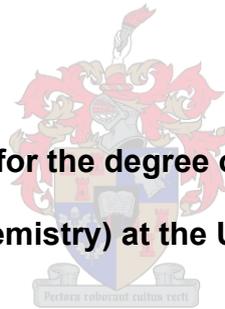


THE TRAFFICKING OF THE *MYCOBACTERIUM TUBERCULOSIS* PE AND PPE PROTEINS

Phetole Walter Mahasha

**Thesis presented for approval for the degree of Masters of Science in Medical
Sciences (Medical Biochemistry) at the University of Stellenbosch.**



**Promoters: Dr. N.C. Gey van Pittius
and Prof. R. M. Warren**

University of Stellenbosch

December 2007

DECLARATION

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

.....

P.W. Mahasha (Student number: 14452715)

.....

Date

SUMMARY

The expansion of the *Mycobacterium tuberculosis* PE and PPE gene families seems to be linked to that of the immunologically-important ESAT-6 (*esx*) gene clusters secretion system, as the ancestral members of these families are found only within the ESAT-6 gene cluster regions. These ancestral members are also the only copies in the earlier mycobacteria like *M. smegmatis*. The later duplications of the PE and PPE families belonging to the PGRS and MPTR subgroups, have been implicated in virulence and are only found within the genomes of the pathogenic mycobacteria closely related to the *M. tuberculosis* complex. The aim of this study was to compare the subcellular localization of the later duplications of the PE and PPE gene families belonging to the PGRS and MPTR subgroups with that of the ancestral PE and PPE proteins found in *M. smegmatis* and to investigate whether the ESX secretion apparatus is involved in the trafficking of these proteins. The PE (Rv3872) and PPE (Rv3873) genes from *M. smegmatis* were PCR amplified with a C-terminal HA tag using *M. smegmatis* genomic DNA as template. Two PPE-MPTR genes, Rv0442c and Rv0878c, and one PE_PGRS gene, Rv2615c, were also PCR amplified with a C-terminal HA tag using *M. tuberculosis* genomic DNA as template. All genes were cloned into the mycobacterial expression vector p19Kpro. Expression and localization was investigated using SDS-PAGE and Western blotting. The PE and PPE genes expressed in *M. smegmatis* were found to be present within the cell wall, membrane, and cytosol fractions, but not in the culture filtrate, indicating no secretion. The PPE-MPTR and PE_PGRS genes expressed in *M. smegmatis*, were also found to be present within the cell wall, membrane and cytosol fractions, but not in the culture filtrate, indicating that they are also not secreted. We hypothesize that their secretion is dependent on ESAT-6 gene cluster region 5, which is absent from the genome of *M. smegmatis*. Ancestral PE and PPE proteins are secreted efficiently in *M. tuberculosis*. The ESAT-6 gene cluster Region 3 and Region 4 of *M. smegmatis* were knocked out, and these knockout mutants could be used in future studies to investigate if the ESAT-6 gene cluster region 1 is involved in the secretion of the ancestral and recent PE and PPE proteins.

OPSOMMING

Die duplisering en uitbreiding van die *Mycobacterium tuberculosis* PE en PPE geenfamilies bly gekoppel te wees aan die van die immunologies-belangrike ESAT-6 (*esx*) geengroep sekresiesisteen, aangesien die vroegste lede van hierdie families slegs aangetref word in die ESAT-6 geengroepe. Hierdie vroegste lede is ook die enigste kopie wat in die genome van die vroeër mikobakterie soos *M. smegmatis* voorkom. Die latere duplisering van die PE en PPE geenfamilies, wat aan die PGRS en MPTR subgroepe behoort, word geimpliseer in virulensie en word slegs aangetref in die genome van die patogeeniese mikobakterie naby-verwant aan die *M. tuberculosis* kompleks. Die doel van hierdie studie was om die subcellulere lokalisering van die latere duplisering van die PE en PPE geenfamilies (behorende tot die PGRS en MPTR subgroepe) te vergelyk met die van die vroegste lede van die PE en PPE geenfamilies wat in *M. smegmatis* voorkom en om te bepaal of die ESX sekresie aparate betrokke is by die lokalisering van hierdie proteïene. Die PE (Rv3872) en PPE (Rv3873) gene van *M. smegmatis* is geamplifiseer met 'n C-terminale HA merker vanaf *M. smegmatis* genomiese DNA deur middel van PCR. Twee PPE-MPTR gene, Rv0442c en Rv0878c, en een PE_PGRS geen, Rv2615c, is ook geamplifiseer met 'n C-terminale HA merker vanaf *M. tuberculosis* genomiese DNA deur middel van PCR. Alle geamplifiseerde gene is in die mikobakteriese uitdrukkingsvektor p19Kpro gekloneer. Uitdrukking en lokalisering is ondersoek deur van SDS-PAGE en Western oordrag gebruik te maak. Die PE en PPE gene wat in *M. smegmatis* uitgedruk is, is aangetoon in die selwand, membraan en sitosol fraksies, maar nie in die kultuurfiltrate nie, wat dui op 'n afwesigheid van sekresie. Die PPE-MPTR en PE_PGRS gene wat in *M. smegmatis* uitgedruk is, is ook aangetoon in die selwand, membraan en sitosol fraksies, maar nie in die kultuurfiltrate nie, wat ook dui op 'n afwesigheid van sekresie. Ons hipotiseer dat die sekresie van hierdie proteïene afhanklik is van ESAT-6 geengroep 5, wat afwesig is van die genoom van *M. smegmatis*. Die vroegste PE en PPE proteïene word effektief gesekreter in *M. tuberculosis*. Die ESAT-6 geengroepe 3 en 4 van *M. smegmatis* is uitgeslaan, en hierdie uitslaanmutante kan gebruik word in toekomstige studies om te ondersoek of die ESAT-6 geengroep 1 betrokke is by die sekresie van die vroegste en mees onlangse PE en PPE proteïene.

DEDICATIONS.

To my mother for being the best mom ever and making sure that I become the best I can be, for the support she gave me all my life, thank you mother, I love you so much.

To my family members (grand mother, Emily, Mmakoma, Tshepo, Theo, Edward, Aggrey, Maria, Mosie, Mapula, Joel, Khutso, Sello, Masuthane, Nkomi, Bongane, Collen) and others for sacrificing their time which we should have shared together and believed in me that I can do it.

To my Love Nommy who supported me throughout, understood that what I was doing was for us and our children, Love you Honey always and forever.

To my beautiful daughter Bonolo and son Tebogo, for being the best gifts that God has given me, daddy for sure will make up for lost times my little angels.

To my grand mother who passed way, Mmatsie Maenetja you will always be in my thoughts grandma, dedicate this work to you, till we meet again.

ACKNOWLEDGEMENTS.

I would like to thank the Almighty God for protecting me against all odds and for His love and grace he had all my life, to you I will forever be grateful.

My sincere gratitude to my supervisors/promoters Dr Gey van Pittius and Prof. Rob Warren for their guidance and support through this journey of my Life in Science.

All the people in Lab 424 and my dear friends who were always encouraging me and supporting me through out the years.

The Medical Research Council of South Africa and National Research Foundation of South Africa for their financial support.

TABLE OF CONTENTS.

	Page numbers
Declaration	ii
Summary	iii
Opsomming	iv
Dedications	v
Acknowledgements	vi
Table of contents	vii-xi
List of abbreviations	x-xii
List of figures	xiii-xiv
CHAPTER 1: BACKGROUND	1
1.1. Tuberculosis (TB)	1-2
1.2. The History of tuberculosis	2-3
1.3. The genome of <i>M. tuberculosis</i>	3-4
1.4. Region of difference (RD1)	4-5
1.5. Secreted antigens-- the ESAT-6 gene family	5
1.6. PE and PPE	5-13
1.7. Targeted gene knock-out	14-15
1.8. Problem statement	15
1.9. Hypothesis	15
1.10. Study Aims and Design	16
CHAPTER 2: MATERIALS AND METHODS	17
2.1. DNA and protein sequence analyses	17
2.2. Bacterial strains	17-18
2.3. Media and Culture conditions	18-19
2.4. Primers and plasmid vectors	19
2.5. PCR amplification and conditions	19
2.6. Molecular cloning and purification of <i>M. tuberculosis</i> and <i>M. smegmatis</i> genes	23-24

2.7. Isolation of cosmid containing selected genomic regions	26-27
2.8. Transformations in <i>M. smegmatis</i>	27
2.9. Ziehl-Nielsen (ZN) Staining	28
2.10. Subfractionation for Protein secretion analyses	28-29
2.11. Bio-Rad Protein assay	30
2.12. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), and Western blotting and detection.	31
2.12.1. SDS-PAGE	31-32
2.12.2. Western Blotting and detection.	32-33
2.13. Deletion of ESAT-6 gene cluster Region 3 (Δ R3) and Region 4 (Δ R4) of <i>M. smegmatis</i> and molecular cloning of the constructs.	33-35
2.14. Delivery of constructs into <i>M. smegmatis</i> .	35-36
2.15. Genomic DNA preparation using CTAB Minipreps.	36
 CHAPTER 3: RESULTS	 37
3.1. The localization of the most ancestral PE and PPE proteins found in <i>M. smegmatis</i> .	38
3.1.1. PCR amplifications	38-39
3.1.2. Isolation of cosmids containing selected genomic regions	39-41
3.1.3. Secretion analyses of the PE and PPE proteins in <i>M. smegmatis</i> .	42
3.2. Cellular localization of the <i>M. tuberculosis</i> PGRS and MPTR proteins in <i>M. smegmatis</i> .	43
3.2.1. PCR amplifications	43
3.2.2. Secretion analyses of the PE_PGRS and PPE_MPTR proteins in <i>M. smegmatis</i> .	43-45
3.3. Knockout of ESAT-6 gene clusters region 3 and region 4 in <i>M. smegmatis</i> .	45
3.3.1. PCR amplifications	45-46
3.3.2. Isolation of mutants.	46-49
 CHAPTER 4: DISCUSSION AND CONCLUSION	 50-52
 CHAPTER 5: FUTURE DIRECTIONS	 53

CHAPTER 6: ADDENDUM A TO E	54
6.1. Addendum A	54-59
6.2. Addendum B	60-62
6.3. Addendum C	63-65
6.4. Addendum D	66-67
6.5. Addendum E	67
CHAPTER 7: LIST OF REFERENCES	68-77

LIST OF ABBREVIATIONS

%	percentage
µg	microgram
µg/ml	micrograms per milliliter
µl	microliter
µM	micromolar
ADC	albumin-dextrose catalase
Amp	Ampicillin
APS	ammonium persulphate
BCG	Bacille de Calmette et Guerin
bp	basepair
BSA	bovin serum albumin
CF	culture filtrate
CFP-10	culture filtrate protein 10
CTAB	centryltrimethyl-ammonium bromide
CW	cell wall
CWL	cell whole lysate
CYT	cytosol
DCO	double crossover
dH₂O	distilled water
DNA	deoxyribonucleic acid
E.	Escherichiae
EDTA	di-sodium ethylenediaminetetra-acetate
EOH	ethanol
ESAT-6	6 kDa early-secreted antigenic target
EtBr	Ethidium bromide
g	gram
g/l	gram per liter
G+C	guanine + cytosine
HA	hemagglutinin
HCl	Hydrochloric acid
HPLC	high performance liquid chromatography

HRPO	horseradish peroxidise
Hyg	Hygromycin
IPTG	Isopropylthio-β-D-galactoside
ISOH	isopropanol
Kan	Kanamycin
kbp	kilobase pairs
KCl	potassium chloride
kDa	kilodaltons
KH₂PO₄	potassium dihydrogen phosphate
KH₂PO₄	potassium di-hydrogen phosphate
L	liter
LB	Luria-Bertani (medium)
M	molar
M.	Mycobacterium
MEM	membrane
mg/ml	micrograms per millilitre
MgCl	magnesium chloride
MgSO₄.7H₂O	magnesium sulphate 7- hydrate
ml	millilitre
mM	millimolar
MPRS	major polymorphic tandem repeat
Na₂HPO₄	di-sodium hydrogen phosphate
Na₂HPO₄	di-sodium hydrogen phosphate
NaCl	Sodium chloride
NaCl	sodium chloride
NaN₃	sodium azide
NaOH	sodium hydroxide
ng	nanogram
nm	nanometer
NP40	Nonidet P 40 substitute
OADC	oleic acid albumin-dextrose catalase
°C	degree Celsius

OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	polymerase chain reaction
PE	proline-glutamic acid
PGRS	polymorphic GC-rich sequence
PPE	proline-proline-glutamic acid
R	resistance
RD	region of difference
rpm	revolutions per minute
s	sensitive
SCO	single crossover
SDS	sodium dodecyl phosphate
TAE	tris/acetic acid/EDTA buffer
TB	tuberculosis
TEMED	N,N,N',N'-Tetra-methylethyenediamine
TIGR	The Institute for Genomic Research
w/v	weight per volume
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactoside
ZN	Ziehl-Nielsen

LIST OF FIGURES AND TABLES.

Figure 1. Schematic diagram showing subgroups of the PE and PPE multi-gene families	6
Figure 2. Figure 2. Phylogenetic tree of all members of the PPE protein family.	8-9
Figure 3. Figure 3. Phylogenetic tree of all members of the PE protein family.	10-11
Figure 4. Schematic representation of the genomic organization of the genes present in the five ESAT-6 gene cluster regions of <i>M. tuberculosis</i> H37Rv	13
Figure 2.1.. Comparing the ESAT-6 region 1 of <i>M. smegmatis</i> and <i>M. tuberculosis</i> .	18
Table 2.1. List of primers used in this study.	20
Table 2.2. Bacterial strains, plasmids, cosmids and probes used in this study.	21
Figure 2.2. Plasmids vectors and cosmid library used in this study.	22
Figure 2.3. Schematic outline of the PCR based site-directed mutagenesis protocol.	24
Figure 2.4. Schematic representation of the main steps in the cloning of the PE35, and PPE68; PE_PGRS (Rv2615c) and PPE-MPTR (Rv0442c and Rv0878c) from the ESAT-6 gene cluster region 1 of <i>M. smegmatis</i> and <i>M. tuberculosis</i> respectively.	25
Figure 2.5. Schematic representation showing the procedure for Colony blotting.	27
Figure 2.6. Schematic representation of the subcellular fractionation of the <i>M. smegmatis</i> cell lysate.	29
Table 2.3. BSA standard preparations.	30
Figure 2.7. Schematic representation of the main steps in the amplification and cloning of the regions Upstream0282 and Downstream0292 of <i>M. smegmatis</i> Region 3 (steps for the cloning of Region 4 is the same).	34
Figure 2.8. Schematic representation of the cloning of pGOAL17 cassette, containing the lacZ and sacB cassette into the PacI site of the p2NIL constructs, i.e. p2NIL:R3/R4.	35
Figure 2.9. Schematic representation of the main steps in the deletion of Region 3 (Δ MSMEG:R3) and Region 4 (Δ MSMEG:R4) in <i>M. smegmatis</i> .	37
Figure 3.1. Colony blotting of a <i>M. smegmatis</i> genomic DNA intergrating cosmid library.	38
Figure 3.2. PCR analysis of the isolated cosmid.	40
Figure 3.3. Genes present within the isolated cosmid P1 insert.	41
Figures 3.4. Subcellular localization of the recombinantly expressed HA tagged PE35 and PPE68 proteins of <i>M. smegmatis</i> .	42
Figure 3.5. Subcellular localization of the recombinantly expressed HA tagged PPE-MPTR	

protein, Rv0442c.	44
Figure 3.6. Subcellular localization of the recombinantly expressed HA tagged PPE-MPTR protein, Rv0878c.	44
Figure 3.7. Subcellular localization of the recombinantly expressed HA tagged PE-PGRS protein, Rv2615c protein analysis.	45
Figure 3.8. Dilution series from 10^{-1} to 10^{-4} (i.e. 50 μ l cells into 450 μ l LB) of the R3 and R4 transformants, i.e. single crossovers spread on plates with and without sucrose.	47
Figure 3.9. Screening of region 3 knockout mutants.	48
Figure 3.10. Screening of region 4 knockout mutants.	49
Figure A.1. PCR amplification of the PPE-MPTR genes (Rv0442c and Rv0878c, respectively).	54
Figure A.2. PCR amplification of the PE_PGRS genes (Rv0978c and Rv2615c, using the outer and inner primers, Nested PCR) and SmegESAT-6 (Rv3875) gene.	55
Figure A.3. PCR amplification of the PE35 and PPE68 genes using the primers in Table 2.1, and the digestion of the cosmid from A to H.	56-58
Figure A.4. PCR amplification of the PE (Rv3872) and PPE (Rv3873).	58
Figure A.5. PCR amplification of the regions upstream and downstream of the <i>M. smegmatis</i> ESAT-6 gene cluster region 1 and region 4.	59
Table C.1. Absorbance of the standards and that of the samples and their concentrations.	63
Table C.2. Absorbance of the standards and that of the samples and their concentrations.	64
Figure C.1. Standard curves for the protein assay, absorbance against protein concentration.	65

CHAPTER 1

1. BACKGROUND

1.1. Tuberculosis (TB)

Tuberculosis (TB) is caused by an organism called *Mycobacterium tuberculosis* that is spread from person to person through the air (as aerosol droplets). TB is a contagious disease that usually attacks the lungs but may spread and affect any other part of the body. *M. tuberculosis* usually enters the alveolar passages of exposed humans in an aerosol droplet, where its first contact is thought to be within resident macrophages, but it is also possible that bacteria can be initially ingested by alveolar epithelial type II pneumocytes (Smith, 2003). *M. tuberculosis* organisms are sometimes called tubercle bacilli. When a person with infectious TB disease coughs or sneezes, droplet nuclei containing tubercle bacilli may be expelled into the air (<http://www.umdj.edu/ntbcweb/cd/tbhistory.htm>). TB is one of the oldest infectious diseases known to mankind (Young, 1998; Stenger and Modlin, 1999; Renshaw, *et al.*, 2002) and remains one of the most significant bacterial diseases of humans, with about one third of the world's population infected and resulting in ~3 million deaths annually (Elhers, 1999).

1.2. *M. tuberculosis* complex

The bacteria responsible for tuberculosis in humans and animals belong to the *M. tuberculosis* complex, which is a group of highly related mycobacteria. The *M. tuberculosis* complex comprises the following members, *M. tuberculosis*, the causative agent in the vast majority of human tuberculosis cases; *M. africanum*, an agent of human tuberculosis in sub-Saharan Africa; *M. microti*, the agent of tuberculosis in voles; *M. bovis*, which infects a wide variety of mammalian species including humans; *M. bovis* BCG (bacilli de Calmette et Guerin), an attenuated variant of *M. bovis*; *M. canettii*, a smooth variant that is very rarely encountered but causes human disease (Cole, 2002); the dassie bacillus an acid-fast microbe causing pulmonary TB in hyraxes (*Procavia capensis*) (Smith, 1960; Wagner, J. C., and Bokkenheuser V. 1961; Huard *et al.*, 2006); the oryx bacillus which causes TB in oryxes (*Oryx leucoryx*, an antelope species) (Gerth *et al.*, 1990; van Soolingen *et al.*, 1994; Huard *et al.*, 2006); *M. pinnipedii*, the cause of endemic TB in various seal species of the southern hemispheric origin (Cousins *et al.*, 2003); and *M. caprae*, named for their preferential goat hosts and they are also known to cause disease in a wide range of domesticated and wild animals as well as humans (Prodingier *et al.*, 2002; Huard *et al.*, 2006).

