTCAGCTTCGACTTCGA  720
TCTGCTGGACGCGACGAAAATCATTCCGGAAGAACAGGTTCCGGTATTGCCGGTGGGCAA  780
GATGGTGTTGAACCGCAACCCCGACAACTTCTTCGCCGAGACCGAGCAGGTCGCTTTTCA  840
CACCGCCAACGTGGTGCCGGGCATCGATTTCACCAACGACCCGTTGCTGCAGTTCCGCAA  900
CTTCTCCTATCTGGACACGCAGCTGATCCGGTTGGGCGGCCCCAACTTCGCGCAGCTGCC  960
GGTCAACCGCCCGGTGGCGCAGGTGCGGACCAACCAGCACGACGGTTACGGGCAGCACAC 1020
GATTCCGCAGGGCCGGTCCAGTTACTTCAAGAACAGCATCGGCGGCGGTTGTCCCGCACT 1080
GGCCGACGAGAACGTGTTCCGGCACTACACCCAGCGGGTGGACGGGCAGACGATCGGCAA 1140
GCGCGCCGAGGCGTTCCAGAACCACTACGGCCAGGCGCGGATGTTCTTCAAGAGCATGTC 1200
GCCGGTGGAGGCCGAACACATCGTGGCCGCCTTCGCCTTCGAACTCGGCAAGGTGGAGAT 1260
GCCCGAAATCCGTTCCGCGGTGGTGGCACAACTCGCCCGCGTCGATGACCAGCTGGCCGC 1320
CCAGGTCGCGGCGAAACTGGGGCTGCCCGAGCCGCCGGAGGAGCAGGTGGACGAGTCGGC 1380
ACCGGTTTCCCCGGCCGTTTCGCAGGTCACCGACGGCGGCGACACCATCGCGTCGCGCCG 1440
GATCGCGGTGCTGGCCCCCGACGGCGTCAACGTGGTGGGCACGCACCGCTTCACCGAGCT 1500
GATGGAGCACCCCGGCGCGGTGGTCAAGGTGCTGGCCCCGGTGCCCGGCCGCACGTTGGC 1560
GGGTGGGTCCGCCCGCAAGCTGCGGGTGGACCGGTCCTTCACGACGATGGCGTCGGTGCT 1620
CTACAACGCGGTGGTGGTGGCGTGCGAACCGCGGTCGGTGTCAACGCTGTCCGAACAGCG 1680
GTACGCCGTGCACTTCGTCACCGAGGCCTACAAACACCTCAAGCCGATCGGCGCCTACGG 1740
GGCCGGTGTCGACCTGCTCCGCAAGGCCGGCATCGACAACCGGCTCGCCGAGGACACCGA 1800
CGTGCTCAACGACCAAGCGGTGGTCACCACCAAGGCCGCCGCCGACGAGCTGCCCGAGCG 1860
CTTCGCCGAGGAATTCGCCGCCGCGCTCGCGCAGCACCGGTGCTGGCAGCGGCGCACCGA 1920
CGCGGTGCCGGCCTGAAAGCCGGCCGAAGACCGCCGAAAGGTTTTCCCGGCCGCCGGGAC 1980
GGGCATCCGGCTCAGAAGGCGTCATCGTGGACAGGAGGACAAGTCATGCCGCGCTCGTCG 2040
ATC-phoA                                                        2043
```

Figure 3: Location of *katE* homolog within the pJEM11-*M. ptb*281 insert. Brown text = *katE* sequence; **TGA** = *katE* stop codon; **GATC** = *Bam* H1/*Sau* 3A vector insert ligation sites.

The *M. ptb katE* homolog covered the nucleotide sequence of pJEM11-*M. ptb*281 from the *Bam* H1/*Sau* 3A (GATC) insert vector ligation site at residue 1 to the TGA *katE* stop codon located 108 bp upstream of *phoA* ending at nucleotide 1936.

The location of *katE*'s stop codon and the knowledge that the gene possessed a single open reading frame (ORF) (53) enabled the search for the ORF responsible for the PhoA$^+$ phenotype (hypothetical *M. ptb*-281 ORF) to be confined to the 107 nucleotides between *phoA* and the 3' region of the *M. ptb katE* homolog. Using this sequence the NCBI BLASTN search was repeated which unexpectedly reported an alignment to *M. avium katE* (Figure 4; A).

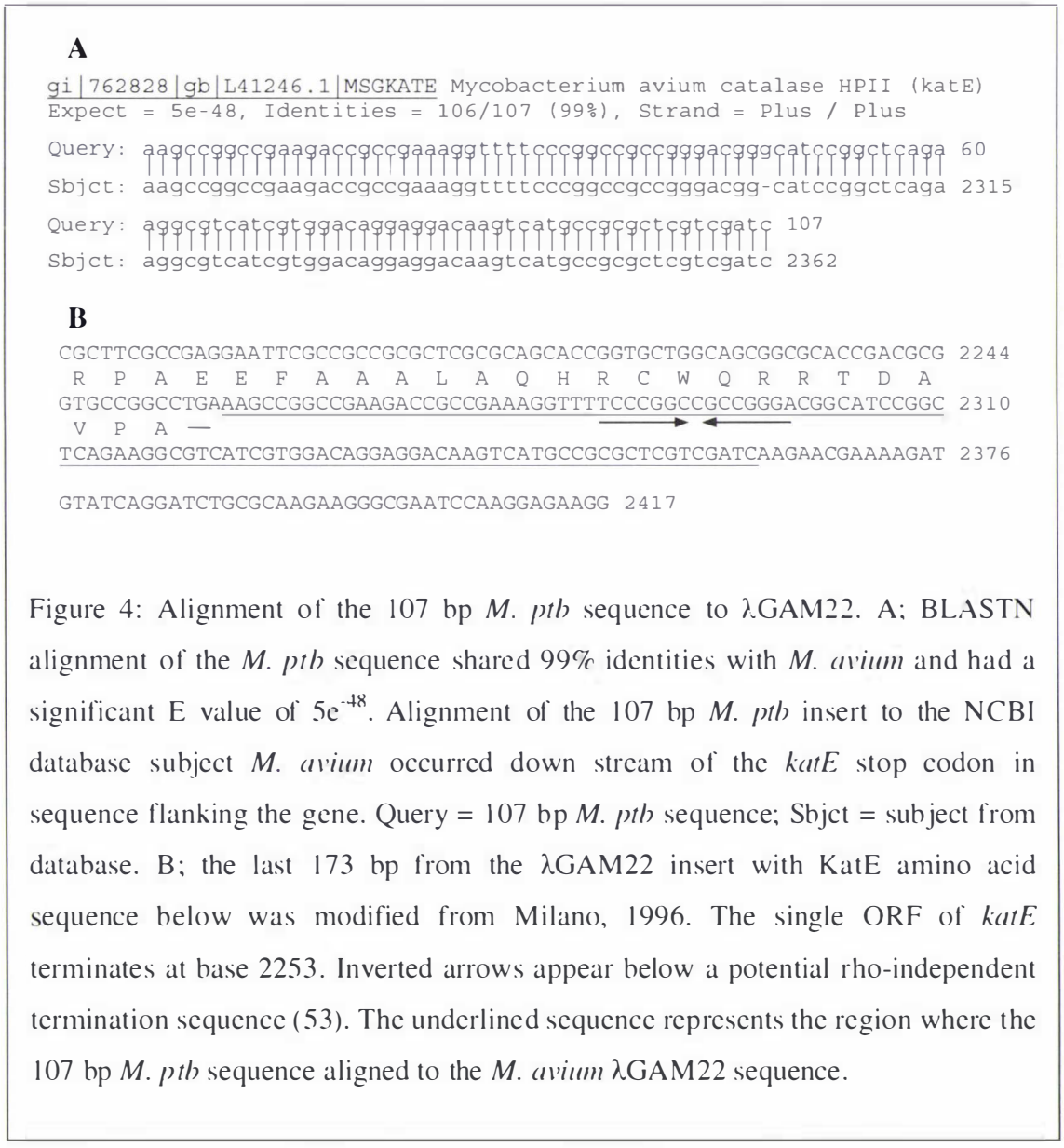


Figure 4: Alignment of the 107 bp *M. ptb* sequence to λGAM22. A; BLASTN alignment of the *M. ptb* sequence shared 99% identities with *M. avium* and had a significant E value of $5e^{-48}$. Alignment of the 107 bp *M. ptb* insert to the NCBI database subject *M. avium* occurred down stream of the *katE* stop codon in sequence flanking the gene. Query = 107 bp *M. ptb* sequence; Sbjct = subject from database. B; the last 173 bp from the λGAM22 insert with KatE amino acid sequence below was modified from Milano, 1996. The single ORF of *katE* terminates at base 2253. Inverted arrows appear below a potential rho-independent termination sequence (53). The underlined sequence represents the region where the 107 bp *M. ptb* sequence aligned to the *M. avium* λGAM22 sequence.

The alignment had an expected (E) value of $5e^{-48}$ with 99% identity and occurred outside the *katE* gene downstream of its C-terminus (Figure 4; B). Alignment of the 107 bp query sequence to *katE* occurred because the original sequence submitted to the GenBank Library, from a λGAM-11 *M. avium* library clone λGAM22 (accession number L41246), carried an insert which included the *katE* gene and its flanking sequences (53).

The BLASTN alignment of the 107 bp *M. ptb* sequence had an extra base at position 47 between the C-terminus of the *M. ptb katE* homolog and *phoA* that was not present in the equivalent *M. avium* sequence from the NCBI database λGAM22 clone (Figure 4). The extra base may be a legitimate difference between these two species or be a sequencing error in either sequence. Of major importance was the absence of any significant alignment between the 107 bp *M. ptb* sequence to members of the MTB complex or related bacteria.

3.4 Preliminary Southern Blot Analysis

To investigate the apparent absence of the hypothetical *M. ptb*281 ORF from member(s) of the MTB complex, the biomolecular based method of Southern blot analysis was used.

3.4.1 Preparation of Probes

phoA-281

The probe *phoA*-281-DIG was designed to screen a Southern blot of *M. ptb* (Neoparasec) and *M. bovis* BCG (BCG) vaccine strains and a *M. ptb* lambda library to isolate the C-terminal encoding region of the *M. ptb*281 ORF. The absence of AP from the genome of mycobacteria (33,34,54,55) made it possible to utilize the pJEM11-*M. ptb*281 vector as a template in a polymerase chain reaction (PCR) with the existing pJEM11 primer, JEM2 and the forward primer MPTB281F1. Using JEM2 as a reverse primer resulted in approximately 100 bp of *phoA* being included in the final *phoA*-281-DIG probe.

With the aid of computer programs Amplify (56) and Mac vector (MacVector, Accelrys Inc), the forward primer MPTB281F1 was designed to anneal to a region within the pJEM11-*M. ptb*281 insert, bases 1448 to 1507 that had a 100% alignment to

M. avium katE on the BLASTN results adding the final 439 bp from the 3' region of *katE* to *phoA*-281-DIG. The absence of *katE* from the genome of *M. tuberculosis* and *M. bovis* (53) suggested that the gene would also be absent from BCG, a derivative of *M. bovis* and the source of genomic material used in the Southern blots. It was therefore anticipated that the 439 bp of *katE* in the probe would not hybridize to BCG causing false positives. Theoretically, MPTB281F1 will not form primer diamers with itself or JEM2 and will anneal specifically to a single site within the pJEM11-*M. ptb*281 insert during a PCR to produce the DNA fragment *pho*A-281 for DIG labelling.

Using the standard PCR protocol with a 65° C annealing temperature, JEM2 and MPTB281F1 primers were used with ~10 ng of pJEM11-*M. ptb*281 template to amplify the ~660 bp *phoA*-281 DNA fragment for DIG labelling. The PCR product was analysed by agarose gel electrophoresis as described in the methods section. Analysis of *phoA*-281 PCR mixture by agarose gel electrophoresis showed the PCR successfully amplified a single band that migrated to the expected distance of ~660 bp as judged by the 100 bp ladder (Figure 5). The remaining PCR mixture was used to gel purify the ~660 bp band as described in section 2.7 in preparation for DIG labelling.

MPB70

As a positive control for BCG on the Southern blot a fragment of the *M. tb* (MTB) complex specific gene, designated MPB70 (mycobacterial protein secreted from BCG) was amplified by PCR for DIG labeling (MPB70-DIG) (57). The primers MPB71 and MPB72 (57) were used to amplify a 396 bp fragment of the MPB70 gene from BCG using standard PCR conditions with a 60° C annealing temperature in the absence of DMSO. Analysis of the MPB70 PCR mixture by agarose gel electrophoresis (Figure 5) showed a single band of approximately 396 bp was successfully amplified from BCG. The remaining PCR mixture was subjected to agarose gel electrophoresis and the ~396 bp MPB70 band excised and purified in preparation for DIG labelling.

DIG labelling of purified *phoA*-281 and MPB70 PCR fragments was performed (section 2.14) and the yield of DIG-labelled DNA estimated using colorimetric detection (section 2.14.1).

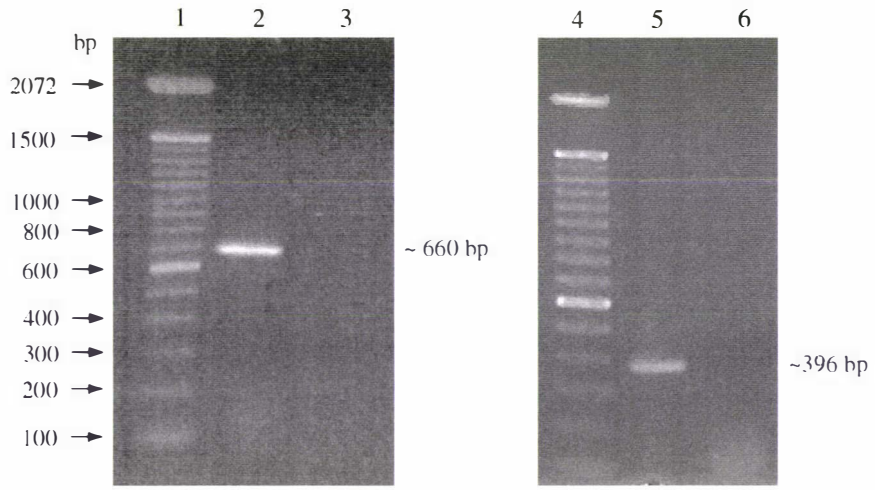


Figure 5: Agarose gel analysis of *phoA*-281 and MPB70 PCR products. Aliquots (4 µL) from *phoA*-281 (left) and MPB70 (right) PCRs were electrophoresed on 1% agarose gels in 1x TAE buffer. The DNA was stained with ethidium bromide and examined under UV light. Both PCRs produced single bands of the size expected (*phoA*-281 = ~660 bp; MPB70 = 396 bp). Lanes: 1 & 4, 100 bp ladder; 2; ~660 bp *phoA*-281 PCR product; 3 & 6, negative controls; 5, ~396 bp MPB70 PCR product.

3.4.2 Neoparasec and BCG *Eco* R1 Southern Blot Set 1

The vaccine strains *M. ptb* 316F (Neoparasec) and *M. bovis* BCG (BCG), derivatives of *M. ptb* and *M. bovis* respectively, were used as the source of *Eco* R1 digested genomic material for Southern blot analysis. The vaccine strains were chosen because they are relatively safe, and faster growing than the parental strains. BCG was grown in 500 mL of 7H9 Middlebrook medium supplemented with albumin-dextrose-catalase (ADC) with consistent shaking at 37° C. DNA was extracted as described by (43). Neoparasec genomic DNA used only for the preliminary Southern blots was kindly provided by Dr Jeremy Rae (Massey University). Aliquots of Neoparasec and BCG genomic DNA were subjected to restriction endonuclease digest with *Eco* R1. Aliquots of native DNA and *Eco* R1 digested samples were loaded onto a 0.7% agarose gel and analysed by electrophoresis to confirm the digests were successful (Figure 6).

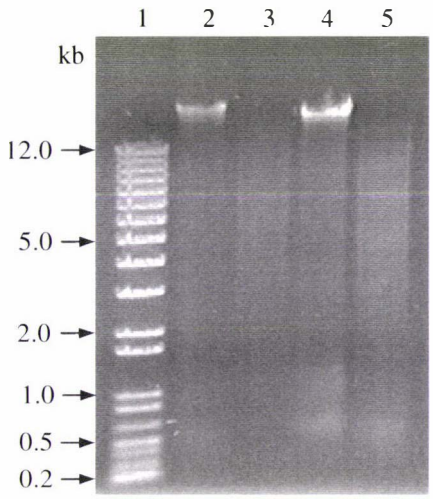


Figure 6: Analysis of *Eco* R1 digested Neoparasec and BCG. Undigested and digested samples were loaded onto a 0.7% agarose gel and subjected to electrophoresis in 1x TAE buffer. Lanes: 1, 1 kb^+ ladder; 2, 6.7 µL undigested Neoparasec genomic DNA; 3, 14.2 µL *Eco* R1 digested Neoparasec DNA; 4, 3 µL undigested BCG genomic DNA; 5, 4 µL *Eco* R1 digested BCG genomic DNA.

Samples of undigested genomic DNA from Neoparasec and BCG were loaded onto lanes 2 and 4 respectively; each contained a single broad band of genomic DNA present above the 12 kb marker. Samples containing approximately equivalent amounts of DNA, relative to the undigested samples, were loaded onto lanes 3 Neoparasec, and 5 BCG, where the absence of the >12 kb band confirmed the digests were complete. Aliquots containing approximately 1 µg of DNA from *Eco* R1 Neoparasec and BCG digests were loaded onto a fresh 0.7% agarose gel and electrophoresed. A reference photo was taken (Figure 7) before the gel was subjected to Southern transfer as described in section 2.13.2. For simplicity, only strips 1 and 2 are shown.

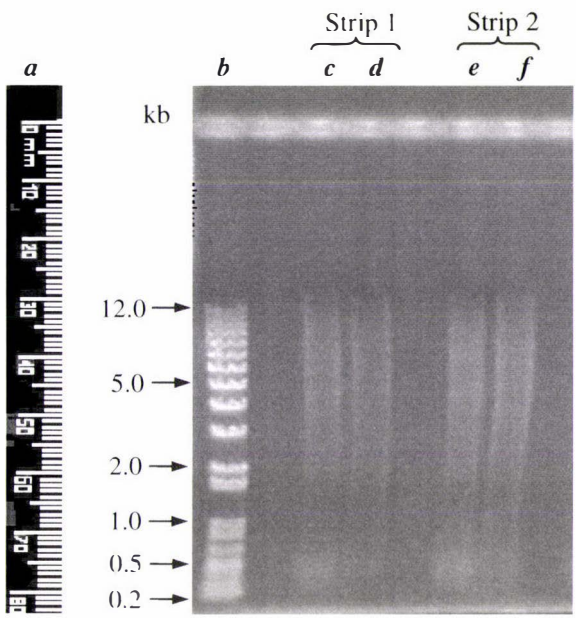


Figure 7: Reference photo of Neoparasec and BCG *Eco* R1 Southern blot set 1, strips 1 & 2. (Strips 3-6 not shown). Aliquots containing approximately 1 µg of *Eco* R1 digested DNA were electrophoresed on a 0.7% agarose gel in 1x TAE buffer. The gel was soaked in dilute ethidium bromide and a reference photo was taken with the gel aligned next to a fluorescent ruler. *a*, UV fluorescent ruler. Lanes: *b*, 1 kb$^+$ ladder; *c* & *e*, *Eco* R1 digested BCG DNA; *d* & *f*, *Eco* R1 digested Neoparasec DNA. Lane *c* & *d*, strip 1; *e* & *f*, strip 2.

Low Stringency Hybridization

The Neoparasec and BCG *Eco* R1 Southern blot set 1, strip 2 was incubated with *phoA*-281-DIG as described in section 2.15 using a 40° C hybridization temperature (low stringency hybridization). Low stringency post-hybridization washes were performed (section 2.15.1), and the blot was prepared for chemiluminescent detection using CSPD (section 2.15.2). The blot was exposed to X-ray film for 1 and 2 hours to get a range of exposures (Figure 8). Under low stringency conditions the *phoA*-281-DIG probe produced a single strong defined band in Neoparasec on the 1 and 2 hour exposures. The 2 hour exposure showed the emergence of background hybridization signals in the Neoparasec lane. Hybridization of *phoA*-281-DIG to BCG DNA produced a smear of hybridization signals in the 1 and 2 hour exposures. The latter exposure also produced two faint but poorly defined bands in fragments approximately 3.4 kb and 1.8 kb that most likely represent areas of low homology to some part of the probe.

The absence of high background signals from Neoparasec and their presence in BCG after 1 and 2 hours of exposure was typical of strips from set 1 under low stringency conditions.

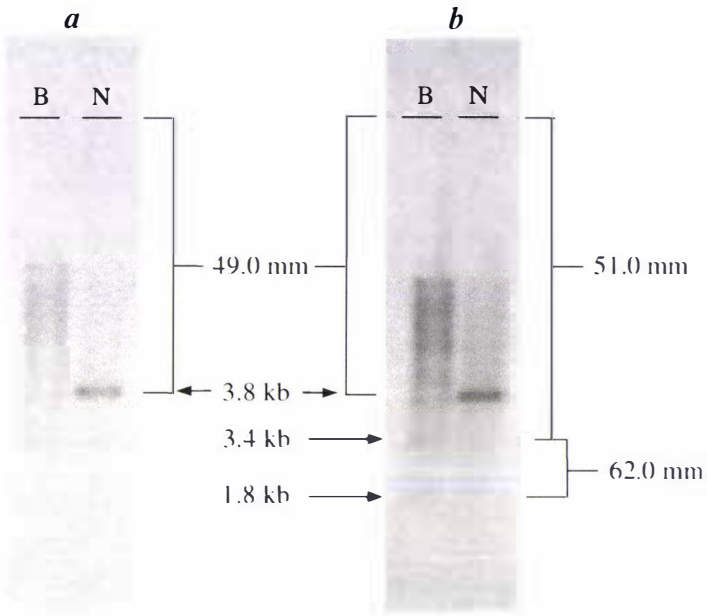


Figure 8: Hybridization of *phoA*-281-DIG using low stringency conditions. Neoparasec and BCG *Eco* R1 Southern blot set 1, strip 2 was probed with *phoA*-281-DIG using low stringency conditions and subjected to chemiluminescent detection with CSPD for 1 and 2 hours. Traveling distance of the bands are shown with the approximate size of the genomic fragments they were located on as judged by a standard curve made from the reference photo. *a*, 1 hr exposure; *b*, 2 hr exposure. B = BCG lane, N = Neoparasec lane.

Marks indicating the position of the wells on the Southern blot were transferred onto the film and the distance between the band and the well in the Neoparasec lane was measured to be ~49 mm. A standard curve of the Neoparasec and BCG *Eco* R1 Southern blot set 1 was made using the 1 kb^+ ladder from the reference photo (Figure 9).

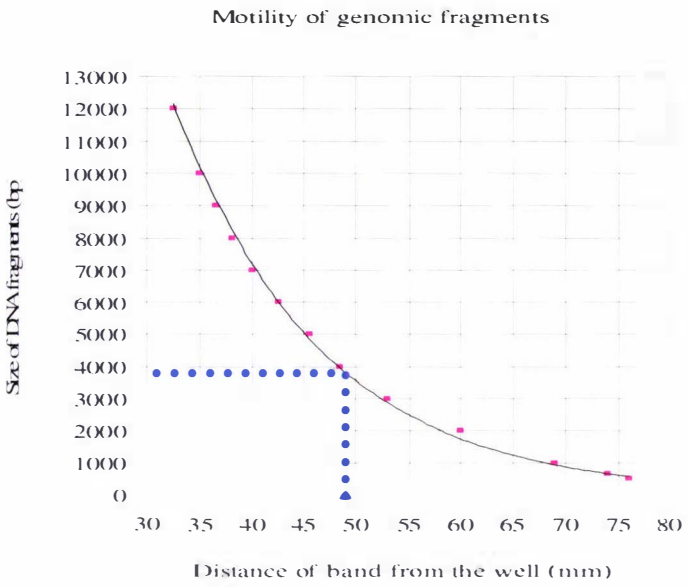


Figure 9: Neoparasec and BCG *Eco* R1 Southern blot, set 1 standard curve. The standard curve was generated from the motility of 1 kb^+ ladder DNA fragments on the reference photo (Figure 7). The band found in Neoparasec 49 mm from the well (shown in blue) corresponds to genomic fragments approximately 3800 bp in length.

Tracing the 49 mm mark onto the standard curve allowed an estimation to be made on the size of the band found in Neoparasec. The band falls within genomic fragments approximately 3800 bp in length.

3.4.3 Neoparasec, BCG *Eco* R1 Southern Blot Set 2

Background signals present in the high molecular weight fragments in BCG on set 1 may have been obscuring hybridization signals of low but significant homology. To reduce the background signals in the high molecular weight fragments so that low homology hybridization signals could be seen, a number of changes were made to the 0.7% agarose gel used to make Neoparasec and BCG *Eco* R1 Southern blot set 2. Firstly, the 0.7% agarose gel was made using a comb that produces 10 mm instead of 5 mm long wells. This was done to allow the high molecular weight fragments to spread out more across the lane to reduce the likelihood of the probe becoming entangled in a dense mat of genomic material. Secondly, the gel was electrophoresed further than in set 1 to better separate the high molecular weight fragments thus improving the

resolution of bands in the high molecular weight fragments. This resulted in the absence of molecular weight fragments from the Southern blot below ~3 kb. The absence of those fragments was not considered important because the focus of this Southern blot was on the high molecular weight fragments where most of the signals were.

A second *Eco* R1 digest of Neoparasec and BCG was prepared and subsequently used to construct the Neoparasec and BCG *Eco* R1 Southern blot set 2. Two 0.7% agarose gels were prepared for Southern transfer each containing multiple samples of *Eco* R1 digested DNA from Neoparasec and BCG. After Southern transfer the membranes were cut into strips containing one lane from each species and numbered in order of their appearance in the reference photo (Figure 10).

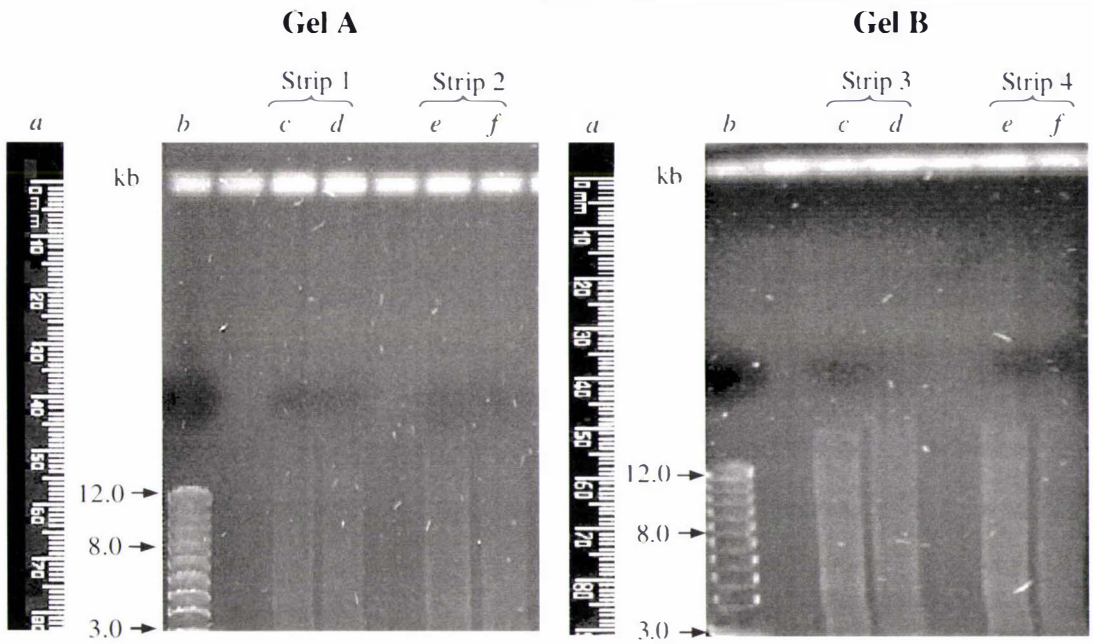


Figure 10: Reference photos of Neoparasec and BCG *Eco* R1 Southern blots, set 2. Aliquots containing approximately 1 µg of *Eco* R1 digested DNA were electrophoresed on two 0.7% agarose gels in 1x TAE buffer. The gels were soaked in dilute ethidium bromide and reference photos were taken with each gel aligned next to a fluorescent ruler. Gel A, strips 1 & 2; Gel B, strips 3 & 4. Lanes: *a*, UV fluorescent ruler; *b*, 1 kb^+ ladder; *c* & *e*, *Eco* R1 digested BCG DNA; *d* & *f*, *Eco* R1 digested Neoparasec DNA.

High Stringency Hybridization

Neoparasec and BCG *Eco* R1 Southern blot set 2, strip 2 was incubated with *phoA*-281-DIG using a 42° C hybridization temperature (high stringency conditions) followed by high stringency post hybridization washes. The blot was prepared for chemiluminescent detection with CDP^{Star} and exposed to film for 1 and 2 hours to get a range of exposures (Figure 11).

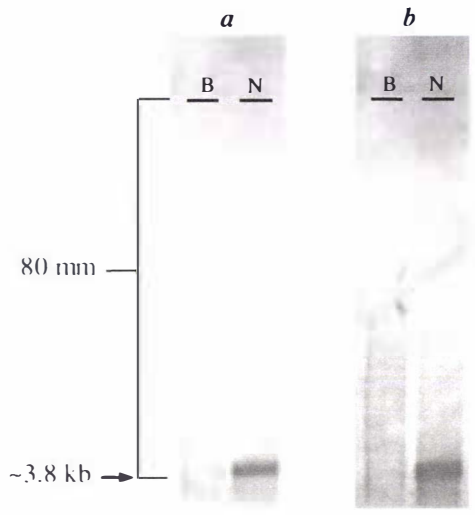


Figure 11: Hybridization of *phoA*-281-DIG using high stringency conditions. Neoparasec and BCG *Eco* R1 Southern blot set 2, strip 2 was probed with *phoA*-281-DIG using high stringency conditions and subjected to chemiluminescent detection with CDP^{Star} and exposed to film for 1 and 2 hours. *a* = 1 hr exposure; *b* = 2 hr exposure; mm = traveling distance of the band from the well; kb = kilobase; arrow indicates approximate size of genomic fragments the band was located on as judged by the standard curve of set 2, strip 2 (data not shown); B = BCG lane, N = Neoparasec lane.

Both 1 and 2 hour exposures revealed the presence of a single strong hybridization signal in Neoparasec while no specific hybridization signals were observed in BCG. The hybridization signal in Neoparasec was located among fragments ~3.8 kb as judged by the standard curve generated from the 1 kb^{+} ladder in the reference photo (data not shown). Background noise emerged in BCG as a faint smear in the high molecular weight fragments after 2 hours exposure (Figure 11) whereas it had been

prominent in both the 1 and 2 hour exposures of the first Southern blot subjected to low stringency conditions (Figure 8).

3.4.4 Hybridization of the BCG Positive Control Probe MPB-DIG

Neoparasec and BCG *Eco* R1 Southern blot set 2 strip 4 was incubated with the MTB complex specific probe MPB70-DIG using a 42° C hybridization temperature and high stringency post hybridization washes. The blot was prepared for chemiluminescent detection with CDP^{Star} and exposed to film for 1 and 2 hours. As expected the probe produced no hybridization signals in Neoparasec after 1 or 2 hour exposures while it produced a single strong hybridization signal in both exposures of BCG (Figure 12).

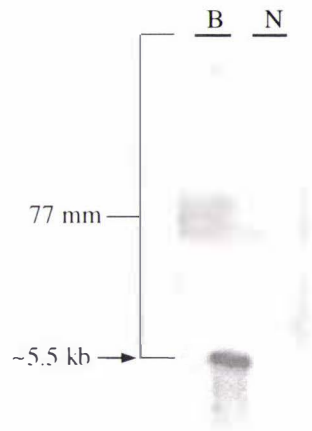


Figure 12: Hybridization of MPB70-DIG using high stringency conditions. Neoparasec and BCG *Eco* R1 Southern blot set 2, strip 4 was hybridized with MPB70-DIG using high stringency conditions and subjected to chemiluminescent detection with CDP^{Star} and exposed for 2 hours. Shown are the travelling distance of the band and the approximate size of the genomic fragments the band was located on as judged by the standard curve doe set 2 (data not shown). B = BCG lane, N = Neoparasec lane.

A standard curve of set 2 strip 4 showed the MPB70-DIG probe annealed to BCG genomic fragments of approximately ~5500 bp (data not shown). The presence of a hybridization signal in BCG showed the DNA used in set 2 was not degraded and

capable of producing a specific hybridization signal. Therefore, the lack of specific hybridization signals from BCG on set 2 strip 2 when hybridized with *pho*A-281-DIG was due to the absence of this sequence in BCG.

3.4.5 Conclusion to Preliminary Analysis of pJEM11-*M. ptb*281

The pJEM11-*M. ptb*281 insert contained the last 1936 bp of an *M. ptb* homolog to the *M. avium* catalase gene, *katE*. Minus its 5' region and present in the wrong reading frame with its stop codon upstream of *phoA*, the *M. ptb katE* homolog was considered unlikely to be the gene responsible for the $PhoA^+$ phenotype. Within the 107 bp sequence between *katE* and *phoA* lay the 5' region of a possible ORF (*M. ptb*281 ORF). The 107 bp sequence was absent from the annotated genome of *M. tuberculosis* indicating the hypothetical *M. ptb*281 ORF may be absent from *M. tuberculosis*. It remains possible that the $PhoA^+$ phenotype of pJEM11-*M. ptb*281 may be a cloning or expression artefact and will require direct evidence such as those from primer extension analysis before the identity of the gene responsible can be assigned.

Strong single hybridization signals were observed in the *M. ptb* derivative Neoparasec and *M. bovis* derivative BCG when Southern blots were probed under high stringency conditions with *phoA*-281-DIG and MPB70-DIG respectively, confirming the integrity of the DNA used to produce these blots. Probing the first Southern blot under low stringency conditions with *phoA*-281-DIG produced a smear of background signals among the high molecular weight fragments of BCG and two faint indistinct bands in fragments below 400 bp. The second Southern blot, focusing on high molecular weight fragments where the bulk of background signal occurred, was subjected to high stringency conditions with *phoA*-281-DIG. This resulted in the reduction of background signals in BCG revealing multiple faint indistinct bands in fragment ranging from ~12 to ~3 kb. While the background signals present in BCG probed with *phoA*-281-DIG most likely represent areas of low homology to some part of the probe, these results do not conclusively prove or disprove the presence of a homolog to the hypothetical *M. ptb*281 ORF in the genome of BCG.

Part B: Identification of the Full Length Hypothetical *M. ptb*281 ORF

3.5 Introduction

Expression and secretion of the PhoA-fusion protein from pJEM11-*M. ptb*281 relies on the fusion of the N-terminal encoding region from the hypothetical secreted protein MPTB281 being fused in frame with and adjacent to *phoA*. Before the immunological properties of the hypothetical MPTB281 protein could be explored, sequence encoding the amino and carboxyl termini was required for the full length ORF to be cloned and expressed.

3.6 MPTB281 C-terminal Search

The insert from pJEM11-*M. ptb*281 contained the putative N-terminal but not the C-terminal encoding region of MPTB281. To identify the C-terminal encoding region of MPTB281 the strategies outlined in the following sections were employed.

3.6.1 *M. ptb* Lambda Library

A *M. ptb* lambda library (58) was originally used to find the 3' region of the hypothetical *M. ptb*281 ORF however, attempts to screen the library using colony plaque hybridization with the *phoA*-281-DIG probe were unsuccessful (data not shown). The lambda library had previously been shown to be incomplete by the absence of the *katE* gene (Christine Dupont, pers. comm.), the gene located upstream of the hypothetical *M. ptb*281 ORF.

3.6.2 TIGR Database

The second strategy utilised sequence data from The Institute for Genomic Research (TIGR) database and the Sanger database which contained the partially completed mycobacterial genome *M. ptb* and its close relative *M. avium*. The *M. ptb* and *M. avium* genome sequencing projects were incomplete at the time of this work and being un-annotated were unable to provide information on hypothetical ORFs. BLAST search results from the *M. ptb* database were limited to an alignment between the query sequence and subjects found in the database, while the *M. avium* database provided in addition to an alignment, the contig to which the most significant alignment belonged.

A contig provided approximately 1000 bp of sequence to each side of the query thus allowing sequence flanking the query to be identified.

The close genealogy of *M. avium* and *M. ptb* meant it was probable that *M. avium* would contain a homolog to the hypothetical *M. ptb*281 ORF responsible for the *phoA* phenotype. Hence, screening the *M. avium* database with sequence containing the N-terminal encoding region of *M. ptb*281 was likely to result in the identification of a *M. avium* homolog and the contig to which it belonged. The *M. avium* contig would contain sequence flanking the *M. avium*281 homolog, including its 3' region, which could then be used to search the *M. ptb* database. This would allow the authentic *M. ptb* C-terminal sequence to be identified in the query/subject alignment. Determination the C-terminus would be achieved by identifying the first downstream stop codon in the *phoA*/*M. ptb*281 reading frame.

M. avium Database Search

The last 119 bp from the pJEM11-*M. ptb*281 insert sequence, containing the 5' region of the hypothetical *M. ptb*281 and the 3' region of *katE*, was used in a BLASTN search of the *M. avium* database (http://tigrblast.tigr.org/ufmg/index.cgi?database=m_avium|seq). The most significant *M. avium* database result was found in contig 14 that contained 20,611 bp. The *M. ptb* query/*M. avium* subject alignment shared 99% identities and positives with an expected E value of $1.1e^{-21}$ (Figure 13). The *M. avium* database also provided approximately 1000 bp of sequence flanking both sides of the database subject from contig 14 (Appendix A).

```
Contig 14, Length = 20,611, Minus Strand:
Expect = 1.1e-21, Identities & Positives = 118/119 (99%), Strand = Minus/Plus

Query:  119 GTGCCGGCCTGAAAGCCGGCCGAAGACCGCCGAAAGGTTTTCCCGGCCGCCGGGACGGGC 60
            ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjet:18098 GTGCCGGCCTGAAAGCCGGCCGAAGACCGCCAAAAGGTTTTCCCGGCCGCCGGGACGGGC 18157

Query:   59 ATCCGGCTCAGAAGGCGTCATCGTGGACAGGAGGACAAGTCATGCCGCGCTCGTCGATC 1
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjet:18158 ATCCGGCTCAGAAGGCGTCATCGTGGACAGGAGGACAAGTCATGCCGCGCTCGTCGATC 18216
```

Figure 13: BLASTN result from the *M. avium* database. Query = 119 bp pJEM11-*M. ptb*281 query sequence; Sbjet (subject) = identities from contig 14 from the *M. avium* database. Brown text = 3' region of *M. ptb katE*; **TGA** = *M. ptb katE* stop codon

M. ptb Database Search

The final 1208 nucleotides from contig 14, including the final 101 bp from the C-terminus of *katE* as a reference point, were used in a BLASTN search of the *M. ptb* database. The most significant alignment was to a subject that shared 99% identities, had no gaps and an expected value of 0.0 in a plus / minus orientation to the first 960 bp of the query sequence (Figure 14).

```
Expect = 0.0, Identities = 954/960 (99%), Strand = Plus / Minus
Query: 1     ccgccgccgacgagctgcccgagcgcttcgccgaggaattcgccgccgcgctcgcgcagc   60
Sbjct: 5484  ccgccgccgacgagctgcccgagcgcttcgccgaggaattcgccgccgcgctcgcgcagc 5544
Query: 61    accggtgctggcagcggcgcaccgacgcggtgccggcctgaaagccggccgaagaccgcc  120
Sbjct: 5545  accggtgctggcagcggcgcaccgacgcggtgccggcctgaaagccggccgaagaccgcc 5604
Query: 121   aaaaggttttcccggccgccgggacgggcatccggctcagaaggcgtcatcgtggacagg  180
Sbjct: 5605  gaaaggttttcccggccgccgggacgggcatccggctcagaaggcgtcatcgtggacagg 5665
Query: 181   aggacaagtcatgccgcgctcgtcgatcaagaacgaaaagatgtatcaggatctgcgcaa  240
Sbjct: 5666  aggacaagtcatgccgcgctcgtcgatcaagaacgaaaagatgtatcaggatctgcgcaa 5725
Query: 241   gaagggcgaatccaaggagaaggccgcgcgcatctccaatgcggccgccggccaaggcaa  300
Sbjct: 5726  gaagggcgaatccaaggagaaggccgcgcgcatctcgaatgcggctgccggccaaggcaa 5785
Query: 301   gtcgtcggtgggccgccgcggcggcaagtccgggtcctatcaggactggaccgtgccgga  360
Sbjct: 5786  gtcgtcggtgggccgccgcggcggcaagtccgggtcctatcaggactggaccgtgccgga 5845
Query: 361   attgaagaagcgggccaaagagcttggcatttccggctattcgggcctgaccaaggacaa  420
Sbjct: 5846  attgaagaagcgggccaaagagcttggcatttccggctattcgggcctgaccaaggacaa 5905
Query: 421   gctggtcgccaaactgcgcaaccactgatccgtcatctcgtcaaccgcagtcgttcggcc  480
Sbjct: 5906  gctggtcgccaaactgcgcaaccactgatccgtcatctcgtcaaccgcagtcgttcggcc 5965
Query: 481   agcaccaggaaggcgaccgcgtagtcgttgtcggcggtctgagcccggatttcgtcccag  540
Sbjct: 5966  agcaccaggaaggcgaccgcgtagtcgttgtcggcggtctgagcccggatttcgtcccag 6025
Query: 541   tccagccgttcgcgcacggcgcgcaccgcgggcagcagcttggcgaagtcgcagtggtgc  600
Sbjct: 6026  tccagccgttcgcgcacggcgcgcaccgcgggcagcagcttggcgaagtcgcagtggtgc 6085
Query: 601   tcgcccagcgagcgcagctgctggacgagcaccatggtgggtggcagcaccggcatgccg  660
Sbjct: 6086  tcgcccagcgagcgcagctgctggacgagcaccatggtgggtggcagcaccggcatgccg 6145
Query: 661   atcgccagcacgtcgtggtgctcggcgcggtccagggtctgcgcatcgacgggcacgccg  720
Sbjct: 6146  atcgccagcacgtcgtggtgctcggcgcggtccagggtctgcgcatcgacgggcacgccg 6205
Query: 721   ttgaggcggtgcagcacatcgacccgcatgtcgccggtgcgggccttgaagagccagtcc  780
Sbjct: 6206  ttgaggcggtgcagcacatcgaccagcatgtcgccggtgcgggccttgaagagccagtcc 6265
```

Figure 14: The most significant *M. ptb* database result. Shown are the first 780 bp of a 960 bp alignment between *M. avium* contig 14 and the *M. ptb* database. The alignment had an expected E value of 0.0 and shared 99% identities. Query = *M. avium* contig 14 sequence; Sbjct = *M. ptb* database identity. Brown text = 3' region of *M. ptb katE*; **TGA** = *M. ptb katE* stop codon

Search for the first downstream Stop Codon

Using the ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) translation tool (http://us.expasy.org/tools/dna.html) the *phoA/M. ptb*281 reading frame was identified (Figure 15; A) and applied to the *M. ptb* database sequence.

A

```
ctgaaagccggccgaagaccgccgaaaggttttcccggccgccgggacgggcatccggct 60
 L  K  A  G  R  R  P  P  K  G  F  P  G  R  R  D  G  H  P  A
cagaaggcgtcatcgtggacaggaggacaagtcatgccgcgctcgtcgatccggatacgt 120
 Q  K  A  S  S  W  T  G  G  Q  V  M  P  R  S  S  I  R  I  R
acggtacctgactcttatacacaagtagcgtcctggacggaacctttcccgttttgccct 180
T  V  P  D  S  Y  T  S  Y  A  S  W  T  E
```

B

```
ctgaaagccggccgaagaccgccgaaaggttttcccggccgccgggacgggcatccggct 5582
 L  K  A  G  R  R  P  P  K  G  F  P  G  R  R  D  G  H  P  A
cagaaggcgtcatcgtggacaggaggacaagtcatgccgcgctcgtcgatcaagaacgaa 5642
 Q  K  A  S  S  W  T  G  G  Q  V  M  P  R  S  S  I  K  N  E
aagatgtatcaggatctgcgcaagaagggcgaatccaaggagaaggccgcgcgcatctcg 5702
 K  M  Y  Q  D  L  R  K  K  G  E  S  K  E  K  A  A  R  I  S
aatgcggctgccggccaaggcaagtcgtcggtgggccgccgcggcggcaagtccgggtcc 5762
 N  A  A  A  G  Q  G  K  S  S  V  G  R  R  G  G  K  S  G  S
tatcaggactggaccgtgccggaattgaagaagcgggccaaagagcttggcatttccggc 5822
 Y  Q  D  W  T  V  P  E  L  K  K  R  A  K  E  L  G  I  S  G
tattcgggcctgaccaaggacaagctggtcgccaaactgcgcaaccactgatccgtcatc 5882
 Y  S  G  L  T  K  D  K  L  V  A  K  L  R  N  H  -  S  V  I
tcgtcaaccgcagtc                                              5956
 S  S  T  A  V
```

Figure 15: The *M. ptb*281 stop codon. A, 180 bp from pJEM11-*M. ptb*281 including *phoA* sequence with translated residues below in the *phoA/M. ptb*281 reading frame. B, *M. ptb* database sequence 5582-5956 with translated residues below in the *M. ptb*281 reading frame. Blue uppercase text = PhoA amino acid residues; bold lower case text = vector sequence; underlined lowercase text = remnant of the vector's multiple cloning site; ***gatc*** = *Bam* H1/*Sau* 3A ligation site; **tga** = *katE* stop codon; **tga** = first downstream stop codon in the *M. ptb*281 reading frame.

The sequence was inspected in the correct reading frame for the stop codons TAG, TGA or TAA. The first in frame stop codon was found 344 bp downstream from the 3' region of *katE* (Figure 15; B). The *M. ptb* and *M. avium* genome sequencing projects were unfinished at the time of this work and could contain errors. Therefore, a section of the *M. ptb* database alignment to *M. avium* contig 14 between the 3' region of

katE and the first down stream stop codon was inspected for gaps and base differences that could potentially conceal the true stop codon (Figure 16).

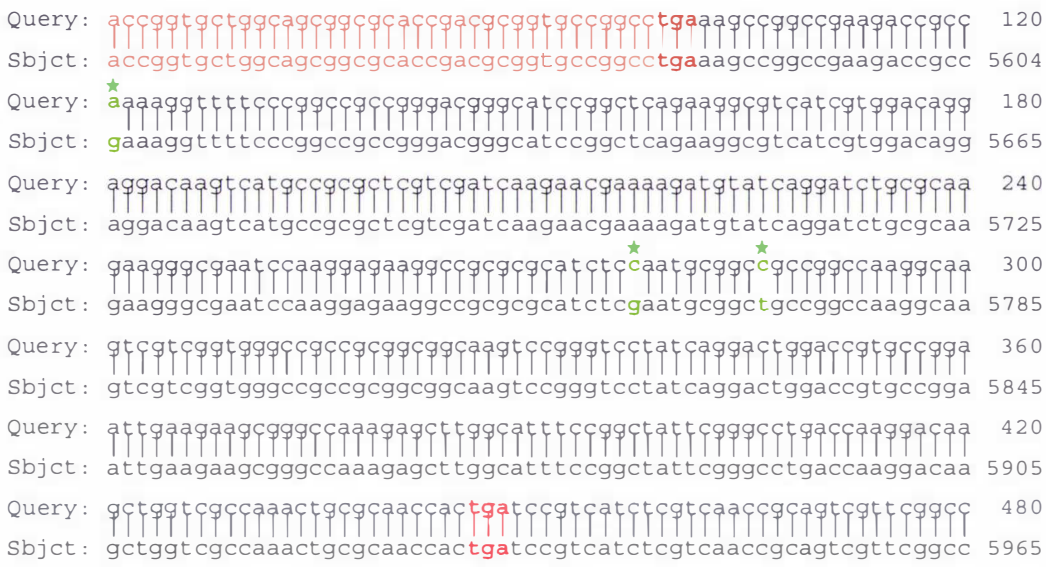


Figure 16: Inspection of *M. ptb* and *M. avium* contig 14 alignment. Sequence between the *katE* stop codon (**tga**) and first downstream stop codon (**tga**) in the *phoA* reading frame were inspected for gap and base differences. Three base differences at nucleotides 5605, 5762 and 5771 were found (indicated in green text below ★). Query = *M. avium* contig 14 sequence; Sbjct = *M. ptb* database identity.

No gap differences were found in these sequences while three base differences were found between the 3' region of the *M. ptb katE* homolog and the hypothetical *M. ptb*281 stop codon at residues 5605, 5762 and 5771 (Figure 16). These bases were all found at the third position of their respective codons and did not involve potential stop codons. Due to codon redundancy these base differences in *M. avium* do not change the translated amino acid sequence from those of the translated *M. ptb* sequence (residues 5605 cca/ccg = proline, 5762 tcc/tcg = serine and 5771 gcc/gct = alanine).

The λGAMM22 sequence differs from the *M. avium* database sequence and pJEM11-*M. ptb*281 insert by the absence of a 'G' down stream of the potential rho-independent termination repeat.

3.7 MPTB281 N-terminus Search

A number of hierarchical elements required for expression and export of a protein can be recognised in and around the gene sequence. RNA polymerase promoter elements -30 and -10 and ribosome binding sites (RBS) can be identified in the DNA sequence upstream of an open reading frame. N-terminal signal peptides, usually required for export across the plasma membrane, can be recognised in the translated sequence of the gene by their characteristic motif. Together these features were used in an attempt to identify the start codon for the hypothetical *M. ptb*281 ORF.

3.7.1 Start of Translation

The database results from Part A identified the single ORF of the *M. avium* catalase gene *katE* in the insert of pJem11-*M. ptb*281. The 5' region of the gene was absent from the insert and it's 3' region was located upstream of, and in the wrong reading frame to the *phoA* gene. Thus, *katE* was considered unlikely to be the gene responsible for the $PhoA^+$ phenotype. The search for the 5' region of the ORF responsible the $PhoA^+$ phenotype focused on the nucleotide sequence between the 3' region of *katE* and the *phoA* gene.

The ExPASy translation tool was used to translate the pJEM11-*M. ptb*281 insert sequence into the *phoA* reading frame to identify potential prokaryote methionine (ATG) and valine (GTG) start of translation codons. Examination of the translated sequence revealed the presence of two potential start codons between *katE*'s stop codon and the *phoA* gene, namely a methionine and a valine residue (Figure 17). In addition, both potential start codons were located downstream of the rho-independent termination repeats reported to be associated with *katE* (53). The valine residue was encoded by a GTC and not the valine start codon GTG, which made it unlikely to be the start codon, leaving the ATG codon as the only other option.

The pJEM11-*M. ptb*281 insert, *M. avium* and *M. ptb* database sequences all contained an extra nucleotide down-stream of the rho-independent termination repeat that was not present in the *M. avium katE* λGAM22 sequence (Figure 4). This base was removed and the pJEM11-*M. ptb*281 sequence translated into the *phoA* reading frame to ensure potential N-terminal residues were not being obscured (data not shown). A

search of this sequence did not reveal any other potential methionine or valine residues between the C-terminus of *katE* and *phoA*, nor did it bring *katE* into the *phoA* reading frame.

3.7.2 Hypothetical Shine-Dalgarno Sequence

Shine-Dalgarno sites are A+G rich sequences approximately 7 bases long followed by an interval of 4-7 bases typically, before an ATG start codon (59). Inspection of the sequence revealed a hypothetical 7 base Shine-Dalgarno sequence (aggagga) 6 bp upstream from the hypothetical ATG start codon (Figure 17).

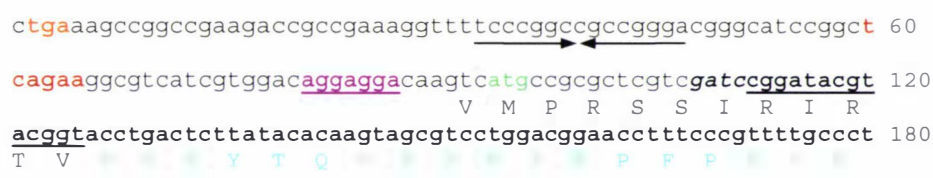


Figure 17: Potential start codon for MPTB281. Nucleotide sequence of pJEM11-*M. ptb*281 (reading strand) shown above with the relevant amino acids below in the *phoA* reading frame. Bold lowercase text = pJEM11 vector sequence; underlined lowercase text = remnant of the vector's multiple cloning site; ***gatc*** = *Bam* H1/*Sau* 3A ligation site; **tga** = *katE* stop codon; **atg** = hypothetical start codon; light blue uppercase text = PhoA amino acid sequence; inverted arrows appear below potential rho-independent termination repeat; **aggagga** = potential Shine-Dalgarno; **tcagaa** = potential promoter.

3.7.3 Hypothetical Promoter Element

The typical *E. coli* A+T rich hexamer promoter elements -10 (TATAAT) and -35 (TTGACA) common to most prokaryotes were absent from sequence up-stream of the hypothetical *M. ptb*281 start codon. However, a potential -10 hexamer promoter element lying between the 3' end of *katE* and the hypothetical ATG start codon was identified based on its high A+T content (Figure 17). No -35 promoter element could be identified based on the *E. coli* consensus or on reported mycobacterial consensus elements (59,60). With a G+C content ranging from approximately 57-65% (*M. leprae* and *M. tb*

respectively) relatively few mycobacterial promoter elements studied to date conform to the *E. coli* consensus, with some exceptions such as the highly conserved heat shock promoters (59). It is possible that there are other as yet unidentified elements enhancing transcription activity up-stream of the hypothetical *M. ptb*281 ATG start codon.

3.7.4 Signal Peptide

Generally, signal peptides range in size from 18 to 30 amino acids with a positively charged N-terminus (N-domain), a hydrophobic core of at least 7 residues (H-domain) followed by neutrally charged residues near the peptidase cleavage site (C-domain) (39,61-63). Inspection of the MPTB281-PhoA hybrid showed the *M. ptb* portion only contributed 6 amino acids to the protein upstream of PhoA; a length considered insufficient for a signal peptide. Starting with the proposed N-terminal methionine, the first 50 amino acids from the MPTB281-PhoA hybrid and the hypothetical MPTB281 protein were sent to the signal peptide prediction servers PSORT and SignalP. Using gram negative and gram positive parameters, neither server predicted a signal peptide for these sequences (data not shown).

3.8 The Hypothetical *M. ptb*281 ORF

The hypothetical *M. ptb*281 ORF (Figure 18) was assembled using the proposed 5' region from the pJEM11-*M. ptb*281 sequence and the proposed 3' region from the *M. ptb* database sequence.