1.3. The history of tuberculosis

Tuberculosis has been present in humans since antiquity. The origins of the disease was first documented in history when a Greek physician Hippocrates identified phthisis (otherwise known as consumption or chronic wasting), around 460 BC as the most widespread disease of the times and noted that it was almost always fatal (<http://www.umdj.edu/ntbcweb/cd/tbhistory.htm>, accessed on the 4th July 2006). *M. tuberculosis* was identified in skeletal remains of prehistoric humans from 4000 BC, and tubercular decay was found in fragments of the spinal column from Egyptian mummies from 3000-2400 BC. There were also references to TB in India around 2000 BC and indications of lung scarring identical to that of modern-day TB sufferers in preserved bodies (such as mummies), suggesting that TB was also present in the Americas from about 2000 BC (<http://www.umdj.edu/ntbcweb/cd/tbhistory.htm>). Salo and co-workers (1994) identified *M. tuberculosis* DNA in 1000-year old lung tissue of a pre-Columbian Peruvian mummy, and spinal deformities in a pre-dynastic Egyptian mummy, shown to be specific to an *M. tuberculosis* complex bacterial infection of the spine (Pott's disease), were identified to be even more ancient at around 5400 years old (Crubezy *et al.*, 1998). The origin of *M. tuberculosis*, the causative agent of TB, has been the subject of much recent investigation, and it is thought that the bacteria in the genus *Mycobacterium*, like other actinomycetes, were initially found in soil and that some species evolved to live in mammals. The domestication of cattle, thought to have occurred between 10,000 and 25,000 years ago, would have allowed the passage of a mycobacterial pathogen from domesticated livestock to humans, and in this adaptation to a new host, the bacterium would have evolved to the closely related *M. tuberculosis* (Smith, 2003). Human tuberculosis for many years was thought to have evolved from the bovine disease by adaptation of an animal pathogen to the human host (Stead, *et al.*, 1995). Brosch *et al* (2002) recently showed that *M. tuberculosis* is the progenitor of *M. bovis*, confirming that TB does not originate from cattle.

Exact pathological and anatomical descriptions of the disease began to appear in the seventeenth century. Tuberculosis was first described by Dr. Sylvius in his *Opera Medica* of 1679, which relates to the tubercles, tuberculous cavities and tuberculous lymph nodes that were associated with the disease (<http://www.umdj.edu/ntbcweb/cd/tbhistory.htm>). An English physician, Benjamin Marten is the one who came with the first great step in the combat of tuberculosis, by recognizing that tuberculosis may be caused by an airborne organism. In 1722, in his publication, *A New Theory of Consumption*, he hypothesized that TB could be caused by "wonderfully minute living creatures", which, once they had gained a foothold in the body, could generate the lesions and symptoms of the

disease (<http://www.umdj.edu/ntbcweb/cd/tbhistory.htm>). A major turning point in the evolution of human medicine occurred with the observation by Robert Koch that *M. tuberculosis* was the causative agent of tuberculosis (Kana and Mizrahi, 2004). Koch was the first to observe the tubercle bacillus under the microscope using specialized staining techniques and in so doing, he began a new era of clinical medicine. The three tenets of Koch's Postulates have formed a foundation upon which infectious diseases research has been based (Kana and Mizrahi, 2004). However, after a century of tuberculosis research, the biology of *M. tuberculosis*, in terms of its growth, metabolism, physiology, and interaction with its human host, remains poorly understood (Kana and Mizrahi, 2004). Koch's second molecular postulate states that the suspected causal organism must be isolated from an infected organism and grown in pure culture, and that when a healthy susceptible host is inoculated with the pathogen from pure culture, symptoms of the original disease must develop. These "wonderfully minute living creatures" were only first isolated more than 100 years later in the late 1800's by the German physician Robert Koch, who subsequently received the Nobel Prize in physiology or medicine for this discovery in 1905, by a staining technique that enabled him to see *M. tuberculosis* bacilli (Koch, 1882). He revealed his discovery of the identification of *M. tuberculosis* as the causative agent of tuberculosis in a historical address to the Berlin Physiological Society on March 24, 1882 (McKinney *et al.*, 1998). The first genuine success in immunizing against tuberculosis was developed from attenuated bovine-strain tuberculosis by Albert Calmette and Camille Guerin in 1906 which is commonly called 'BCG' (Bacillus of Calmette and Guerin). The BCG vaccine was first used on humans on July 18, 1921 in France. It wasn't until after World War II that BCG received widespread acceptance in the USA, Great Britain, and Germany (<http://www.umdj.edu/ntbcweb/cd/tbhistory.htm>). It is disturbing that after more than a century of research on the biochemistry and physiology of *M. tuberculosis*, the disease processes are still poorly understood and we still have no clear indications of what differentiates this from the lesser virulent and avirulent mycobacterial species (Ehlers, 1993).

1.4. The genome of *M. tuberculosis*

A major breakthrough in the research of tuberculosis came with the complete sequencing of the genome sequence of the *M. tuberculosis* reference laboratory strain H37Rv in 1998 (Cole *et al.*, 1998). Cole and co-workers revealed that the genome of *M. tuberculosis* is around 4,411,529 base pairs (bp) and contains around 4000 genes distributed evenly on both strands. These genes account for >91% of the potential coding capacity and has a G+C content of 65.6%. Roughly 40% of the 3924 protein-

coding genes were identified to be of known function, 20% of vague function, and 40% of unknown function, although over half of these belong to the class known as conserved hypotheticals (Cole, 1999). The sequencing of the genome of *M. tuberculosis* revealed a large number of previously unknown genes with the potential to be involved in the pathogenesis of the organism (Cole, 1998). Interestingly, about 51% of the genome originated due to gene duplication or domain shuffling (Takaia *et al.*, 1999). The genome of *M. tuberculosis* contains five duplicate copies of the immunologically important ESAT-6 gene clusters (Gey van Pittius *et al.*, 2001). Each gene cluster encodes proteins involved in energy provision for active transport, membrane pore formation and protease processing and could thus assemble to form a dedicated biosynthesis, transport and processing system for the secretion of the potent T-cell antigens belonging to the ESAT-6 protein family (Tekaia *et al.*, 1999; Gey van Pittius *et al.*, 2001; Pallen 2002; Pym *et al.*, 2003). Interestingly, in addition to these genes, there are two families of genes present within these clusters which seem to be anomalous, named the PE and PPE gene (Cole *et al.*, 1998; Gey van Pittius *et al.*, 2001). These novel gene families (representing areas of the genome with an exceptionally high G+C content of more than 80%), were found to make up around 10% of the genome (Cole, 1999). The names PE and PPE signify the presence of characteristic Pro-Glu (positions 8-9) and Pro-Pro-Glu (positions 8-10) motifs respectively, in the well-conserved NH₂-terminal domains of the proteins and these generally precede repetitive domains of variable length.

1.5. Region of difference 1 (RD1)

The 9455-bp region of difference 1 (RD1) deletion region has been identified as a major region of difference between the attenuated *M. tuberculosis* complex strains (such as *M. bovis* BCG and *M. microti*), compared to the pathogenic species such as *M. bovis* and *M. tuberculosis* (Cole, 1998; Brosch *et al.*, 2002; Pym *et al.*, 2002; Daugelat *et al.*, 2003; Lewis, 2003; Demangel, 2004). The RD1 endpoints are not identical in BCG and *M. microti* (Brodin *et al.*, 2002), although these deletions have removed from both vaccine strains a core cluster of six genes (Rv3871–Rv3876) that are part of the ESAT-6 locus (Pym *et al.*, 2002). *M. bovis* BCG contains a DNA segment of 10 kb (equivalent to kb 4350.3 to 4359.7 of the *M. tuberculosis* H37Rv genome) that partly overlaps RD1 of *M. microti*. *M. microti* contains a DNA segment of 14 kb (equivalent to kb 4340.4 to 4354.5 of the *M. tuberculosis* H37Rv genome) that partly overlaps RD1 of *M. bovis* BCG. In *M. bovis* BCG, ORFs Rv3871 and Rv3879 are truncated and ORFs Rv3872 to Rv3878 are absent (Pym *et al.*, 2002). In *M. microti*, ORFs Rv3864 and Rv3876 are truncated and ORFs Rv3865 to Rv3875 are absent. The RD1^{mic} and the

RD1^{bcg} deletions are responsible for the loss of a large portion of the conserved ESAT-6 gene cluster core region (Takaia *et al.*, 1999), including the genes coding for major T-cell antigens ESAT-6 and CFP-10 (Berthet *et al.*, 1998; and Harboe *et al.*, 1998). RD1 gene products are interesting for various reasons: they represent potential virulence factors, due to their absence in the attenuated strains, and may also be useful for diagnostic purposes due to their absence in the vaccine strain BCG (Daugelat *et al.*, 2003).

1.6. Secreted antigens- the ESAT-6 gene family

The best studied RD1 gene products are CFP-10 and ESAT-6, both of which elicits strong T- and B-cell responses in experimental animals and humans (Colangeli *et al.*, 2000; Dillon *et al.*, 2000; Brusasca *et al.*, 2001; and Daugelat, 2003). Both ESAT-6 and CFP-10 are low molecular mass proteins that belong to the larger ESAT-6 protein family, which has 23 members in the genome of *M. tuberculosis*. The *esat-6* (*esx*) and *lhp* genes (encoding ESAT-6 and CFP-10, respectively) lie next to each other in an operon like-structure (Berthet *et al.*, 1998; Daugelat *et al.*, 2003), their gene products are co-transcribed and the proteins form a tight 1:1 protein complex (Renshaw *et al.*, 2002). These genes, as well as other RD1 genes, are part of a cluster called the “ESAT-6 gene cluster”, which is found in multiple copies in several mycobacterial genomes (Gey van Pittius *et al.*, 2001) and encode a novel secretion apparatus (Pallen, 2002; Daugelet *et al.*, 2003). When *M. tuberculosis* is grown in broth culture, these proteins are released into the surrounding medium (Andersen *et al.*, 1995; Berthet *et al.*, 1998; Okkels and Anderson, 2004), through a Sec-independent pathway (since none of the proteins contain common signal peptides) (Okkels *et al.*, 2003).

1.7. PE and PPE

One of the major findings from the genome sequence of *M. tuberculosis* was the presence of the PE and PPE families of genes (Cole *et al.*, 1998; Chakhaiyar and Hasnain, 2004; Tundup *et al.*, 2006). These two large glycine-rich gene families, together account for about 10% of the coding capacity of the genome (Cole *et al.*, 1998). As mentioned previously, the names PE and PPE are derived from the motifs Pro-Glu (PE) and Pro-Pro-Glu (PPE) found in most cases near the N-terminus of the proteins (Figure 1). All the 99 members of the PE family have a highly conserved N-terminal domain of 110 amino acids (Chaitra *et al.*, 2005). Similarly, the PPE family consists of 68 members and has a highly conserved N-terminal domain of around 180 amino acid residues with varying carboxy terminal

domains (Cole *et al.*, 1998). The sequences of the PE and PPE N-terminal domains are conserved within each family, with very little homology between the two families. The C-terminal domains of both proteins are of variable size and sequence and frequently contain repeat sequences of different copy numbers (Gordon *et al.*, 1999; Gey van Pittius *et al.*, 2006). The PE and PPE protein families can be divided into subgroups according to the homology and presence of characteristic motifs in their C-terminal domains (Gordon *et al.*, 1999; Gey van Pittius *et al.*, 2006). The polymorphic GC-rich sequence (PGRS) subgroup of the PE family is the largest subgroup and contains proteins with multiple tandem repeats of a glycine-glycine-alanine (Gly-Gly-Ala) or a glycine-glycine-asparagine (Gly-Gly-Asn) motif in the C-terminal domain. The other subgroup consists of proteins with N-terminal domains of low homology. The PPE family can be broadly divided into four subgroups (Gordon *et al.*, 1999) of which the PPE-SVP subgroup is the largest, comprising of 24 members (Gey van Pittius *et al.*, 2006). The PPE-SVP subgroup is characterized by the motif GlyXXSerValProXXTrp at position 300 and 350 in the amino acid sequence. The second largest PPE subgroup is the major polymorphic tandem repeat (MPTR) comprising of 23 members. This subgroup contains multiple C-terminal repeats of the motif AsnXGlyXGlyAsnXGly, encoded by a consensus repeat sequence GCCGGTGTTG, separated by 5 bp spacers (Hermans *et al.*, 1992; Cole and Barrell, 1998; Gey van Pittius, *et al.*, 2006). Adindla and Guruprasad (2003) identified the third subgroup of the PPE subfamily comprising of 10 members. This subgroup is characterized by a conserved 44 amino acid residue region in the C-terminus comprising of highly conserved GlyPheXGlyThr and ProXXProXXTrp sequence motifs. The last PPE subgroup comprising of 12 members, consists of proteins with a low percentage of homology at the C-terminus. Figure 2 and 3 shows the evolutionary duplication of all the members of the PE and PPE protein families (Gey van Pittius *et al.*, 2006). Even though an early paper by Doran and co-workers (1992) had suggested that the members of the PPE-MPTR family were likely to be cell wall associated, until recently there was no evidence available for the subcellular localization of the members of the PE and PPE proteins. Brennan *et al.*, (2001) and Banu *et al.*, (2002) showed that certain PE_PGRS proteins are cell-surface constituents that influence the interactions of the organism with other cells. In addition, studies by Sampson *et al.*, (2001) demonstrated the PPE-MPTR protein Rv1917c to be situated in the mycobacterial cell wall and to be at least partly exposed on the cell surface.

Figure 1. Schematic diagram showing subgroups of the PE and PPE multi-gene families. (A) In addition to the open reading frames (ORFs) encoding simple PE proteins, these sequences are present as N-terminal domains of more complex genes. The predominant subfamily consists of the PE_PGRS proteins in which the PGRS domain can vary in size but is typically composed of Gly-Ala repetitive sequences. A few PE_PGRS proteins contain atypical sequences at the C-terminus, and some of the PE genes are also linked to other non-PGRS sequences, including one, Rv3097c that encodes a putative lipase (Cole et al., 1998). (B) The PPE family consists of four subfamilies of which the PPE-SVP subfamily is the largest (24 members). The proteins of this subfamily are characterized by the motif Gly-X-X-Ser-Val-Pro-X-X-Trp between position 300 and 350 in the amino acid sequence. The major polymorphic tandem repeat (MPTR) PPE subfamily is the second largest (23 members) and contains multiple C-terminal repeats of the motif Asn-X-Gly-X-Gly-Asn-X-Gly, encoded by a consensus repeat sequence GCCGGTGTG, separated by 5 bp spacers. The third subfamily (10 members) recently identified by Adindla and Guruprasad, is characterized by a conserved 44 amino acid residue region in the C-terminus comprising of highly conserved Gly-Phe-X-Gly-Thr and Pro-X-X-Pro-X-X-Trp sequence motifs named the "PPE-PPW" subfamily. The last PPE subfamily (12 members) consists of proteins with a low percentage of homology at the C-terminus. Reproduced from Gey van Pittius et al., 2006 with kind permission of the authors.

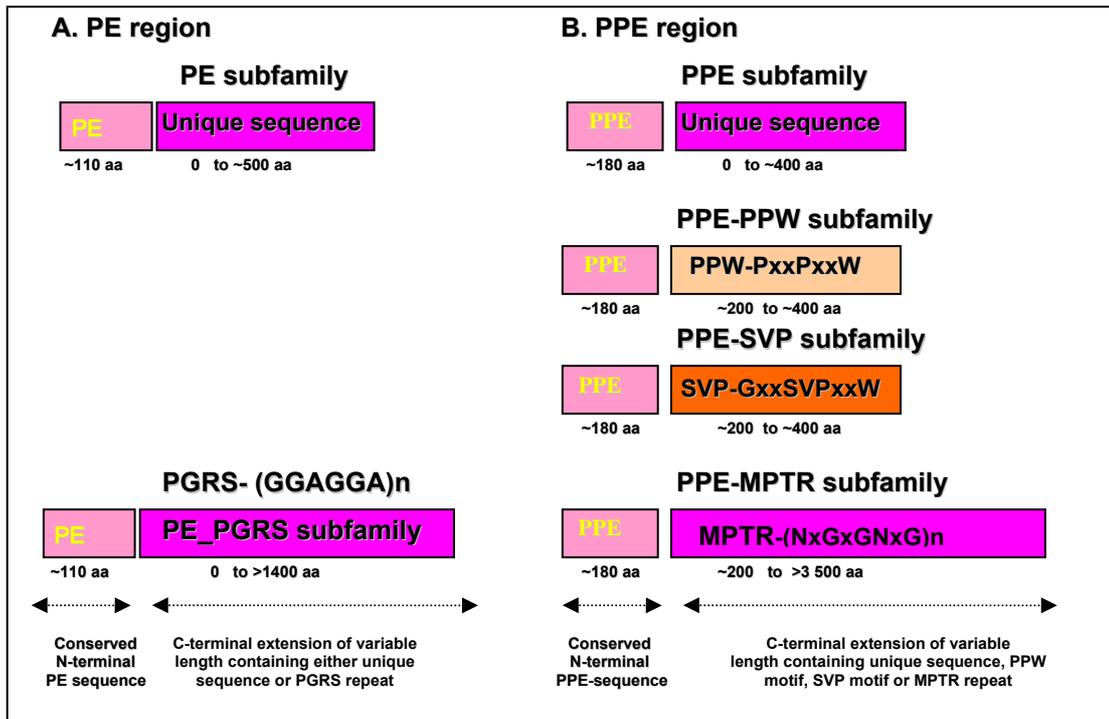


Figure 2. Phylogenetic tree of all members of the PPE protein family. The tree was constructed from the phylogenetic analyses done on the 180 amino acid N-terminal domains of the PPE proteins. The tree was rooted to the outgroup, which was chosen as Rv3873. This gene has been shown previously to be the first PPE insertion into the ESAT-6 gene cluster (region 1). The gene highlighted in purple is present in the ESAT-6 gene cluster region 1, genes highlighted in green are present in or have been previously shown to be duplicated from ESAT-6 gene cluster region 3 (Gey van Pittius et al., 2001), gene highlighted in blue is present in ESAT-6 gene cluster region 2, genes highlighted in red are present in or have been previously shown to be duplicated from ESAT-6 gene cluster region 5 (Gey van Pittius et al., 2001) and genes highlighted in yellow are members of the MPTR subgroup of the PPE family. Arrows indicate genes that are identified to be present within the genome sequence of *M. smegmatis*. Five sublineages (including the PPE-SVP and PPE-MPTR subgroups) are indicated by Roman numerals. Reproduced from Gey van Pittius et al., 2006, with kind permission of the authors. Genes highlighted in red were used in this study.