```
atgccgcgctcgtcgatcaagaacgaaaagatgtatcaggatctgcgcaagaagggcgaa 60
 M  P  R  S  S  I  K  N  E  K  M  Y  Q  D  L  R  K  K  G  E
tccaaggagaaggccgcgcgcatctcgaatgcggctgccggccaaggcaagtcgtcggtg 120
 S  K  E  K  A  A  R  I  S  N  A  A  A  G  Q  G  K  S  S  V
ggccgccgcggcggcaagtccgggtcctatcaggactggaccgtgccggaattgaagaag 180
 G  R  R  G  G  K  S  G  S  Y  Q  D  W  T  V  P  E  L  K  K
cgggccaaagagcttggcatttccggctattcgggcctgaccaaggacaagctggtcgcc 240
 R  A  K  E  L  G  I  S  G  Y  S  G  L  T  K  D  K  L  V  A
aaactgcgcaaccactga 258
 K  L  R  N  H  -
```

Figure 18: The full hypothetical *M. ptb*281 ORF and translated sequence. Nucleotide (above) and translated (below) sequences were assembled from the pJem11-*M. ptb*281 sequence (bold text) and *M. ptb* database sequence.

The hypothetical gene comprised 258 nucleotides with a G+C content of 61.24% that translates into 85 amino acids and was predicted by Bionavigator (http://www.bionavigator.com) to have a molecular mass of 10.81 kDa.

The *M. pth* database sequence used to locate the 3' region of *M. pth*281, differed at positions 87 and 96 (Figure 15; residues 5762 and 5771) from its equivalent in the *M. avium* contig 14. These base differences occur at the third position of their respective codons and may be sequencing errors, or alternatively may be authentic differences between *M. pth* and *M. avium*. The presence of these base differences does not change the amino acid sequence predicted for MPTB281 and the *M. avium* homolog due to codon redundency.

3.9 Conclusion to Identification of Hypothetical *M. pth*281 ORF

The hypothetical N-terminal methionine start codon was selected on the basis that A; it was the only ATG start of translation codon between the 3' region of *katE* and the vector/*phoA* sequence that was in frame with *phoA* and B; it had a potential Shine-Dalgarno sequence 6 bp upstream. The only other possible N-terminal residue was a valine found adjacent to and upstream of the hypothetical methionine ATG start. The valine was encoded by GTC and not the valine start codon GTG so was considered unlikely to be the N-terminal residue. However, direct experimental evidence will be required to identify the true start codon.

The translated hypothetical *M. pth*281 ORF and the putative MPT281-PhoA fusion protein had no detectable signal sequence. In fact the MPT281 portion of the PhoA fusion protein, predicted from the proposed ATG start codon, carried only 6 amino acids upstream of PhoA; a length considered insufficient to act as a signal sequence. Hence, the possibility that the PhoA phenotype was a cloning or expression artefact can not be eliminated until direct experimental evidence has been gathered. Exported leaderless proteins, however, have been reported for *M. tb* (polyphosphate glucokinase) and *M. smegmatis* (acetamidase) fuelling speculation that an as yet unknown protein secretion pathway may exist in mycobacteria (64,65).

Part C: Cloning the Hypothetical *M. ptb*281 ORF

3.10 Introduction to Directional Cloning

Directional cloning describes the process where a gene is cloned into an expression vector in an orientation that maintains the correct translational reading frame. This process requires the use of two different restriction endonuclease recognition sites that are absent from the gene of interest but present in the multiple cloning site (MCS) of the expression vector. Those restriction sites are incorporated into oligonucleotide primers that are designed to flank and amplify the gene of interest by PCR. The first of these restriction sites to occur upstream in the MCS is incorporated into the 5` end of the forward primer while the second site, located downstream of the first, is incorporated into the 5` end of the reverse primer.

Following amplification by PCR the insert and expression vector were digested with the appropriate restriction endonucleases. The double digest creates two different oligonucleotide protrusions (cohesive ends) at the ends of the vector and insert. These cohesive ends have different sequences which prevent intramolecular ligation of vector and insert during cloning. The ligation process involves pairing of the cohesive ends from the vector with its complement on the insert thereby providing directionality during cloning and ensuring the correct reading frame of the insert is maintained in the expression vector.

3.10.1 Prokaryote Expression Vectors

The hypothetical *M. ptb*281 gene was directionally cloned into the *E. coli* and mycobacterial expression vectors pPROEX-HTb and pMIP12 respectively; both of which carry sequence encoding a poly-histidine affinity tag (His_6). A gene cloned into the MCS of these vectors will be expressed fused to the His_6-tag which allows purification of the recombinant protein by immobilized metal affinity chromatography (IMAC).

pMIP12

The *E. coli*/mycobacterial shuttle vector pMIP12 (supplied by the Pasteur Institute) allows mycobacterial proteins to be expressed in a mycobacterial system,

hence enabling possible immunogenic post-translational modifications specific to mycobacteria to be preserved (Appendix B).

pPROEX-Htb

The commercial *E. coli* expression vector pPROEX-HTb contains the powerful *E. coli* promoter *Trc* (*pTrc*) for high-level expression and the *lac*I^q repressor for regulated expression with IPTG (isopropyl β-D-thiogalactoside) (Appendix B). The presence of a recognition sequence for rTEV protease between the His_6-tag sequence and the extensive MCS, allows removal of the N-terminal affinity tag from the recombinant protein after purification.

3.11 *M. ptb*281 Insert Preparation

The hypothetical *M. ptb*281 ORF was analysed by MacVector; a computer programme that predicts the presence or absence of restriction endonuclease sites from a specified nucleotide sequence. Analysis showed the restriction endonuclease sites *Bam* HI and *Kpn* I were absent from *M. ptb*281. These sites were present in the MCSs of pPROEX-Htb and pMIP12 with *Bam* HI being located upstream of *Kpn* I in both cases. Using the *Bam* HI and *Kpn* I sites to clone *M. ptb*281 into pMIP12 and pPROEX-Htb would results in recombinant MPTB281 proteins larger than the 10.81 kDa predicted for the *M. ptb*281 ORF due to the addition of the affinity His_6-tag. Recombinant MTPB281 proteins from pMIP12 (M-MPTB281) and pPROEX-Htb (X-MPTB281) would be 11.59 kDa and 14.39 kDa respectively, with the latter being larger due to the presence of a spacer region and protease cleavage site between the His_6 tag and recombinant protein.

Two oligonucleotide primers were designed to flank and amplify the *M. ptb*281 gene by PCR from ***M. ptb***. Mac Vector Analysis (software; version I) showed both these primers annealed specifically to their complementary sequence in the *M. ptb*281 gene and did not form primer-dimers. To facilitate directional cloning of *M. ptb*281 a *Bam* HI site (*GGATCC*) was incorporated into the 5' end of the forward primer in a manner that would ensure the correct reading frame was expressed in the recombinant protein. A *Kpn* I site (*GGTACC*) was incorporated into the 5' end of the reverse primer and the *M. ptb*281 stop codon removed in a way as to allow the translation and fusion of the C-terminal histidine tag to the recombinant protein expressed by pMIP12. Inclusion of restriction endonuclease sites in the primers results in a PCR product of 273 bp

instead of the 258 bp predicted for the *M. ptb*281 ORF. Double restriction endonuclease digest removes the *Bam* H1 and *Kpn* 1 sites from the 273 bp PCR product to yield a 257 bp cloning fragment with *Bam* H1/*Kpn* I cohesive ends. To ensure the primers amplified the desired product and not a product of a single primer, a PCR check was performed using standard PCR conditions (section 2.16). Analysis of the PCR by agarose gel electrophoresis showed the primers MPTB281F2 & MPTB281R1 amplified a single product of approximately the expected size (273 bp) from *M. ptb* genomic DNA (Figure 19).

Individually, only the forward primer was able to produce an amplified product from *M. ptb* that ranged in size from 400 to 1650 bp with no products apparent below 300 bp. Using the high fidelity enzyme PLATINUM *Pfx* DNA polymerase the PCR was repeated with an anneal temperature of 61° C and subjected to agarose gel electrophoresis (data not shown).

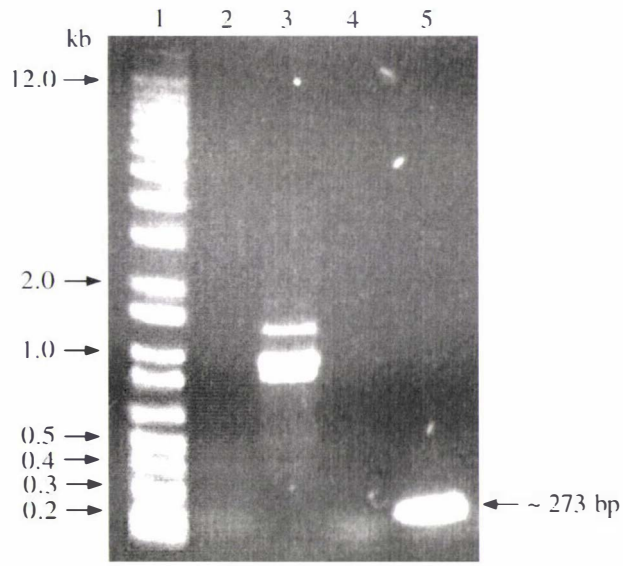


Figure 19: MPTB281F2 & MPTB281R1 primer check. Aliquots (7μL) from the primer check PCR and controls were loaded onto a 1% agarose gel and subjected to electrophoresis in 1x TAE buffer. The DNA was stained with dilute ethidium bromide and examined under UV light. Lanes: 1, 1 kb$^+$ ladder; 2, MPTB281F2 + MPTB281R1 negative control (no template); 3, forward primer MPTB281F2; 4, reverse primer MPTB281R1; 5, MPTB281F2 + MPTB281R1.

The PCR produced a single 273 bp band which was excised and purified by gel extraction (section 2.7) in preparation for digestion with restriction endonucleases. A *Kpn* I/*Bam* H1 double digest of the 273 bp PCR product was performed as described in the methods section (section 2.17.2).

3.12 pMIP12 and pProEX Preparation

Initial attempts to clone the *M. ptb*281 insert into pMIP12 and pPROEX using chemically competent *E. coli* DH5α cells manufactured in the laboratory failed, as did attempts to transform and harvest pUC18 from these cells (data not shown).

To improve the transformation efficiency of further cloning attempts, commercial electrocompetent *E. coli* DH10B cells were used instead of the chemically competent *E. coli* DH5α cells. In addition, pMIP12 and pPROEX vectors supplied by Chris Dupont that had previously been confirmed by DNA sequencing to have the MK35 insert ligated into their *Kpn* I/*Bam* H1 sites (unpublished data), were used as the source of vector for cloning. Their use enabled the excision of the insert to be monitored by agarose gel electrophoresis thus confirming the presence of *Kpn* I and *Bam* H1 cohesive ends.

E. coli cells containing pMIP12-MK35 and pPROEX-MK35 were grown and harvested as described in the methods section (section 2.10). Samples of pMIP12-MK35 and pPROEX-MK35 were subjected to double digest with *Kpn* I/*Bam* H1 (section 2.17.2) and purified by gel extraction (section 2.7). Samples from the *Kpn* I and *Kpn* I/*Ba*m H1 digests along with native samples were analysed by agarose gel electrophoresis to confirm the excision of the ~560 bp MK35 insert (Figure 20). Both pMIP12-MK35 and pPROEX-MK35 were rendered linear by the *Kpn* I digest as was evident by the absence of circular forms of vector on the gel. In addition, pMIP-MK35 and pPROEX-MK35 vectors migrated to ~7320 bp and ~5430 bp respectively; the size expected for these clones after a single site digest. Both the pMIP-MK35 and pPROEX-MK35 *Kpn* I/*Ba*m H1 digests produced two bands; one corresponding in size to the vector and the other to the ~560 bp insert, confirming the excision of MK35 and the presence of *Kpn* I/*Ba*m H1 cohesive ends. The remaining *Kpn* I/*Bam* H1 vector digests were subjected to agarose gel electrophoresis, purified by gel extraction and used to clone the *M. ptb*281 insert.

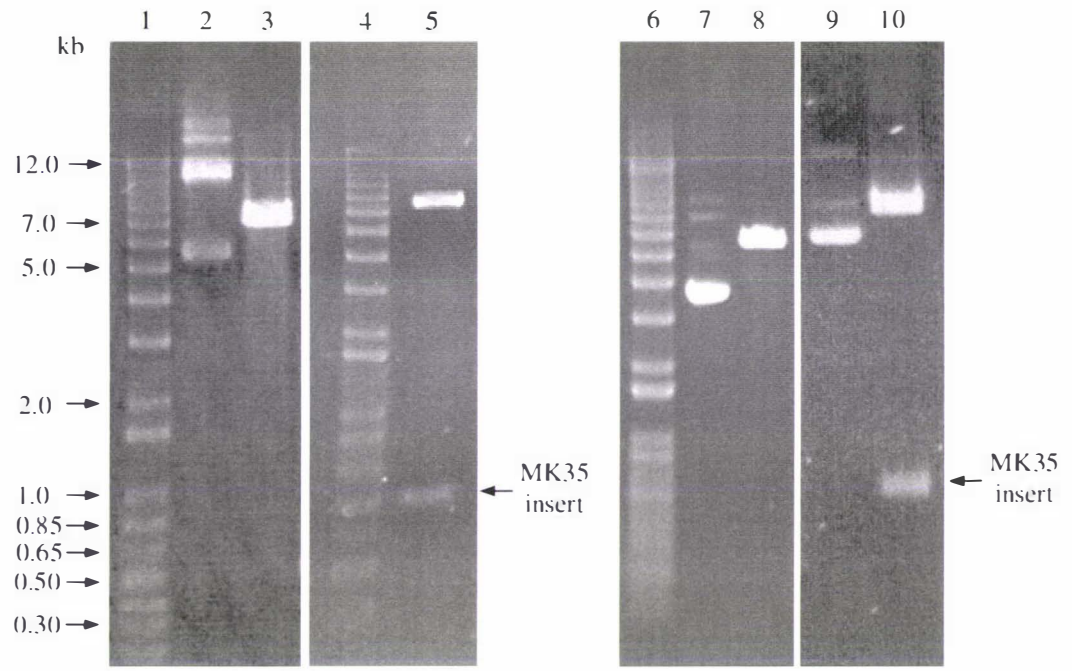


Figure 20: Preparation of vectors for cloning. Samples of native, *Kpn* 1 and *Kpn* 1/*Bam* H1 digests of pMIP-MK35 and pPROEX-MK35 were analysed by agarose gel electrophoresis soaked in ethidium bromide and visualized by UV light. Lanes: 1, 4 & 6, 1 kb^{+} ladder; 2, native pMIP-MK35; 3, pMIP-MK35 *Kpn* 1 digest; 5, pMIP-MK35 *Kpn* 1/*Bam* H1 digest; 7, native pPROEX-MK35; 8, pPROEX-MK35 *Kpn* 1 digest; 9, native pPROEX-MK35; 10, pPROEX-MK35 *Kpn* 1/*Bam* H1 digest.

3.13 Ligation and Transformation

The *Bam* H1/*Kpn* 1 digested *M. ptb*281 insert (section 3.11) and the *Bam* H1/*Kpn* 1 digested vectors pMIP12 and pPROEX (section 3.12) were ligated and transformed as described in the methods section (section 2.18 & 2.19). Briefly, the *M. ptb*281 insert was ligated into pMIP12 and pPROEX in separate reactions using T4 DNA Ligase at room temperature overnight. A sample from each ligation was dialysed, electroporated into *E. coli* DH10B cells and incubated at 37° C for 2 hours in 200 µL LB broth. The entire pMIP-*M. ptb*281 transformation mixture was spread onto an LB agar plate containing kanamycin (kan) while 15 µL of the pPROEX-*M. ptb*281 transformation mixture was spread onto an LB agar plate containing ampicillin (amp) and incubated at 37° C overnight. After incubation, the LB/kan plate contained approximately 2000 pMIP-*M. ptb*281 *E. coli* DH10B colony forming units (CFU). From this plate ~90

individual colonies were selected and streaked in a reference grid pattern on a fresh LB/kan plate and incubated at 37° C overnight in preparation for a PCR check.

The LB/amp plate containing pPROEX-*M. ptb*281 *E. coli* DH10B was confluent with what appeared to be small background colonies growing among larger CFUs. Samples were taken from the larger colonies and streaked onto a second LB/amp plate to separate individual colonies and incubated as above. After the incubation single colonies were selected and the process repeated on a third plate. Single colonies were transferred from the third plate to a fresh LB/amp plate in a reference grid pattern in preparation for a PCR check.

3.14 Screening of Recombinant Clones

To identify colonies carrying the cloned *M. ptb*281 insert, a standard PCR was performed with the primers MPTB281F2 & MPTB281R1 using colonies from the grid reference plates as template. Of the first 17 pMIP *M. ptb*281 *E. coli* DH10B colonies tested, 16 were positive for the *M. ptb*281 insert while 3 of the first 14 pPROEX-*M. ptb*281 *E. coli* DH10B colonies were positive (Figure 21). The presence of a high number of ampicillin resistant colonies on the LB/amp reference grid plate that do not appear to carry the *M. ptb*281 insert may indicate a recombination event took place between the vector and the host's genome that eliminated the insert or that the vector re-ligated to itself.

Single colonies from the PCR positive clones pMIP-*M. ptb*281 *E. coli* DH10B CFU #1 and pPROEX-*M. ptb*281 *E. coli* DH10B CFU #12 were streaked onto fresh LB agar plates containing the appropriate antibiotic and incubated at 37° C overnight.

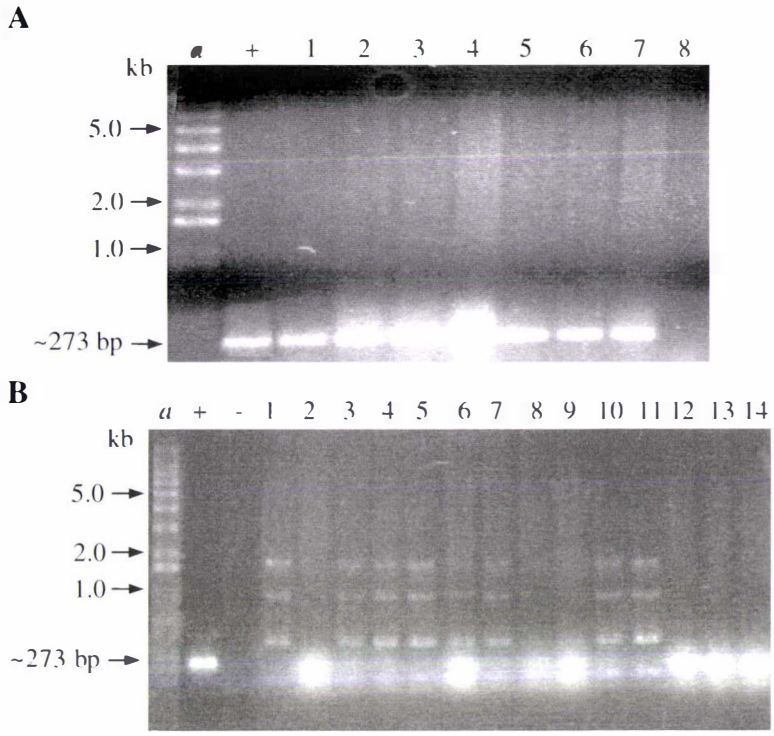


Figure 21: PCR screen to identify *M. ptb*281 positive clones. Using the primers MPTB281F2 and MPTB281R1 a standard PCR was performed using colonies from the grid reference plates as template. Aliquots (7 µL) from each PCR was analysed by agarose gel electrophoresis, soaked in dilute ethidium bromide and visualized by UV light. A, pMIP-*M. ptb*281 *E. coli* DH10B CFUs #1-8. B, pPROEX-*M. ptb*281 *E. coli* DH10B CFUs #1-14. Lanes: *a*, 1kb$^+$ ladder; +, positive control (template = 0.1 µL of purified *M. ptb*281 PCR product); -, negative control (no template).

After incubation a single colony was taken from each plate and separately incubated in 10 mL of LB broth with the appropriate antibiotic overnight at 37° C. A sample of each culture was taken and stored in glycerol at -70° C (section 2.11) for future use; while the remainder was used to harvest the plasmid vectors (section 2.10). Aliquots of the purified plasmids pMIP-*M. ptb*281 and pPROEX-*M. ptb*281 were subjected to *Kpn* I and *Kpn* I/*Bam* H1 digests and analysed by agarose gel electrophoresis to confirm the excision of the *M. ptb*281 insert (Figure 22).

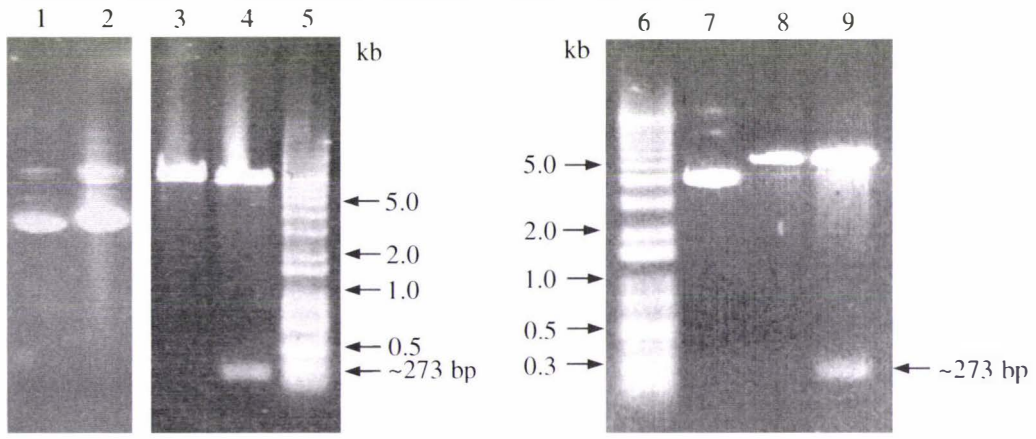


Figure 22: PMIP12-*M. ptb*281 and pPROEX-*M. ptb*281. Vectors were harvested from pMIP12-*M. ptb*281 *E. coli* DH10B CFU #1 and pPROEX-*M. ptb*281 *E. coli* DH10B CFU #12 and digested with *Kpn* 1 and *Kpn* 1/*Bam* H1. Aliquots were analysed by agarose gel electrophoresis, soaked in dilute ethidium bromide and visualized by UV light. Lanes: 1, native pMIP12; 2, native pMIP-*M. ptb*281; 3, pMIP-*M. ptb*281 *Kpn* 1 digest; 4, pMIP-*M. ptb*281 *Kpn* 1/*Bam* H1 digest; 5 & 6, $1kb^{+}$ ladder; 7, native pPROEX-*M. ptb*281; 8, pPROEX-*M. ptb*281 *Kpn* 1 digest; 9, pPROEX-*M. ptb*281 *Kpn* 1/*Bam* H1 digest.

The *Kpn* 1 digest rendered both pMIP12-*M. ptb*281 and pPROEX-*M. ptb*281 linear as was evident by the absence of circular forms on the gel. Both pMIP-*M. ptb*281 and pPROEX-*M. ptb*281 *Kpn* 1/*Bam* H1 digests produced two bands; one corresponding in size to the vector and the other to the ~273 bp insert. A sample of purified pPROEX-*M. ptb*281 was sent for DNA sequencing with the sequencing primer M13/pUC (section 2.3). Sequencing confirmed the *M. ptb*281 insert had been cloned into *Bam* H1 and *Kpn* 1 restriction sites of pPROEX and was in the correct reading frame (Figure 23).

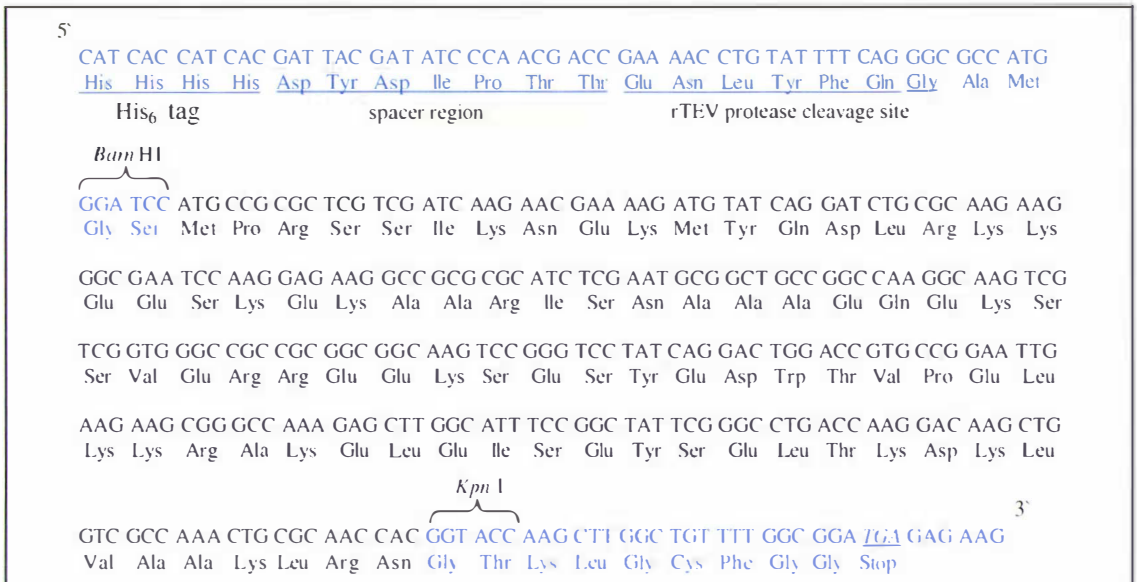


Figure 23: Nucleotide and translated sequence of pPROEX-*M. ptb*281. The plasmid vector from pPROEX-*M. ptb*281 *E. coli* DH10B CFU #12 was harvested and sent for DNA sequencing with the primer M13/pUC18. Blue text = pPROEX nucleotide (above) and amino acid sequence (below). Black text = *M. ptb*281 nucleotide (above) and amino acid sequence (below). *Bam* H1 and *Kpn* 1 restriction sites are indicated above by brackets. Underlined text indicates His_6-tag, spacer region, protease cleavage site and stop codon as stated.

3.15 Transformation of *M. smegmatis* mc^{2}155 with pMIP-*M. ptb*281

A sample of the plasmid vector purified from pMIP-*M. ptb*281 *E. coli* DH10B CFU #1 was transformed, as previously described, into *M. smegmatis* mc^{2}155 (section 2.20). The transformation mixture was incubated with 200 µL of LB broth at 37° C for ~2 hours before being spread onto a fresh LB/kan plate and incubated for 72 hours at 37° C. Of the ~200 CFUs that grew, ~90 individual colonies were streaked onto a fresh LB/kan plate and incubated as above. Two individual pMIP-*M. ptb*281 *M. smegmatis* mc^{2}155 colonies were selected and streaked onto a fresh LB/kan plate and incubated as before. A single colony was selected, incubated in 5 mL of Sauton's minimal media (section 2.9) supplemented with kan and incubated with agitation at 37° C for 72 hours. The *M. smegmatis* mc^{2}155 culture developed a turbid, clumping appearance that was typical for mycobacteria grown in liquid media. A sample of culture was taken and stored in glycerol at -70° C for future use while another was used to prepare a template for PCR (section 2.21). A standard PCR was performed using the primers BlaF2 and

TermR2 with 5 µL of pMIP-*M. ptb*281 *M. smegmatis* mc^{2}155 template to prepare a fragment for DNA sequencing. Analysis of the PCR by agarose gel electrophoresis showed three bands; a major product of the size expected (~460 bp) and two minor products below 300 bp (data not shown). The ~460 bp band was excised, purified and sent for DNA sequencing with the forward primer BlaF2. The nucleotide sequence confirmed the *M. ptb*281 insert was cloned into *Bam* H1 and *Kpn* 1 sites of pMIP12-*M. ptb*281 and in the correct reading frame (Figure 24). In addition, the nucleotide sequence of the *M. ptb*281 insert from pMIP12-*M. ptb*281 was 100% identical to its equivalent in pPROEX-*M. ptb*281.

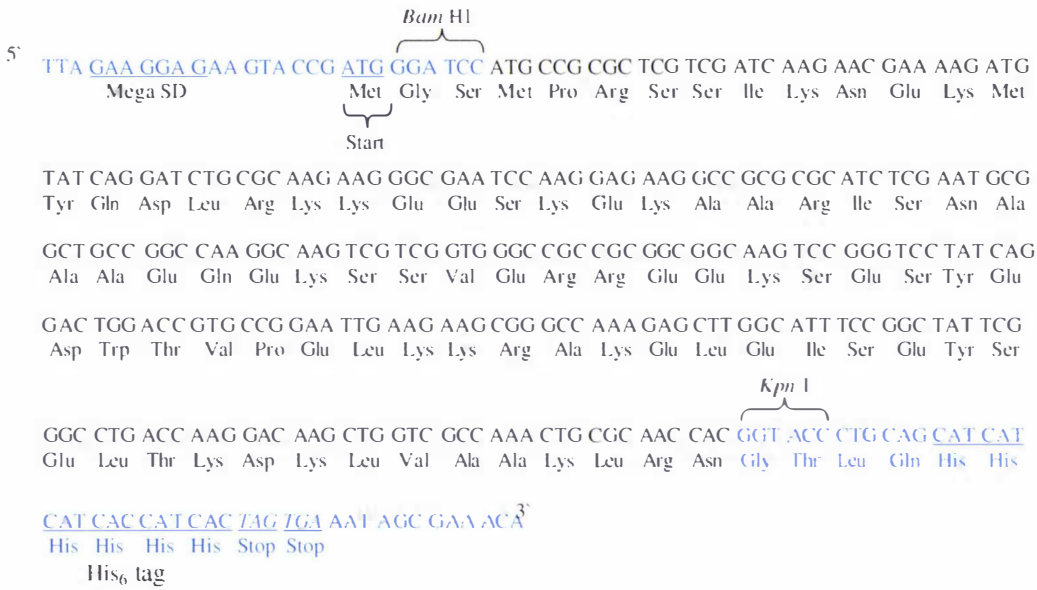


Figure 24: DNA and translated sequence of pMIP-*M. ptb*281. The plasmid vector from pMIP-*M. ptb*281 *M. smegmatis* mc^{2}155 CFU #1 was harvested and its DNA sequence (shown above with translated amino acids below) determined using the sequencing primer BlaF2. Blue text = pMIP sequence. Black text = *M. ptb*281. Endonuclease restriction sites, and start codon are indicated with brackets. His_6 tag, stop codons and Mega SD (mega Shine-Dalgarno) are underlined where indicated.

3.16 Conclusion to Cloning the Hypothetical *M. ptb*281 Gene

DNA sequencing confirmed the *M. ptb*281 insert was cloned into the *Kpn* I/*Bam* HI site in the correct reading frames of both pMIP-*M. ptb*281 *M. smegmatis* mc^2155 CFU #1 and pPROEX-*M. ptb*281 *E. coli* DH10B CFU #12. The *M. ptb*281 insert sequences from both these clones were 100% identical to the *M. ptb* database sequence at the nucleotide level.

Part D: Protein Expression and Western Analysis

3.17 Introduction

E. coli and mycobacterial expression systems pPROEX-Htb and pMIP12 respectively (section 3.10.1) produce histidine tagged (His_6) recombinant proteins. Recovery of His_6-tagged recombinant proteins from host bacteria can be facilitated by sonication, a method that uses high frequency ultrasonic waves to disrupt the cell membrane releasing the cells contents. Soluble recombinant proteins can then be semi-purified from cellular debris by immobilized metal affinity chromatography (IMAC). IMAC utilizes the strong and natural affinity of histidine for divalent ions immobilized to a solid inert support. The His_6-tagged proteins are bound to the solid support via the ion/histidine interaction allowing cellular debris and impurities to be removed in a series of washes. Disruption of the ion/histidine interaction can be achieved by changing the pH or by the using the histidine analog imidazole allowing His_6-tagged recombinant proteins to be eluted from the solid support.

Crude sonicated or semi-purified proteins can be analysed by sodium dodecyl sulphate polyacrylamide (SDS-PAGE) gel electrophoresis. This involves denaturing a protein sample in the presence of a reducing agent prior to electrophoresis. Anions of SDS bind to protein giving it a negative charge. The proteins move in the presence of an electric current through the polyacrylamide molecular sieve toward the anode separating the proteins based on size. The gel can then be stained to observe the protein banding pattern or subjected to western blotting to immunodetect recombinant proteins.

Western blotting, a process that transfers proteins by electrophoresis from an acrylamide gel to a membrane is a convenient robust method for identifying specific proteins by immunodetection. This technique utilizes the strong and specific binding affinity of an antibody for its antigen which is located on the blot. Conjugated to either a primary or secondary antibody is an assayable label allowing the antibody/antigen complex to be located on the blot.

Recombinant MPTB281 protein expressed by pMIP-*M. ptb*281 (M-MPTB281) and pPROEX-*M. ptb*281 (X-MPTB281) were predicted to be 11.59 kDa and 14.39 kDa respectively with the latter appearing larger due to the spacer region and rTEV protease cleavage site located between the His_6 tag and the recombinant protein.

3.18 Protein Expression from *E. coli* DH10B-pPROEX-*M. ptb*281

Small scale expression of the recombinant protein X-MPTB281 from *E. coli* DH10B-pPROEX-*M. ptb*281 was initially carried out as described in the methods section (section 2.22.1) to ascertain the optimal incubation time and concentration of the inducing agent IPTG. Pre-induced and induced samples were subjected to 8 x 30 sec sonication on ice (section 2.23) and aliquots of the supernatants and pellets subjected to 15% PAGE gel electrophoresis (section 2.24). Comparison of the protein pattern from the pre and post induced samples of pPROEX-*M. ptb*281 showed no obvious expression of the 14.39 kDa X-MPTB281 after 4 hrs or overnight incubation with 0.2 mM or 1.0 mM IPTG and appeared similar to the protein pattern of the pPROEX negative control (data not shown). Conversely the pPROEX-MK35 positive control showed strong induction of the 30 kDa MK35 protein after 4 hr and overnight incubations with no observable differences between 0.2 mM or 1.0 mM IPTG induction (data not shown).

A number of explanations could account for the apparent absence of X-MPTB281 from the cellular fraction of the culture including physical or biological proteolysis during preparation of the samples. Observation that the *E. coli* DH10B-pPROEX-*M. ptb*281 and control cultures shared similar OD_{600nm} readings over the incubation period suggests cell toxicity was not the cause of the low cytosolic concentration. Identified as a secreted protein by the pJEM11-*M. ptb* library but having no identifiable *E. coli* or mycobacterial signal sequence, export of X-MPTB281 from the cytosol into the media by a Sec (secretion) independent pathway remained one possibility. Alternatively, X-MPTB281 could have an adverse effect on its own expression in *E. coli* or be subject to high rates of mRNA or protein degradation resulting in low cytosolic concentrations.

With the assumption that X-MPTB281 was present but at low cytosolic concentrations, large scale 0.2 mM IPTG induced overnight cultures of pPROEX-*M. ptb*281 and pPROEX were grown (section 2.22.2). Pellet samples (0.1 g) from each were re-suspended in 6 mL start buffer (20 mM Na_2HPO_4, 0.5 M NaCl, 15 mM imidzole, pH 7.4) and sonicated on ice for the shorter duration of 30 sec to reduce the possibility of mechanical proteolysis. The sonicates were then subjected to a series of centrifugations at 4° C for 5 mins at 14000 g to remove cellular debris and the supernatants applied to a nickel affinity chromatography column to enrich for the His_6-tagged recombinant protein (section 2.25). Aliquots of selected fractions were subjected to 15% SDS PAGE gel electrophoresis. Comparison of the fractions showed the presence of a very faint

band (~15 kDa) in the 250 mM imidazole fraction of pPROEX-*M. ptb*281 that was absent from the equivalent negative control fraction and was of the approximate size expected for X-MPTB281 (14.39 kDa) (data not shown).

To increase the concentration of the ~15 kDa protein the sonication step was repeated using 0.2 g of pellet re-suspended in 3 mL start buffer. Selected column fractions containing ~9.4 µL of sample were subjected to 15% SDS-PAGE analysis (Figure 25). A strong distinct band was observed in the 250 mM imidazole fraction of pPROEX-*M. ptb*281 that was absent from the equivalent negative control fraction. The band migrated to ~15 kDa slightly larger than the 14.4 kDa expected for X-MPTB281.

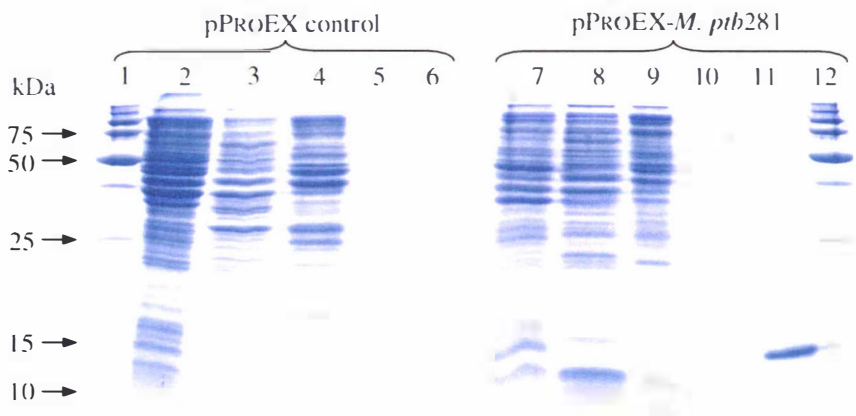


Figure 25: Enriched recombinant X-MPTB281. Pellet samples from the overnight 0.2 mM IPTG induced pPROEX-*M. ptb*281 and pPROEX control cultures were sonicated and the supernatant applied to a nickel affinity chromatography column. Selected fractions were analysed by 15% SDS PAGE gel electrophoresis and stained with Coomassie blue. Lanes: 1 & 12, Precision Protein Standard; 2-6, pPROEX negative control fractions; 7-11, pPROEX-*M. ptb*281 fractions; 2 & 7, sonicated insoluble cellular debris; 3 & 8, unbound protein flow through fraction; 4 & 9, unbound protein fraction after start buffer wash; 5 & 10, 100 mM imidazole fraction; 6 & 11, 250 mM imidazole fraction.

The size discrepancy may be an artefact of the SDS-PAGE gel or an indication that post-translational modification occurred making the recombinant protein migrate further than expected. In addition high molecular weight bands were observed migrating

to ~75 kDa and 50 kDa. A strong ~12 kDa band was observed in lane 8 of pPROEX-*M. ptb*281. While a band of approximately the same size was present in the equivalent control fraction it was not of the same relative intensity. N-terminal sequencing could be undertaken to indicate the nature of this protein. Further attempts to purify X-MPTB281 were unsuccessful and may have been due to loss of the protein during an additional step that involved concentrating the sonicated supernatant prior to nickel affinity chromatography (data not shown).

3.19 Protein Expression from *M. smegmatis* mc^{2}155-pMIP-*M. ptb*281

Sauton's cultures (50 mL) of *M. smegmatis* mc^{2}155-pMIP-12 (negative control) and *M. smegmatis* mc^{2}155-pMIP-*M. ptb*281 were grown as described in the methods section (section 2.22.3). The pellets were resuspended in 4 mL start buffer, sonicated for 40 sec on ice and subjected to a series of centrifugations at 4° C for 5 mins at 14000 rpm to remove cellular debris. The supernatants were applied to a nickel affinity chromatography column to enrich for the constitutively expressed ~11.59 kDa recombinant protein M-MPTB281 and selected fractions subjected to 15% SDS PAGE gel electrophoresis. A faint ~15 kDa band was observed in the 250 mM imidazole fraction of *M. smegmatis* mc^{2}155-pMIP-*M. ptb*281 that was absent from the equivalent control fraction (data not shown). The 250 mM imidazole *M. smegmatis* mc^{2}155-pMIP-*M. ptb*281 fraction was concentrated using a Centricon Centrifugal Filter Device with a 3,000 molecular weight (MW) cut-off (section 2.26). A sample of flow through and concentrate (9.4 µL) from the filter was analysed by 15% SDS PAGE gel electrophoresis (Figure 26). A faint ~60 kDa band was present in both the flow through and concentrated fraction of the gel that cannot be seen in the scanned image shown here. A strong defined ~15 kDa band was present only in the concentrated 250 mM imidazole fraction. Although the image shown does not include the 10 kDa molecular weight marker no bands were observed below 15 kDa when SDS-PAGE gel analysis of the concentrated fraction was repeated (data not shown). The absence of the ~15 kDa band from the control culture and its presence in *M. smegmatis* mc^{2}155-pMIP-*M. ptb*281 culture suggests the band represents M-MPTB281, however, N-terminal amino acid sequencing would be required to confirm this. The size discrepancy of 3.4 kDa between the predicted M-MPTB281 and the semi-purified protein may be an artefact or an indication of post-translational modification.

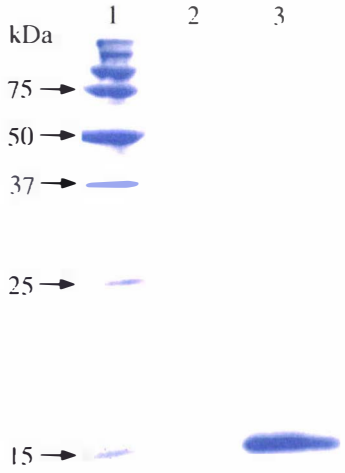


Figure 26: Enriched recombinant M-MPTB281. The pellet from a 7 day 50 mL Sauton's culture of *M. smegmatis* mc^{2}155-pMIP-*M. ptb*281 was sonicated in 4 mL start buffer and the supernatant applied to a nickel affinity chromatography column. The 250 mM imidazole fraction was concentrated 20 fold using a 3,000 MW cut-off centrifugal filter device. Aliquots containing ~9.4 µL of the flow through and concentrated fractions were analysed by 15% SDS PAGE gel electrophoresis and stained with coomassie blue. Lanes: 1, Precision Protein Standard; 2, flow through fraction; 3, concentrated fraction

Further attempts to purify M-MPTB281 were unsuccessful and hampered by the inability to reproduce *M. smegmatis* mc^{2}155 cultures with sufficient growth. As previously mentioned the inclusion of an additional step that concentrated the sonicated supernatant prior to nickel affinity chromatography may have resulted in loss of the protein (data not shown).

3.20 Preparation of Western Blots

A western blot containing ~1.1 µL/lane of semi-purified X-MPTB281 from the 250 mM imidazole nickel column fraction was prepared as described in the methods section (section 2.27). After transfer the gel was stained with Coomassie blue and the membrane stained with Ponceau S to visualize the degree of protein transfer from the gel to the membrane (Figure 27).

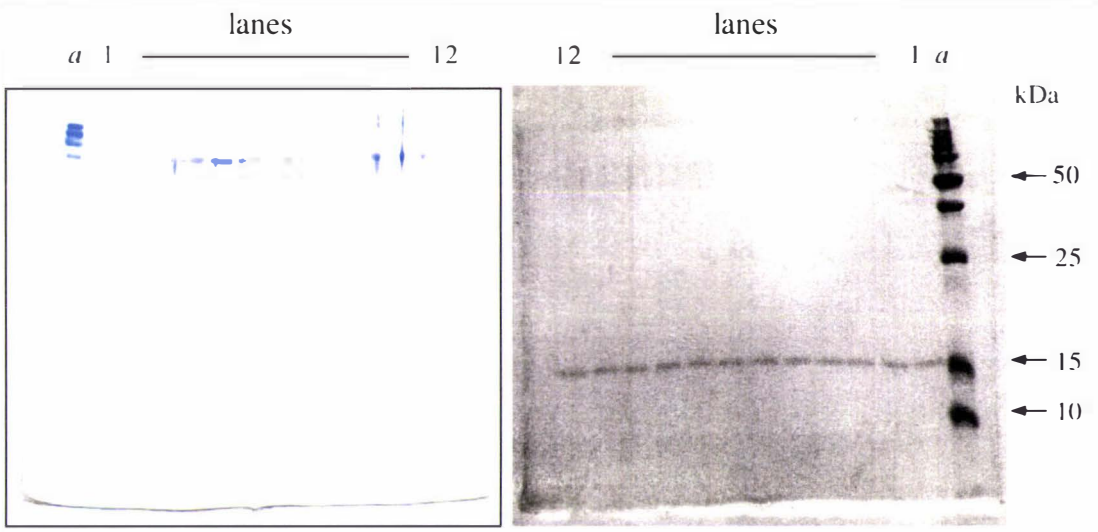


Figure 27: Western transfer of X-MPTB281 recombinant protein. A 15% SDS-PAGE gel was electrophoresed with X-MPTB281 and the protein transferred to nitrocellulose membrane by electroblotting. The gel was stained with Coomassie blue and the membrane with Ponceau S to confirm protein transfer occurred. Left, Coomassie blue stained PAGE gel after electroblotting. Right, Ponceau S stained membrane after electroblotting. Lanes: *a*, Prestained Precision Protein Standard; 1 to12 (equivalent to strips 1 to 12).

The absence of the ~15 kDa protein from the gel and its presence on the membrane showed transfer of the recombinant protein was successful. In addition, high molecular weight impurities transferred poorly as they were pronounced on the gel while not visible on the membrane. A second western blot containing ~1.5 µL/lane of semi-purified M-MPTB281 protein from the concentrated 250 mM imidazole nickel column fraction was prepared as above. The post-transfer Coomassie blue and Ponceau S stains of the gel and membrane confirmed the ~15 kDa recombinant protein transferred successfully, in contrast to the high molecular weight proteins that were pronounced on the gel while not visible on the membrane (data not shown).

3.21 Immunodetection

One strip from each of the X-MPTB281 and M-MPTB281 western blots prepared above were subjected to immunodetection with mouse monoclonal Anti-His_6-Peroxidase Antibody (anti-His_6-antibody) and the antibody-antigen complex visualized

with the chemiluminescent substrate SuperSignal West Femto (Figure 28) as described in the methods section (section 2.27.1). Briefly, the membranes were blocked (1 hr) and incubated (1 hr) in anti-His_6-antibody at room temperature. After a series of washes and the addition of substrate, the membranes were exposed to film for 1 sec, 10 sec, 1 min, 30 min, and overnight. After 30 min exposure a chemiluminescent signal was detected on the X-MPTB281 western blot but not the M-MPTB281 western blot. Overnight exposures produced a single chemiluminescent signal from X-MPTB281 and M-MPTB281 blots that correspond to ~15 kDa as judged by alignment of the appropriate reference ladders to each blot (Figure 28).

These signals correspond to the semi-purified ~15 kDa proteins from the 250 mM imidazole nickel column fractions from *E. coli* DH10B-pPROEX-*M. ptb*281 and *M. smegmatis* $mc^2$155-pMIP-*M. ptb*281

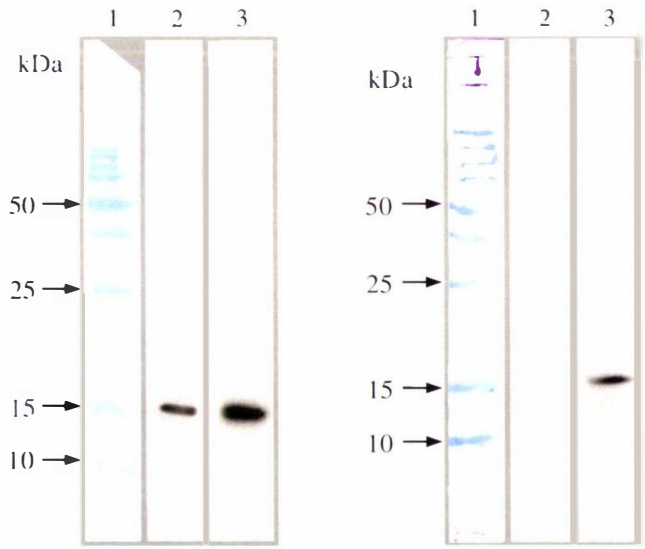


Figure 28: Immunodection of X-MPTB281 and M-MPTB281 western blots. Membranes were blocked 1 hr and incubated in anti-His_6-antibody 1 hr, washed chemiluminescent substrate added and exposed to X-ray film. Reference ladders were aligned with the outline of the blots on the film to allow estimation of the protein size that the signal emanated from. Left, X-MPTB281 western blot. Right, M-MPTB281 western blot. Lanes: 1, Prestained Precision Protein Standard; 2, 30 min exposure; 3, over night exposure.

3.22 Conclusion to Protein Expression and western Analysis

Affinity chromatography appears to have semi-purified recombinant proteins X-MPTB281 and M-MPTB281 as judged by the absence of these bands from control cultures on SDS-PAGE gels. The size predicted for X-MPTB281 and M-MPTB281 was ~14.39 kDa and ~11.6 kDa respectively, however, both semi-purified proteins migrated to ~15 kDa. This size discrepancy may be an artefact or an indication of post-translational modification. Western blot analysis of the semi-purified recombinant proteins using specific anti-His_6-tag antibodies appears to confirm the presence of a His_6 affinity tag on these proteins. These results lend strong support to the conclusion that X-MPTB281 and M-MPTB281 have been semi-purified. However, definitive confirmation of the identity of these proteins can not be assigned until they have been further purified and their amino acid sequences deduced.

Part E: Species Distribution of the Hypothetical *M. ptb*281 ORF

3.23 Introduction

To further investigate the species distribution of *M. ptb*281, the full length hypothetical *M. ptb*281 gene was used in database screens, Southern blot hybridizations and PCR analyses of multiple mycobacterial species.

3.24 Database Search

NCBI, TIGR, Sanger and Center for Computational Genomics and Bioinformatics (CCGB) databases were searched with both the nucleotide and translated sequences of the full length hypothetical *M. ptb*281 gene using BLAST programs. Using these databases no significant alignments were found to MTB complex members *M. tb* (strains H37Rv and CDC1551), *M. bovis* (strain AF2122/97, spoligotype 9) or *M. leprae* at the protein or nucleotide level.

The CCGB database reported a significant alignment to *M. ptb* (cattle strain K-10) of 100% identities at the nucleotide level (Figure 29; A, TBLASTN) and 81% identities at the amino acid level (Figure 29; B, TBLASTX). Closer inspection of MPTB281 sequence and the amino acid database subject showed the two sequences shared 100% identities (Figure 29; C, manual alignment). The database automatically substituted query sequence of low complexity with the letter "X" to reduce the prevalence of artifactual hits and excluded that sequence from the percentage calculation resulting in the incorrect report of 85% amino acid identity where there was 100%.

A