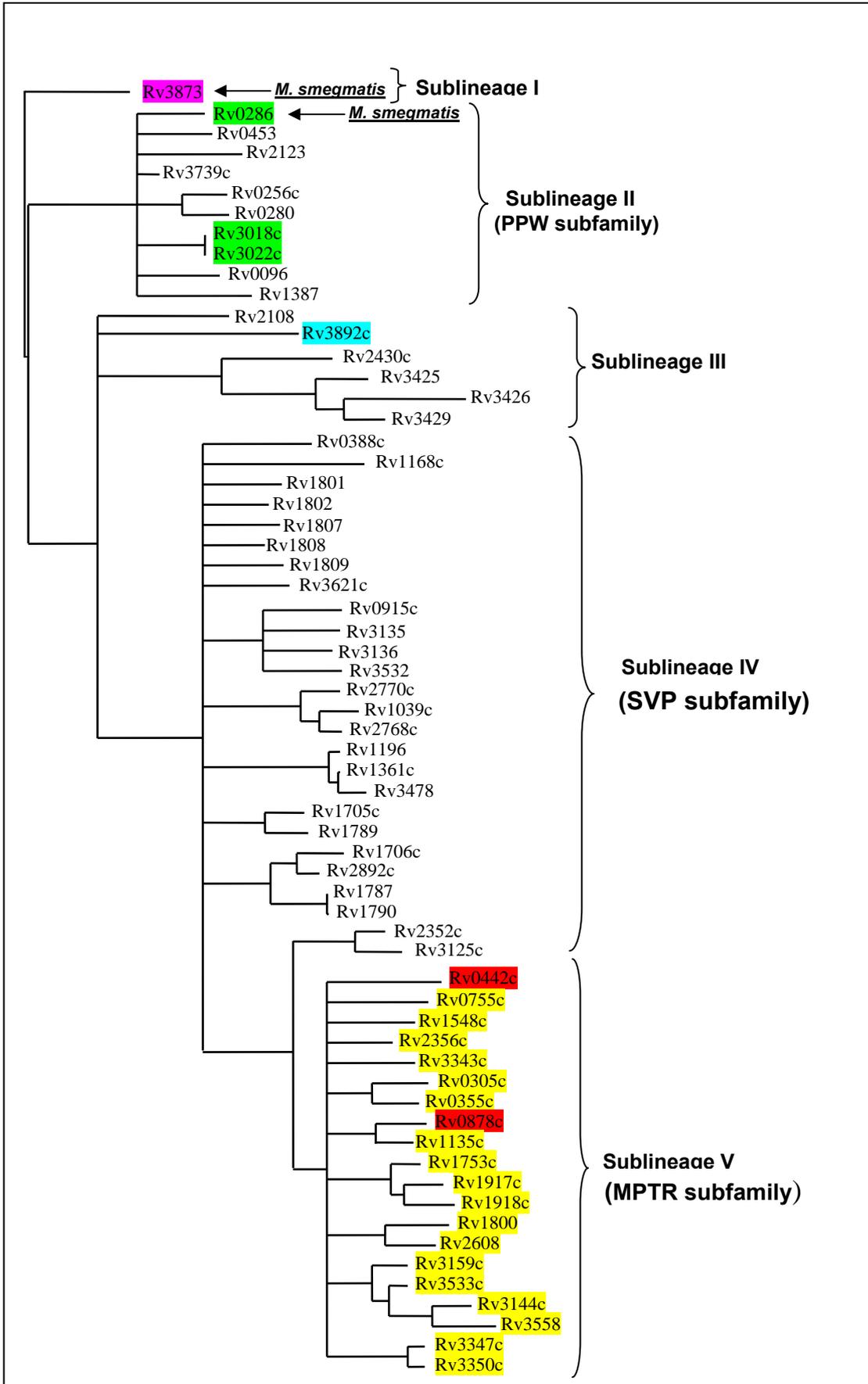
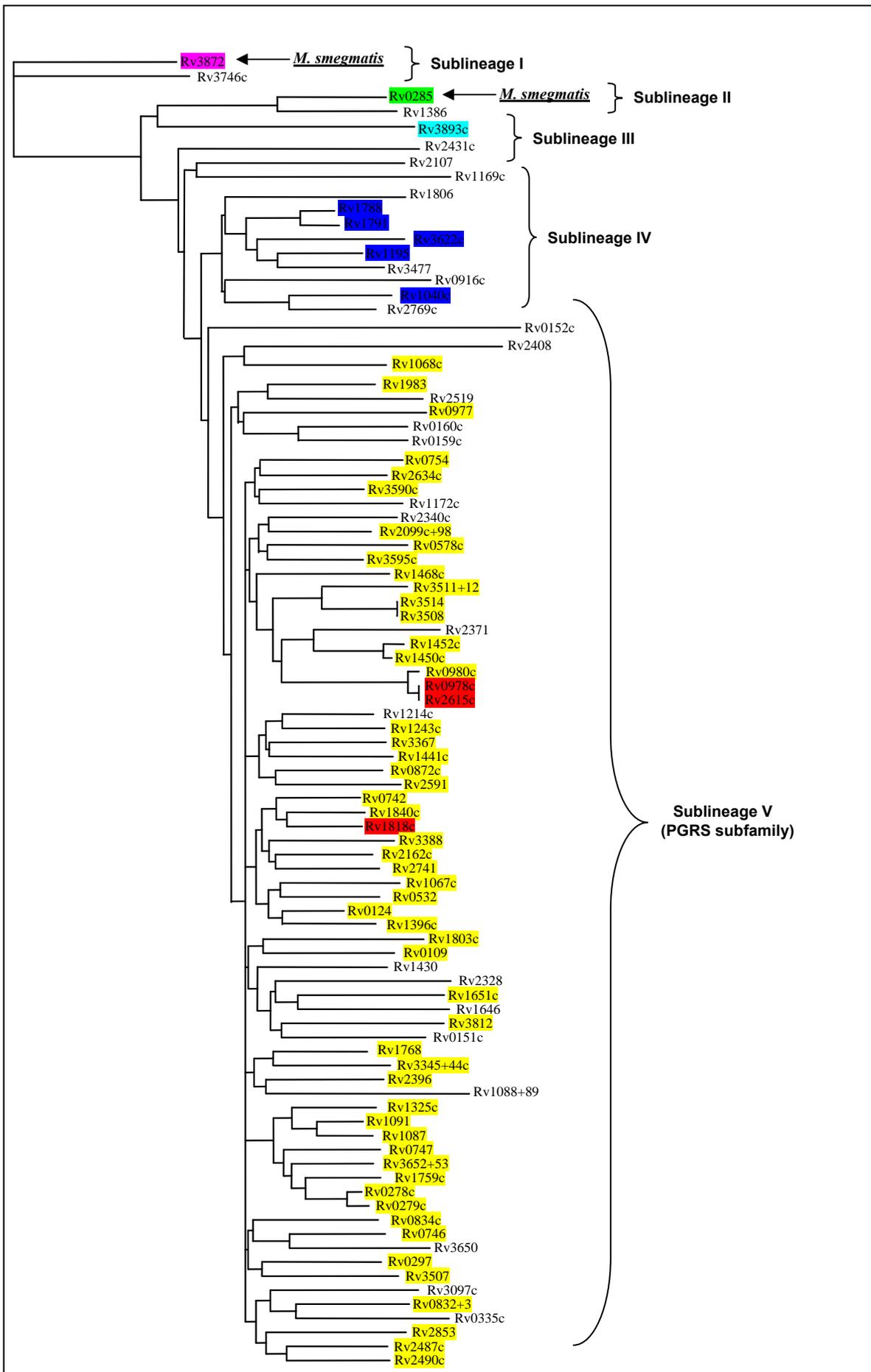


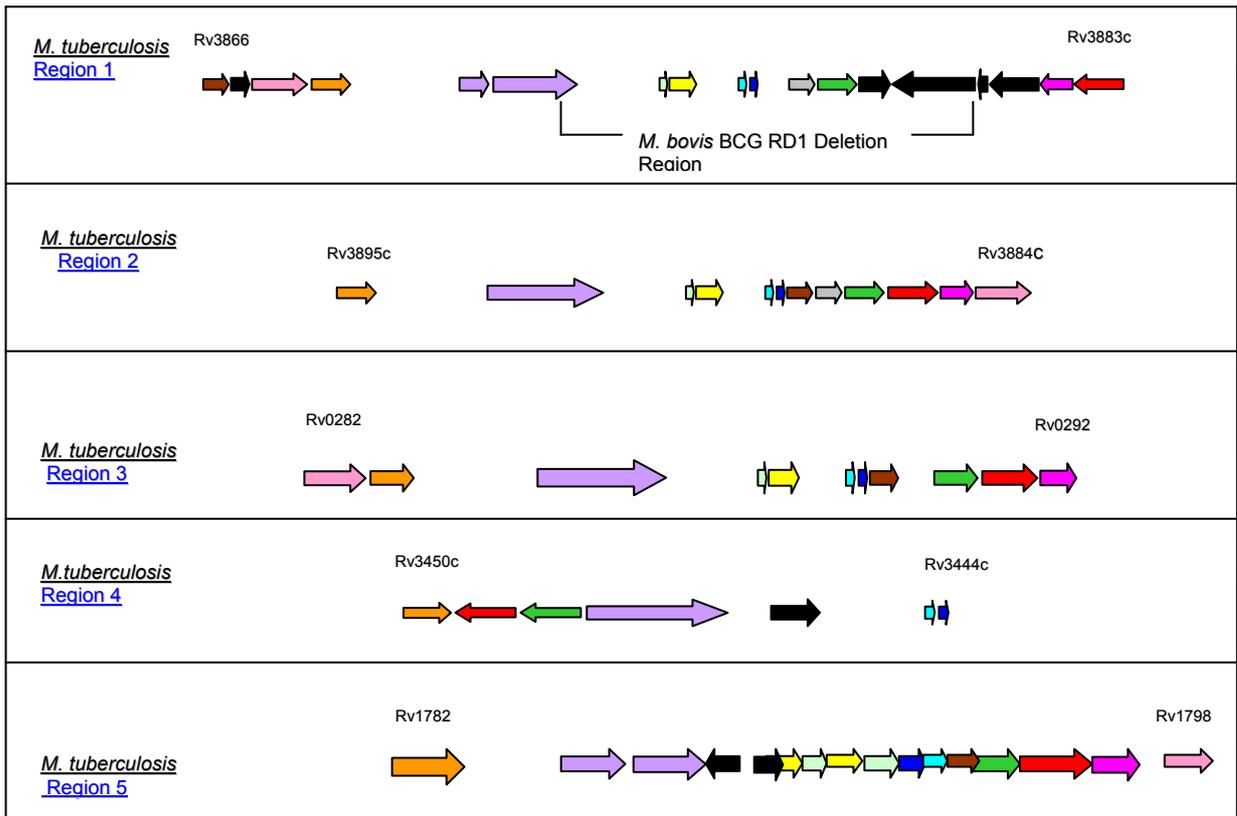
Figure 3. Phylogenetic tree of all members of the PE protein family. The tree was rooted to the out-group, which was chosen as Rv3872. This gene has been previously shown to be the first PE insertion into the ESAT-6 gene cluster (region 1). The gene highlighted in purple is present in the ESAT-6 gene cluster region 1, the gene highlighted in green is present in ESAT-6 gene cluster region 3, the gene highlighted in blue is present in ESAT-6 gene cluster region 2, genes highlighted in red are present in or have been previously shown to be duplicated from the EAST-6 gene cluster region 5 (Gey van Pittius et al., 2001) and genes highlighted in yellow are members of the PGRS subgroup of the PE family. Arrows indicate genes that are identified to be present within the genome sequence of *M. smegmatis*. Five sublineages (including the PE_PGRS subgroup) are indicated by Roman numerals. Reproduced from Gey van Pittius et al., 2006, with permission from the authors. Genes highlighted in red were used in this study.



Cole *et al.* (1998) and Tekaiia *et al.* (1999) hypothesized that these two multigene families are of potential interest from an immune response point of view, since they could function as a source of antigenic variation for *M. tuberculosis* in order to evade the host immune response (Chaitra *et al.*, 2005), and act as cell surface antigens (Banu *et al.*, 2002; Chaitra *et al.*, 2005). One PE gene, Rv3097c, contains a sequence at the 3' end which encodes a polypeptide with a predicted esterase/lipase activity (Cole *et al.*, 1998). This gene is the only PE gene with a predicted function, and this unique gene may have been created by a fusion event between a PE gene and a gene encoding an enzyme with this activity (Brennan *et al.*, 2005). Some of the PE and PPE proteins have been shown to be potent B and T cell antigens (Espitia *et al.*, 1999; Singh *et al.*, 2001). Two proteins from the PE_PGRS subfamily, Rv1759c and Rv3367 are expressed during infection and show antibody responses in humans and rabbits, respectively (Chaitra *et al.*, 2005). Dillion *et al.* (1999) and Skeiky *et al.* (2000) showed that Rv1196 and Rv0915 from the PPE family are T cell antigens (Chaitra *et al.*, 2005). It was shown by Delogu and Brennan (2001) that the PE domain of the PE_PGRS protein Rv1818c induced good cell mediated immune response upon immunization into mice, whereas the PGRS domain is responsible for good humoral response (Chaitra *et al.*, 2005).

Many PE/PPE gene pairs occur in the *M. tuberculosis* genome, and these pairs are frequently associated with ESAT-6 gene clusters (Gey van Pittius *et al.*, 2003 and Strong *et al.*, 2003). The duplication order of the ESAT-6 gene clusters within the genome of *M. tuberculosis* was previously predicted by systematic phylogenetic analyses of the constituent genes. This duplication order was shown to extend from the ancestral region named region 4 (Rv3444c-Rv3450c) to region 1 (Rv3866-Rv3883c), 3 (Rv0282-Rv0292), 2 (Rv3884c-Rv3895c), and lastly region 5 (Rv1782-Rv1798)(Gey van Pittius *et al.*, 2001) (Figure 4). The absence of a pair of PE and PPE proteins within the ancestral region 4, indicates that these genes may have been integrated into the first duplicate of this region (region 1), and have subsequently been co-duplicated together with the rest of the genes within the regions. Support for the published duplication order of the ESAT-6 gene clusters also comes from the fact that the last duplicate (region 5) includes multiple separate duplications of the PE and PPE genes within the region (Gey van Pittius *et al.*, 2006).

Figure 4. Schematic representation of the genomic organization of the genes present in the five ESAT-6 gene cluster regions of *M. tuberculosis* H37Rv. ORFs are represented as blocked arrows showing the direction of transcription, with the different colours reflecting the specific gene family and length of the arrow reflecting the relative lengths of the genes. Annotations of *M. tuberculosis* H37Rv genes are according to Cole et al., (1998). Reproduced from Gey van Pittius et al., 2001, with kind permission of the authors.



1.8. Targeted gene knock-out

Targeted gene knockout techniques have been developed to study the function of proteins in *M. tuberculosis*. To test whether the gene in fact encodes the proposed function, expression of its functional protein must be abolished, and the function must be restored by re-introduction of a functional gene, thereby fulfilling the second requirement of the Molecular Koch's postulates (Machowski *et al.*, 2005). The alternative approach to knockout mutagenesis is random generation of mutant libraries where the phenotype is used as the determining factor to identify genes. Snapper *et al.* (1990), described a mutant of *M. smegmatis* mc²6 (ATCC607) with increased electro-transformability; (mc²155). The availability of this strain meant that rare events were more likely to be detected and thus greatly facilitated the identification of knockout mutants in *M. smegmatis*. Work by Husson *et al.* (1990) showed that *pyrF* mutants could be readily obtained by transforming *M. smegmatis* with a non-replicating plasmid carrying *pyrF* disrupted by a kanamycin resistance gene (*aph*). The *pyrF* gene was chosen as the target for inactivation in this proof-of-principle study because, depending on the supplement included in the medium, it can be selected either for or against (Husson *et al.* 1990).

Initial attempts at targeted mutagenesis in mycobacteria of the *M. tuberculosis* complex reported high frequencies of illegitimate recombination, homologous recombination by single-crossover only, or unstable double-crossover clones (DCO) (Kalpana *et al.*, 1991; Aldovini *et al.*, 1993; and Norman *et al.*, 1995). Reytrat *et al.* (1995) were the first to report of targeted knockout mutagenesis in the *M. tuberculosis* complex and they targeted the *ureC* gene of *M. bovis* BCG. Out of the 50 clones screened, they found two that displayed a urease negative phenotype, and were indeed DCO recombinants. Azad *et al.* (1996 and 1997) showed a strategy in which part of the *mas* gene and *pps* gene cluster of *M. bovis* were replaced with the hygromycin cassette, with the resulting mutants predictably incapable of synthesizing mycocerosic acids and phthiocerol dimycocerosates, respectively.

The lack of a phenotypic screen required a Southern blot genotypic screen for identification of these DCO homologous recombinants. In 1996, Balasubramanian *et al.* argued that large linear DNA molecules should be more recombinogenic than closed circular plasmid, and devised a strategy by which they could electroporate DNA fragments of up to 50 kb into *M. tuberculosis*. Because DNA of this size is difficult to manipulate by molecular cloning techniques, requiring endonuclease restriction digestion, the disrupted allele was introduced into the cosmid molecule by interplasmid recombination in *E. coli*. Electroporation of the linearised cosmid molecule yielded *M. tuberculosis* kanamycin resistant

clones, of which 6% were leucine auxotrophs with a disrupted *leuD* allele. Since then a lot of research has been done in trying to fulfil the second requirement of the Molecular Koch's postulates by gene knockout strategy and other molecular techniques.

1.9. Problem statement

The expansion of the *Mycobacterium tuberculosis* PE and PPE families appears to be linked to that of the immunologically-important ESX gene cluster secretion system, as the ancestral members of these families are found only within the ESX regions. These ancestral members are also the only copies in the earlier mycobacteria like *M. smegmatis* (Gey van Pittius *et al.*, 2006). Later duplications of the PE and PPE families belonging to the PGRS and MPTR subgroups, respectively, have been implicated in virulence and are mostly found within the pathogenic mycobacteria closely related to the *M. tuberculosis* complex. It is unknown whether there exists a function/relationship between the ESAT-6 gene cluster secretion system and the PE and PPE gene families, and if so, whether this is true for the later duplications belonging to the PGRS and MPTR subfamilies.

1.10. Hypothesis

We hypothesise that the ancestral PE and PPE proteins uses the ESAT-6 gene cluster regions, which form a multi-competent system for secretion of the ESAT-6 protein family to reach the extracellular milieu. We also hypothesise that the PGRS and MPTR subfamilies use the same system to localize to the cell wall, but are incapable of being secreted into the medium.

1.11. Study Aims and Design:

The aims of this study can be divided in three major parts, summarized as follows:

1. To investigate the localization of the most ancestral PE and PPE proteins translated from genes located within the ESAT-6 gene cluster region 1 of *M. smegmatis*, i.e. to elucidate whether they partition in the culture filtrate, cell wall, cell membrane or in the cytosol. This will be done with the aim of determining whether the PE and PPE proteins are secreted, in a similar manner to the ESAT-6 antigens situated adjacently in the ESAT-6 gene cluster regions.

The localization of the most ancestral PE and PPE proteins found in *M. smegmatis* were investigated by:

- ❖ Design, PCR amplification, cloning and expression of HA tagged PE (Rv3872) and PPE (Rv3873) proteins from the ESAT-6 gene cluster region 1 of *M. smegmatis*.

❖ Determination of the localization of these proteins in the cell fractions of *M. smegmatis* using fractionation, SDS-PAGE and Western blotting.

2. To investigate the localization of the more recently evolved members of the PE and PPE families (belonging to the PGRS and MPTR subgroups) and to determine whether they display similar localization to the most ancestral PE and PPE members.

The localization of the PGRS and MPTR subfamilies found in *M. tuberculosis* were investigated by:

❖ Design, PCR amplification, cloning and expression of *M. tuberculosis* HA tagged PGRS (Rv1818c, Rv0978c and Rv2615c) and MPTR (Rv0442c and Rv0878c) genes in *M. smegmatis*

❖ Determination of the localization of these proteins, in the cell fractions of *M. smegmatis* by fractionation, SDS-PAGE and Western blotting.

3. To construct knockouts which could be used in future to investigate whether the ESAT-6 secretion apparatus (or parts thereof) are involved in the trafficking of both the ancestral and recent PE and PPE proteins.

Knockouts of ESAT-6 gene clusters region 3 and 4 in *M. smegmatis* were constructed by:

❖ Designing primers flanking (~800 bp) the regions upstream and downstream of the two target ESAT-6 gene cluster regions of *M. smegmatis*, i.e. ESAT-6 gene cluster region 3 and region 4.

❖ Cloning, and generating knockouts of the ESAT-6 gene cluster region 3 and 4 in *M. smegmatis* by homologous recombination *in vivo*.

CHAPTER 2. MATERIALS AND METHODS

2.1. DNA and protein sequence analyses

The DNA and protein sequence information for both *M. tuberculosis* H37Rv and that of *M. smegmatis* mc²155 were obtained from publicly available completed genome sequence databases at the Pasteur Institute and the Institute for Genomic Research (TIGR) websites (<http://genolist.pasteur.fr/TubercuList/> for *M. tuberculosis* H37Rv and <http://cmr.tigr.org/tigr-scripts/CMR/GenomePage.cgi?org=gms> for *M. smegmatis* mc²155), as well as from previously published sequence analyses (Tekaiia *et al.*, 1999, Gey van Pittius *et al.*, 2001).

2.2. Bacterial strains

E. coli XL-1 Blue (Stratagene) and/or *E. coli* DH5 α (Promega) were used as host strains for propagation of the vectors and clones. To determine the potential secretion function of the ancestral PE, and ancestral PPE, and the more recently duplicated PE_PGERS and PPE-MPTR genes, a suitable host had to be identified. The avirulent, fast-growing species *M. smegmatis* was chosen as a model organism in this study, due to its high efficiency during transformations and the fact that it has been used widely in the study of *M. tuberculosis* proteins (Snapper *et al.*, 1990). *M. smegmatis* is a member of the genus *Mycobacterium* and its cell wall structure is similar to that of other mycobacteria in the genus, which is an advantage as another bacterial host with a different cell wall structure would not necessarily display the same secretion characteristics. The genome of *M. smegmatis* contains three of the ESAT-6 gene clusters (region 1, 3 and 4), and it contains two each of the PE and PPE genes in its genome. These genes have been shown to display the same organization as in *M. tuberculosis*, with a high percentage of homology shared between the gene sequences of the two organisms (see Figure 2.1., Gey van Pittius *et al.*, 2001). All cloning was done in *E. coli* XL-1 Blue (Stratagene), unless otherwise stated.

g Yeast extract (Merck) per liter) containing appropriate antibiotics (Hygromycin B, 100 µg/ml, Kanamycin, 50 µg/ml or Ampicillin, 50 µg/ml) when required. For transformant selection on solid media, *M. smegmatis* was grown on Middlebrook 7H11 agar (Difco) supplemented with 1/10 volume filter-sterilized 10% oleic acid-albumin-dextrose catalase (OADC), (0.005% oleic acid, 0.5% BSA, 0.2% glucose, 0.015% catalase), 0.2% glycerol, and 0.05% Tween-80 (Sigma). To obtain culture filtrate proteins in a detergent/protein free culture medium for secretion analyses, recombinant clones were grown at 37°C in 150 ml Kirchner's broth (3 g/l Na₂HPO₄, 4 g/l KH₂PO₄, 1.07 g/l MgSO₄·7H₂O, 2.5 g/l tri-sodium citrate, 20% glycerol, 5 g/l asparagine) to an optical density (OD) of 0.3 at 600 nm (early log phase), containing the appropriate antibiotic (Hygromycin - 100 µg/ml, Roche, Kanamycin - 50 µg/ml, Roche and/or Ampicillin - 50 µg/ml, Roche).

For expression of recombinant proteins, *M. smegmatis* was grown in Middlebrook 7H9 medium (Difco) supplemented with 1/10 volume of 10% albumin-dextrose catalase (ADC), 0.05% Tween-80 (Sigma), and 0.2% glycerol. Hygromycin B (100 µg/ml, Roche), Kanamycin (50 µg/ml, Roche) or Ampicillin (50 µg/ml, Roche) was added to bacterial cultures when antibiotic selection was required. Whenever mycobacteria were double transformed, i.e. with integrating cosmids and plasmids, the culture was grown in both Hygromycin and Kanamycin to select for both transforming structures.

2.5. Primers and plasmid vectors

All the primers used in this study are listed in Table 2.1. The plasmid vectors used in this study to construct the recombinants are listed in Table 2.2 and shown in Figure 2.2. All constructs were confirmed by DNA sequencing using an ABI 3130X sequencing machine (Applied Biosystems).

2.6. PCR amplification and conditions

PCR was performed in 0.2 ml thin-walled tubes for the amplification of genes (using the Applied Biosystems GeneAmp PCR system 2400). PCR was carried out for one initial activation cycle of 15 minutes at 95°C, followed by 10 cycles (94°C for 1 minute, denaturation; 54°C for 2 minutes, annealing; 72°C for 1 minute, extension), and then followed by 30 cycles (94°C for 1 minute, denaturation; at an annealing temperature of between 57°C and 62°C depending on the primer pair for 2 minutes; an extension step at 72°C for 1 minute) and one final extension cycle at 72°C for 10 minutes using HotStar Taq polymerase kit (Qiagen). When using Expand Long Template PCR system (Roche), PCR was

carried out for one initial activation cycle of 2 minutes at 94°C, followed by 10 cycles (94°C for 30 seconds, denaturation; 54°C for 30 seconds, annealing; 68°C for 2 minutes, extension), and then followed by 25 cycles (94°C for 30 seconds, denaturation; at an annealing temperature of between 57°C and 62°C depending on the primer pair for 30 seconds; an extension step at 68°C for 2 minute) and one final extension cycle at 68°C for 7 minutes. When using Advantage™-GC 2 PCR kit (BD Biosciences), PCR was carried out for one initial activation cycle of 2 minutes at 94°C, followed by 10 cycles (94°C for 30 seconds, denaturation; 54°C for 30 seconds, annealing; 68°C for 2 minutes, extension), and then followed by 25 cycles (94°C for 30 seconds, denaturation; at an annealing temperature of between 57°C and 62°C depending on the primer pair for 30 seconds; an extension step at 68°C for 2 minute) and one final extension cycle at 68°C for 7 minutes. The PCR amplification products were electrophoretically fractionated in 1% agarose 1xTAE (tris/acetic acid/EDTA buffer) containing ethidium bromide (EtBr) at 80 V/cm for 1 hour. All PCR amplification products were purified using the Wizard®SV Gel and PCR Clean-Up System, (Promega), and ligated into the pGEM-T Easy vector (Promega) according to the manufacturer's instructions. Forty microliters (40 µl) TE buffer (1 M Tris, pH 8.0 and 0.5 M Na₂EDTA, pH 8.0) was added to the ligation, which was purified using the Wizard®SV Gel and PCR Clean-Up System, (Promega). PCR results are shown in Addendum A.

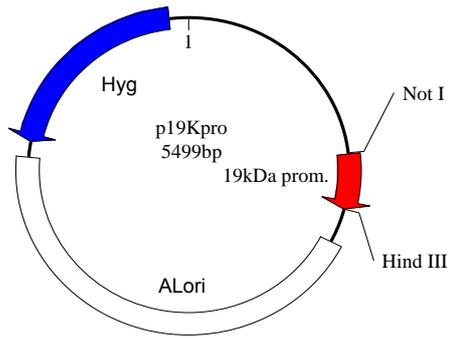
Table 2.1. List of primers used in this study. Primers longer than 25 bases were HPLC purified due to their length.

Primer ID	Sequence (5' → 3')
Primers for Aim 1.	
SmegESAT-6 f	5' ATGAGTTAACAGGTATGGAATTTTCG 3' (25 bases)
SmegESAT-6 r	5' GGATCCGGCAAACATTCCCGTGAC 3' (24 bases)
SmegRv3866 f	5' ATGCCCTCGACGAAACCATCC 3' (21 bases)
SmegRv3866 r	5' TCAGTGCTGGTAGGCGTTGTC 3' (21 bases)
SmegRv3883c f	5' ATGCCCTCGACGAAACCATCC 3' (21 bases)
SmegRv3883c r	5' TCAGTGCTGGTAGGCGTTGTC 3' (21 bases)
PE35 f	5' AAGGGGAGTACGCAAATG TACCCGTACGACGTGCCGGACTACGCCCA ACCGATGACACACAATC3' (64 bases)
PE35 r	5' ATCTGCCTCGAAGAACTA GGCGTAGTCCGGCACGTCGTACGGGTAC ATCGTGGCCGCGTTGGC 3' (63 bases)
PPE68 f	5' TAGGGACACCAACCGATG TACCCGTACGACGTGCCGGACTACGCC GTTGCGGCCCGCCTATC 3' (63 bases)
PPE68 r	5' TTCTTGTGCGAGTGTTC AGCGTAGTCCGGCACGTCGTACGGGTAC CACTCGTCGTCTTCATC 3' (63 bases)
Sall f	5' CGTGCTGCTG GTCGAC GACTGGAC 3' (25 bases)
BclI r	5' GCCACG TGATCA TCCGCATCG 3' (21 bases)
BstXI r	5' CTC CAAATCCATGG CGAGTCTCC 3' (26 bases)
SphI f	5' GGCAAGAACATGCAGCAG GCATGC 3' (25 bases)
BglII r	5' GAGGTGTGGATGTTGCC AGATCT 3' (25 bases)
PstI f	5' GCGGTCACCGCATTGGTGC CTGCAG 3' (25 bases)
PstI r	5' GCCGAGACCTCGTCAGCG CTGCAG 3' (25 bases)
PE f	5' GGATCC ATGCAACCGATGACACACAATCCGG 3' (31 bases)
PE r	5' AAGCTT TCAG GGCGTAGTCCGGCACGTCGTACGGGTAC ATCGTGGCCGCGTTGGC 3' (54 bases)
PPE f	5' GGATCC GTGTGGATCGCTCTACCACCGAGA 3' (31 bases)
PPE r	5' AAGCTT TCAG GGCGTAGTCCGGCACGTCGTACGGGTAC CACTCGTCGTCTTCATC 3' (54 bases)
Primers for Aim 2.	
Rv0978c f	5' GGATCC ATGTCGTTTGTCAACGTGGCC 3' (27 bases)
Rv0978c r	5' AAGCTT TCAG GGCGTAGTCCGGCACGTCGTACGGGTAG CTGATTACCGA CACCGTG 3' (55 bases)
Rv0978c outer f	5' GGCAGGGATCGTCCGAATAA 3' (20 bases)
Rv0978c outer r	5' AGCGGGCGTTCTCGATTGT 3' (19 bases)
Rv1818c f	5' GGATCC ATGTCATTGTGGTACGATCCCGG 3' (31 bases)
Rv1818c r	5' AAGCTT TCAG GGCGTAGTCCGGCACGTCGTACGGGTAC GGTAACCGTT CATCCCGT 3' (56 bases)
Rv1818c outer f	5' GCCACAGGGTGTGACGAGGTG 3' (22 bases)
Rv1818c outer r	5' GCCGACAAGGCGTCGTGAGA 3' (20 bases)
Rv2615c f	5' GGATCC ATGTCGTTTGTCAACGTGGCCCCAC 3' (31 bases)
Rv2615c r	5' AAGCTT TCAG GGCGTAGTCCGGCACGTCGTACGGGTAG CCGTCGGCTCCGTTGG 3' (53 bases)
Rv2615c outer f	5' CGTGGCGTACGAGGATTT 3' (20 bases)
Rv2615c outer r	5' GCTCGATGAGCCAAAGGATGT 3' (22 bases)
Rv0442c f	5' GGATCC GTGACAAGCCCGCATTTTGCGTGGT 3' (31 bases)
Rv0442c r	5' AAGCTT TCAG GGCGTAGTCCGGCACGTCGTACGGGTACT CCGAACCGACCGGTGCC 3' (56 bases)
Rv0878c f	5' GGATCC ATGAATTCATGGTGTGCCCGCG 3' (31 bases)
Rv0878c r	5' AAGCTT TCAG GGCGTAGTCCGGCACGTCGTACGGGTAC CCGCTGTTCC TACTTTTT 3' (56 bases)
Primers for Aim 3.	
Upmsmeg0605 f	5' GG GGTACC GGAGCATCCGCTGCAGACC 3' (27 bases)
Upmsmeg0605 r	5' GGGG AGATCT CTCTCCCTTATGTATGCC 3' (28 bases)
Downmsmeg0616 f	5' GGGG AGATCT CGATCCCAGTGCCTCCACA 3' (29 bases)
Downmsmeg0616 r	5' GGGG AAGCTT CCCGAGCGATCCTTTCC 3' (27 bases)
Upmsmeg1532 f	5' GGGG GTCGAC GACGAGGAAGAGGCG 3' (25 bases)
Upmsmeg1532 r	5' GGGG TTCGAA CCGTGCTGTGAACGAAACC 3' (29 bases)
Downmsmeg1538 f	5' GGGG TTCGAA CCGACGACTCGTTGGAGCG 3' (29 bases)
Downmsmeg1538 r	5' GGGG AAGCTT CCGCGCAGTCGCCGT 3' (26 bases)
Region 3 F1-f	5' GCAGTGGTTCTCCGAGCGTGG 3' (21 bases)
Region 3 F1-r	5' ACGACGTCCGACCAGCGTTGG 3' (21 bases)
Region 3 F2-f	5' CGCTGAACCCGTTGTGACCG 3' (21 bases)
Region 3 F2-r	5' GTTCGGTGATGGACCATCTGG 3' (21 bases)
Region 4 F1-f	5' GAGATCGCGATGGCCATCGCC 3' (21 bases)
Region 4 F1-r	5' CGCGTAACCGTTGCCGTGCC 3' (21 bases)
Region 4 F2-f	5' CCAACGCCTGGGGCGTCTGAT 3' (21 bases)
Region 4 F2-r	5' CTGCACGGCAACCTGTCGCAG 3' (21 bases)

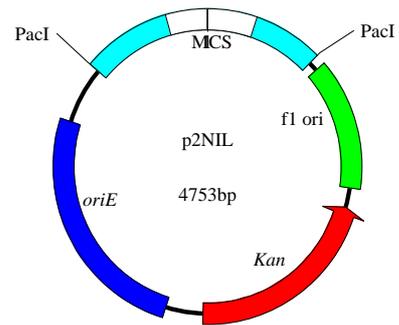
The Hemagglutinin (HA) tag is indicated in red, and restriction sites are highlighted in pink. *Sall*-**GTCGAC**, *BclI*-**TGATCA**, *BstXI*-**CAAATCCATGG**, *SphI*-**GCATGC**, *BglII*-**AGATCT**, *PstI*-**CTGCAG**, *BamHI*-**GGATCC**, *HindIII*-**AAGCTT**, *KpnI*-**GGTACC**, and *SfuI*-**TTCGAA**.

Table 2.2. Bacterial strains, plasmids, cosmids and probes used in this study. Plasmids carrying ampicillin, hygromycin, and kanamycin resistance markers are indicated by amp^R , hyg^R and kan^R , respectively

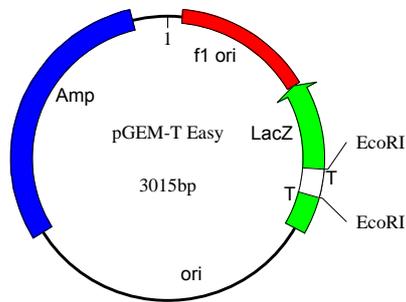
Strains or Plasmids	Description	Reference
Strains		
<i>E. coli</i> XL-1 Blue	Cloning host	Stratagene
<i>E. coli</i> DH5 α	Cloning host	Promega
<i>M. smegmatis</i> mc ² 155	Mycobacterial host strain	Snapper <i>et al.</i> , 1990
<i>M. tuberculosis</i> H37Rv	Laboratory strain of <i>M. tuberculosis</i>	ATCC 27294
Plasmids		
pGEM-T Easy	TA cloning vector (amp^R)	Promega
p19Kpro	Mycobacterial expression vector (hyg^R)	Gift from Koen de Smet
p2NIL	Gene manipulation vector (Cloning vector), (kan^R)	Parish, and Stoker, 2000, gift from V. Mizrahi
pGOAL 17	Plasmid carrying <i>lacZ</i> and <i>sacB</i> genes as a <i>PacI</i> cassette (amp^R)	Parish, and Stoker, 2000, gift from V. Mizrahi
pBluescript	Cloning vector (amp^R)	Stratagene
pYUB412 library	Single copy cosmid library containing <i>M. smegmatis</i> genomic DNA fragment inserts of approximately 40 000 bp, (Hyg^R and Amp^R)	Gift from Franz- Christoph Bange
Probes		
P1	288 bp PCR product of <i>M. smegmatis</i> ESAT-6 gene	This study



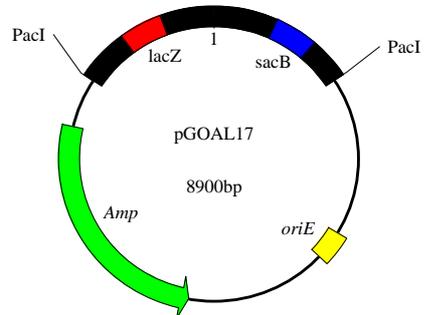
1. p19Kpro vector



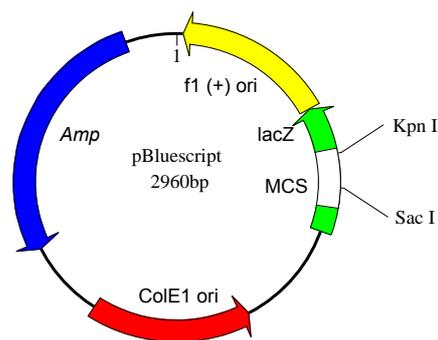
2. p2NIL vector



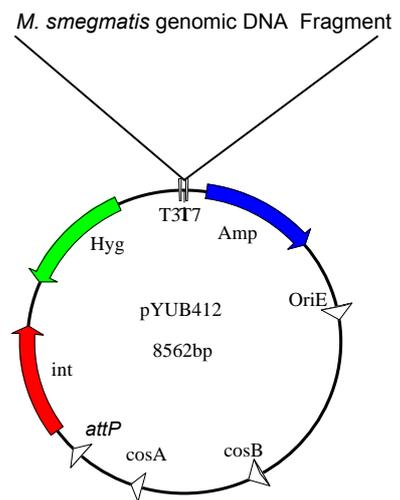
3. pGEM-T Easy vector



4. pGOAL17 vector



5. pBluescript vector



6. pYUB412 cosmid library

Figure 2.2. Plasmids vectors and cosmid library used in this study.

2.7. Molecular cloning and purification of *M. tuberculosis* and *M. smegmatis* genes

The ORFs of the *M. smegmatis* PE35 (Rv3872) and PPE68 (Rv3873) genes, respectively, were amplified using the “Megaprimer” method of PCR-based site-directed mutagenesis (SDM) (Smith and Klugman, 1997) in order to insert an N- or C- terminal hemagglutinin (HA) tag into the gene sequences of these genes in the genome of *M. smegmatis*. This method makes use of three primers and two amplification steps, as illustrated in Figure 2.3 (Smith and Klugman, 1997; and Ke and Madison, 1997). The Primer Premier 5 software was used to design the primers. The *M. smegmatis* mc²155 genome was used as template with the primers shown in Table 2.1., bearing restriction sites as highlighted in pink. The HotStar Taq polymerase PCR kit (Qiagen), Expand long template PCR system (Roche), and AdvantageTM-GC 2 PCR kit (BD Biosciences), were used for the amplification of the two genes depending on the length and GC content of each gene. In addition, another set of primers was designed for both genes which were C-terminally HA tagged, bearing restriction sites *Bam*HI and *Hind*III as highlighted in pink in Table 2.1. The two genes were PCR amplified using HotStar Taq polymerase PCR kit (Qiagen), and *M. smegmatis* DNA as the template. Five of the recently duplicated subfamilies of the PE and PPE genes; i.e. three PGRS genes, Rv1818c, Rv0978c, and Rv2615c and two MPTR genes, Rv0442c and Rv0878c genes were also selected, and were amplified by polymerase chain reaction (PCR) using *M. tuberculosis* H37Rv chromosomal DNA as template, in combination with the HotStar Taq polymerase kit (Qiagen). Specific primer pairs bearing *Bam*HI and *Hind*III restriction sites on their 5' end are as highlighted in pink were used in this study, and are listed in Table 2.1. A Hemagglutinin (HA) tag was incorporated in the 3' end of the primers. The amplicons were sequenced on an automated sequencer using SP6 and T7 sequencing primers to confirm the sequences. The restriction enzymes *Bam*HI (Roche) and *Hind*III (Roche) were used to digest the inserts cloned into pGEM-T Easy vector and the inserts were purified. The fragments corresponding to the amplified genes were ligated into a mycobacterial expression vector (p19Kpro), which was digested with *Bam*HI and *Hind*III, by incubating overnight at 4°C. Forty microliters (40 µl) TE buffer was added to the ligation reaction which was purified using the Wizard[®]SV Gel and PCR Clean-Up System (Promega). The insert sequence was confirmed by sequencing, and once the correct orientation of the insert had been determined by restriction analysis, the constructs were transformed and allowed to express in *M. smegmatis*, illustrated in Figure 2.4.