```
Expect = e-143, Identities = 258/258 (100%)

Query: atgccgcgctcgtcgatcaagaacgaaaagatgtatcaggatctgcgcaagaagggcgaa 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: atgccgcgctcgtcgatcaagaacgaaaagatgtatcaggatctgcgcaagaagggcgaa 516244

Query: tccaaggagaaggccgcgcgcatctcgaatgcggctgccggccaaggcaagtcgtcggtg 120
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: tccaaggagaaggccgcgcgcatctcgaatgcggctgccggccaaggcaagtcgtcggtg 516304

Query: ggccgccgcggcggcaagtccgggtcctatcaggactggaccgtgccggaattgaagaag 180
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: ggccgccgcggcggcaagtccgggtcctatcaggactggaccgtgccggaattgaagaag 516364

Query: cgggccaaagagcttggcatttccggctattcgggcctgaccaaggacaagctggtcgcc 240
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: cgggccaaagagcttggcatttccggctattcgggcctgaccaaggacaagctggtcgcc 516424

Query: aaactgcgcaaccactga 258
       ||||||||||||||||||
Sbjct: aaactgcgcaaccactga 516442
```

B

```
Expect = 2e-40, Identities = 70/86 (81%)

Query 1   MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAXXXXXXXXXXXXXXXXXXYQDWTVPELKK 178
          MPRSSIKNEKMYQDLRKKGESKEKAARISNAAA                  YQDWTVPELKK
Sbjct     MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKK

Query 179 RAKELGISGYSGLTKDKLVAKLRNH* 254
          RAKELGISGYSGLTKDKLVAKLRNH*
Sbjct     RAKELGISGYSGLTKDKLVAKLRNH*
```

C

```
MPTB281 vs M. ptb database subject; Amino acid Identities = 100%

MPTB281: 1  MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKK 60
            MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKK
Sbjct:      MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKK