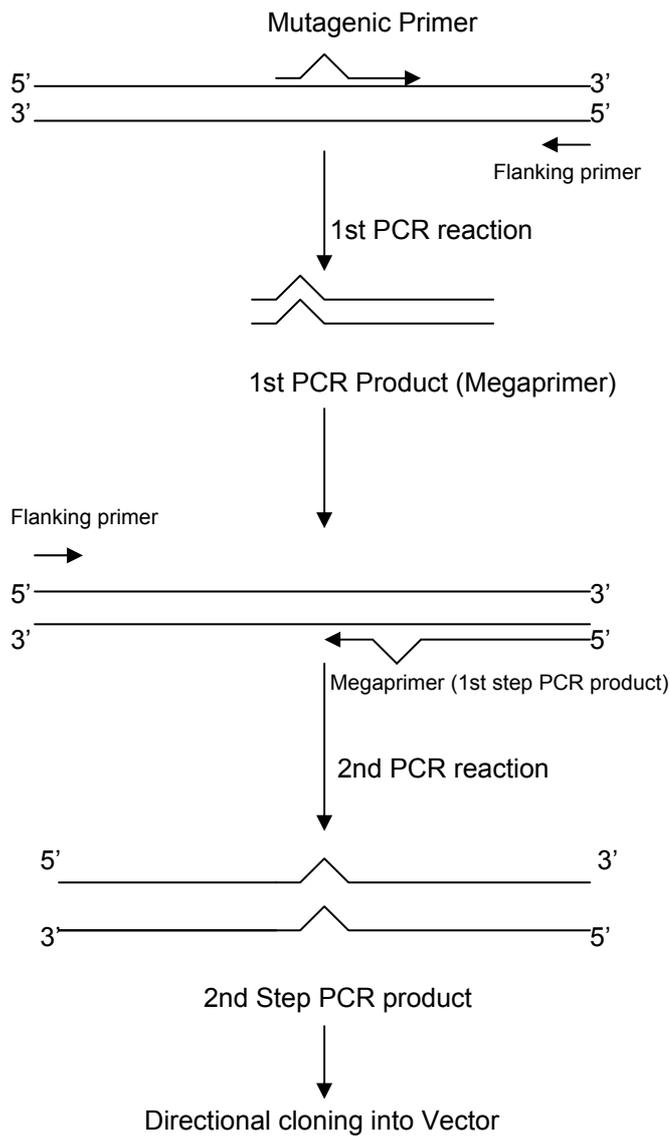


Figure 2.3. Schematic outline of the PCR based site-directed mutagenesis protocol. In the first step, the mutant antisense oligomer and a flanking reverse primer were used to amplify and incorporate the mutation in a desired site of DNA. The product of the 1st step (megaprimer) was used as either a reverse primer or forward primer in the second step with a flanking primer. The product of the 2nd PCR step was obtained, flanked with restriction sites to allow directional cloning. Reproduced from Ke and Madison (1997).

***M. tuberculosis* H37Rv genomic DNA/*M. smegmatis* genomic DNA**

PCR amplified the following genes:

PE 35, PPE68, PE_PGRS: Rv2615c, Rv1818c, and Rv0978c, and PPE-MPTR: Rv0878c, and Rv0442c

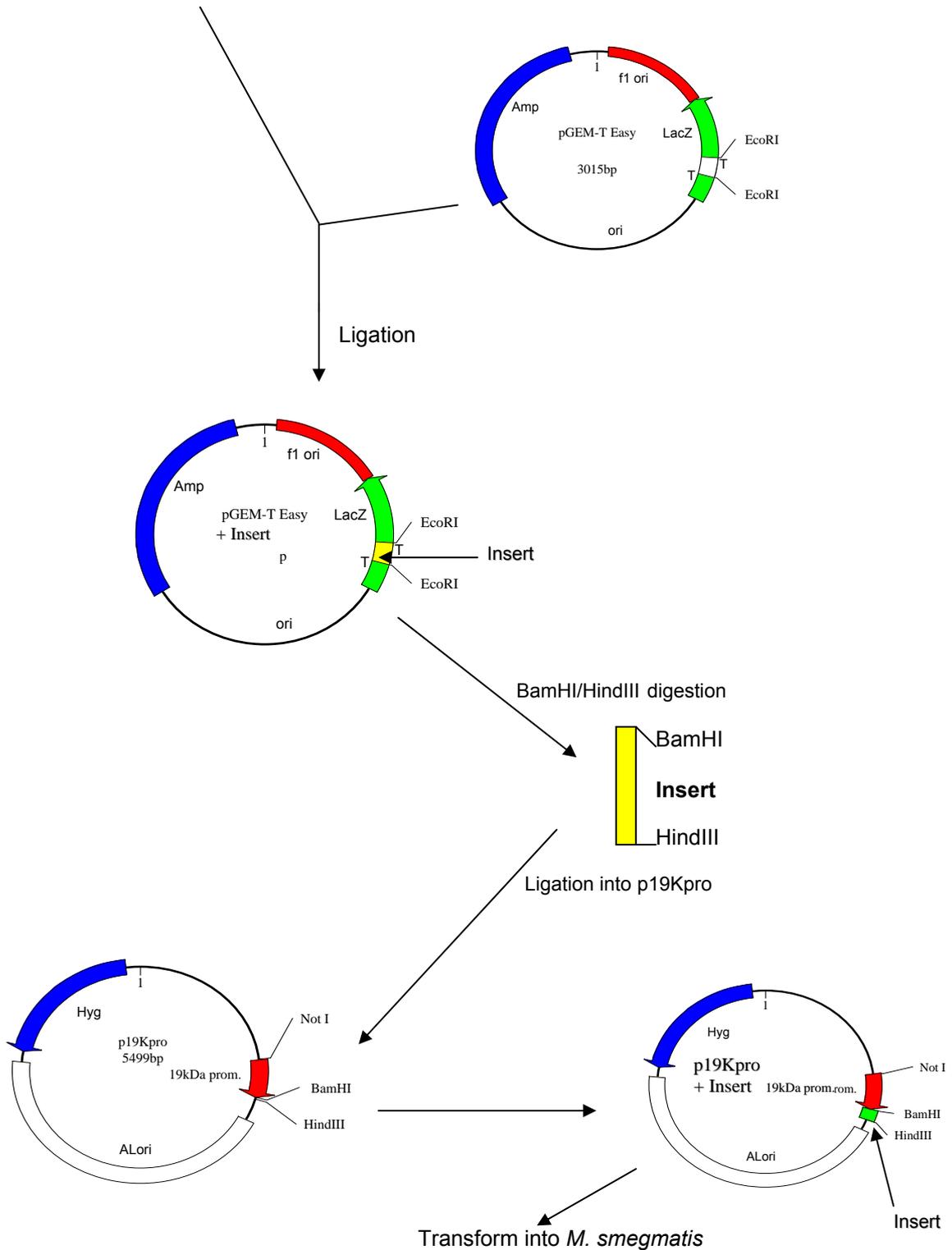
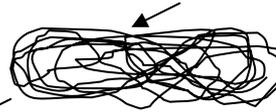
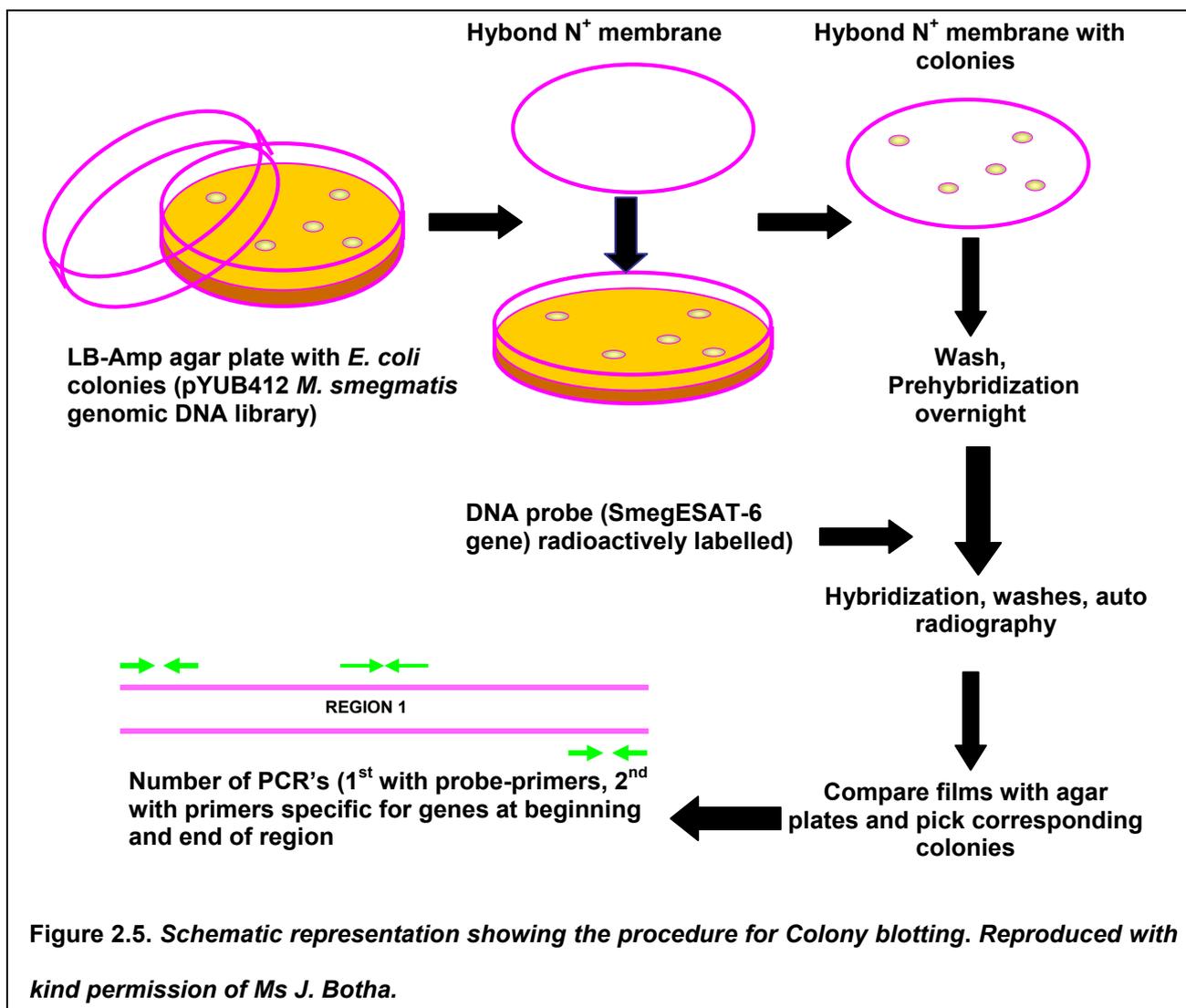


Figure 2.4. Schematic representation of the main steps in the cloning of the PE35, and PPE68; PE_PGRS (Rv2615c) and PPE-MPTR (Rv0442c and Rv0878c) from the ESAT-6 gene cluster region 1 of *M. smegmatis* and *M. tuberculosis* respectively.

2.8. Isolation of cosmids containing selected genomic regions

Cosmids containing selected *M. smegmatis* mc²155 ESAT-6 gene cluster region 1 were isolated by colony blotting, as described by Sambrook *et al* (1989), shown diagrammatically in Figure 2.5. Briefly, approximately 2000 clones of the *M. smegmatis* genomic DNA cosmid library (in pYUB412) were spread onto LB plates containing Ampicillin (50 µg/ml, Roche) and allowed to grow at 37°C for 16 hours. Bacterial colonies were transferred to Hybond-N⁺ membrane filters (Amersham Biosciences) by placing the membrane onto the colonies and incubating for one minute. The membrane was then subjected to several steps of denaturing, neutralization, and washing with denaturing solution, neutralizing solution and 2x SSC, respectively after which the filters were baked for 2 hours at 80°C in a vacuum oven. After baking, the membranes were washed and hybridized overnight at 42°C using [α -³²P] dCTP-radioactively labeled probe, which was complementary to an internal region of a gene in the ESAT-6 gene cluster region. For probing the ESAT-6 gene cluster region 1, the *M. smegmatis* ESAT-6 gene Rv3875 was chosen (see Table 2.1 for primers and Table 2.2 for probe). Labeling of the radioactive probe was done using the commercial Prime-It RmT Random Primer Labeling Kit (Stratagene) according to the manufacturer's instructions. Labeled probe was purified from unincorporated radioactive nucleotides using a G50M Sephadex desalting spin column. Positive clones were visualized by autoradiography, and corresponding colonies were subsequently picked (as the cosmid sizes are around 40 000 kb, positive clones were obtained with a frequency of \pm 1:300). Positive clones were confirmed by PCR of the gene used as probe (*M. smegmatis* ESAT-6 from region1), using the SmegESAT primers in Table 2.1., and as well as the primers of the genes present in the first and last positions of the ESAT-6 gene cluster region 1, the *M. smegmatis* orthologues of Rv3866 (MSMEG_0057) and Rv3883c (MSMEG_0083), respectively. PCR was done using the HotStar Taq system (Qiagen), according to the manufacturer's conditions. The cosmids were purified using the Wizard[®]SV Gel and PCR Clean-Up System (Promega). Finally, the inserts of the isolated cosmids were sequenced on an automated sequencer using T3 and T7 flanking sequencing primers to confirm correct DNA region as well as to determine exact start and stop of insert DNA. Radioactive labelling of the probe, hybridization and detection is as discussed in Addendum B.



2.9. Transformation in *M. smegmatis*

M. smegmatis mc²155 competent cells were used for the transformation of the plasmid constructs by electroporation as described previously (Jacobs *et al.*, 1991). The transformed cells were plated onto Middlebrook's 7H11 agar supplemented with 1/10 volume filter-sterilized 10% oleic acid-albumin-dextrose catalase (OADC) (0.005 oleic acid, 0.5% BSA, 0.2% glucose, 0.015% catalase) (Bactlab Systems); and containing 0.2% glycerol. Liquid cultures were grown in Middlebrook 7H9 medium supplemented with filter-sterile 10% ADC (0.05% BSA, 0.2% glucose, 0.015 catalase) and containing 0.05% Tween-80 (Sigma) and 0.2% glycerol with appropriate antibiotics (Hygromycin- 50 µg/ml, Roche; Kanamycin-50 µg/ml, Roche and/or Ampicillin- 50 µg/ml, Roche).

2.10. Ziehl-Nielsen (ZN) Staining

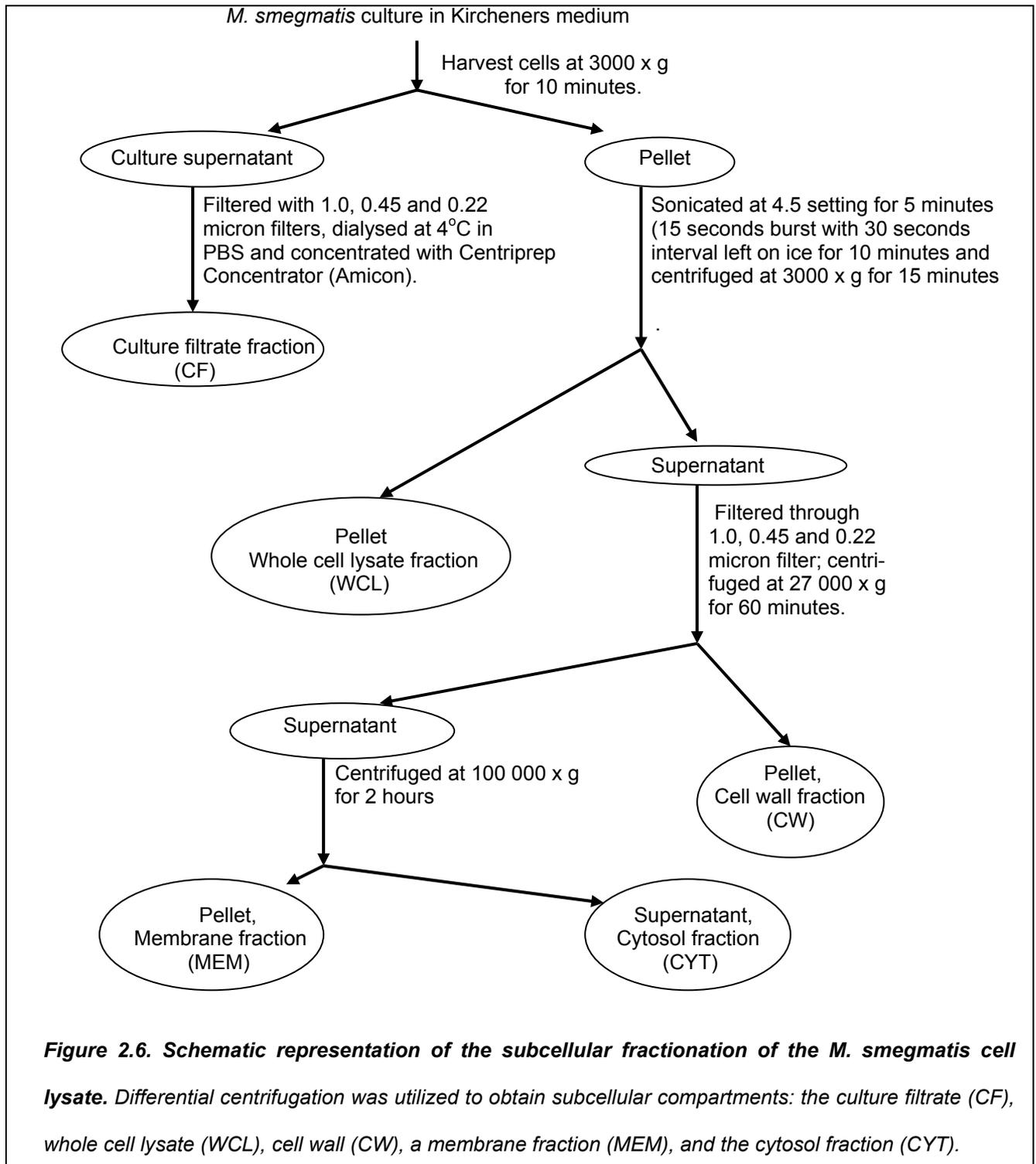
To confirm that there was no contamination of the *M. smegmatis* culture, a Ziehl-Nielsen (ZN) staining was made from the culture. A smear of the culture was prepared by fixing a drop of culture onto a microscopic slide and allowing to air dry. The cells were heat fixed by passing the slide through the blue cone of a Bunsen flame three or four times. The slide was flooded with ZN carbol fuchsin and was heated intermittently for 2-5 minutes (until plenty of steam comes off the slide). The slide was washed with tap water (without blotting). Acid-alcohol 3% was used to decolourize until no more stain came off the slide, and the slide was washed with tap water (without blotting). The slide was counter-stained with methylene blue for 1-2 minutes, after which it was washed with tap water and blot dried. The slide was examined under the microscope using the oil-immersion objective, 100 X magnification. Mycobacteria appeared as red/pink- stained rods or coccobacilli, and non- acid fast contaminants were identified by a blue colour.

2.11. Subfractionation for Protein secretion analyses

To monitor the expression of recombinant genes, recombinant *M. smegmatis* clones were grown under conditions as specified in section 2.6. The cells were harvested by centrifugation at 3000 x g for 10 minutes, and washed twice in 10 ml phosphate buffered saline (PBS) (8 g NaCl, 0.2 g KCl, 1.44 g Na₂HPO₄, and 0.24 g KH₂PO₄, pH 7.4, per liter), and then resuspended in 10 ml Kircheners medium (3 g Na₂HPO₄, 4 g KH₂PO₄, 1,07 g MgSO₄.7H₂O, 2.5 g C₆H₅Na₃O₇.2H₂O, 20 ml Glycerol, 5 g Asparagine in 980 ml dH₂O). For the protein secretion assays in liquid media, *M. smegmatis* cultures were grown to an optical density of 0.3 at 600 nm (early log phase), after which the cells were precipitated by centrifugation at 3000 x g for 10 min. The fractionation method is illustrated in Figure 2.6. When grown to an OD₆₀₀ = 0.3, 150 ml of culture filtrate produced approximately 0.5 – 1 mg of concentrated culture filtrate proteins. The culture supernatant was kept as it contained the culture medium and secreted proteins. It was filtered through 1.0, 0.45 and 0.22 micron filters, dialyzed at 4°C overnight against PBS and concentrated with a Centriprep concentrator (Amicon), resulting in the culture filtrate fraction (CF).

The cell pellet was washed twice in 10 ml PBS by centrifugation at 3000 x g and resuspended in 5 ml of PBS containing 1% NP40. Protease inhibitors were added prior to disruption by sonication at 4.5 setting using a cup sonicator on ice for a total of 5 minutes (15-second bursts with 30-second intervals). After standing for 10 minutes on ice, the samples were centrifuged to clear the unlysed cells and cell

residue at 3000 x g for 15 min, twice. NP40 (0.33%), was added to the pellet, which contained the whole cell lysate (WCL) fraction. The supernatant was filter-sterilized through 1.0 μ M, 0.45 μ M, and 0.22 μ M filters, and centrifuged twice for 1 hour at 27 000 x g, whereafter the pellet contained the cell wall fraction (CW). The pellet was resuspended in PBS containing 0.33% NP40. The supernatant was centrifuged further at 100 000 x g for 2 hours and the resulting pellet contained the membrane fraction (MEM), while the resulting supernatant contained the cytosol fraction (CYT). The pellet was resuspended in PBS containing 0.33% NP40. NP40 (0.33%) was also added to the cytosol fraction.