MPTB281: 61 RAKELGISGYSGLTKDKLVAKLRNH- 87
            RAKELGISGYSGLTKDKLVAKLRNH-
Sbjct:      RAKELGISGYSGLTKDKLVAKLRNH-
```

Figure 29: CCGB database alignment to *M. ptb*. Alignment of *M. ptb*281 to the CCGB database subject *M. ptb* strain K10. A, BLASTN alignment; B, TBLASTX alignment; C, manual alignment of MPTB281 and amino acid subject. Query = *M. ptb*281 nucleotide sequence; Sbjct = *M. ptb* identity in database; X = residue omitted from search by low complexity filter.

The Sanger database gave significant BLASTN and TBLASTX alignments to *M. marinum* (strain M) with E values of $1.1e^{-32}$ and $2.8e^{-30}$ respectively (Figure 30).

A

```
mar502b11.q1k: Expect = 1.1e-32, Identities & Positives = 209/259 (80%)

Query:   1 ATGCCGCGCTCGTCGATCAAGAACGAAAAGATGTATCAGGATCTGCGCAA-GAAGGGCGA 59
           ||||||   | | ||||||||||||||| ||| | || | ||||||||||||| |||||  |
Sbjct: 268 ATGCCGAACCCCTCGATCAAGAACGAGAAGCTATACCGGGATCTGCGCAAAGAAGGCAGC 327

Query:  60 ATCCAAGGAGAAGGCCGCGCGCATCTCGAATGCGGCTGCCGGCCAAGGCAAGTCGTCGGT 119
           ||||||| | ||| ||||| || || ||||||| || | ||| ||||  || |||||
Sbjct: 328 -TCCAAGGAAAGGGCAGCGCGGATTTCCAATGCGGCAGCTGCCCAGGGCACTTCCTCGGT 386

Query: 120 GGGCCGCCGCGGCGGCAAGTCCGGGTCCTATCAGGACTGGACCGTGCCGGAATTGAAGAA 179
           |||||||   |||||||| || |||||||| |||||||| ||||| ||| |||||||
Sbjct: 387 GGGCCGCAAAGGCGGCAAGGCCCGGTCCTATCCGGACTGGACGGTGCCAGAACTGAAGAA 446

Query: 180 GCGGGGCCAAAGAGCTTGGCATTTCCGGCTATTCGGGCCTGACCAAGGACAAGCTGGTCGC 239
           |||||||||||||||||||||| ||| | ||||| |||||||| || ||| || || ||
Sbjct: 447 ACGGGCCAAAGAGCTTGGCATGTCCAGGTATTCCGGCCTGACAAAAGACGAGTTGATCAG 506

Query: 240 CAAACTGCGCAACCACTGA 258
           ||| |||||||||| ||||
Sbjct: 507 CAAGCTGCGCAACCGCTGA 525
```

B

```
mar502b11.q1k: Expect = 2.8e-30, Identities & Positives = 77/86 (89%)

Query:   1 MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKK 180
           MP  SIKNEK+Y+DLRK+G SKE+AARISNAAA QG SSVGR+GGK+ SY DWTVPELKK
Sbjct: 268 MPNPSIKNEKLYRDLRKEGSSKERAARISNAAAAQGTSSVGRKGGKARSYPDWTVPELKK 447

Query: 181 RAKELGISGYSGLTKDKLVAKLRNH* 258
           RAKELG+S YSGLTKD+L++KLRN *
Sbjct: 448 RAKELGMSRYSGLTKDELISKLRNR* 525
```

Figure 30: Sanger *M. marinum* database alignment. A, BLASTN alignment of *M. ptb*281 to the *M. marinum* subject. B, TBLASTX alignment of translated *M. ptb*281 to a *M. marinum* database subject. Query = MPTB281 nucleotide or amino acid sequence; target = *M. marinum* database subject; + = conservative amino acid substitution.

The TIGR database gave significant alignments to *M. avium* strain 104 at the nucleotide and amino acid level of 99% (Figure 31; A, BLASTN) and 100% (Figure 31 TBLASTX) respectively.

A

```
Expect = 4.6e-53, Identities = 257/258 (99%)

Query: ATGCCGCGCTCGTCGATCAAGAACGAAAAGATGTATCAGGATCTGCGCAAGAAGGGCGAA 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: ATGCCGCGCTCGTCGATCAAGAACGAAAAGATGTATCAGGATCTGCGCAAGAAGGGCGAA 4193448

Query: TCCAAGGAGAAGGCCGCGCGCATCTCGAATGCGGCCGCCGGCCAAGGCAAGTCGTCGGTG 120
       |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct: TCCAAGGAGAAGGCCGCGCGCATCTCCAATGCGGCCGCCGGCCAAGGCAAGTCGTCGGTG 4193508

Query: GGCCGCCGCGGCGGCAAGTCCGGGTCCTATCAGGACTGGACCGTGCCGGAATTGAAGAAG 180
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: GGCCGCCGCGGCGGCAAGTCCGGGTCCTATCAGGACTGGACCGTGCCGGAATTGAAGAAG 4193568

Query: CGGGCCAAAGAGCTTGGCATTTCCGGCTATTCGGGCCTGACCAAGGACAAGCTGGTCGCC 240
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: CGGGCCAAAGAGCTTGGCATTTCCGGCTATTCGGGCCTGACCAAGGACAAGCTGGTCGCC 4193628

Query: AAACTGCGCAACCACTGA 258
       ||||||||||||||||||
Sbjct: AAACTGCGCAACCACTGA 4193646
```

B

```
Expect = 1.4e-40, Identities = 85/85 (100%)

Query: MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKK 60
       MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKK
Sbjct: MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKK 4193568

Query: RAKELGISGYSGLTKDKLVAKLRNH 85
       RAKELGISGYSGLTKDKLVAKLRNH
Sbjct: RAKELGISGYSGLTKDKLVAKLRNH 4193643
```

Figure 31: TIGR database alignment to *M. avium*. Alignment of *M. ptb*281 to the TIGR database subject *M. avium* strain 104. A, BLASTN alignment; B, TBLASTN alignment. Query = *M. ptb*281; Sbjct = *M. avium* identity in database.

The database also gave significant alignments to *M. smegmatis* strain mc^2155 at the nucleotide and amino acid level of 75% (Figure 32; A, BLASTN) and 85% (Figure 32; B, TBLASTN) respectively. Being un-annotated at the time of this work the *M. ptb*, *M. avium*, *M. smegmatis* and *M. marinum* databases gave no additional information with their alignments.

A

```
Expect = 2.6e-26, Identities = 195/260 (75%)

Query: ATGCCGCGCTCGTCGATCAAGAACGAAAAGATGTATCAGGATCTGCGCAAGAAGGGCGAA 60
       ||||||    || |||||||||| |||| ||| | || |||| |||||||||| |||||||
Sbjct: ATGCCGAATTCATCGATCAAGGACGAGAAGCTCTACCAGGACCTGCGCAAGCAGGGCGAC 1365648

Query: TCCAAGGAGAAGGCCGCGCGCATCTCGAATGCGGCCGCCGGCCAAGGCAAGTCGTCGGTG 120
       || |||||||||||||||||||||||| || |||||  ||    || |||| |||  |||
Sbjct: TCGAAGGAGAAGGCCGCGCGCATCTCCAACGCGGCGGCGTCCCGAGGCCGGTCCAAGGTC 1365708

Query: GGCCGC-CGCGGCGGCAAGTCCGGGTCCTATCAGGACTGGACCGTGCCGGAATTGAAGAA 179
       || ||| || ||||||||||||||| ||||||||||||||||||| || ||  ||||||
Sbjct: GGGCGCTCG-GGCGGCAAGTCCGGATCCTATGAGGACTGGACCGTGTCCGACCTGC-GCA 1365766

Query: G-CGGGCCAAAGAGCTTGGCATTTCCGGCTATTCGGGCCTGACCAAGGACAAGCTGGTCG 238
       | || || ||||| ||||| || |||| |||||| | || |||||  || ||||
Sbjct: GTCGTGCGAAAGAACTTGGGATCACCGGATATTCCGACAAGAACAAGGGTGAGTTGGTGA 1365826

Query: CCAAACTGCGCAACCACTGA 258
       |  ||  | ||||||
Sbjct: AGATGCTCAGGAATCACTGA 1365846
```

B

```
Expect = 4.9e-29, Identities = 62/85 (72%), similarities = 73/85 (85%)

Query: MPRSSIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKK 60
       MP SSIK+EK+YQDLRK+G+SKEKAARISNAAA +G+S VGR GGKSGSY+DWTV +L+
Sbjct: MPNSSIKDEKLYQDLRKQGDSKEKAARISNAAASRGRSKVGRSGGKSGSYEDWTVSDLRS 1365768

Query: RAKELGISGYSGLTKDKLVAKLRNH 85
       RAKELGI+GYS   K +LV  LRNH
Sbjct: RAKELGITGYSDKNKGELVKMLRNH 1365843
```

Figure 32: TIGR database alignment to *M. smegmatis*. Alignment of *M. ptb*281 to TIGR database subject *M. smegmatis* strain mc^2155. A, BLASTN alignment; B, TBLASTN alignment. Query = *M. ptb*281; Sbjct = *M. smegmatis* identity in database; + indicates conservative substitutions.

The NCBI database reported two significant BLASTP alignments to two fully sequenced and annotated genomes (Figure 33). The first was to a hypothetical protein of *Deinococcus radiodurans* (*D. radiodurans*) strain R1 (ATCC 13939) (66) that had an E value of $6e^{-04}$ with 51% similarity (Figure 33; A). The second alignment was to a hypothetical protein from *Corynebacterium efficiens* (*C. efficiens*) strain YS-314 (67) which had an E value of $5e^{-20}$ with 75% similarity (Figure 33; B).

A

```
gi|15807544|ref|NP_296280.1| hypothetical protein [D. radiodurans]
gi|7472660|pir||F75258 hypothetical protein D. radiodurans (strain R1)
gi|6460391|gb|AAF12105.1|AE002085_8   hypothetical protein [D. radiodurans]
Expect = 6e-04, Identities = 31/95 (32%), Similarities = 49/95 (51%),
Gaps = 13/95 (13%)

Query: 1  MPRS-SIKNEKMYQDLR----KKGESKEKAARISNAAAGQGKSSVGRR--------GGKS 47
          MP++ S K+E+ Y+ ++    K+GES ++A  I+     + +   GR         G
Sbjct: 1  MPKAWSNKDERQYEHVKDSEVKRGESPDRAEEIAARTVNKSRREEGRTPNKRTQGTGNPD 60

Query: 48 GSYQDWTVPELKKRAKELGISGYSGLTKDKLVAKL 82
            +  D T  EL  RAKE GI+G S ++K +LV  L
Sbjct: 61 AALSDLTRDELYNRAKEKGIAGRSRMSKAELVRAL 95
```

B

```
gi|25029165|ref|NP_739219.1| hypothetical protein [C. efficiens YS-314]
gi|23494453|dbj|BAC19419.1| hypothetical protein [C. efficiens YS-314]
Expect = 5e-20, Identities = 45/81 (55%), Similarities = 61/81 (75%)

Query: 5  SIKNEKMYQDLRKKGESKEKAARISNAAAGQGKSSVGRRGGKSGSYQDWTVPELKKRAKE 64
          S+K+ ++Y++LR+ G SKEKAARI+NA A    +  VG +GGK+GSY+DWTV EL+ RA E
Sbjct: 18 SVKDGELYEELREDGASKEKAARIANATANTSRGEVGEKGGKAGSYEDWTVEELRTRAAE 77

Query: 65 LGISGYSGLTKDKLVAKLRNH 85
          L I G S + KD+L+  LRNH
Sbjct: 78 LDIDGRSKMKKDELIDALRNH 98
```

Figure 33: NCBI database alignments. A, BLASTP alignment of MPTB281 to database subject *D. radiodurans*; B, BLASTP alignment of MPTB281 to database subject *C. efficiens* strain YS-314. Query = MPTB281; Sbjct = identity in database; + indicates conservative substitutions; - = gap in alignment.

The completed and analysed genomes of *C. diphtheriae* (strain NCTC13129) (68), a known pathogen and *C. glutamicum* (ATCC 13032) (69) used for the commercial production of amino acids, are both close relatives of *C. efficiens*. Their genomes were subjected to BLAST searches through the Sanger and PEDANT (Protein Extraction, Description, and ANalysis Tool) databases using the nucleotide and translated sequence of *M. ptb*281. The search results reported no significant alignments.

3.25 Southern Blot Analysis of Selected Mycobacterial Species

The use of mycobacterial database searches to determine the species distribution of *M. ptb*281 was limited due to the small number of fully sequenced genomes. To obtain more information, genomic DNA from selected mycobacterial species were subjected to Southern blot analysis and probed with the full length *M. ptb*281 gene.

3.25.1 Preparation of Probes

*M. ptb*281-DIG

The cloning primers MPTB281F2 and MPTB281R1 were used in a standard PCR to amplify the insert from pMIP-*M. ptb*281. The resultant 273 bp fragment was purified and DIG labelled as previously described (section 3.4.1).

16S-DIG

The 1020 bp fragment specific to the 16S ribosomal subunit of mycobacteria was amplified from BCG genomic DNA in a standard PCR with the primers 246 and 264. The fragment was purified and DIG labelled for use as a positive control (data not shown).

3.25.2 *Bam* H1 Southern Blot #3

A selection of mycobacterial species was grown and their genomic DNA extracted as described in the methods section (section 2.5). Samples of DNA from each species were digested with *Bam* H1 (section 2.17.1) and aliquots containing approximately 1 µg of DNA were subjected to 0.7% agarose gel electrophoresis. A reference photo was taken of the gel aligned next to an ultra-violet reflective ruler (Figure 34) and the DNA pattern subjected to Southern transfer.

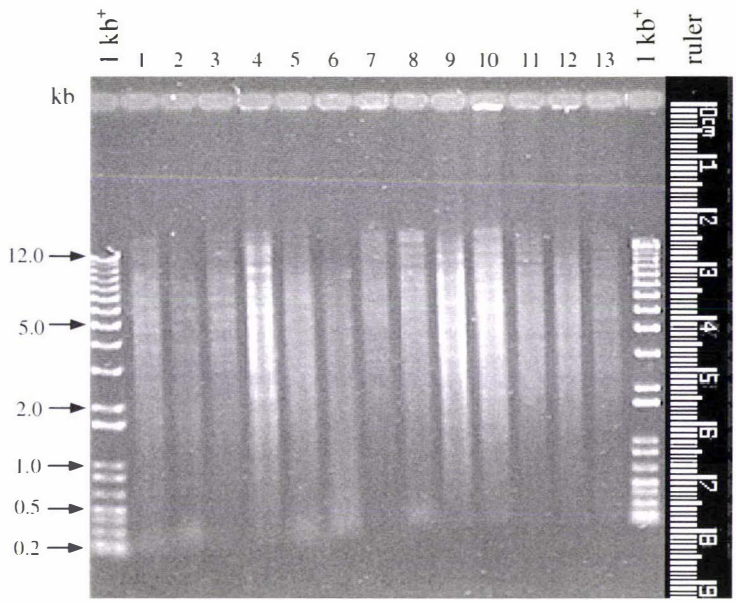


Figure 34: Reference photo of *Bam* H1 digested mycobacterial DNA. Aliquots containing approximately 1 µg of *Bam* H1 digested DNA from selected mycobacterial species were electrophoresed on a 0.7% agarose gel in 1x TAE buffer. The gel was soaked in dilute ethidium bromide and a photo taken of the gel aligned next to fluorescent rulers. Ruler; ultra violet florescent ruler. 1 kb^+ = DNA ladder; Lanes: 1, BCG; 2, *M. tb*; 3, *M. bovis*; 4, *M. intracellulare*; 5, Neoparasec; 6, *M. scrofulaceum*; 7, *M. gordonae*; 8, *M. kansasii*; 9, *M. phlei*; 10, *M. marinum*; 11, *M. terrae*; 12, *M. fortuitum*; 13, *M. ptb*.

3.25.3 High Stringency Conditions

Bam H1 Southern blot #3 made from the gel in Figure 34 was incubated with *M. ptb*281-DIG as previously described (section 3.4.3) using a 42° C hybridization temperature (high stringency hybridization). High stringency post-hybridization washes were performed and the blot prepared for chemiluminescent detection with CDP^{Star}. The blot was exposed to X-ray film for ½ hr, 3 hrs and overnight to get a range of exposures (Figure 35).

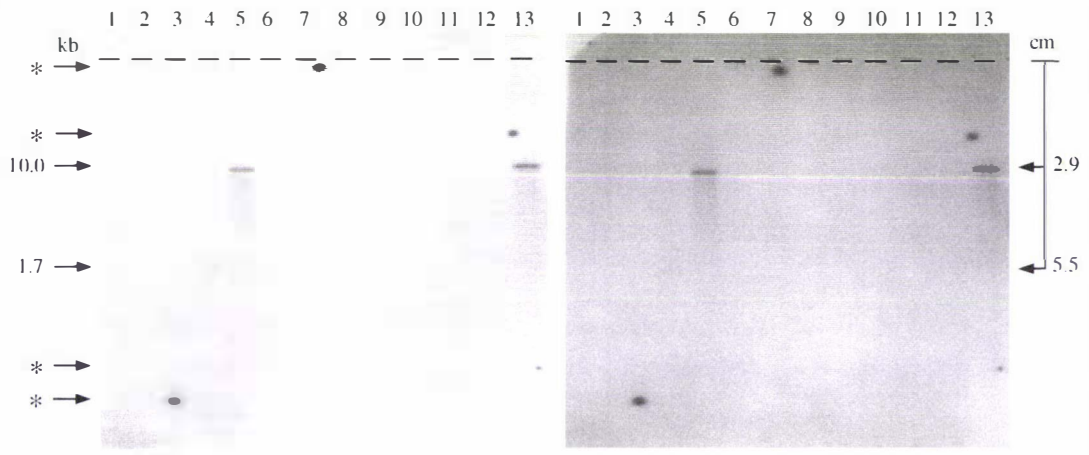


Figure 35: High stringency hybridization of *M. ptb*281-DIG. *Bam* H1 Southern blot #3 was hybridized with *M. ptb*281-DIG using high stringency conditions, subjected to chemiluminescent detection with CDP^{Star} and exposed to film for ½ hr, 3 hrs (left) and overnight (right). Traveling distances of the bands and the size of the genomic fragments they were located on are indicated. Lanes: 1, BCG; 2, *M. tb*; 3, *M. bovis*; 4, *M. intracellulare*; 5, Neoparasec; 6, *M. scrofulaceum*; 7, *M. gordonae*; 8, *M. kansasii*; 9, *M. phlei*; 10, *M. marinum*; 11, *M. terrae*; 12, *M. fortuitum*; 13, *M. ptb*. * = artefact

*M. ptb*281-DIG produced a single hybridization signal in *M. ptb* and its derivative Neoparsec after 3 hr and overnight exposure as expected. The signals were among genomic fragments approximately 10.0 kb as judged by a standard curve prepared from the reference photo (data not shown). In addition, a single faint hybridization signal was observed in *M. intracellulare* among fragments approximately 1.7 kb on the 3 hr and overnight exposures. Since high stringency conditions allow areas of target DNA with a high degree of sequence homology to that of the probe to be detected, these results suggest that *M. intracellulare* carries a single *M. ptb*281 homolog. Conversely, no hybridization signals were observed in any other species on the Southern blot. Another *Bam* H1 Southern blot (*Bam* H1 Southern blot #4, data not shown) was hybridized with *M. ptb*281-DIG under high stringency conditions which shared the same banding pattern as *Bam* H1 Southern blot #3. Spots occurring in lanes 3, 7 and 13 of *Bam* H1 Southern blot #3 were not present on *Bam* H1 Southern blot #4 confirming their presence as artefacts.

3.25.4 Low Stringency Conditions

To detect areas of low identity *Bam* H1 Southern blot #3 was probed with *M. ptb*281-DIG using low stringency conditions as previously described (section 3.4.2). Briefly, the hybridization was performed overnight at 40° C followed by low stringency post-hybridization washes as follows; 2x 5 min in excess 2x SSC, 0.1% SDS at room temperature, 2x 15 min in excess 1x SSC, 0.1% SDS at 68° C and 2x 15 min in excess 0.5x SSC, 0.1% SDS at 68° C. The blot was subjected to chemiluminescent detection with CDP^{Star} and exposed to film for 2 ½ hr and overnight. Both exposures shared the same banding pattern and low background as produced under high stringency conditions (data not shown).

In an attempt to identify sequences with low similarity to the probe the procedure was repeated with the stringency conditions reduced to a 37° C overnight hybridization followed by low stringency post-hybridization washes of 4x 15 min with 5x SSC, 0.1% SDS at room temperature. The blot was subjected to chemiluminescent detection and exposed to X-ray film for 1 hour and overnight (Figure 36). Hybridization signals emerged in the 1 kb^{+} ladder on both exposures among fragments derived from pUC and lambda DNA suggesting these conditions were able to stabilize weak hybridization. These conditions also increased the level of background noise and artifacts making interpretation more difficult. The three hybridization signals present on the high stringency exposures (~1.7 kb *M. intracellulare*, ~10.0 kb *M. ptb* and Neoparasec) were also present under these conditions. The appearance of a second band in *M. intracellulare* among the ~4.9 kb fragments appears at the same point as a horizontal artefact that stretches across lanes 3 to 8 making interpretation difficult. The second band in *M. ptb* among the ~7.5 kb fragments appears to be an artefact poorly represented in the scanned image.

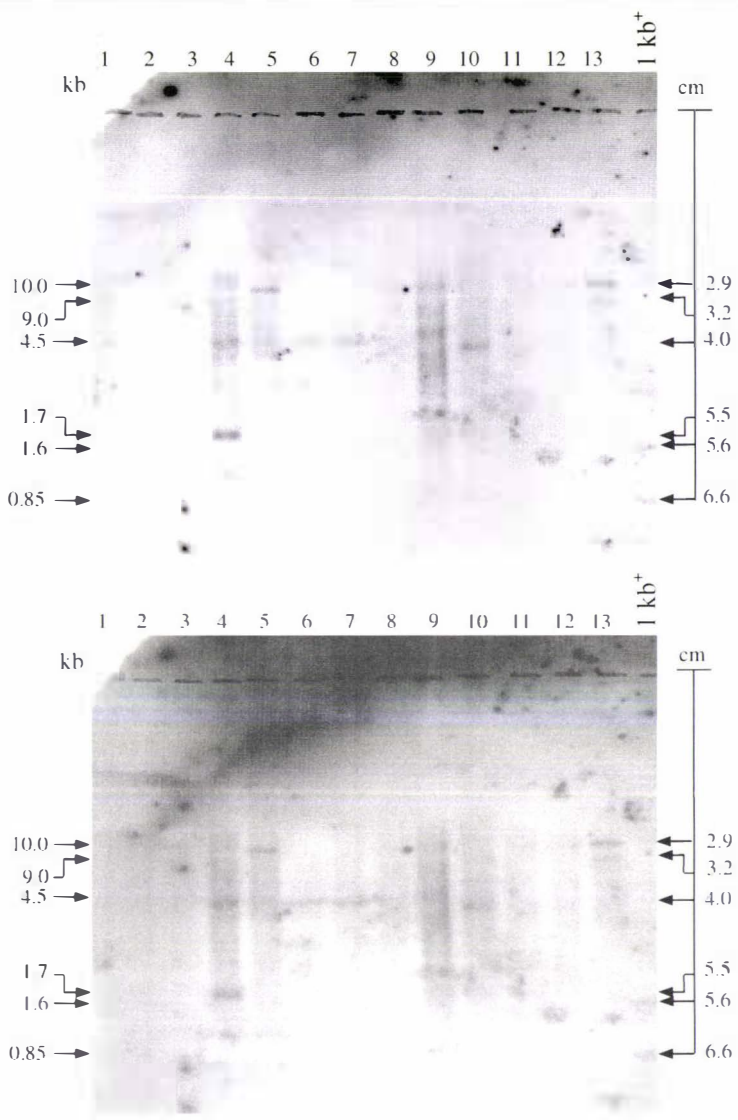


Figure 36: Low stringency hybridization of *M. ptb*281-DIG. *Bam* H1 Southern blot #3 was hybridized with *M. ptb*281-DIG at 37° C and subjected to 4x 15 min washes with 5x SSC, 0.1% SDS at room temperature. The blot was subjected to chemiluminescent detection with CDPStar and exposed to film for 1 hour (upper) and overnight (lower). Approximate traveling distances of the bands and their sizes are indicated. Lanes: 1, BCG; 2, *M. tb*; 3, *M. bovis*; 4, *M. intracellulare*; 5, Neoparasec; 6, *M. scrofulaceum*; 7, *M. gordonae*; 8, *M. kansasii*; 9, *M. phlei*; 10, *M. marinum*; 11, *M. terrae*; 12, *M. fortuitum*; 13, *M. ptb*.

The reference photo shows *M. intracellulare*, *M. phlei* and *M. marinum* had excess genomic DNA in their lanes which may have contributed to the high background in those species. The presence of high background made interpretation of poorly defined

bands in those species difficult. Thick bands of genomic DNA in the high molecular weight fragments of *M. marinum* appear to correspond to faint hybridization signals among fragments ~6.2 kb to ~11 kb on the blot making their interpretation difficult. A strong hybridization signal among ~4.5 kb fragments of *M. marinum* appears to be below the bands of genomic DNA and may correspond to the nucleotide sequence of the *M. marinum* database alignment. Conversely there were no hybridization signals in *M. fortuitum*, *M. terrae*, *M. kansasii*, *M. scrofulaceum*, *M. gordonae*. Of major importance was the absence of bands from the MTB complex representatives BCG, *M. tb* and *M. bovis*.

3.25.5 16S Positive Control

Bam H1 Southern blot #3 was subjected to high stringency hybridization with the 16S-DIG probe to test the ability of all species on the blot to produce specific hybridization signals. High stringency post-hybridization washes were performed and the blot prepared for chemiluminescent detection with CDP^{Star}. After a 1 hr exposure all species on the blot produced 2-4 strong distinct bands confirming the ability of these species to produce specific hybridization signals (Figure 37).

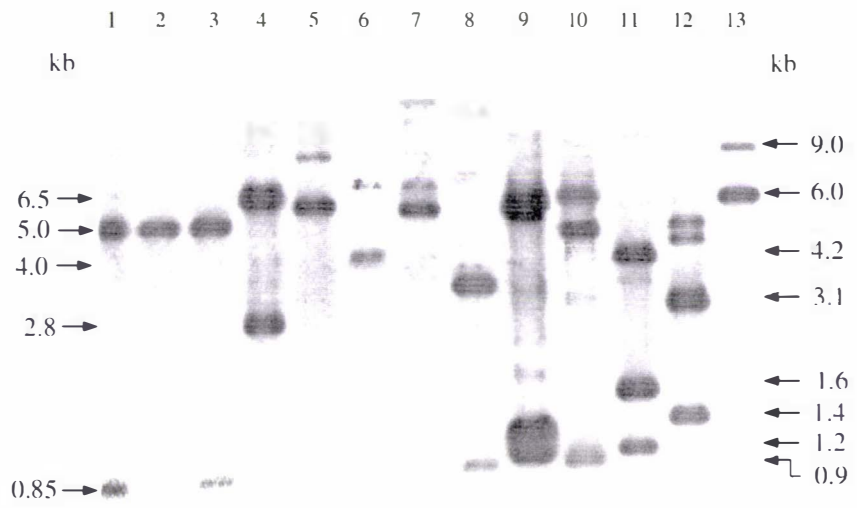


Figure 37: 16S control probe. Multi-*M. sp. Bam* H1 Southern blot #1 was hybridized with 16S-DIG using high stringency conditions, subjected to chemiluminescent detection with CDP^{Star} and exposed to film for 1 hour. Lanes: 1, BCG; 2, *M. tb*; 3, *M. bovis*; 4, *M. intracellulare*; 5, Neoparasec; 6, *M. scrofulaceum*; 7, *M. gordonae*; 8, *M. kansasii*; 9, *M. phlei*; 10, *M. marinum*; 11, *M. terrae*; 12, *M. fortuitum*; 13, *M. ptb*.

The presence of two hybridization signals from most species on the blot was likely due to the conservation of an internal *Bam* H1 site known to occur in the 16S operon of BCG (70). The appearance of more than two hybridization signals could be due to the presence of a second 16S allele notably but not exclusively found in fast growing mycobacteria such as *M. fortuitum* (71) or to an additional *Bam* H1 site.

3.26 PCR Analysis of Selected Mycobacterial Species

Using a combination of PCR anneal temperatures and DNA polymerases, PCR analysis was performed against selected mycobacterial species and *M. ptb* isolates to search for sequences with similarity to *M. ptb*281.

A standard PCR was performed using the cloning primers MPTB281F2 and MPTB281R1 with template from the mycobacterial species. Samples were subjected to 1% agarose gel electrophoresis and examined for the presence of a 273 bp band, the size expected for *M. ptb*281 amplified with these primers (Figure 38).

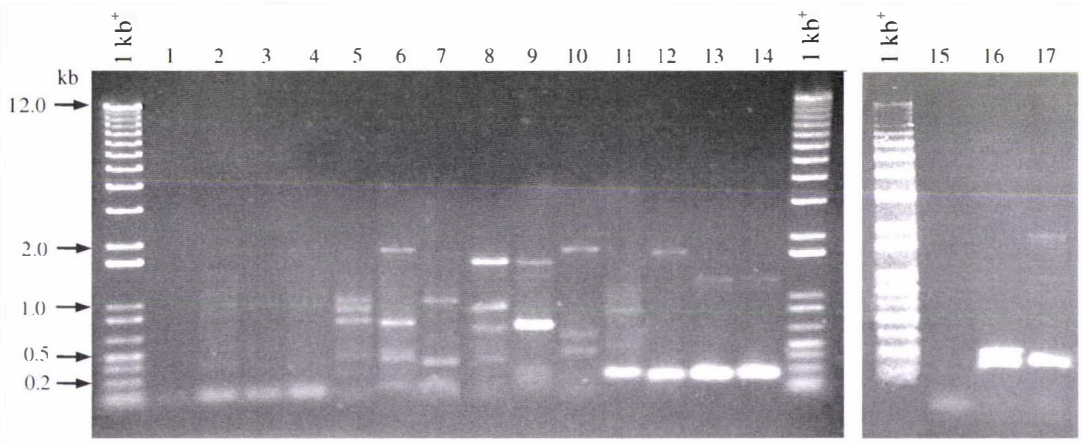


Figure 38: PCR screen of selected mycobacterial species. *M. ptb*281 cloning primers were used in a standard PCR with template from selected mycobacterial species. Aliquots (2 μL) were subjected to 1% agarose gel electrophoresis, soaked in ethidium bromide and visualised with UV light. 1 kb$^+$ = DNA ladder. Lanes: 1, negative control; 2, BCG; 3, *M. bovis*; 4, *M. tb*; 5, *M. fortuitum*; 6, *M. gordonae*; 7, *M. kansasii*; 8, *M. marinum*; 9, *M. terrae*; 10, *M. smegmatis*; 11, *M. intracellulare*; 12 & 17, *M. scrofulaceum*; 13, *M. ptb*; 14, Neoparasec; 15, negative control; 16, *M. phlei*. (16S controls not shown).

The PCR produced strong distinct bands of approximately 270 bp in *M. intracellulare*, *M. scrofulaceum*, *M. ptb*, Neoparasec and *M. phlei* with the latter appearing as two closely migrating bands (doublet).

In addition non-specific bands of varying size were present in most species including a strong distinct ~700 bp band in *M. terrae* and an indistinct faint ~270 bp band in *M. terrae*; results that were typically found in subsequent PCRs using the same conditions (data not shown).

To increase the specificity of the reaction, the PCR was repeated using an annealing temperature of 62° C resulting in a reduction in the level of non-specific bands (Figure 39). These conditions produced strong distinct ~270 bp bands in *M. ptb* and Neoparasec, a distinct ~700 bp band in *M. terrae* and faint ~270 bp bands in *M. intracellulare*, *M. marinum*. *M. terrae* and *M. phlei* the latter being present as a doublet. The faint appearance of the ~270 bp bands in the latter four species suggests there may be significant differences in the nucleotide sequence of the primers to that of the target sequence in those species (Figure 39).

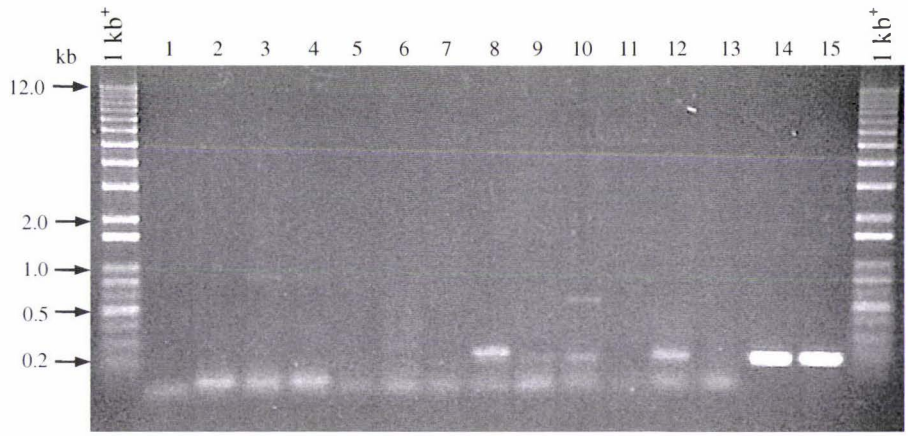


Figure 39: The 62° C PCR screen of selected mycobacterial species. *M. ptb*281 cloning primers were used in a 62° C PCR with template from selected mycobacterial species. Aliquots (2 µL) were subjected to 1% agarose gel electrophoresis, soaked in ethidium bromide and visualised with UV light. 1 kb^+ = DNA ladder. Lanes: 1, negative control; 2, BCG; 3, *M. bovis*; 4, *M. tb*; 5, *M. fortuitum*; 6, *M. gordonae*; 7, *M. kansasii*; 8, *M. phlei*; 9, *M. marinum*; 10, *M. terrae*; 11, *M. smegmatis*; 12, *M. intracellulare*; 13, *M. scrofulaceum*; 14, *M. ptb*; 15, Neoparasec. (16S controls not shown).

The 55° C and 62° C PCRs were repeated using the high fidelity enzyme PLATINUM *Pfx* DNA polymerase instead of *Taq* DNA polymerase. Samples were subjected to agarose gel electrophoresis and examined for the presence of a ~270 bp band (Figure 40). The 55° C *Pfx* PCR produced distinct ~270 bp bands in *M. intracellulare*, *M. scrofulaceum*, *M. ptb*, Neoparasec, and a faint band in *M. phlei*. The 62° C *Pfx* PCR produced distinct ~270 bp bands in *M. intracellulare*, *M. scrofulaceum*, *M. ptb*, Neoparasec, *M. phlei* and *M. kansasii* with the latter two appearing slightly larger than 270 bp. The 55° C and 62° C Pfx DNA polymerase PCRs had high background making interpretation of bands > 270 bp difficult.

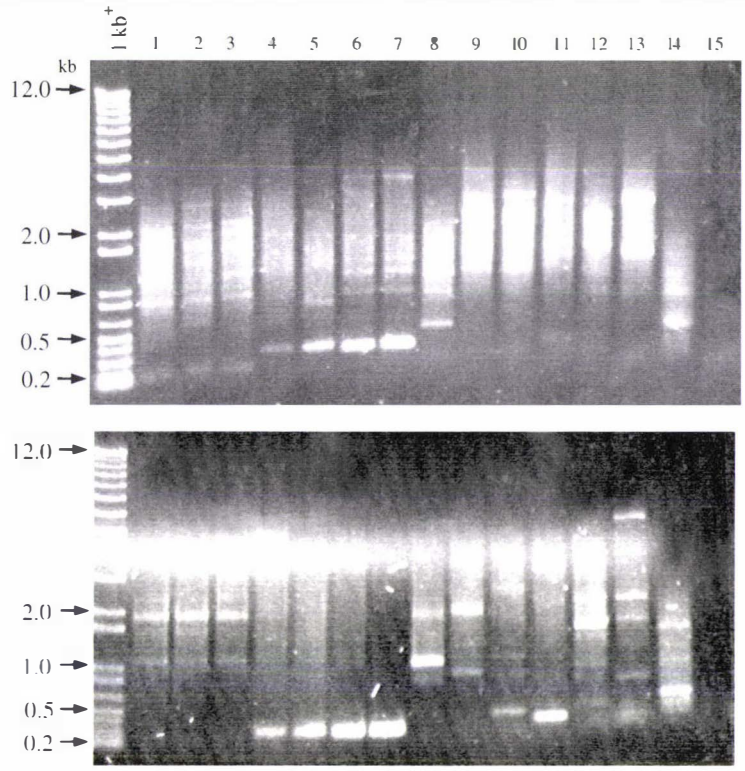


Figure 40: PLATINUM *Pfx* PCR screens. PCRs were performed with PLATINUM *Pfx* using anneal temperatures of 55° C (upper) and 62° C (below) with template from selected mycobacterial species. Samples (3 μL) were electrophoresed on 1% agarose gels, soaked in ethidium bromide and visualised with UV light. 1 kb^+ = DNA ladder. Lanes: 1, BCG; 2, *M. bovis*; 3, *M. tb*; 4, *M. intracellulare*; 5, *M. scrofulaceum*; 6, *M. ptb*; 7, Neoparasec; 8, *M. fortuitum*; 9, *M. gordonae*; 10, *M. kansasii*; 11, *M. phlei*; 12, *M. marinum*; 13, *M. terrae*; 14, *M. smegmatis*; 15, negative control. (Positive 16S controls not shown).

While the ~270 bp bands from these species are approximately the size expected for *M. ptb*281, their true identity can not be assigned until nucleotide sequencing has been performed.

3.27 PCR Analysis of Selected *M. ptb* Isolates

The final PCR was performed to determine the distribution of the hypothetical *M. ptb*281 gene within a group of *M. ptb* isolates collected from cattle, sheep and deer from New Zealand and abroad. Samples of genomic DNA from *M. ptb* isolates donated by Dr

Des Collins (AgResearch, Wallaceville Animal Research Centre) were subjected to a standard PCR with the primers MPTB281F2 and MPTB281R1 (Figure 41).

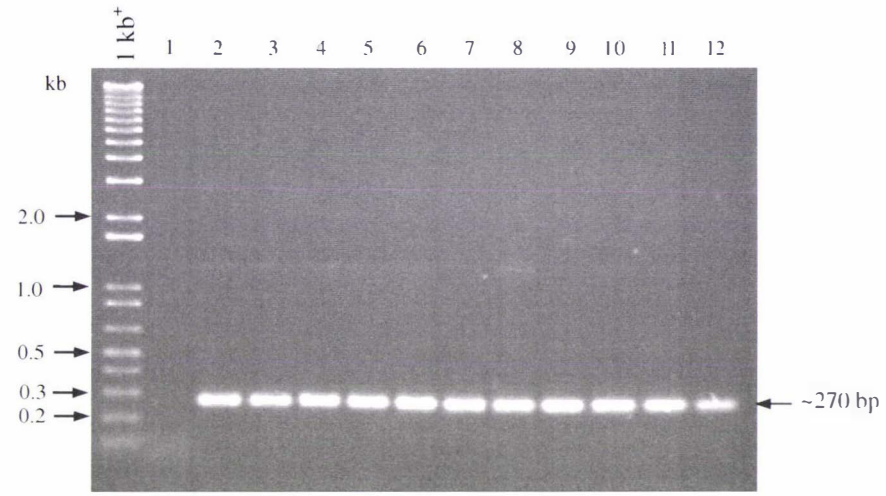


Figure 41: PCR screen of multiple *M. ptb* isolates. Primers MPTB281F2 and MPTB281R1 were used in a standard PCR with templates from multiple *M. ptb* isolates (all from New Zealand unless stated otherwise). Aliquots (2 µL) were subjected to 1% agarose gel electrophoresis, soaked in ethidium bromide and visualised with UV light. 1 kb^+ = DNA ladder. Lanes: 1, negative control; 3, 4 & 5, Deer isolates 6770A, 6771A & 6665; 6, 7, 8, & 9, Sheep isolates 6760A, 6756B, 6758A, & 6759B; 10 & 11, South African sheep isolates 469A & 506C; 12, Cattle isolate 6349; 13, Faeroe Island sheep isolate P133/79. (Neoparasec, *M. ptb* (ATCC 53950) and 16S positive controls not shown).

Aliquots were analysed by 1% agarose gel electrophoresis and examined for the presence of a ~270 bp band, the size expected for *M. ptb*281 (Figure 41). All *M. ptb* isolates were positive for a ~270 bp band.

3.28 Conclusion to Species Distribution

Database search results, Southern blot and PCR analyses all appear to confirm the absence of a *M. ptb*281 homolog from the MTBC members *M. tb*, *M. bovis* and *M. bovis* BCG (Table 2). In addition, no *M. ptb*281 homolog was reported from the *M. leprae* database. *M. ptb*, *M. avium*, *M. marinum* and *M. smegmatis* databases reported significant similarities at both the nucleotide and amino acid levels to *M. ptb*281. This result confirmed the presence of the hypothetical *M. ptb*281 ORF in the *M. ptb* database isolate and indicated the presence of a homolog in *M. avium*, *M. marinum* and *M. smegmatis*.

The *M. marinum* and *M. smegmatis* homologs, reported by database searches to share significant similarities at the nucleotide level to *M. ptb*281, were not evident in PCR analysis. The absence of a ~270 bp band from these species indicates the sequence homology between the primers and their target was insufficient to amplify the homolog under these conditions, a factor that could also be true for other species that were negative for the ~270 bp band.

Database searches revealed *M. ptb*281 homologs in *C. efficiens* and *D. radiodurans* at the amino acid level but not at the nucleotide level indicating these homologs may differ significantly at the nucleotide level from *M. ptb*281.

M. intracellulare, a member of the MAIS complex to which *M. ptb* and *M. avium* both belong, was the only mycobacterial species to give significant results in both Southern blot and PCR analyses. These results suggest that *M. intracellulare* carries a *M. ptb*281 homolog with significant sequence similarity at the nucleotide level, not only within the gene but at its termini as well. Nucleotide sequencing of the ~270 bp band will be required for confirmation of its identity

High background on the low stringency Southern blot exposures and the non-specific banding pattern of the PCRs made assigning significance to some of these results difficult. However, the possibility remains that some of these species may carry *M. ptb*281 homologs that, while not detected due to significant deviations in the nucleotide sequence, share significant similarity at the amino acid level due to codon redundancy.

PCR analysis of multiple *M. ptb* isolates collected from cattle, sheep, and deer from a number of geographical regions were all positive for a ~270 bp band. Nucleotide sequencing of these bands will be required before their identity can be confirmed. These

results indicate *M. ptb*281 is a single copy gene present in members of the MAIS complex, *M. marinum*, *M. smegmatis* and likely other mycobacteria. In addition it is present in *C. efficiens* but not its closest relatives and *D. radiodurans* both non-pathogenic members of the Actinomycetales order to which mycobacteria belong.

Table 2: Summary of *M. ptb*281 species distribution

Species	Database % similarity nuc	Database % similarity aa	Southern blot analysis HS	Southern blot analysis LS	PCR analysis *Taq* 62° C	PCR analysis *Taq* 55° C	PCR analysis *Pfx* 62° C	PCR analysis *Pfx* 55° C
M. tuberculosis	X		x	x	x	x	x	x
M. bovis	X		x	x	x	x	x	x
M. bovis BCG	x		x	x	x	x	x	x
M. leprae	x		-	-	-	-	-	-
M. ptb	100%	100%	✓	✓	✓	✓	✓	✓
M. avium	99%	100%	-	-	-	-	-	-
M. intracellulare	x		✓	✓	✓	✓	✓	✓
M. scrofulaceum	x		x	x	x	✓	✓	✓
M. marinum	80%	89%	x	✓	✓	x	x	x
M. smegmatis	75%	85%	-	-	x	x	x	x
M. phlei	x		x	?	✓	✓	✓	✓
M. terrae	x		x	x	✓	✓	?	x
C. efficiens	-	75%	-	-	-	-	-	-
C. diphtheriae	x		-	-	-	-	-	-
C. glutamicum	x		-	-	-	-	-	-
D. radiodurans	-	51%	-	-	-	-	-	-

Significant results from database searches, Southern analyses and PCR analyses. x = no significant result; ✓ = significant result; - = action not performed; ? = result not interpretable; nuc = nucleotide similarity; aa = amino acid similarity.

Chapter 4: Discussion and Conclusions

4.1 Hypothetical *M. ptb*281 ORF

There are at present no reliable rapid methods suitable for detecting animals in the preclinical stages of Johne's disease in whole herds. Preclinical animals can take 2 to 5 years to develop clinical symptoms. During this time they intermittently shed the bacilli from their digestive tracts, spreading the organism to herd mates through faecal contaminated pastures and drinking water causing serious economic loss to farmers (18). Secreted proteins from pathogenic mycobacteria have been found to be important for the development of protective immunity, namely a CMI response (49). The development of reliable differential diagnostic tests to detect animals in the preclinical stages of Johne's disease will require the use of species-specific secreted protein antigens and the CMI response.

The hypothetical secreted protein MPTB281, from the *M. ptb* secreted protein library clone pJEM11-*M. ptb*281, may be a candidate for use in a CMI based differential diagnostic test or for use in a sub-unit vaccine for the control of Johne's disease. The ORF of *M. ptb*281 was not determined experimentally but was interpreted from the nucleotide sequence of the $PhoA^+$ clone. The genomic insert from pJEM11-*M. ptb*281 contained a significant portion of a *M. ptb* homolog to the *M. avium* catalase gene *katE* (53). The 5' region of the gene was absent from the insert and it's stop codon was 107 bp upstream of and in the wrong reading frame to the reporter gene, all of which made *katE* unlikely to be the gene responsible for the $PhoA^+$ phenotype.

Inverted repeats

A potential rho-independent termination repeat was located downstream of the *M. ptb katE* gene (53). Rho-independent termination repeats form a characteristic hairpin motif in the 3' region of mRNA and have been shown experimentally to stop elongation of the transcript. The typical motif consists of a hairpin structure with a stem length of 4 to 20 nucleotide pairs, a loop of 3 to 10 nucleotide and a poly-U tail of 3 or more 'U's in the 3' region no more than 5 nucleotides downstream of the stem (72). The repeats downstream of *katE* have not been shown experimentally to terminate transcription and do not conform to the typical hairpin motif by having 1 base in the loop and no 3' poly-

U tail. Bioinformatic analysis of complete bacterial genomes led to suggestions that bacteria may employ alternative transcription termination structures to the conventional stem loop poly-U tail (72,73). Devoid of a poly-U tail, the inverted repeats downstream of threonine dehydrogenase (*tdh*) from *Xanthomonas campestris* pv. *campestris* (*X. campestris*), were shown experimentally to exhibit bi-directional transcription termination activity in *E. coli* and *X. campestris*. This demonstrates alternative structures for transcription termination exist in bacteria (74).

Evidence exists that prokaryotic genes can overlap at their termini by 1 or 4 nucleotides, however, it is considered extremely unlikely that complete overlaps between genes occur (75,76). For this reason it is considered highly unlikely that the inverted repeat between *M. ptb katE* and *M. ptb*281 are associated with an ORF on the reverse complement.

Inverted repeats in the 5' untranslated region of mRNA transcripts have been shown experimentally to increase the stability of bacterial transcripts (77,78). Direct experimental evidence has shown that the hairpin in the 5' untranslated region of mRNA gyrase from *M. smegmatis* increased the stability of the transcript significantly under nutrient deprived conditions (79). The stabilization of mRNA transcripts has been proposed as a resource conservation mechanism allowing slow growing pathogenic mycobacteria to survive, especially during a state of dormancy or nutrient deprivation such as might be experienced during infection. Since the inverted repeats between *M. ptb katE* and the hypothetical *M. ptb*281 ORF have not been determined experimentally to function as a transcription terminator, they may be associated with the 5' untranslated region of *M. ptb*281 mRNA, providing stability for the transcript.

Promoter regions

The absence of *E. coli* consensus -10 and -35 promoter elements upstream of the *M. ptb*281 ORF is consistent with other mycobacterial genes (64,65,80-82). Furthermore, mycobacterial promoters are often different in structure from the *E. coli* consensus (83) which may explain the observation that many mycobacterial promoters function poorly in *E. coli* (80,83-85). Bashyam studied randomly isolated *M. tb* and *M. smegmatis* promoters and found the -10 promoters had similarity with the *E. coli* consensus in contrast to the -35 promoter which was highly variable (83).

Promoter elements from *Streptomycetes*, a member of the Actinomycetales order to which mycobacteria belong, are diverse with some having little resemblance to the *E. coli* consensus (86) especially in the -35 region (87). In addition *Streptomycetes sp.* have been described as being better hosts for expression of mycobacterial genes than *E. coli* (85). The sigma subunit of the holoenzyme RNA polymerase, provides most of the promoter recognition factors and initiates transcription (88: Borukhov, 2003 #2368). Multiple sigma factors with different or overlapping specificities have been found in *Streptomyces sp.* which enables them to tolerate the high degree of sequence flexibility in the -35 promoter region (83,89). Eight putative sigma factors have been identified in *M. tb*, of which four have been examined (59). The two constitutively expressed sigma factors (MysA and MysB) found in *M. smegmatis*, *M. tb*, and *M. leprae* corresponding to the genes *mysA* and *mysB* have been cloned and sequenced (90). The regions responsible for recognition of the -10 and -35 promoters, regions 2.4 and 4.2 respectively, from MysA were compared to their equivalent in *Streptomyces* (HrdB) and *E. coli* (RpoD). MysA and HrdB were identical in region 2.4, while RpoD differed by three amino acids. Region 4.2 showed the greatest variability with MysA and HrdB differing from each other by three amino acids while RpoD differed by 14 from both MysA and HrdB (83). Similarities between mycobacteria, *Streptomyces* and their transcriptional mechanisms have led to the assumption that mycobacterial tolerance for diverse -35 promoters is due to multiple sigma factors.

A number of bacterial promoters lacking a -35 region have been found with a 5' extended -10 promoter produced by a TGn motif. This motif has been identified in mycobacteria and shown in vitro to enhance transcription (59,91,92). The significance of a 'TGG' sequence between the 3' region of *katE* and the proposed *M. ptb*281 start codon can not be assigned until mRNA mapping has established the location of promoter and Shine-Delgarno elements on the *M. ptb*281 transcript.

An A+T rich element upstream of the *M. tb katG* promoter has been shown to increase promoter activity. Similar sequences upstream of promoters in *E. coli*, known as UP elements, increase promoter activity by enhancing the initial association of RNA polymerase with the DNA in a manner independent of the sigma factor. An A+T rich sequence of 20 bp has been found upstream of *M. ptb*281 located 5' to the inverted repeat. While all UP elements are A+T rich no consensus sequence has been determined for comparison (59,93). The presence of these nucleotides upstream of the *M. ptb*281 ORF may be unrelated and a subject that only further analysis can determine. It is

possible that other as yet unknown mechanisms regulate the transcription activity of *M. ptb*281.

Signal sequence

Four types of protein secretion pathways have been identified in prokaryotes of which only type II is Sec-dependent requiring an N-terminal signal sequence (63). Examination of the translated *M. ptb*281 ORF did not identify an N-terminal signal sequence or membrane spanning domain expected for export by the Sec dependent pathway.

The small secreted *M. tb* protein ESAT6 has no signal sequence (94). Export of this protein has been suggested to be by the type 1 protein secretion pathway using ABC transporter proteins (95). The presence of putative ABC transporter genes in the genome of *M. tb* (96) and *M. smegmatis* (97), plus the close genealogy of mycobacteria indicates that *M. ptb* could also contain ABC transporter genes. Export by this mechanism is believed to involve a C-terminal signal sequence of ~60 bp preceded by a glycine rich domain. PTB281 has a cluster of glycine residues ~40 bp from its C-terminus although their significance in this context has not been determined.

Type III and IV Sec independent pathways involve large groups of genes for which no homologs have been identified in the genome of *M. tb* (95) making it unlikely that homologs to those genes exist in *M. ptb*.

The small *Staphylococcus aureus* nuclease protein has been used successfully as a reporter gene to identify secreted products from *Lactococcus lactis* (98). This reporter gene system was used in *M. smegmatis* to detect *M. tb* secreted proteins, however, *M. smegmatis* exported the construct with or without a leader peptide suggesting this reporter system should not be used in mycobacteria (95). Results showed that secretion of the nuclease by *M. smegmatis* was not due to leakage or autolysis. This group concluded that correct folding of the nuclease was necessary for efficient export, and suggest its export may be via a redundant secretion pathway or by components of an alternative Sec pathway (95). Extensive mutational studies, however, were unable to identify the mechanism responsible (95). MPTB281 may be exported by such a pathway but until demonstrated otherwise, it is possible that the $PhoA^+$ phenotype of pJEM11-*M. ptb*281 was an artifact, due to *M. smegmatis* interpreting sequences within the insert as elements for transcription, translation and export.

Recombinant MPTB281

Initially, Sauton's and LB cultures of pPROEX-*M. ptb*281, pMIP12-*M. ptb*281 and controls grew well and were used to analyze the cloned inserts and for glycerol stocks. In addition, cultures of pPROEX-*M. ptb*281, pMIP12-*M. ptb*281 and controls were grown to purify and analyze the putative His_6-tagged recombinant proteins. These cultures were the only ones from which the putative His_6-tagged recombinant proteins were recovered, as judged by SDS-PAGE and western blot analyses. All subsequent pPROEX-*M. ptb*281, pMIP12-*M. ptb*281 and controls cultures grew poorly and nickel affinity chromatography failed to enrich for recombinant His_6-tagged proteins. This may have been due to the inclusion of an extra step prior to purification as mentioned in the results section.

The poor and inconsistent growth of cultures was found to be independent of the *M. ptb*281 insert or the bacterial species used, was not due to glassware, its preparation for use, water source or antibiotic stock (data not shown). The presence of contaminants in the stocks used to make these cultures was not examined and may have been the cause of the poor growth. The inability to grow cultures of sufficient quality and quantity prevented the putative His_6-tagged recombinant proteins from being recovered in sufficient quantities for use. Hence, the identity and immunogenicity of the ~15 kDa proteins from the 250 mM imidazole nickel affinity chromatography fractions from pPROEX-*M. ptb*281 and pMIP12-*M. ptb*281 cultures were not established in the time remaining.

4.2 Species distribution

The number of mycobacterial species screened in the work presented here was not comprehensive, however, the *M. ptb*281 ORF appears to be absent from 3 *M. tb* isolates, 2 *M. bovis* isolates and *M. bovis* BCG Pasteur, all members of the MTB complex. Further examination of all MTB complex members (*M. tb*, *M. bovis*, *M. bovis* BCG, *M africanum*, *M. microti*, and *M. canetti*) will be required before the absence of *M. ptb*281 from this group can be confirmed. *M. ptb*281 homologs were present in a number of mycobacterial species including all members of the MAIS complex (*M. avium*, *M. ptb*, *M. intracellulare* and *M. scrofulaceum*), *M. marinum* and *M. smegmatis*. Other suspected *M. ptb*281 homologs identified by PCR or Southern blot analyses have yet to be confirmed by DNA sequencing. Biomolecular methods such as those used here to

detect homologs at the nucleotide level can not identify all homologs that can exist at the amino acid level due to codon redundancy. Hence, *M. ptb*281 homologs may exist at the protein level in species analysed that were not detected by these methods.

C. efficiens and *D. radiodurans* are both non-pathogenic members of the Actomycetes order to which mycobacterial belong. The *M. ptb*281 homologs reported in these species correspond to hypothetical proteins predicted by glimmer version 2.0, a software program designed to identify potential ORFs in bacterial genomes. Previously the presence of hypothetical genes in multiple species has been taken as good evidence that the gene is functional (99) though this may not always be true and has yet to be proven in the case of *M. ptb*281.

4.3 Potential use for MPTB281

While the CMI response for MPTB281 has yet to be determined, the use of a single antigen in an immunological diagnostic test would not provide sufficient sensitivity. This is due to genetic variation between individuals which causes some individuals to react to certain antigens and not to others (100). A number of studies have investigated the use of secreted proteins from *M. tb* as antigens for a DTH test (50,101,102). Results from such tests have shown that a combination of recombinant proteins (protein cocktail) from *M. tb* were able to stimulate a far greater T-cell response than could a single protein antigen. The protein cocktail proved to be highly sensitive and specific compared to conventional PPDs used in DTH tests. Hence, PPD may be replaced by designer protein cocktails as the antigen of choice in future CMI based tests to diagnose *M. tb* infections. A similar strategy could be employed to improve the sensitivity of diagnostic antigens used to detect Johne's disease.

*M. ptb*281 appears to be conserved in a subsection of mycobacteria and in two closely related non-pathogenic species leading one to speculate that *M. ptb*281 may not play a direct role in pathogenesis. Used as a diagnostic antigen, MPTB281 could likely differentiate between MTB and MAIS complex infections but not between members of the MAIS complex. As such MPTB281 could not distinguish between Johne's disease and animals transiently infected with *M. avium*, a species commonly found in the environment. Alternatively, MPTB281 could be used with other antigens absent from the MTB complex as components of a sub-unit vaccine for Johne's disease. A vaccine

such as this would not interfere with bovine tuberculosis herd surveillance which is one limitation of current vaccines.

DNA sequencing of clones from the pJEM11-*M. ptb* secreted protein library and database searches identified five clones with no significant sequence homology to *M. tb* or *M. bovis* [Dupont, 2002 #1567]. Further examination of these clones could identify secreted proteins that may also be absent from *M. avium*. Proteins such as these may have immunological activity and be suitable for use as differential diagnostic antigens.

Completion of the *M. avium* and *M. ptb* genome sequencing projects will greatly enhance the search for diagnostic antigens. Once completed bioinformatic tools could be used to perform subtractive comparative analyses of the *M. ptb*, *M. avium*, *M. tb* and *M. bovis* genomes to identify regions specific to *M. ptb*. Investigation of such regions may reveal the ORF's to *M. ptb* secreted proteins. If immunogenic, these proteins would be useful as components of a differential diagnostic protein cocktail for use in a CMI based diagnostic test for the identification of animals in the preclinical stages of Johne's disease.

4.4 Future Work

Results described here appear to place the hypothetical *M. ptb*281 ORF outside the MTB complex. The effectiveness of this hypothetical protein as a differential diagnostic antigen or component of a sub-unit vaccine will depend on its absence from all MTB complex members and on its ability to stimulate an immune response. Future experiments required to ascertain the suitability of the hypothetical *M. ptb*281 protein for such purposes are described below.

- Determine the species distribution of the *M. ptb*281 ORF across a comprehensive range of mycobacterial and related species (especially MTB complex members). This will be achieved using a combination of database searches (where applicable), PCR and Southern blot analyses.
- Perform DNA sequencing on the ~270 bp bands from the *M. ptb*281 PCR screens of *M. phlei*, *M. intracellulare*, *M. scrofulaceum* and *M. terrae* to determine their degree of sequence homology to *M. ptb*281.
- Perform RT-PCR or primer extension analysis on mRNA isolated from *M. ptb* to confirm the transcription of the ORF and to identify the start of transcription.
- Repeat the M-MPTB281 and X-MPTB281 western blots with the appropriate pPROEX and pMIP nickel column control fractions to ensure the chemiluminescent signals were specific to the His_6-tagged recombinant proteins.
- Perform amino acid sequencing of the recombinant proteins X-MPTB281 and M-MPTB281 to confirm expression of the cloned hypothetical *M. ptb*281 gene. High molecular weight impurities should be removed from both protein samples using size exclusion chromatography and rTEV protease and nickel chromatography should be performed on X-MTPB281 to remove the N-terminal His-tag prior to sequencing.
- Perform western blot analysis on X-MPTB281 and M-MPTB281 using sera from Neoparasec vaccinated and BCG vaccinated animals to determine their ability to stimulate a humoral immune response.
- Use recombinant proteins X-MPTB281 and M-MPTB281 as antigens in a Bovigam assay using whole blood from Neoparsec vaccinated and BCG vaccinated animals to test the ability of these proteins to stimulate a γINF response in T-cells.

- Use recombinant proteins X-MPTB281 and M-MPTB281 as antigens in a cutaneous test of vaccinated and non-vaccinated animals to test the ability of these proteins to produce a DTH reaction.
- Remove the His_6-tag from X-MPTB281 and raise poly-clonal antibodies to it (anti-x-MPTB281). Using suitable controls for bacterial lysis, conduct western blot analyses on culture filtrate and sonicated cellular pellets from Neoparasec and/or *M. ptb* cultures using anti-x-MPTB281 to locate the native protein and assess its status as a secreted or cytosolic protein.
- Use anti-x-MPTB281 to semi purify native MPTB281 from Neoparasec and/or *M. ptb* cultures. Send purified protein for amino acid sequencing to confirm the expression of MPTB281 in vivo.
- Produce a *M. ptb* 281 gene knockout in *M. ptb* and/or Neoparasec and *M. smegmatis* and if viable, test its pathogenic abilities in macrophages from Johne's vaccinated and non-vaccinated animals.
- Perform northern blot analysis on mRNA isolated from *M. ptb* to see if the *M. ptb*281 ORF is transcribed as part of an operon which may give a clue to its function and explain the absence of promoter elements upstream of the ORF.
- Perform quantitative mRNA analyses using a nuclease protection assay or RT-PCR to ascertain the level of *M. ptb*281 transcription during stress conditions such as those that would be experienced during infection. Information from such assays could yield information important to the discovery of the proteins function during infection.

References

1 Collins, M. T. 1997. *Mycobacterium paratuberculosis*: a potential food-borne pathogen? *J Dairy Sci* **80**:3445-8.

2 Clarke, C. J. 1997. Paratuberculosis and molecular biology. *Vet J* **153**:245-7.

3 Cousins, D. V., Whittington, R., Marsh, I., Masters, A., Evans, R. J., and Kluver, P. 1999. Mycobacteria distinct from *Mycobacterium avium subsp. paratuberculosis* isolated from the faeces of ruminants possess IS900-like sequences detectable by IS900 polymerase chain reaction: implications for diagnosis. *Mol Cell Probes* **13**:431-442.

4 Clarke, C. J. 1997. The pathology and pathogenesis of paratuberculosis in ruminants and other species. *J Comp Pathol* **116**:217-61.

5 Cocito, C., Gilot, P., Coene, M., de Kesel, M., Poupart, P., and Vannuffel, P. 1994. Paratuberculosis. *Clin Microbiol Rev* **7**:328-45.

6 Sockett, D. C., Conrad, T. A., Thomas, C. B., and Collins, M. T. 1992. Evaluation of four serological tests for bovine paratuberculosis. *J Clin Microbiol* **30**:1134-9.

7 Muskens, J., Mars, M. H., Elbers, A. R., Van Maanen, K., and Bakker, D. 2003. The results of using faecal culture as confirmation test of paratuberculosis-seropositive dairy cattle. *J Vet Med B Infect Dis Vet Public Health* **50**:231-4.

8 Harris, N. B. and Barletta, R. G. 2001. *Mycobacterium avium subsp. paratuberculosis* in Veterinary Medicine. *Clin Microbiol Rev* **14**:489-512.

9 Stabel, J. R. 1998. Johne's disease: a hidden threat. *J Dairy Sci* **81**:283-8.

10 Van Kruiningen, H. J. 1999. Lack of support for a common etiology in Johne's disease of animals and Crohn's disease in humans. *Inflamm Bowel Dis* **5**:183-91.

11 Travis, S. P. 1995. Mycobacteria on trial: guilty or innocent in the pathogenesis of Crohn's disease? *Eur J Gastroenterol Hepatol* **7**:1173-6.

12 Engstrand, L. 1995. *Mycobacterium paratuberculosis* and Crohn's disease. *Scand J Infect Dis Suppl* **98**:27-9.

13 Sanderson, J. D., Moss, M. T., Tizard, M. L., and Hermon-Taylor, J. 1992. *Mycobacterium paratuberculosis* DNA in Crohn's disease tissue. *Gut* **33**:890-6.

14 Brett, E. 1998. Johne's disease: an economic evaluation of control options for the New Zealand livestock industries. Agriculture New Zealand.

15 Collins, D., Hilbink, F., West, D., Hosie, B., Cooke, M., and de Lisle, G. 1993. Investigation of *Mycobacterium paratuberculosis* in sheep by faecal culture, DNA characterisation and the polymerase chain reaction. *Vet Res* **133**:599-600.

16 Valentin-Weigand, P. and Goethe, R. 1999. Pathogenesis of *Mycobacterium avium subspecies paratuberculosis* infections in ruminants: still more questions than answers. *Microbes Infect* **1**:1121-1127.

17 Orme, I. M. 1988. Induction of nonspecific acquired resistance and delayed-type hypersensitivity, but not specific acquired resistance in mice inoculated with killed mycobacterial vaccines. *Infect Immun* **56**:3310-2.

18 Collins, M. T. 1996. Diagnosis of paratuberculosis. *Vet Clin North Am Food Anim Pract* **12**:357-71.

19 Kiehn, T. E. 1993. The diagnostic mycobacteriology laboratory of the 1990s. *Clin Infect Dis* **17: Suppl 2**:S447-54.

20 Kalis, C. H. J., Hesslink, J. W., and Barkema, H. W. 1999. Comparison of Culture of Individual and Strategically Pooled Bovine Fecal Samples for *Mycobacterium avium subsp. paratuberculosis*. Manning, E. J. B. and Collins, M. T., eds., *Proceedings of the Sixth International Colloquium on Paratuberculosis*, p. 345-349. International Association for Paratuberculosis, Melbourne, Australia.

21 Whittington, R. J., Marsh, I., Turner, M. J., McAllister, S., Choy, E., Eamens, G. J., Marshall, D. J., and Ottaway, S. 1998. Rapid detection of *Mycobacterium paratuberculosis* in clinical samples from ruminants and in spiked environmental samples by modified BACTEC 12B radiometric culture and direct confirmation by IS900 PCR. *J Clin Microbiol* **36**:701-7.

22 Stabel, J. R. 1997. Johne's Disease: A Hidden Threat. *J Dairy Res* **81**:283-288.

23 Abbas, B., Riemann, H., and Behymer, D. E. 1983. Evaluation of the fluorescent antibody test for diagnosis of paratuberculosis. *Am J Vet Res* **44**:720-1.

24 Cousings, D., Whittington, R., Masters, A., Marsh, I., Evens, R., and Kluver, P. 1999. Investigation of False-positives in the IS900 PCR for identification of *Mycobacterium avium subsp. paratuberculosis*. Manning, E. J. B. and Collins, M. T., eds., *Proceedings of the Sixth International Colloquium on Paratuberculosis*, p. 259-264. International Association for Paratuberculosis, Mebourne, Australia.

25 Gwozdz, J., Thompson, K., Murray, A., Reichel, M., Manktelow, B., and West, D. 2000. Comparison of three serological tests and an interferon-gamma assay for

the diagnosis of paratuberculosis in experimentally infected sheep. *Aust Vet J.* **Nov; 78**:779-83.

26 Geysen, H. M., Meloen, R. H., and Barteling, S. J. 1984. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc Natl Acad Sci U S A* **81**:3998-4002.

27 Whitlock, R. H., Wells, S. J., Sweeney, R. W., and Van Tiem, J. 1999. ELISA and fecal culture: sensitivity and specificity of each method. Manning, E. J. B. and Collins, M. T., eds., *Proceedings of the Sixth International Colloquium on Paratuberculosis*, p. 353-362. International Association for Paratuberculosis, Melbourne, Australia.

28 Rothel, J. S., Jonees, S. L., Corner, L. A., Cos, J. C., and Wood, P. R. 1992. The gamma-interferon assay for diagnosis of bovine tuberculosis in cattle: conditions affecting the production of gamma-interferon in whole blood culture. *Aust Vet J* **69**:1-4.

29 Fraser, C. M. and Mays, A. 1986. *The Merck Veterinary Manual*, 6 edn. Merck & Co. Inc., Rahway U.S.A.

30 Kormendy, B. 1988. Diagnostic value of mammalian, avian and johnin PPD tuberculins in cattle herds infected by *Mycobacterium paratuberculosis. Acta Vet. Hung* **36**:177-183.

31 Wayne, L. G. 1982. Microbiology of tubercle bacilli. *Am Rev Respir Dis* **125**:31-41.

32 van Soolingen, D., Hoogenboezem, T., de Haas, P. E., Hermans, P. W., Koedam, M. A., Teppema, K. S., Brennan, P. J., Besra, G. S., Portaels, F., Top, J., Schouls, L. M., and van Embden, J. D. 1997. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol* **47**:1236-45.

33 Dupont, C. and Murray, A. 2001. Identification, cloning and expression of sodC from an alkaline phosphatase gene fusion library of *Mycobacterium avium subspecies paratuberculosis. Microbios* **106**:7-19.

34 Lim, E. M., Rauzier, J., Timm, J., Torrea, G., Murray, A., Gicquel, B., and Portnoi, D. 1995. Identification of *Mycobacterium tuberculosis* DNA sequences encoding exported proteins by using phoA gene fusions. *J Bacteriol* **177**:59-65.

35 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol* **215**:403-10.

36 Pearson, W. R. and Lipman, D. J. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85**:2444-8.

37 Karlin, S. and Altschul, S. F. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc Natl Acad Sci U S A* **87**:2264-8.

38 Nakai, K. and Kanehisa, M. 1991. Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins* **11**:95-110.

39 von Heijne, G. 1985. Signal sequences. The limits of variation. *J Mol Biol* **184**:99-105.

40 Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* **8**:581-99.

41 Nielsen, H. and Krogh, A. 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol* **6**:122-30.

42 Nielsen, H., Brunak, S., and von Heijne, G. 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng* **12**:3-9.

43 Borich, S. M. 1997. A genetic approach to identify *Mycobacterium bovis* exported protein antigens. Massey, Palmerston North.

44 Vogelstein, B. and Gillespie, D. 1979. Preparative and analytical purification of DNA from agarose. *Proc Natl Acad Sci U S A* **76**:615-9.

45 Sauton, B. 1912. Sur la nutrition minérale du bacille tuberculeux. *Comptes rendus hebdomadaires des Séances de l'Académie des Sciences* **155**:860-861.

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

47 Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T., and Jacobs, W. R., Jr. 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**:1911-9.

48 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. 1993. Preparation and analysis of DNA. Chanda, V. B., ed., *Current Protocols in Molecular Biology*. John Wiley & Sons, Massachusetts.

49 Orme, I. M. 1993. Immunity to mycobacteria. *Curr Opin Immunol* **5**:497-502.

50 Weldingh, K. and Andersen, P. 1999. Immunological evaluation of novel *Mycobacterium tuberculosis* culture filtrate proteins. *FEMS Immunol Med Microbiol* **23**:159-64.

51 Derman, A. I. and Beckwith, J. 1991. *Escherichia coli* alkaline phosphatase fails to acquire disulfide bonds when retained in the cytoplasm. *J Bacteriol* **173**:7719-22.

52 Dupont, C. 2002. Identification and characterisation of an exported immunogenic protein of *Mycobacterium avium subspecies paratuberculosis* IVABS. Massey, Palmerston North.

53 Milano, A., De Rossi, E., Gusberti, L., Heym, B., Marone, P., and Riccardi, G. 1996. The katE gene, which encodes the catalase HPII of *Mycobacterium avium*. *Mol Microbiol* **19**:113-23.

54 Timm, J., Perilli, M. G., Duez, C., Trias, J., Orefici, G., Fattorini, L., Amicosante, G., Oratore, A., Joris, B., Frere, J. M., and et al. 1994. Transcription and expression analysis, using lacZ and phoA gene fusions, of *Mycobacterium fortuitum* beta-lactamase genes cloned from a natural isolate and a high-level beta-lactamase producer. *Mol Microbiol* **12**:491-504.

55 Carroll, J. D., Wallace, R. C., Keane, J., Remold, H. G., and Arbeit, R. D. 2000. Identification of *Mycobacterium avium* DNA sequences that encode exported proteins by using phoA gene fusions. *Tuber Lung Dis* **80**:117-30.

56 Engels, W. R. 1993. Contributing software to the internet: the amplify program. *Trends Biochem Sci* **18**:448-50.

57 Terasaka, K., Yamaguchi, R., Matsuo, K., Yamazaki, A., Nagai, S., and Yamada, T. 1989. Complete nucleotide sequence of immunogenic protein MPB70 from *Mycobacterium bovis* BCG. *FEMS Microbiol Lett* **49**:273-6.

58 Murray, A., Moriarty, K. M., and Scott, D. B. 1989. A cloned DNA probe for the detection of *Mycobacterium paratuberculosis*. *NZ Vet J* **37**:47-50.

59 Mulder, M. A., Zappe, H., and Steyn, L. M. 1997. Mycobacterial promoters. *Tuber Lung Dis* **78**:211-23.

60 Bannantine, J. P., Barletta, R. G., Thoen, C. O., and Andrews, R. E., Jr. 1997. Identification of *Mycobacterium paratuberculosis* gene expression signals. *Microbiology* **143**:921-8.

61 von Heijne, G. 1983. Patterns of amino acids near signal-sequence cleavage sites. *Eur J Biochem* **133**:17-21.

62 Watson, M. E. 1984. Compilation of published signal sequences. *Nucleic Acids Res* **12**:5145-64.

63 Fekkes, P. and Driessen, A. J. 1999. Protein targeting to the bacterial cytoplasmic membrane. *Microbiol Mol Biol Rev* **63**:161-73.

64 Hsieh, P. C., Shenoy, B. C., Samols, D., and Phillips, N. F. 1996. Cloning, expression, and characterization of polyphosphate glucokinase from *Mycobacterium tuberculosis*. *J Biol Chem* **271**:4909-15.

65 Mahenthiralingam, E., Draper, P., Davis, E. O., and Colston, M. J. 1993. Cloning and sequencing of the gene which encodes the highly inducible acetamidase of *Mycobacterium smegmatis*. *J Gen Microbiol* **139**:575-83.

66 White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D., Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L., Moffat, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J. J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K. S., Aravind, L., Daly, M. J., Fraser, C. M., and et al. 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286**:1571-7.

67 Nishio, Y., Nakamura, Y., Kawarabayasi, Y., Usuda, Y., Kimura, E., Sugimoto, S., Matsui, K., Yamagishi, A., Kikuchi, H., Ikeo, K., and Gojobori, T. 2003. Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. *Genome Res* **13**:1572-9.

68 Cerdeno-Tarraga, A. M., Efstratiou, A., Dover, L. G., Holden, M. T., Pallen, M., Bentley, S. D., Besra, G. S., Churcher, C., James, K. D., De Zoysa, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M. A., Rabbinowitsch, E., Rutherford, K. M., Thomson, N. R., Unwin, L., Whitehead, S., Barrell, B. G., and Parkhill, J. 2003. The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res* **31**:6516-23.

69 Ikeda, M. and Nakagawa, S. 2003. The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Appl Microbiol Biotechnol* **62**:99-109.

70 Yasuhiko Suzuki, Akihisa Nagata, Yasuko Ono, and Yamada, T. 1988. Complete Nucleotide Sequence of the 16S rRNA Gene of *Mycobacterium bovis* BCG. *J Bacteriol* **170**:2886-2889.

71 Menendez, M. C., Garcia, M. J., Navarro, M. C., Gonzalez-y-Merchand, J. A., Rivera-Gutierrez, S., Garcia-Sanchez, L., and Cox, R. A. 2002. Characterization of an rRNA Operon (*rrnB*) of *Mycobacterium fortuitum* and other mycobacterial species: implications for the classification of mycobacteria. *J Bacteriol* **184**:1078-1088.

72 Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O., and Salzberg, S. L. 2000. Prediction of transcription terminators in bacterial genomes. *J Mol Biol* **301**:27-33.

73 Lillo, F., Basile, S., and Mantegna, R. N. 2002. Comparative genomics study of inverted repeats in bacteria. *Bioinformatics* **18**:971-9.

74 Weng, S. F., Liu, Y. S., Lin, J. W., and Tseng, Y. H. 1997. Transcriptional analysis of the threonine dehydrogenase gene of *Xanthomonas campestris*. *Biochem Biophys Res Commun* **240**:523-9.

75 Lukashin, A. V. and Borodovsky, M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26**:1107-15.

76 Shmatkov, A. M., Melikyan, A. A., Chernousko, F. L., and Borodovsky, M. 1999. Finding prokaryotic genes by the 'frame-by-frame' algorithm: targeting gene starts and overlapping genes. *Bioinformatics* **15**:874-86.

77 Emory, S. A., Bouvet, P., and Belasco, J. G. 1992. A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes Dev* **6**:135-48.

78 Hansen, M. J., Chen, L. H., Fejzo, M. L., and Belasco, J. G. 1994. The ompA 5' untranslated region impedes a major pathway for mRNA degradation in *Escherichia coli*. *Mol Microbiol* **12**:707-16.

79 Unniraman, S., Chatterji, M., and Nagaraja, V. 2002. A hairpin near the 5' end stabilises the DNA gyrase mRNA in *Mycobacterium smegmatis*. *Nucleic Acids Res* **30**:5376-81.

80 Kremer, L., Baulard, A., Estaquier, J., Content, J., Capron, A., and Locht, C. 1995. Analysis of the *Mycobacterium tuberculosis* 85A antigen promoter region. *J Bacteriol* **177**:642-53.

81 Kong, T. H., Coates, A. R., Butcher, P. D., Hickman, C. J., and Shinnick, T. M. 1993. *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. *Proc Natl Acad Sci U S A* **90**:2608-12.

82 Cirillo, J. D., Weisbrod, T. R., Pascopella, L., Bloom, B. R., and Jacobs, W. R., Jr. 1994. Isolation and characterization of the aspartokinase and aspartate

semialdehyde dehydrogenase operon from mycobacteria. *Mol Microbiol* **11**:629-39.

83 Bashyam, M. D., Kaushal, D., Dasgupta, S. K., and Tyagi, A. K. 1996. A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. *J Bacteriol* **178**:4847-53.

84 Clark-Curtiss, J. E., Jacobs, W. R., Docherty, M. A., Ritchie, L. R., and Curtiss, R., 3rd. 1985. Molecular analysis of DNA and construction of genomic libraries of *Mycobacterium leprae*. *J Bacteriol* **161**:1093-102.

85 Kieser, T., Moss, M. T., Dale, J. W., and Hopwood, D. A. 1986. Cloning and expression of *Mycobacterium bovis* BCG DNA in "*Streptomyces lividans*". *J Bacteriol* **168**:72-80.

86 Murali, D., Bashyam, M. B., Kaushal, D., Dasgupta, S. K., and Tyagi, A. K. 1996. A study of the mycobacterial transcriptional apparatus: identification of novel features in promoter elements. *J Bacteriol* **178**:4847-4853.

87 Strohl, W. R. 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Res* **20**:961-74.

88 Gruber, T. M. and Gross, C. A. 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**:441-66.

89 Westpheling, J., Ranes, M., and Losick, R. 1985. RNA polymerase heterogeneity in *Streptomyces coelicolor*. *Nature* **313**:22-7.

90 Predich, M., Doukhan, L., Nair, G., and Smith, I. 1995. Characterization of RNA polymerase and two sigma-factor genes from *Mycobacterium smegmatis*. *Mol Microbiol* **15**:355-66.

91 Bashyam, M. D. and Tyagi, A. K. 1998. Identification and analysis of "extended-10" promoters from mycobacteria. *J Bacteriol* **180**:2568-73.

92 Agarwal, N. and Tyagi, A. K. 2003. Role of 5'-TGN-3' motif in the interaction of mycobacterial RNA polymerase with a promoter of 'extended-10' class. *FEMS Microbiol Lett* **225**:75-83.

93 Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**:1407-13.

94 Sorensen, A. L., Nagai, S., Houen, G., Andersen, P., and Andersen, A. B. 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect Immun* **63**:1710-7.

95 Recchi, C., Rauzier, J., Gicquel, B., and Reyrat, J. M. 2002. Signal-sequence-independent secretion of the staphylococcal nuclease in *Mycobacterium smegmatis*. *Microbiology* **148**:529-36.

96 Braibant, M., Gilot, P., and Content, J. 2000. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev* **24**:449-67.

97 Barsom, E. K. and Hatfull, G. F. 1997. A putative ABC-transport operon of *Mycobacterium smegmatis*. *Gene* **185**:127-32.

98 Poquet, I., Ehrlich, S. D., and Gruss, A. 1998. An export-specific reporter designed for gram-positive bacteria: application to *Lactococcus lactis*. *J Bacteriol* **180**:1904-12.

99 Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**:4636-41.

100 Lyashchenko, K., Colangeli, R., Houde, M., Al Jahdali, H., Menzies, D., and Gennaro, M. L. 1998. Heterogeneous antibody responses in tuberculosis. *Infect Immun* **66**:3936-40.

101 Lyashchenko, K., Manca, C., Colangeli, R., Heijbel, A., Williams, A., and Gennaro, M. L. 1998. Use of *Mycobacterium tuberculosis* complex-specific antigen cocktails for a skin test specific for tuberculosis. *Infect Immun* **66**:3606-10.

102 van Pinxteren, L. A., Ravn, P., Agger, E. M., Pollock, J., and Andersen, P. 2000. Diagnosis of tuberculosis based on the two specific antigens ESAT-6 and CFP10. *Clin Diagn Lab Immunol* **7**:155-60.

103 Boddinghaus, B., Rogall, T., Flohr, T., Blocker, H., and Bottger, E. C. 1990. Detection and identification of mycobacteria by amplification of rRNA. *J Clin Microbiol* **28**:1751-9.

Appendix A: Nucleotide sequence.

pJEM11-*M. ptb*281 insert

(Reading strand)