2.12. Bio-Rad Protein assay

The Bio-Rad Protein assay (Bio-Rad laboratories, *GmbH*) was used to determine the concentrations of the protein samples (supplementary data in Addendum C). Bovine Serum Albumin (BSA) (Bio-Rad laboratories, *GmbH*) was used as a standard (dissolved in water). A 2 mg/ml stock was made and 3-5 dilutions of standards, e.g. 0, 0.2, 0.4, 0.5, 0.6, 0.8 mg/ml (see below). Ten microliters (10 μ l) of the standard and 10 μ l of Kircheners medium were added together. Samples were diluted with the Kircheners medium, and from these 10 μ l of sample and 10 μ l of water were added together in a clean Eppendorf tube. The Bio-Rad reagent was diluted with water 1:4 (i.e. 1 part of the Bio-Rad reagent added to 4 parts of distilled water). The diluted reagent was added to each standard and samples (making a final volume of 20 μ l) including the blank (containing lysis buffer and water). The samples were left at room temperature for 5 minutes, and the absorbance was measured at 595 nm. A standard curve was made with standards and the protein concentrations of samples were read from the standard curve.

Table 2.3. BSA standard preparations

BSA dilutions (mg/ml)	2 mg/ml BSA stock (μ l)	H ₂ O (μ l)
0	0 μ l	10 μ l
0.2	1 μ l	9 μ l
0.4	2 μ l	8 μ l
0.5	2.5 μ l	7.5 μ l
0.6	3 μ l	7 μ l
0.8	4 μ l	6 μ l
1.0	5 μ l	5 μ l

2.13. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), and Western blotting and detection.

2.13.1. SDS-PAGE

The gel apparatus was assembled as per manufacturer's instructions (Bio-Rad), the two duplicate separating gels were prepared as follows:

(a) 10% separating gel:

Components	Volume
Acrylamide/bis-acrylamide, 30:0.8 (% w/v)	3.33 ml
1.5 M Tris-HCl, pH 8.8	2.5 ml
10% SDS	0.1 ml
H ₂ O	4.0 ml
TEMED	10 ul
10% ammonium persulphate (APS)	50 ul

The separating gel was poured between the glass plates. Water was added to the top of the gel to allow acrylamide to polymerize. After setting, the water was poured off, and the stacking gel was added on top of the set separating gel. The stacking gel was prepared as follows:

Components	Volume
Acrylamide/bis-acrylamide, 30%:0.8 (% w/v)	1.3 ml
1M Tris-HCl pH 6.8	1.25 ml
10% SDS	0.1 ml
H ₂ O	7.4 ml
TEMED	20 ul
10% ammonium persulphate (APS)	50 ul

An equal volume of 2 x SDS Sample Buffer (1 ml β -mercaptoethanol, 3.4 ml Tris-HCl (pH 6.8), 3 ml SDS (20%), 2ml glycerol, 500 μ l Bromophenol blue (0.75%) and 200 μ l EDTA (0.5 M)) was added to the protein samples (5-10 μ g), and then mixed, after which they were placed in a boiling water bath for 5 min. Running buffer was poured into the upper and lower chambers of the Western blot apparatus so that the top and bottom of the gel and both electrodes were covered. The samples were loaded into the wells (approximately 15 - 20 μ l), as well as a protein molecular weight marker in one lane, for the

determination of the protein sizes. The proteins were fractionated by electrophoresis at 200V (constant voltage) on SDS-PAGE (10% separating gel, and 3% stacking gel) until the dye front reached the end of the gel (for approximately 1 hour). After electrophoresis, one gel was stained with Coomassie Blue staining solution (2.5 g Coomassie Blue R250, 450ml methanol, 100 ml acetic acid, and 450 ml distilled water) for 1 hour, and were destained overnight. The other gel was used for Western Blotting. SDS-PAGE buffers and solutions were prepared as in Addendum D.

2.13.2. Western Blotting and detection.

After electrophoresis, the proteins were transferred to nitrocellulose membrane by Western blotting. Two Western blot pads, two Whatmann 3M sheets, and one Hybond ECL nitrocellulose membrane (Amersham Biosciences) were pre soaked in a large tray filled with Transfer Buffer (3.03 g, 25 mM Tris; 14.4 g, 192 mM Glycine; 200 ml, 20% Methanol and 800 ml distilled water per liter). The nitrocellulose was cut to a suitable size and floated on top of the blotting buffer before totally immersing it to prevent air entrapment. The membranes were sandwiched, with the gel lying squarely on top of the membrane, between two pieces of Whatman 3M sheets, which were inserted between two Western blot pads and positioned in the sandwich of the Western blotting apparatus. The proteins were transferred to the nitrocellulose membrane in the transfer buffer for 60 minutes (1 hr) at 200 V, in the cold room. After transfer the nitrocellulose membranes were stained in Ponceau S solution (1.0 g Ponceau S Red dye and 1 ml acetic acid, dissolved in 100 ml of distilled water) for 1 minute. The positions of any visible marker bands were marked and the nitrocellulose membranes were blocked overnight in 100 ml 10% Blocking buffer (0.1% Tween), containing sodium azide (NaN_3) to prevent microbial growth.

Thereafter, the membranes were rinsed well in Tris-buffered saline-Tween (TBS-T) buffer (50 ml Tris (1 M), pH 7.6; 137 ml NaCl (5 M), 5 ml Tween 20 and make up to 5 liters). Mouse anti-HA monoclonal antibodies (HA.11, Clone 16B12, Covance) was used at a dilution of 1/1000 diluted in TBS-T to detect the HA-tagged recombinantly expressed *M. tuberculosis* and *M. smegmatis* proteins (expressed from p19Kpro). This primary antibody was incubated for 60 minutes at room temperature with shaking. Anti-HA antibodies have a high specificity to allow unambiguous identification of the influenza hemagglutinin epitope (YPYDVPDYA). After incubation with the primary antibody the membranes were washed twice in TBS-T buffer, and then washed three times for 10 minutes in TBS-T buffer with shaking. Goat anti-mouse IgG (H+L) horse radish peroxidase (HRPO)-conjugate antibodies (Caltag Laboratories) combined with peroxidase-conjugated anti-rabbit antibody (ICN, Cappel) were

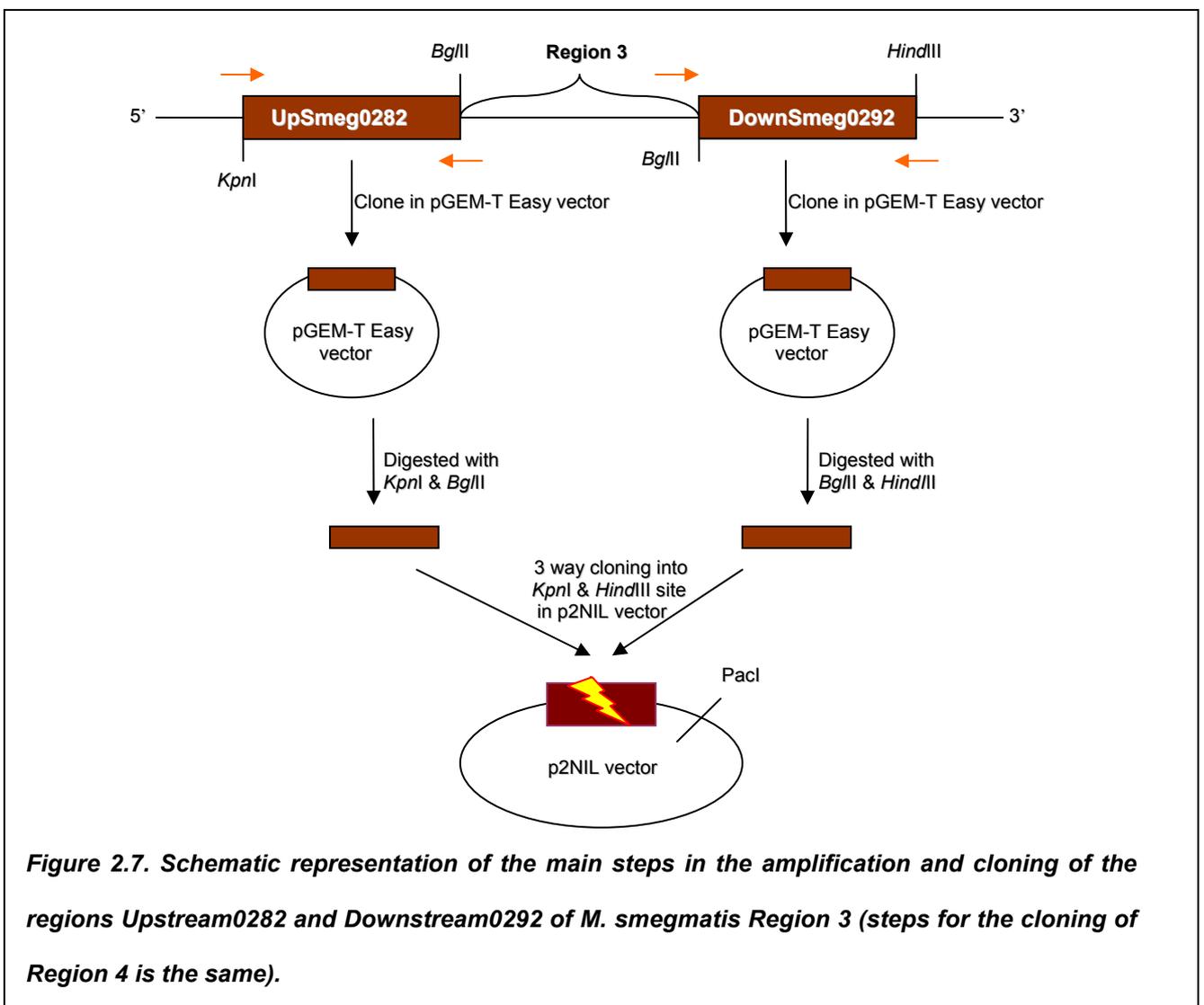
used as secondary antibodies at a concentration of 1/10 000. Chemiluminescent signal was initiated with the addition of the detection reagents from the ECL Plus kit (Amersham Pharmacia Biotech) and the signal was recorded on X-ray film (Hyperfilm Amersham, Pharmacia Biotech). Western blotting buffers were prepared as in Addendum E.

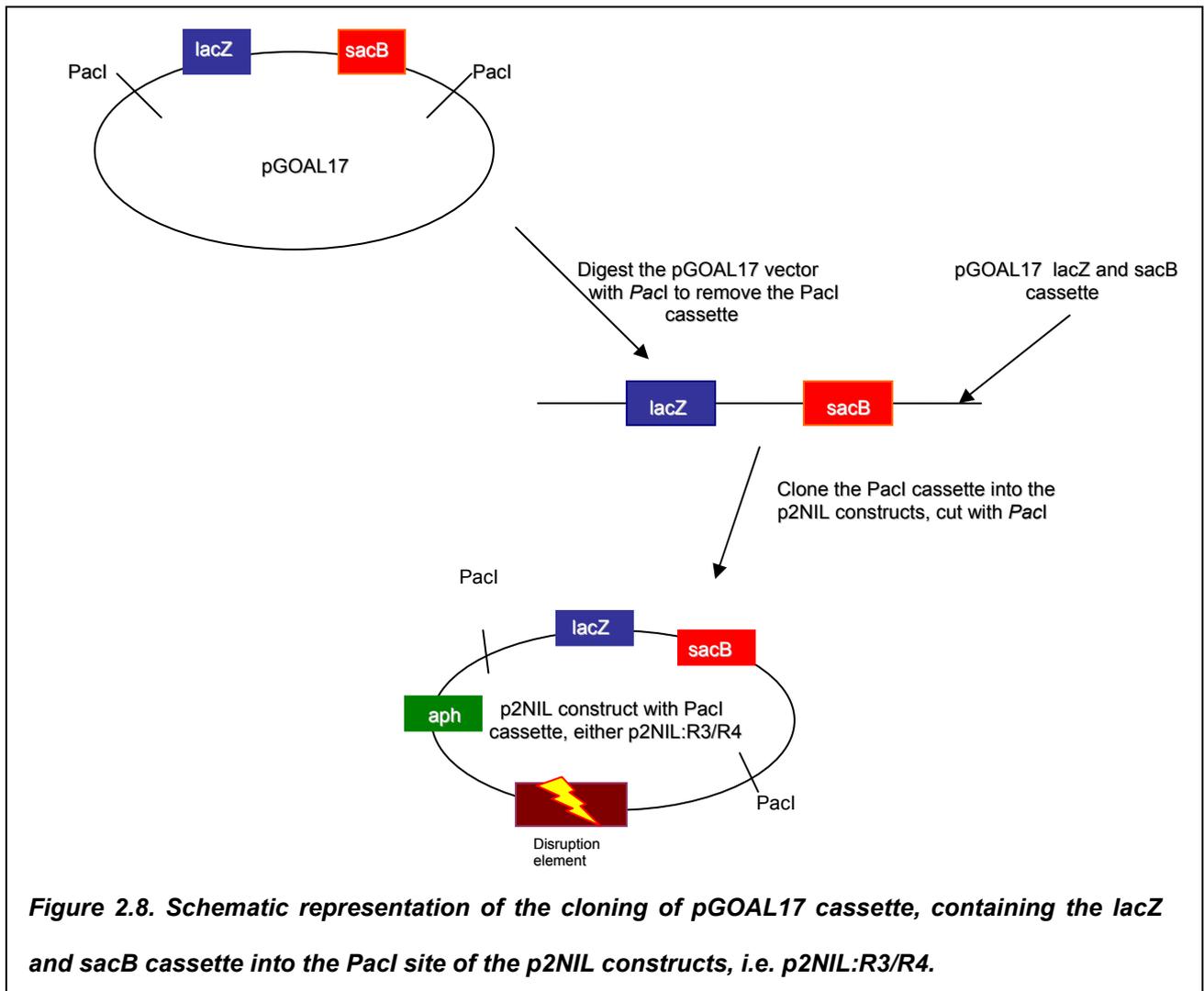
2.14. Knockouts of ESAT-6 gene cluster Region 3 ($\Delta R3$) and Region 4 ($\Delta R4$) of *M. smegmatis* and molecular cloning of the constructs.

For generation of knockouts of *M. smegmatis* ESAT-6 gene cluster Region 3 (R3) and Region 4 (R4), genomic regions (~800 bp each) flanking the N- and C- termini of the two regions (named UpstreamMSMEG0604, and DownstreamMSMEG0617; and UpstreamMSMEG1531, and DownstreamMSMEG1539; respectively) were PCR amplified, using the primers UpMSMEG0604f&r, DownMSMEG0617f&r, UpMSMEG1531f&r, and DownMSMEG1539f&r. The primers contained the restriction sites KpnI, and BglII; and BglII, and HindIII upstream and downstream of R3; and Sall, and SfuI; and SfuI, and HindIII upstream and downstream of R4, respectively. The amplified regions were cloned separately into the T vector, pGEM-T Easy, and were confirmed by sequencing. For Region3, UpstreamMSMEG0604:pGEM-T Easy was digested with KpnI and BglII and DownstreamMSMEG0617:pGEM-T Easy was digested by BglII and HindIII. For Region 4, UpstreamMSMEG3450:pGEM-T Easy was digested with Sall and SfuI and DownstreamMSMEG3444:pGEM-T Easy was digested with SfuI and HindIII (see Figure 2.7.) The resultant fragments were subcloned in the corresponding sites within the mycobacterial suicide vector p2NIL, by three way cloning.

The two constructs, p2NIL: $\Delta R3$ and p2NIL: $\Delta R4$ were both digested with *PacI*. The pGOAL17 vector was also digested with *PacI*, generating two bands of 2.3 kb (the vector band) and 6.3 kb (the *PacI* cassette with *lacZ* and *sacB* genes), as described by Parish and Stoker (2000). The 6.3 kb *PacI* cassette fragment from pGOAL17 carries the *lacZ* gene for the selection of single crossover intermediate clones by blue colour, and the *sacB* gene used for killing of single crossover intermediate clones. The pGOAL17 cassette was cloned into the *PacI* site of p2NIL: $\Delta R3$ and p2NIL: $\Delta R4$ (see Figure 2.8), respectively. The constructs were transformed by electroporation into *E. coli* XL-1 Blue (Stratagene) and incubated for 1 hour at 37°C in an LM-530 Orbital shaker incubator. After incubation, the cultures were spread onto LB-Kanamycin plates containing 50 µg/ml X-gal (5-bromo-4-chloro-3-

indolyl- β -D-galactopyranoside, Roche) and were incubated at 30°C. Colonies containing the PacI cassette were identified by blue colour. Thereafter blue colonies were patched onto duplicate plates, i.e. LB plates containing 2.5 - 5% sucrose and LB plates without sucrose, and incubated at 30°C to check for sensitivity of *sacB* cassette. The clones that showed no growth (died) in the presence of sucrose were then picked from the duplicate plate where no sucrose was added. These colonies were inoculated into 150 ml of LB medium containing 25 μ g/ml Kanamycin and incubated overnight at 37 °C. The constructs were purified from the large cultures using Nucleobond AX anion exchange columns (Macherey-Nagel, Germany).





2.15. Delivery of constructs into *M. smegmatis*.

Four micrograms of each of the p2NIL constructs were electroporated into *M. smegmatis*, at 2.5 kV, 1000 Ohm, 25 uF and allowed to grow for 3 hours at 37°C after which it was spread onto LB plates containing 50 µg/ml Kanamycin and 20 µg/ml X-gal. The plates containing the constructs were incubated at 37°C for 3-4 days. The blue colonies were identified as single crossovers (SCO), and they were picked and inoculated into 500 µl LB. From the inoculum, 20 µl were inoculated into five tubes, four of which contained 2 ml of LB + Tween 80 (0.05%), and a fifth to which kanamycin was added. The tubes were incubated for 3-4 days with shaking at 37°C. After 3-4 days of incubation serial dilutions were made from the four tubes without kanamycin by pipetting out 50 µl into 450 µl LB containing Tween 80. One hundred microliters from the 10⁻³ to the 10⁻⁶ dilution was spread onto X-Gal plates and from the undiluted culture 200 µl and the neat was spread onto X-Gal plus 2.5% sucrose plates. The plates were incubated for 3-4 days at 37°C. Each of the white colonies which grew on the X-gal/sucrose plates could be a double crossover (DCO) or a revertant to wild type (Figure 2.9). Two colonies were picked from X-Gal and

sucrose plates with a yellow tip and inoculated into 50 µl of distilled water (dH₂O) used as controls. Using the same tip, the remaining cells were inoculated into 50 µl LB +Tween 80. The 50 µl LB +Tween 80 tubes were incubated at room temperature for 3-4 days after which they were inoculated into 2 ml LB + Tween 80 and grown for Cetyltrimethyl-ammonium bromide (CTAB) Miniprep.

2.16. Genomic DNA preparation using CTAB Method

One and half microliters (1.5 ml) of culture was pipetted into an Eppendorf tube from a culture grown for 3-4 days in LB containing Tween 80 in the absence of Kanamycin. The cells were heat killed at 80°C for 30 minutes. Thereafter the cells were precipitated by centrifugation at 9000 x g in microcentrifuge tubes for 30 seconds and the supernatant was discarded. Five hundred microliters of TE buffer containing 50 µl of lysozyme (10 mg/ml) was added into the pellet, and the pellet was incubated at 37°C for 60 minutes. Seventy microliters of 10% SDS containing 6 µl of proteinase K (10 mg/ml) was added into the mixture and the samples were further incubated at 65°C for 120 minutes. One hundred microliters of 5 M NaCl and 80 µl of CTAB/NaCl solution (10% CTAB prepared in 0.7 M NaCl) were added to the samples after incubation, and the mixture was incubated at 65°C for 10 minutes. Eighty microliters of 10% CTAB containing NaCl solution was added into the mixture and mixed. This was incubated at 65°C for 10 minutes, and thereafter chloroform isoamylalcohol (CIAA), at a ratio of 24:1 was added, and mixed. The mixture was centrifuged for 5 minutes at a speed of 9 000 x g. The aqueous phase was transferred into a clean 1.5 ml Eppendorf tube and 450 µl of Isopropanol was added to this. The mixture was incubated on ice for 30 minutes, and the DNA was precipitated by centrifugation for 20 minutes at 9 000 x g. The pellet was washed with 1 ml 70% ethanol, dried in the spin vacuum at 65°C and resuspended in 20 µl of dH₂O. The product of the CTAB Minipreps was used as template for PCR analysis (FastStart Taq DNA Polymerase kit, Roche) using primers flanking the deletion regions and primers within the wild type to confirm whether a DCO allele or a deletion occurred, or if there was reversion to wild type. *M. smegmatis* mc²155 was used as positive control for R4 and water was used as negative control for R3. Results were confirmed by sequencing.