```
GATCACCCACTTCGACCACGAGCGCATCCCGGAGCGTGTGGTGCATGCGCGCGGGGCCGG   60
CGCCTACGGCTATTTCGAACCGTACGACGACCGGTTGGCGCAGTACACGGCGGCGAAATT  120
TCTGACCTCGCCGGGCACCAGGACGCCGGTGTTCGTGCGTTTTTCGACGGTCGCCGGATC  180
GCGCGGTTCGGCCGACACCGTCCGCGACGTGCGCGGGTTCGCCACCAAGTTCTATACCGA  240
ACAAGGCAATTACGACTTGGTGGGCAACAACTTCCCGGTGTTCTTCATCCAGGACGGCAT  300
CAAGTTCCCCGACTTCGTGCACGCGGTGAAACCCGAGCCGCACAACGAGATTCCGCAGGC  360
GCAGTCGGCGCACGACACGCTGTGGGACTTCGTGTCGCTGCAGCCCGAGACGTTGCACGC  420
CATCATGTGGCTGATGTCGGACCGGGCGCTGCCGCGCAGCTACCGCATGATGCAGGGGTT  480
CGGGGTGCACACCTTCCGGCTGGTGAACGCCCGCGGCCGAGGGACTTTCGTGAAGTTCCA  540
CTGGAAGCCCCGACTCGGCGTGCACTCGCTGATCTGGGACGAATGCCAGAAGATCGCCGG  600
CAAAGACCCCGATTACAACCGCCGCGACCTGTGGGAGGCCATCGAATCCCGCCAGTACCC  660
GGAGTGGGAGCTGGGCGTGCAGCTGGTCGCCGAGGACGACGAGTTCAGCTTCGACTTCGA  720
TCTGCTGGACGCGACGAAAATCATTCCGGAAGAACAGGTTCCGGTATTGCCGGTGGGCAA  780
GATGGTGTTGAACCGCAACCCCGACAACTTCTTCGCCGAGACCGAGCAGGTCGCTTTTCA  840
CACCGCCAACGTGGTGCCGGGCATCGATTTCACCAACGACCCGTTGCTGCAGTTCCGCAA  900
CTTCTCCTATCTGGACACGCAGCTGATCCGGTTGGGCGGCCCCAACTTCGCGCAGCTGCC  960
GGTCAACCGCCCGGTGGCGCAGGTGCGGACCAACCAGCACGACGGTTACGGGCAGCACAC 1020
GATTCCGCAGGGCCGGTCCAGTTACTTCAAGAACAGCATCGGCGGCGGTTGTCCCGCACT 1080
GGCCGACGAGAACGTGTTCCGGCACTACACCCAGCGGGTGGACGGGCAGACGATCGGCAA 1140
GCGCGCCGAGGCGTTCCAGAACCACTACGGCCAGGCGCGGATGTTCTTCAAGAGCATGTC 1200
GCCGGTGGAGGCCGAACACATCGTGGCCGCCTTCGCCTTCGAACTCGGCAAGGTGGAGAT 1260
GCCCGAAATCCGTTCCGCGGTGGTGGCACAACTCGCCCGCGTCGATGACCAGCTGGCCGC 1320
CCAGGTCGCGGCGAAACTGGGGCTGCCCGAGCCGCCGGAGGAGCAGGTGGACGAGTCGGC 1380
ACCGGTTTCCCCGGCCGTTTCGCAGGTCACCGACGGCGGCGACACCATCGCGTCGCGCCG 1440
GATCGCGGTGCTGGCCCCCGACGGCGTCAACGTGGTGGGCACGCACCGCTTCACCGAGCT 1500
GATGGAGCACCCCGGCGCGGTGGTCAAGGTGCTGGCCCCGGTGCCCGGCCGCACGTTGGC 1560
GGGTGGGTCCGCCCGCAAGCTGCGGGTGGACCGGTCCTTCACGACGATGGCGTCGGTGCT 1620
CTACAACGCGGTGGTGGTGGCGTGCGAACCGCGGTCGGTGTCAACGCTGTCCGAACAGCG 1680
GTACGCCGTGCACTTCGTCACCGAGGCCTACAAACACCTCAAGCCGATCGGCGCCTACGG 1740
GGCCGGTGTCGACCTGCTCCGCAAGGCCGGCATCGACAACCGGCTCGCCGAGGACACCGA 1800
CGTGCTCAACGACCAAGCGGTGGTCACCACCAAGGCCGCCGCCGACGAGCTGCCCGAGCG 1860
CTTCGCCGAGGAATTCGCCGCCGCGCTCGCGCAGCACCGGTGCTGGCAGCGGCGCACCGA 1920
CGCGGTGCCGGCCTGAAAGCCGGCCGAAGACCGCCGAAAGGTTTTCCCGGCCGCCGGGAC 1980
GGGCATCCGGCTCAGAAGGCGTCATCGTGGACAGGAGGACAAGTCATGCCGCGCTCGTCG 2040
ATC-phoA                                                        2043
```

TGA = *M. avium katE* stop codon

Brown text = *M. avium katE* gene

ATG = hypothetical *M. ptb*281 start codon

Green text = hypothetical *M. ptb*281 ORF

TGA = hypothetical *M. ptb*281 stop codon

NCBI BLASTN alignment of pJEM11-*M. ptb*281 insert to *M. avium katE*

gi|762828|gb|L41246.1|MSGKATE Mycobacterium avium catalase HPII(katE) gene,
Expect = 0.0, Identities = 2007/2044 (98%),