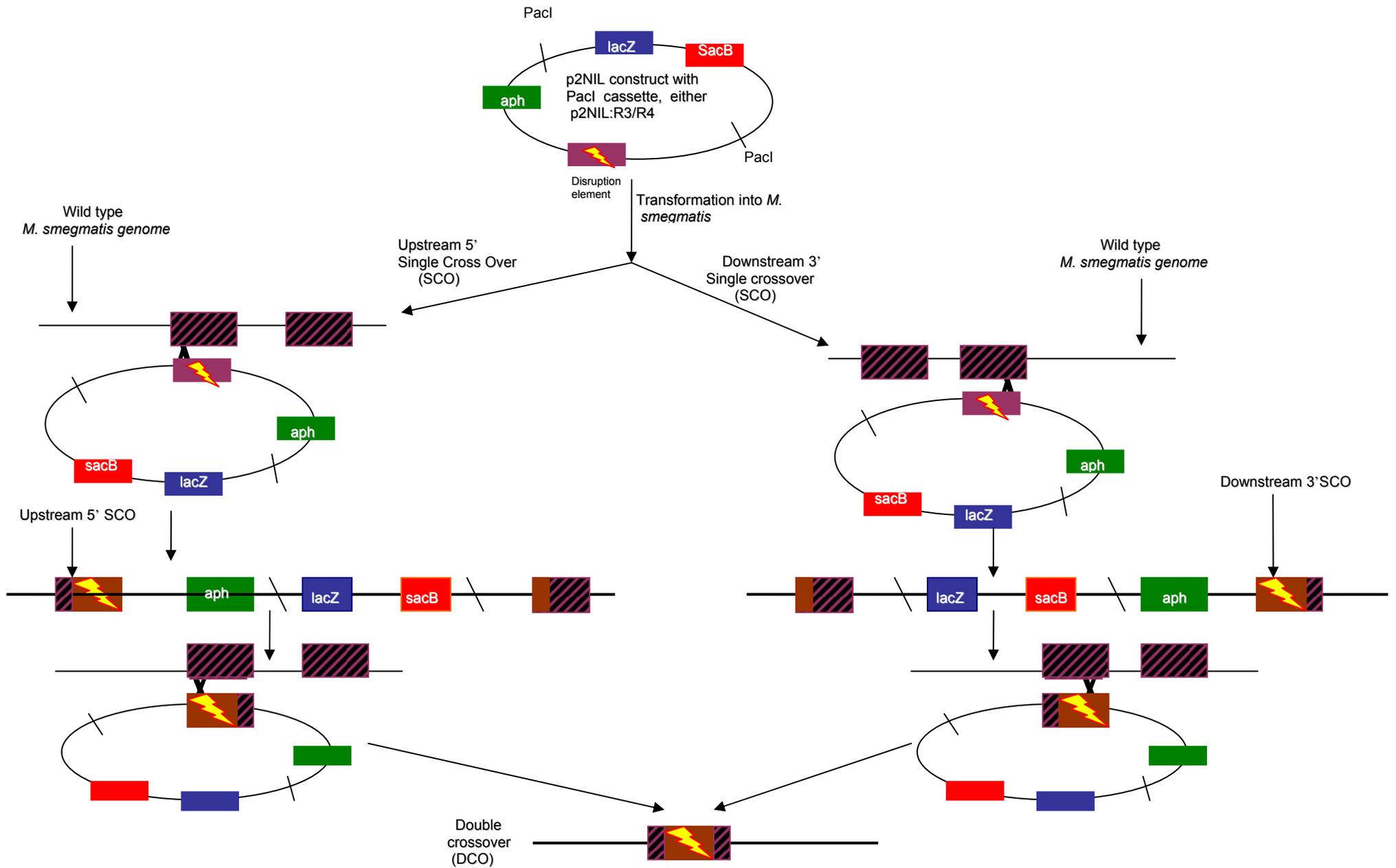


Figure 2.9. Schematic representation of the main steps in the deletion of Region 3 (Δ MSMEG:R3) and Region 4 (Δ MSMEG:R4) in *M. smegmatis*.

CHAPTER 3. RESULTS

3.1. The localization of the most ancestral PE and PPE proteins found in *M. smegmatis*.

3.1.1. PCR amplifications

The *M. smegmatis* orthologue of the PE35 (Rv3872) gene was amplified using the primers PE35f and *Bcl*Ir to obtain the megaprimer. To obtain the N-terminally HA- tagged PE35, the product of the first PCR reaction, PE35f megaprimer and *Sa*llr were used in order to engineer the restriction site on both ends of the amplification product to allow directional cloning. The PE35r megaprimer was obtained using the primers PE35r and *Sa*llf. To obtain the C-terminally HA-tagged PE35, the PE35r megaprimer and *Bc*llf were used and there was no amplification obtained. The mutagenised product of PE35-N was cloned into the pGEM-T Easy vector and the product was confirmed by sequencing. The *M. smegmatis* orthologue of the PPE68 (Rv3873) gene, was also amplified using the “Megaprimer” method. *M. smegmatis* mc²155 genomic DNA was used as template. The PPE68f and PPE68r megaprimers were amplified using the primers PPE68f, and *Bst*Xlr; and PPE68r, and *Sa*llf respectively. When using the PPE68f and PPE68r megaprimers in order to engineer the restriction site on both ends of the amplification products, no amplification was obtained. The reason why we failed to obtain the desired result might be because the expected products were very long. The other reason could be that they might have had very GC-rich secondary structure since we failed to obtain amplification after adding reagents such as DMSO and the Q buffer which helps to overcome GC-rich structures.

A second set of primers for both genes were designed, which had smaller products as compared to the first strategy. We then PCR amplified PE35f, PE35r, PPE68f and PPE68r to obtain the megaprimers using primers bearing the restriction site. Amplification was achieved using the HotStar PCR kit (Qiagen). The megaprimer of PE35-N was obtained using the primers PE35f and *Pst*lr generating an amplicon of size, 158 bp. When using the megaprimer in the second PCR step as one of the primers in order to engineer another restriction site on the other end, no amplification was obtained. The megaprimer of the N terminally tagged PE35 was obtained using the primers PE35f and *Bc*llr and the PE35f megaprimer and *Sa*llr were used in order to engineer the restriction site on both ends of the amplification product to allow directional cloning. The mutagenised product of PE35-N was cloned into the pGEM-T Easy vector and the product was confirmed by sequencing. When

trying to replace the untagged gene with the tagged PE35-N through homologous recombination in *M. smegmatis*, could not succeed. The megaprimer of PE35-C was obtained using the primers PE35r and *Pst*I, generating an amplicon of 237 bp. To obtain the C-terminally HA-tagged PE35 the product of the first PCR reaction, PE35r megaprimer and *Sa*I were used to generate an amplicon of size, 802 bp in order to engineer the *Sa*I site on the other end. When cloning the 802 bp amplicon of PE35-C into p2NIL vector for further analysis, could not succeed. PPE68-N was PCR amplified using primers PPE68f and *Bc*I to obtain the megaprimer, generating an amplicon size of 262 bp. When using the megaprimer and *Sa*I to engineer the restriction site on the other end to allow directional cloning, there was no amplification obtained. The PPE68-C megaprimer was obtained using the primers PPE68r and *Sph*I generating an amplicon of 585 bp. There was no amplification of the desired product when using the megaprimer of PPE68-C and *Bg*I to engineer the *Bg*I restriction site (Figure 3.1. showing the restriction sites and the expected product of the megaprimers). Due to the negative results obtained with both the engineering of the restriction sites and the cloning for the megaprimer, new pairs of primers were designed for both the PE35 and PPE68 of *M. smegmatis*. These primers both contained C-terminally HA tags. The PE35 and PPE68 genes were amplified from *M. smegmatis* DNA using the HotStar Taq polymerase kit (Qiagen), generating amplicon sizes of 333 bp and 1335 bp, respectively. The amplicons were sequenced in order to confirm the base sequence, cloned into the pGEM-T Easy vector and subcloned into the mycobacterial expression vector, p19Kpro. These were transformed and expressed in *M. smegmatis*. The results for the amplification of all the genes are shown in Addendum A, Figure A1 to A4.

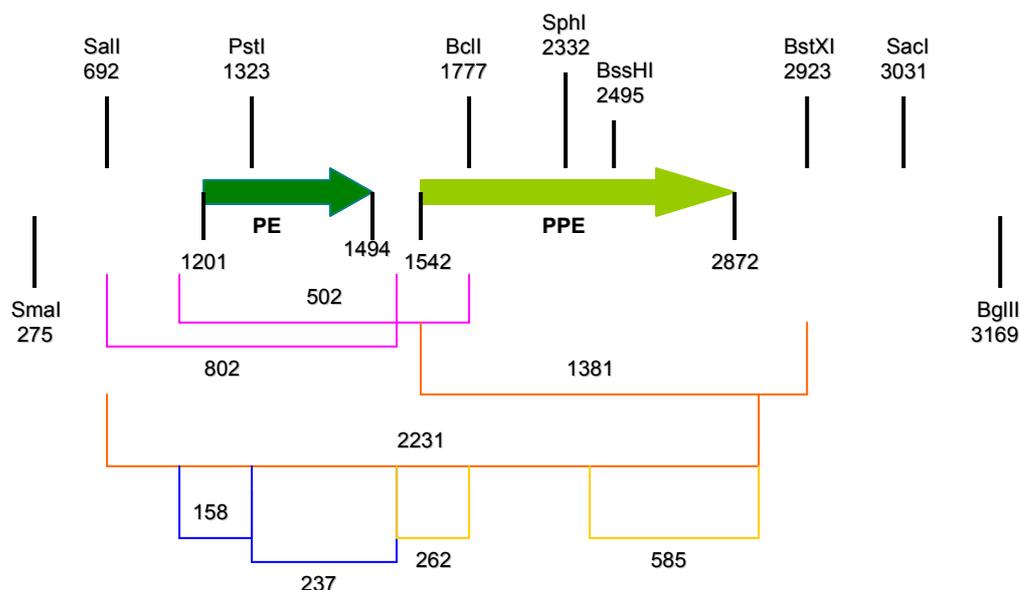


Figure 3.1. A figure illustrating the positions of the restriction sites and position of the primers. Indicated in pink and orange are the product of the PE35N, PE35C, PPE68N, and PPE68C using the first designed pair of primers, respectively. Indicated in blue and gold is the product of the PE35N, PE35C, PPE68N, and PPE68C using the second designed pair of primers, respectively.

3.1.2. Isolation of cosmids containing selected genomic regions

Cosmids containing the *M. smegmatis* genomic DNA encoding ESAT-6 gene cluster region 1 were isolated using colony blotting (Figure 3.2.).



Figure 3.2. Colony blotting of a *M. smegmatis* genomic DNA intergrating cosmid library. *M. smegmatis* genomic DNA cosmid library plated on LB plates were transferred to Hybond-N⁺ membrane filters, and probed with a [α -³²P] dCTP-labeled probe complementary to an internal region of the ESAT-6 gene sequence (Rv3875). Arrow indicate the colony with insert size of 33 000-35000 bp. No controls were done for colony blotting because it was not necessary.

Purified cosmids were analysed by PCR in order to confirm the presence of the ESAT-6 (MSMEG_0066) gene as well as the first gene Rv3866 (MSMEG_0057) and the last gene Rv3883c (MSMEG_0083) of the ESAT-6 gene cluster (Figure 3.3.). DNA sequencing using the T3 and T7 sequencing primers situated at the 5' and 3' ends of the insert was finally used to confirm the exact start and stop of the genome sequence fragment present in the cosmid. A cosmid containing the complete ESAT-6 gene cluster region 1 of *M. smegmatis* was isolated and named P1 (Figure 3.4.).

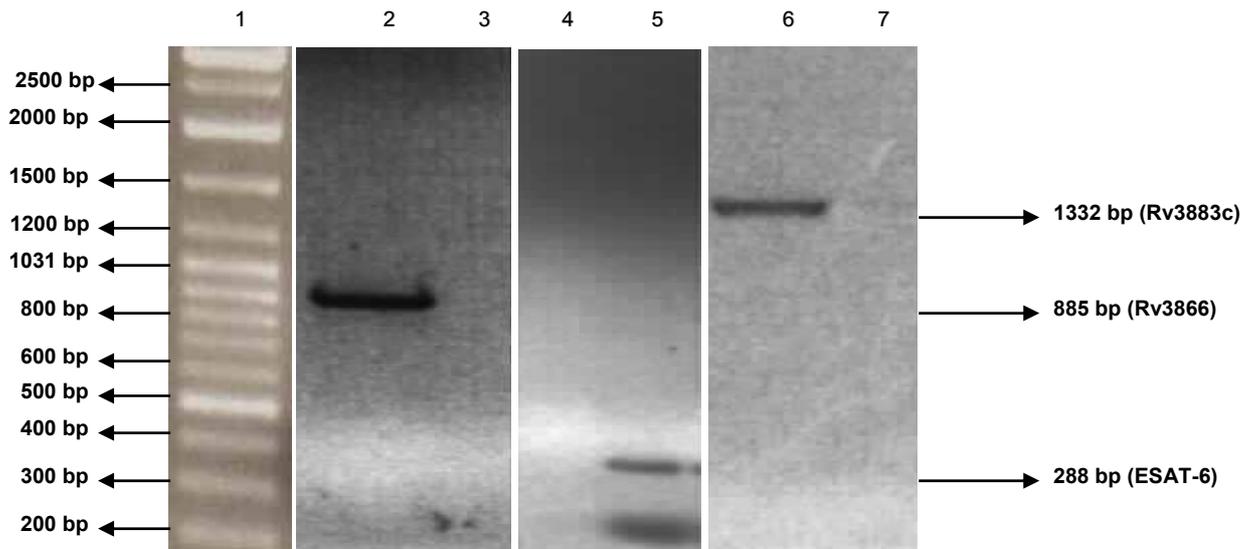


Figure 3.3. PCR analysis of the isolated cosmid. PCR analysis was used to confirm the presence of the ESAT-6 (MSMEG_0066) gene and that of the first Rv3866 (MSMEG_0057) and the last Rv3883c (MSMEG_0083) genes of the ESAT-6 gene cluster region 1 of *M. smegmatis* in the isolated cosmid. Lane 1: molecular weight marker, Lane 2: Rv3866 (MSMEG_0057)—885 bp, Lane 3: water as control for Rv3866 (MSMEG_0083), Lane 4: water as control for ESAT-6 (MSMEG_0066), lane 5: MSMEG_0066—288 bp, Lane 6: Rv3883c—1332 bp, and Lane 7: water as control for Rv3883c. The fragment sizes (bp) are indicated with arrows.

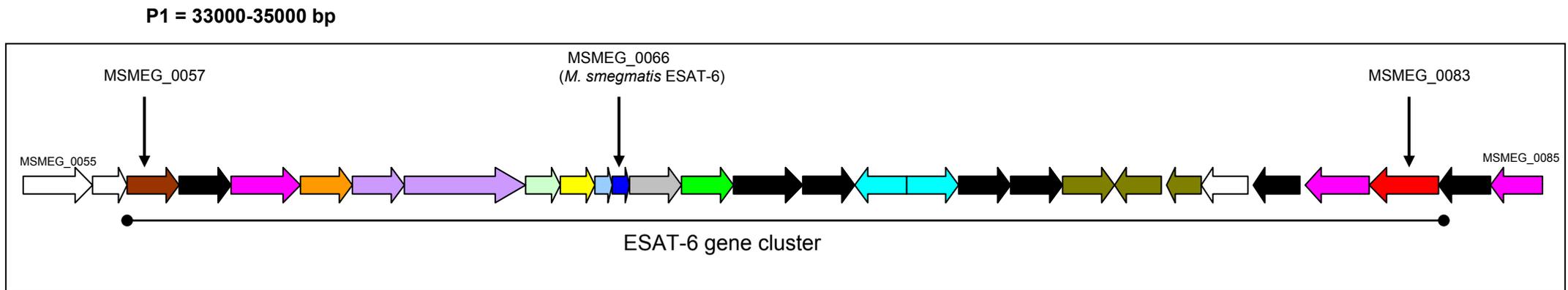
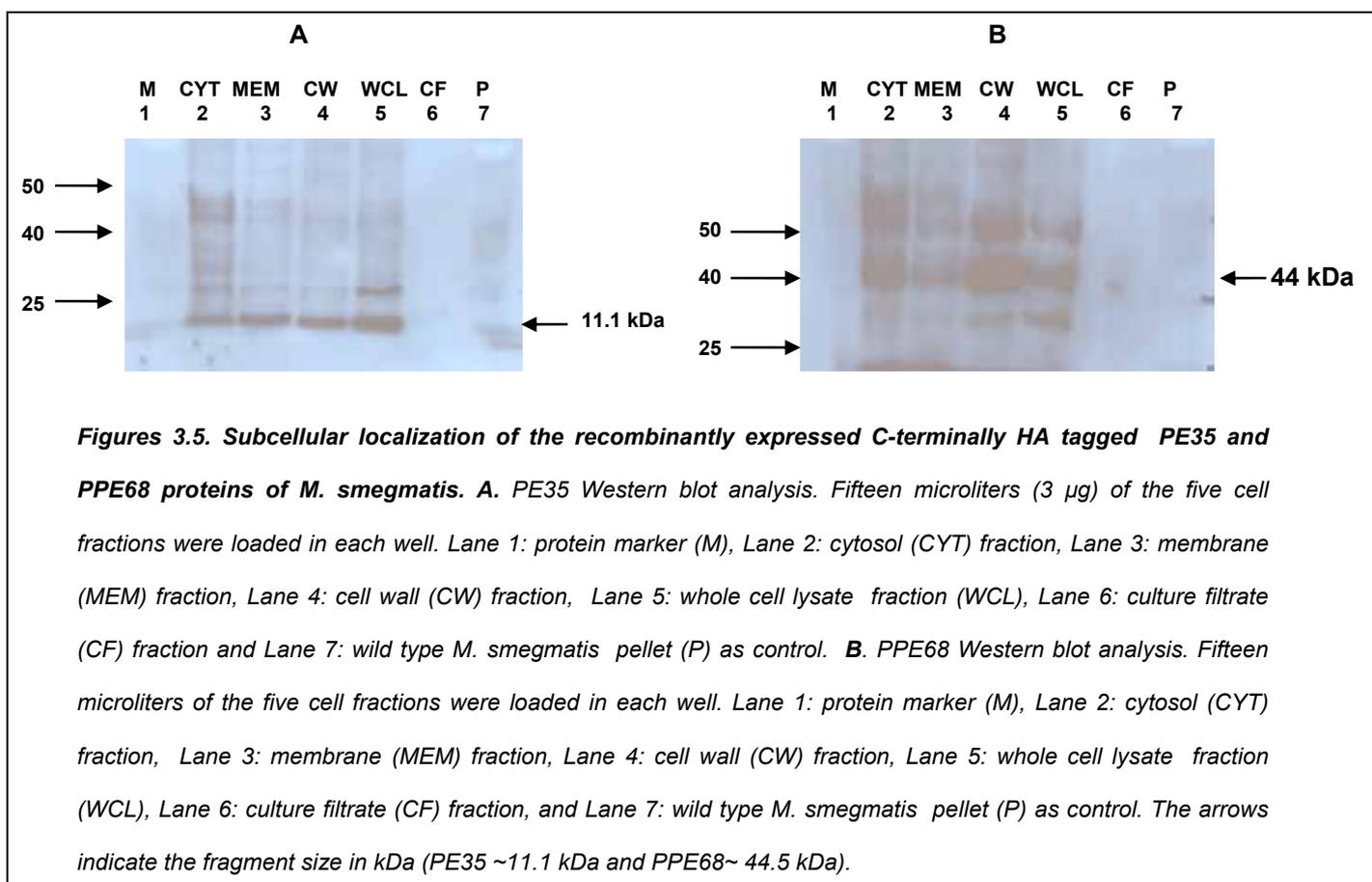


Figure 3.4. Genes present within the isolated cosmid P1 insert. The *M. smegmatis* ORF numbers (as determined by T3 and T7 sequencing) is shown at the termini of the insert. Genes confirmed to be present by PCR are indicated by vertical arrows. The ESAT-6 gene cluster region 1 is indicated by a horizontal line.