```
Query: 1    gatcacccacttcgaccacgagcgcatcccggagcgtgtggtgcatgcgcgcggggccgg 60
Sbjct: 321  gatcacccacttcgaccacgagcgcatcccggagcgtgtggtgcatgcgcgcggggccgg 380

Query: 61   cgcctacggctatttcgaaccgtacgacgaccggttggcgcagtacacggcggcgaaatt 120
Sbjct: 381  cgcctacggctatttcgaaccgtacgacgaccggttggcgcagtacacggcggcgaaatt 440

Query: 121  tctgacctcgccgggcaccaggacgccggtgttcgtgcgtttttcgacggtcgccggatc 180
Sbjct: 441  tctgacctcgccgggcaccaggacgccggtgttcgtgcgtttttcgacggtggccggatc 500

Query: 181  gcgcggttcggccgacaccgtccgcgacgtgcgcgggttcgccaccaagttctataccga 240
Sbjct: 501  gcgcggttcggccgacaccgtccgcgacgtgcgcgggttcgccaccaagttctataccga 560

Query: 241  acaaggcaattacgacttggtgggcaacaacttcccggtgttcttcatccaggacggcat 300
Sbjct: 561  acaaggcaattacgacttggtgggcaacaacttcccggtgttcttcatccaggacggcat 620

Query: 301  caagttccccg-acttcgtgcacgcggtgaaacccgagccgcacaacgagattccgcagg 359
Sbjct: 621  caagttccccgtactt-gtgcacgcggtgaaacccgagccgcacaacgagattccgcagg 679

Query: 360  cgcagtcggcgcacgacacgctgtgggacttcgtgtcgctgcagcccgagacgttgcacg 419
Sbjct: 680  cgcagtcggcgcacgacacgctgtgggacttcgtgtcgctgcagcccgagacgttgcacg 739

Query: 420  ccatcatgtggctgatgtcggaccgggcgctgccgcgcagctaccgcatgatgcaggggt 479
Sbjct: 740  ccatcatgtggctgatgtcggaccgggcgctgccgcgcagctaccgcatgatgcaggggt 799

Query: 480  tcggggtgcacaccttccggctggtgaacgcccgcggccgagggactttcgtgaagttcc 539
Sbjct: 800  tcggggtgcacaccttccggctggtgaacgcccgcggccgagggactttcgtgaagttcc 859

Query: 540  actggaagccccgactcggcgtgcactcgctgatctgggacgaatgccagaagatcgccg 599
Sbjct: 860  actggaagccccgactcggcgtgcactcgctgatctgggacgaatgccagaagatcgccg 919

Query: 600  gcaaagaccccgattacaaccgccgcgacctgtgggaggccatcgaatcccgccagtacc 659
Sbjct: 920  gcaaagaccccgattacaaccgccgcgacctgtgggaggccatcgaatccggccagtacc 979

Query: 660  cggagtgggagctgggcgtgcagctggtcgccgaggacgacgagttcagcttcgacttcg 719
Sbjct: 980  cggagtgggagctgggcgtgcagctggtcgccgaggacgacgagttcagcttcgacttcg 1039

Query: 720  atctgctggacgcgacgaaaatcattccggaagaacaggttccggtattgccggtgggca 779
Sbjct: 1040 acctgctggacgcgacgaaaatcattccggaagaacaggttccggtattgccggtgggca 1099

Query: 780  agatggtgttgaaccgcaaccccgacaacttcttcgccgagaccgagcaggtcgctttc 839
Sbjct: 1100 agatggtgttgaaccgcaaccccgacaacttcttcgccgagaccgagcaggtcgcgttcc 1159

Query: 840  acaccgccaacgtggtgccgggcatcgatttcaccaacgacccgttgctgcagtttcgca 899
Sbjct: 1160 acaccgccaacgtggtgccgggcatcgatttcaccaacgacccgttgctgcagtttcgca 1219

Query: 900  acttctcctatctggacacgcagctgatccggttgggcggccccaacttcgcgcagctgc 959
Sbjct: 1220 acttctcctacctggacacgcagctgatccggttgggcggccccaacttcgcgcagctgc 1279

Query: 960  cggtcaaccgcccggtggcgcaggtgcggaccaaccagcacgacggttacgggcagcaca 1019
Sbjct: 1280 cggtcaaccgcccggtggcgcaggtgcggaccaaccagcacgacggttacgcgcagcacg 1339
```

The alignment of pJEM11- *M. ptb*281 insert sequence to NCBI subject *M. avium katE*. Query = *M. ptb* query sequence. Sbjct = (subject) *M. avium katE* sequence from the NCBI database. Bold text = *M. avium katE*. Continued over page.

Bold text = *M. avium katE*; **tga** = 3' region of *katE*

Inverted arrows below text indicate the location of a potential rho-independent termination repeat.

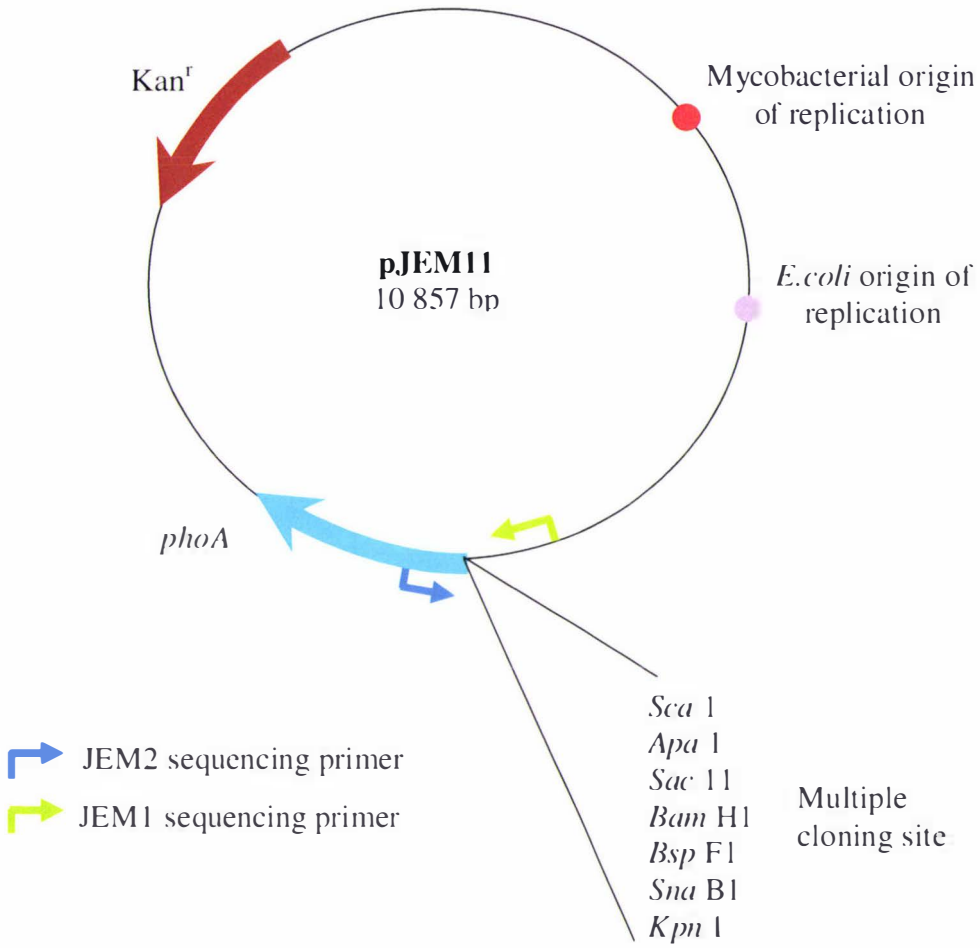
M. avium contig 14.

```
GCGGCCCCAACTTCGCGCAGCTGCCGGTCAACCGCCCGGTGGCGCAGGTGCGGACCAACC     60
AGCACGACGGTTACGGGCAGCACGCGATTCCGCAGGGCCGGTCCAGTTACTTCAAGAACA    120
GCATCGGCGGCGGTTGTCCCGCACTGGCCGACGAGGACGTGTTCCGGCACTACACCCAGC    180
GGGTGGACGGGCAGACGATGCGCAAGCGCGCCGAGGCGTTCCAGAACCACTACGGCCAGG    240
CGCGGATGTTCTTCAAGAGCATGTCGCCGGTGGAGGCCGAACACATCGTGGCTGCCTTCG    300
CCTTCGAACTCGGCAAGGTGGAGATGCCCGAAATCCGTTCCGCGGTGGTGGCACAACTCG    360
CCCGCGTCGATGACCAGCTGGCCGCCCAGGTCGCGGCGAAACTGGGGCTGCCCGAGCCGC    420
CCGAGGAGCAGGTGGACGAATCGGCACCGGTTTCCCCGGCGCTTTCGCAGGTCACCGACG    480
GCGGCGACACCATCGCGTCGCGCCGGATCGCGGTGCTGGCCGCCGACGGCGTCGACGTGG    540
TGGGCACGCAACGCTTCACCGAGCTGATGGAGCAGCGCGGCGCGGTGGTCGAGGTGCTGG    600
CCCCGGTGGCCGGCGGCACGCTGGCGGGTGGGTCCGGCGGCGAGCTGCGGGTGGACCGGT    660
CCTTCACGACGATGGCGTCGGTGCTCTACGACGCGGTGGTGGTGGCGTGCGGACCGCGGT    720
CGGTGTCGACGCTGTCCGACGACGGCTACGCCGTGCACTTCGTCACCGAGGCCTACAAAC    780
ACCTCAAGCCGATCGGCGCCTACGGGGCCGGTGTCGACCTGCTCCGCAAGGCCGGCATCG    840
GCAACCGGCTCGCCGAGGACACCGACGTGCTCAACGACCAAGCGGTGGTCACCACCAAGG    900
CCGCCGCCGACGAGCTGCCCGAGCGCTTCGCCGAGGAATTCGCCGCCGCGCTCGCGCAGC    960
ACCGGTGCTGGCAGCGGCGCACCGACGCGGTGCCGGCCTGAAAGCCGGCCGAAGACCGCC   1020
AAAAGGTTTTCCCGGCCGCCGGGACGGGCATCCGGCTCAGAAGGCGTCATCGTGGACAGG   1080
AGGACAAGTCATGCCGCGCTCGTCGATCAAGAACGAAAAGATGTATCAGGATCTGCGCAA   1140
GAAGGGCGAATCCAAGGAGAAGGCCGCGCGCATCTCCAATGCGGCCGCCGGCCAAGGCAA   1200
GTCGTCGGTGGGCCGCCGCGGCGGCAAGTCCGGGTCCTATCAGGACTGGACCGTGCCGGA   1260
ATTGAAGAAGCGGGCCAAAGAGCTTGGCATTTCCGGCTATTCGGGCCTGACCAAGGACAA   1320
GCTGGTCGCCAAACTGCGCAACCACTGATCCGTCATCTCGTCAACCGCAGTCGTTCGGCC   1380
AGCACCAGGAAGGCGACCGCGTAGTCGTTGTCGGCGGTCTGAGCCCGGATTTCGTCCCAG   1440
TCCAGCCGTTCGCGCACGGCGCGCACCGCGGGCAGCAGCTTGGCGAAGTCGCAGTGGTGC   1500
TCGCCCAGCGAGCGCAGCTGCTGGACGAGCACCATGGTGGGTGGCAGCACCGGCATGCCG   1560
ATCGCCAGCACGTCGTGGTGCTCGGCGCGGTCCAGGGTCTGCGCATCGACGGGCACGCCG   1620
TTGAGGCGGTGCAGCACATCGACCCGCATGTCGCCGGTGCGGGCCTTGAAGAGCCAGTCC   1680
TCCGGCGGGCGTTCGATTTCGAACCCGGCGCTGGCCAGGGTGGTCACCGCGGTCTCCACG   1740
TCCGCCTCGGCCACCACCAGGTCGGCGTCGTGGCTGGGTTCCGGTGCGCCGTAAGCCCAT   1800
AACGCGTAGCTGCCGGCCAGGGCGAAACGCGGGCCCTTCTCCTTGAGCGCCGATGCCGCC   1860
GCGCGCAGCGCCTCCCGCAGCCGCGGGGCGGCCGCGGCGGTGTCGGCGCGCGCGTCGCCG   1920
GATGTCATGGTGCATAGTCATACCCCGCCCGCAGGGCGGCCAACCCGTTGACCCCTGGTC   1980
ACGGCGCCGCGCGGTTGTCAAGACCGTTTAACTTGCGGCCGATTGGGAAATCGTGGGTTA   2040
TGTCAACTCCTGTGCCATGCGCCTATGATGCTGTCGCCGTGGTGATTCCCGCGCACAACG   2100
AGGCAGCC                                                       2108
```

Appendix B: Plasmid Maps.

pJEM11.

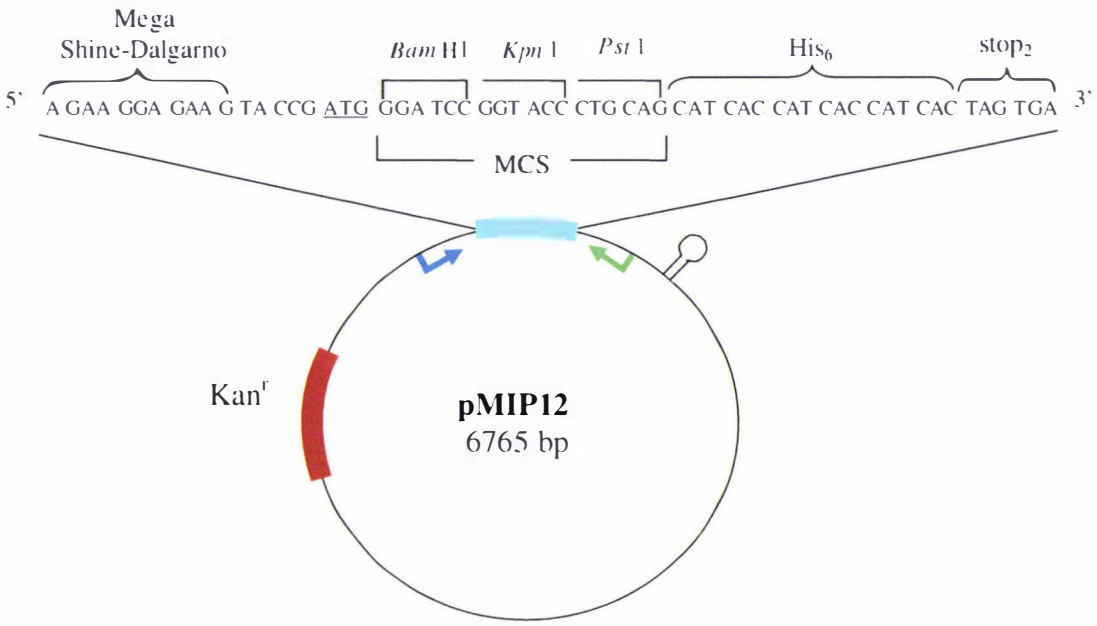
E. coli – mycobacterial shuttle vector.



Modified from Lim, *et al* (34).

pMIP12

Mycobacterial expression vector



Mega Shine-Dalgarno = ribosome binding site

ATG = pMIP start codon

MCS = multiple cloning site

His_6 = poly-histidine affinity tag

$stop_2$ = two stop codons

italics = restriction endonuclease sites

BlaF2 sequencing primer

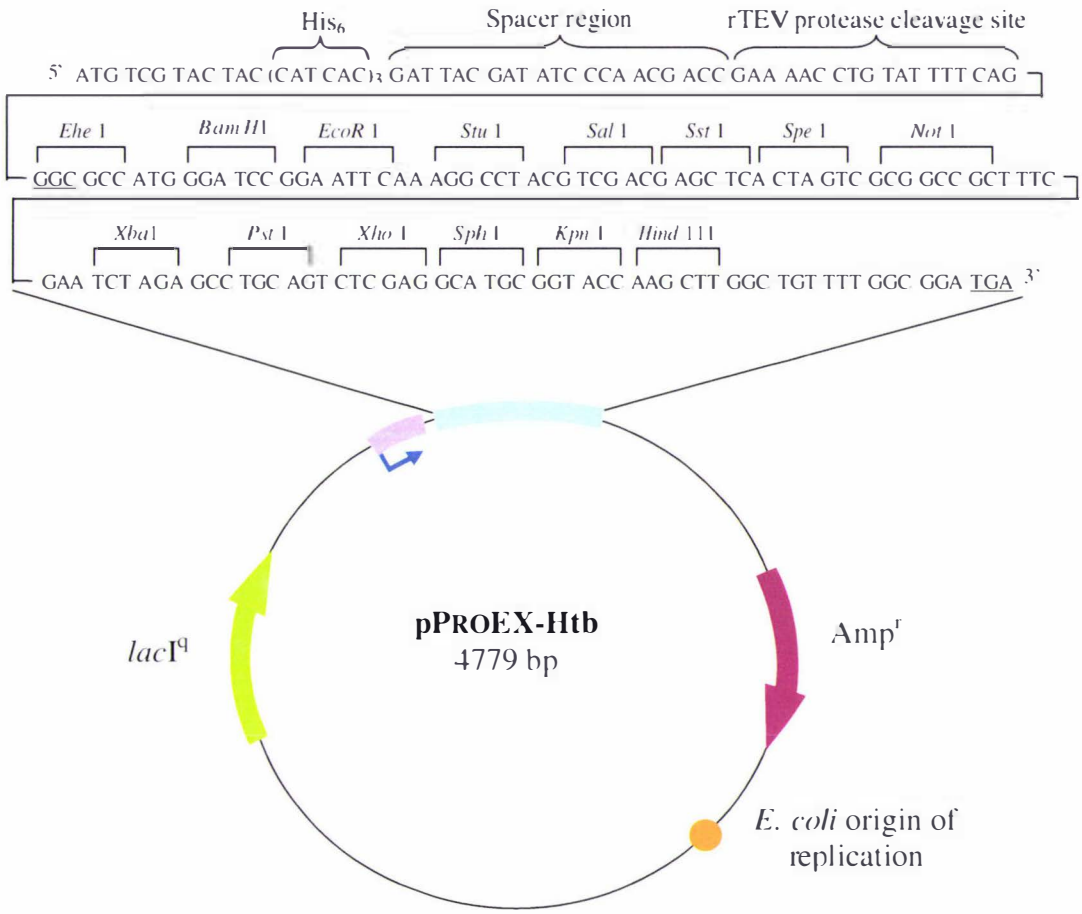
TermR2 sequencing primer

ESAT 6 transcription terminator

Modified from Pasteur Institute information.

pPROEX-Htb

E. coli Expression vector.



ATG = pPROEX start codon

His6 = poly-histidine affinity tag

TGA = pPROEX stop codon

ribosome binding site

M13/pUC sequencing primer

italics = restriction endonuclease sites

Modified from Invitrogen product information booklet.

Appendix C: Primers.

pJEM11 sequencing primers.

Primer	Sequence (5'-3')	Reference
JEM1	CGA GCT GCAGTG GGA TGA CC	Gift of Professor Brigitte Gicquel, Pasteur Institute
JEM2	TCG CCC TGA GCA GCC CGG TT	"

pJEM11-*M. ptb*281 internal sequencing primers

Primer	Sequence (5'-3')	Reference
KatE-i1	GGGCCAGCACCTTGACC	(52)
KatE-i2	GACCGGCCCTGCGGAATCGTGT	"
KatE-i3	GGCTTCCAGTGGAACTTCAC	"
KatE-i4	CGCTCGTCGTACGGTTC	"

pPROEX-Htb sequencing primer

Primer	Sequence (5'-3')	Reference
M13/pUC	AGCGGATAACAATTTCACACAGG	Invitrogen Inc, USA

pMIP12 sequencing primers

Primer	Sequence (5'-3')	Reference
BlaF3	TCGCGGGACTACGGTGCC	Gift of Professor Brigitte Gicquel, Pasteur Institute
TermR2	TCGAACTCGCCCGATCCC	"

***M. ptb*281 primers.**

Primer	Sequence (5'-3')	Reference
MPTB281F1	GCT TCA CCG AGC TGA TGG AG	This study
*MPTB281F2	GGA GGA TCC ATG CCG CGC TCG TCG AT	"
*MPTB281R1	GAC GGT ACC GTG GTT GCG CAG TTT GG	"

*Endonuclease restriction sites underlined.

MPB70 primers.

Primer	Sequence (5'-3')	Reference
MPB71	GAA CAA TCC GGA GTT GAC AA	(57)
MPB72	AGC ACG CTG TCA ATC ATG TA	"

MPB70 primers MPB71/MPB72 produce a 396 bp DNA fragment.

16S primers.

Primer	Sequence (5'-3')	Reference
16S 246	AGAGTTTGATCCTGGCTCAG	(103)
16S 264	TGCACACAGGCCACAAGGGA	"

16S primers 246/264 product a 1030 bp DNA fragment.