3.1.3. Secretion analyses of the PE and PPE proteins in *M. smegmatis*.

To determine if wild-type *M. smegmatis* was capable of secreting recombinant *M. smegmatis* PE and PPE proteins, the isolated cosmid containing the ESAT-6 gene cluster region 1 (P1) was used as template for the amplification of the two genes. When using the megaprimer technique of PCR-based site directed mutagenesis it was not possible to amplify the desired products of the PE35 and PPE68 genes. Unfortunately the primers could not be moved to obtain smaller amplicons, due to the size of the genes. We then designed a new set of primers for both the PE35 and PPE68 which were both C-terminally HA tagged in order to analyse the secretion of these two ancestral proteins by episomal expression. These recombinant genes were cloned into the mycobacterial expression vector p19Kpro (Figure 2.4.) and were transformed into *M. smegmatis* mc²155. A culture of the cells was subfractionated into culture filtrate, whole cell lysate, cell wall, membrane, and cytosol. The protein concentrations of the recombinant proteins were determined using the Bio-Rad Protein assay and results shown in Addendum B). The resulting proteins were separated on PAGE and Western blotted and probed with anti-HA antibodies. The results of this analysis indicates that although very high levels of the PE35 and PPE68 proteins were expressed in recombinant *M. smegmatis* cell wall, whole cell lysate, membrane and in the cytosol, no PE35 and PPE68 proteins could be detected in the culture filtrate (Figure 3.5. A and B).



3.2. Cellular localization of the *M. tuberculosis* PGRS and MPTR proteins in *M. smegmatis*.

3.2.1. PCR amplifications

Three PE_PGRS genes, Rv1818c, Rv0978c, and Rv2615c and two PPE-MPTR genes, Rv0442c and Rv0878c were chosen for this study and they were C-terminally HA-tagged. Of the three PE_PGRS genes, only Rv2615c could be amplified using nested-PCR, with both the outer and inner primers as indicated in Table 2.1. Rv0978c was amplified using the outer primers, but when the inner primers were used with the product of the outer primers as template, there was no amplification. This was also the case with the other PE_PGRS, Rv1818c. Both the PPE-MPTR genes (Rv0442c and Rv0878c) were amplified using the primers in Table 2.1. All the amplified genes (Rv2615c, Rv0442c and Rv0878c) were cloned into the pGEM-T Easy and subcloned into p19Kpro. The gene sequences were confirmed by sequencing, and the constructs were transformed and expressed in *M. smegmatis*. The gel pictures for the amplification of all the genes are shown in Addendum A, Figure A3 to A4.

3.2.2. Secretion analyses of the PE_PGRS and PPE-MPTR proteins in *M. smegmatis*.

To determine whether wild-type *M. smegmatis* was capable of secreting the recombinant C-terminally HA-tagged *M. tuberculosis* PE_PGRS protein (Rv2615c) and the two recombinant C-terminally HA-tagged PPE-MPTR proteins (Rv0442c and Rv0878c), the amplified genes were cloned into the mycobacterial expression vector, p19Kpro. Thereafter, constructs were transformed into *M. smegmatis* mc²155, and large cultures were prepared for expression. The cell wall, whole cell lysate, membrane, cytosol, and culture filtrate proteins were fractionated by PAGE, Western blotted and probed with anti-HA antibodies. The results indicated that although very high levels of the HA-tagged PE_PGRS and PPE-MPTR proteins were expressed in recombinant *M. smegmatis* cell wall, whole cell lysate, membrane and in the cytosol, no HA-tagged PE_PGRS and PPE-MPTR proteins could be detected in the culture filtrate (Figure 3.6 and 3.7, lane 1 to lane 7; respectively) and (Figure 3.8, lane 1 to lane 7). From these results we conclude that the PE_PGRS and PPE-MPTR proteins are not able to be secreted by *M. smegmatis*, although the presence in the cell wall/membrane indicates a mechanism for the proteins translocating to the extracellular locations. This suggests that the native *M. smegmatis* ESAT-6 gene clusters (Region 1, 3 and 4) are not able to secrete the PE_PGRS and PPE-MPTR subfamilies of proteins.

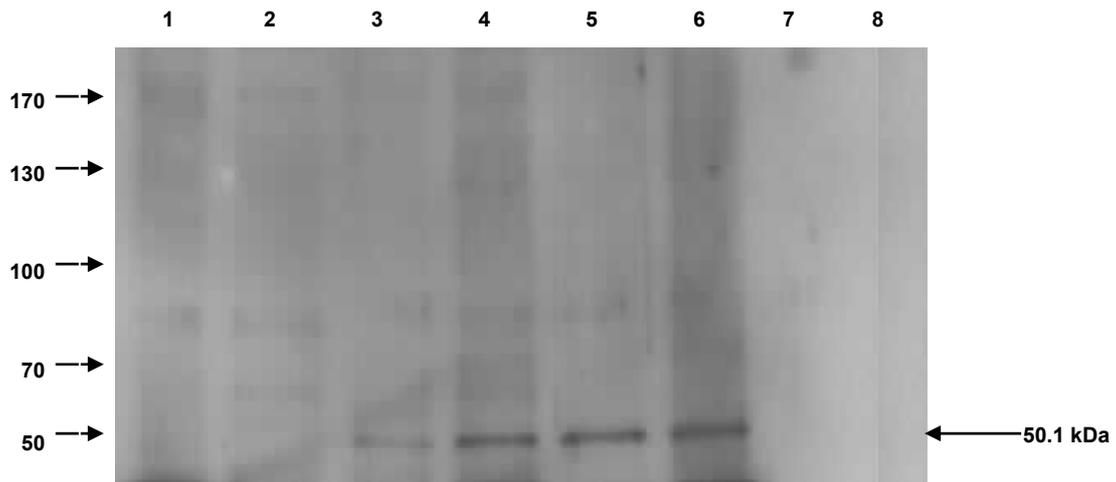


Figure 3.6. Subcellular localization of the recombinantly expressed C-terminally HA tagged PPE-MPTR protein, Rv0442c. Western blot analysis of Rv0442c. Fifteen microliters (3 μ g) of the five cell fractions were loaded in each well. Lane 1: Wild type *M. smegmatis* control pellet, Lane 2: Wild type *M. smegmatis* supernatant, Lane 3: the cytosol (CYT) fraction, Lane 4: membrane (MEM) fraction, Lane 5: cell wall (CW) fraction, Lane 6: whole cell lysate (WCL) fraction, Lane 7: culture filtrate (CF) fraction of the HA-tagged PPE-MPTR (Rv0442c) expressed in *M. smegmatis*, and Lane 8: protein marker. The arrows indicate the fragment size in kilodaltons.

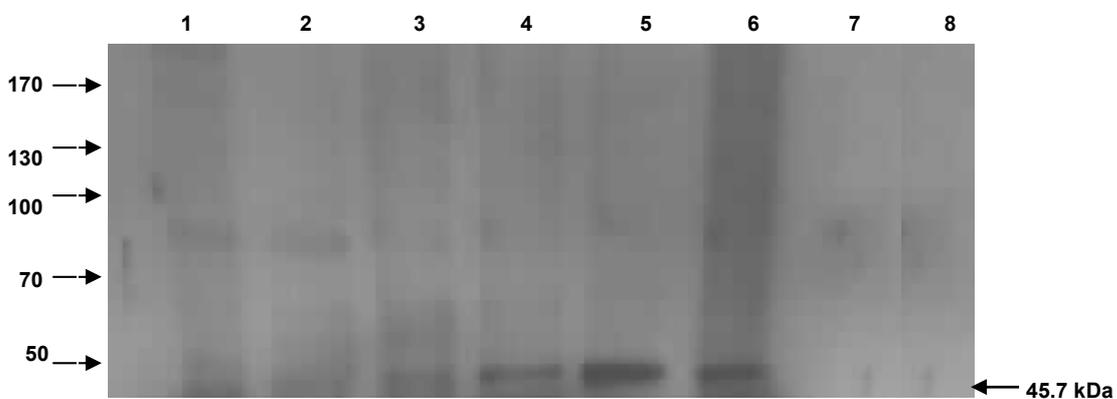


Figure 3.7. Subcellular localization of the recombinantly expressed C-terminally HA tagged PPE-MPTR protein, Rv0878c. Western blot analysis of Rv0878c. Fifteen microliters (3 μ g) of the five cell fractions were loaded in each well. Lane 1: Wild type *M. smegmatis* control pellet, Lane 2: Wild type *M. smegmatis* supernatant, Lane 3: the cytosol (CYT) fraction, Lane 4: membrane (MEM) fraction, Lane 5: cell wall (CW) fraction, Lane 6: whole cell lysate (WCL) fraction, Lane 7: culture filtrate (CF) fraction of the HA-tagged PPE-MPTR (Rv2615c) expressed in *M. smegmatis*, and Lane 8: protein marker. The arrows indicate the fragment size in kilodaltons.

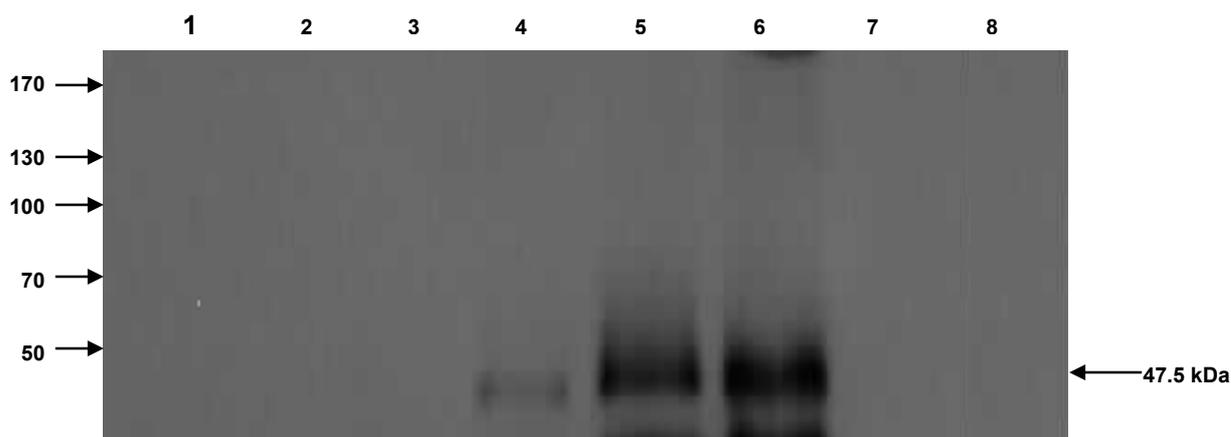


Figure 3.8. Subcellular localization of the recombinantly expressed C-terminally HA tagged PE-PGRS protein, Rv2615c protein analysis. Western blot analysis of Rv2615c. Fifteen microliters ($3 \mu\text{g}$) of the five cell fractions and wild type *M. smegmatis* were loaded in each well. Lane 1: Wild type *M. smegmatis* control pellet, Lane 2: Wild type *M. smegmatis* supernatant, Lane 3: the cytosol (CYT) fraction, Lane 4: membrane (MEM) fraction, Lane 5: cell wall (CW) fraction, Lane 6: whole cell lysate (WCL) fraction, Lane 7: culture filtrate (CF) fraction of the HA-tagged PPE-MPTR expressed in *M. smegmatis*, and Lane 8: protein marker. The arrows indicate the fragment size in kilodaltons.

3.3. Knockout of ESAT-6 gene clusters region 3 and region 4 in *M. smegmatis*

3.3.1. PCR amplifications

Primers flanking the ESAT-6 gene clusters region 3 and region 4 (shown in Table 2.1) in *M. smegmatis* were used to PCR amplify products of ~ 800 bp upstream and downstream of the two regions in order to generate recombinant vectors to generate knockouts of these two regions in *M. smegmatis*. All the genomic regions i.e. upstream of Rv0282 (UpMSMEG_0605), downstream of Rv0292 (DownMSMEG_0616), upstream of Rv3450 (UpMSMEGS_1532) and downstream of Rv3444 (DownMSMEG_1538) were amplified using *M. smegmatis* genomic DNA as template and the Hotstar Taq Polymerase kit (Qiagen). The amplified regions were cloned separately into the T vector, pGEM-T Easy, and the products were confirmed by sequencing. The resultant products were subcloned into the corresponding sites within the mycobacterial vector p2NIL, by three way cloning, i.e. both genomic regions upstream and downstream of each of the regions to be knocked out were ligated in a reaction mixture at the same time with the p2NIL vector. The following constructs were generated p2NIL: Δ R3 and

p2NIL:ΔR4, respectively. After successful cloning of both the regions into p2NIL, the pGOAL17 cassette was digested from the pGOAL17 and cloned into the p2NIL constructs to generate the p2NIL:pGOAL17:ΔR3 and p2NIL:pGOAL17:ΔR4 constructs containing the selection cassette. The results for the amplification of all the regions are shown in Addendum A, Figure A5 B.

3.3.2. Isolation of mutants.

Both the p2NIL constructs, p2NIL:pGOAL17:ΔR3 and p2NIL:pGOAL17:ΔR4, were electroporated separately into *M. smegmatis* and were allowed to grow for 3 hours at 37°C. Thereafter, they were spread onto plates containing kanamycin and X-gal, in order to select the single crossover (SCO) mutants by the presence of blue colonies. For the selection of the double crossover mutant alleles, blue SCO colonies were inoculated into LB containing Tween 80 and were allowed to grow for 3-4 days at 37°C. Thereafter, cells were spread onto plates containing X-gal and 2.5% sucrose and were allowed to grow for 3-4 days at 37°C, in order to isolate white sucrose-resistant colonies (i.e. colonies containing the double crossover mutant alleles which lack the plasmid-borne *lacZ* gene). Dilution series were done for both transformants per colony and patched on sucrose plates (shown in Figure 3.9. A to D). The *lacZ* gene was used to distinguish the SCO from double crossovers (DCO) and in the absence of adequate expression, we used kanamycin sensitivity as a screen. One SCO transformant was picked from each plate and streaked out onto plates lacking antibiotics. Following growth, a loopful of cells was resuspended in liquid medium and plated onto sucrose plates to select for cells which had lost the integrated plasmid through a SCO. A reduction in colony forming units (cfu) was seen on plates containing sucrose as compared to plates containing no sucrose. Sucrose resistance (suc^R) colonies were spread onto plates with and without kanamycin and scored for growth after 3-4 days to distinguish double crossovers from single crossovers which had acquired spontaneous resistance to sucrose. DNA was prepared from kanamycin sensitive (kan^S) colonies and analysed by PCR in order to screen for DCO mutant. For the region 3 knockout, colonies which had the phenotype $\text{kan}^S \text{suc}^R$, (i.e. DCO mutants) were screened by PCR and only one had the expected mutant double cross-over genotype, shown in Figure 3.10. (A), (B) and (C). For the region 4 knockout, colonies which had the phenotype $\text{kan}^S \text{suc}^R$ (i.e. double crossover mutants). Five colonies which were suc^R colonies and kan^S were screened for the deletion allele by PCR and they all had the expected mutant double cross-over genotype, shown in Figure 3.11. (A), (B) and (C). Only PCR was used to confirm the deletion mutants, did not used Southern blotting and the efficiency of the mutagenesis procedure was 100% for both mutants. Results were confirmed by sequencing using the primers flanking the deletion alleles.

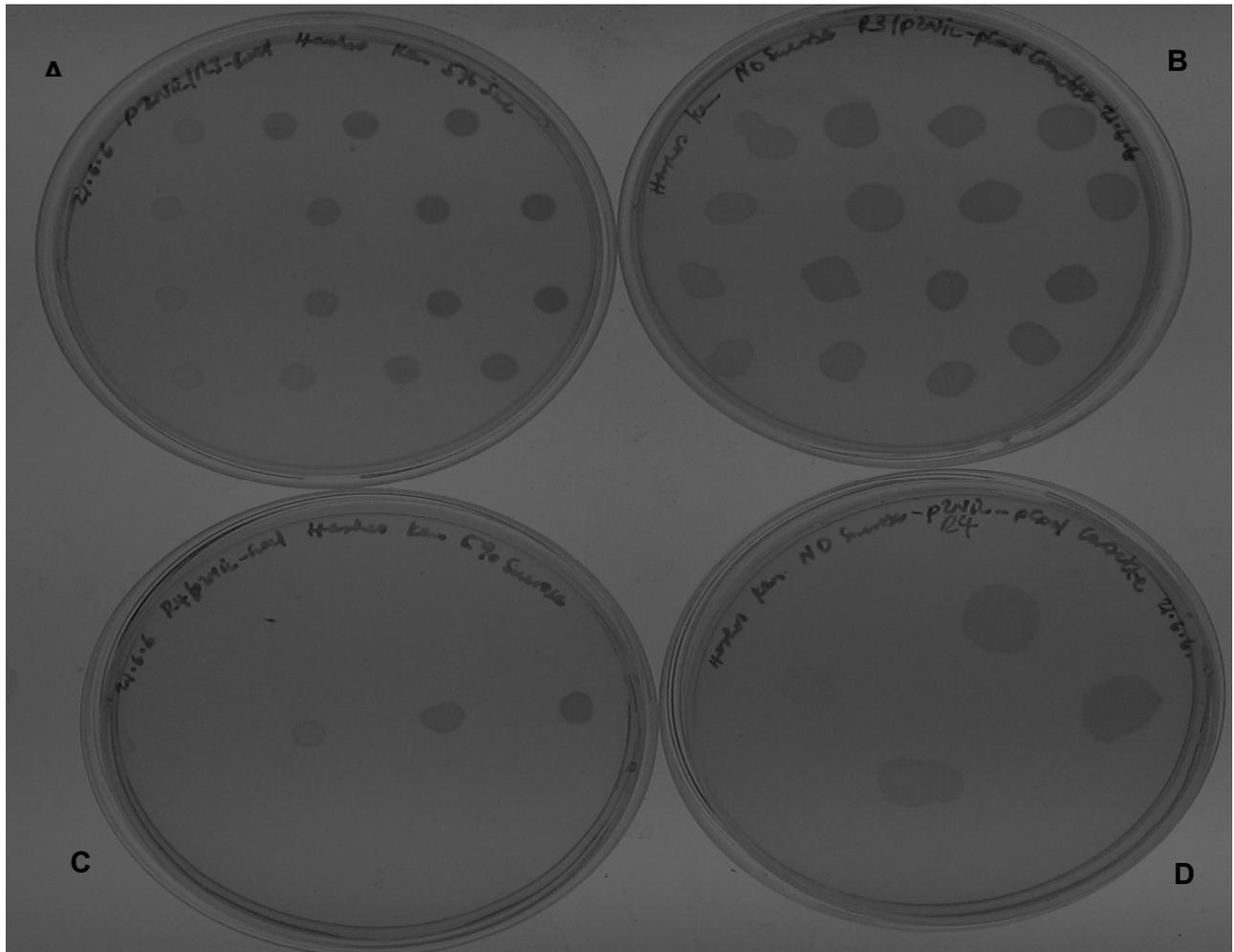


Figure 3.9. Dilution series from 10^{-1} to 10^{-4} (i.e. 50 μ l cells into 450 μ l LB) of the R3 and R4 transformants, i.e. single crossovers spread on plates with and without sucrose. (A) Plate containing region 3 transformants patched on sucrose, (B) Plate containing region 3 transformant patched on a plate without sucrose (control), (C) Plate containing region 4 transformant patched on sucrose and (D). Plate containing region 4 transformants patched on a plate without sucrose (control). Plate (A) and (C), shows the effect of sucrose on the two transformants, respectively.

From Figure 3.9 above, shows dilution series from 10^{-1} to 10^{-4} from the left to the right showing four single crossover transformants picked for region 3 (plate A to B) and with one single crossover transformant picked for region 4 (plate C and D). Plate A and C contains region 3 and region 4 transformants patched on plates containing kanamycin and 5% sucrose and also with plate B and D containing transformants patched on plates containing kanamycin and without sucrose. Plate A and C containing transformants patched on 5% sucrose, shows sensitivity to sucrose as compared to plate B and D.

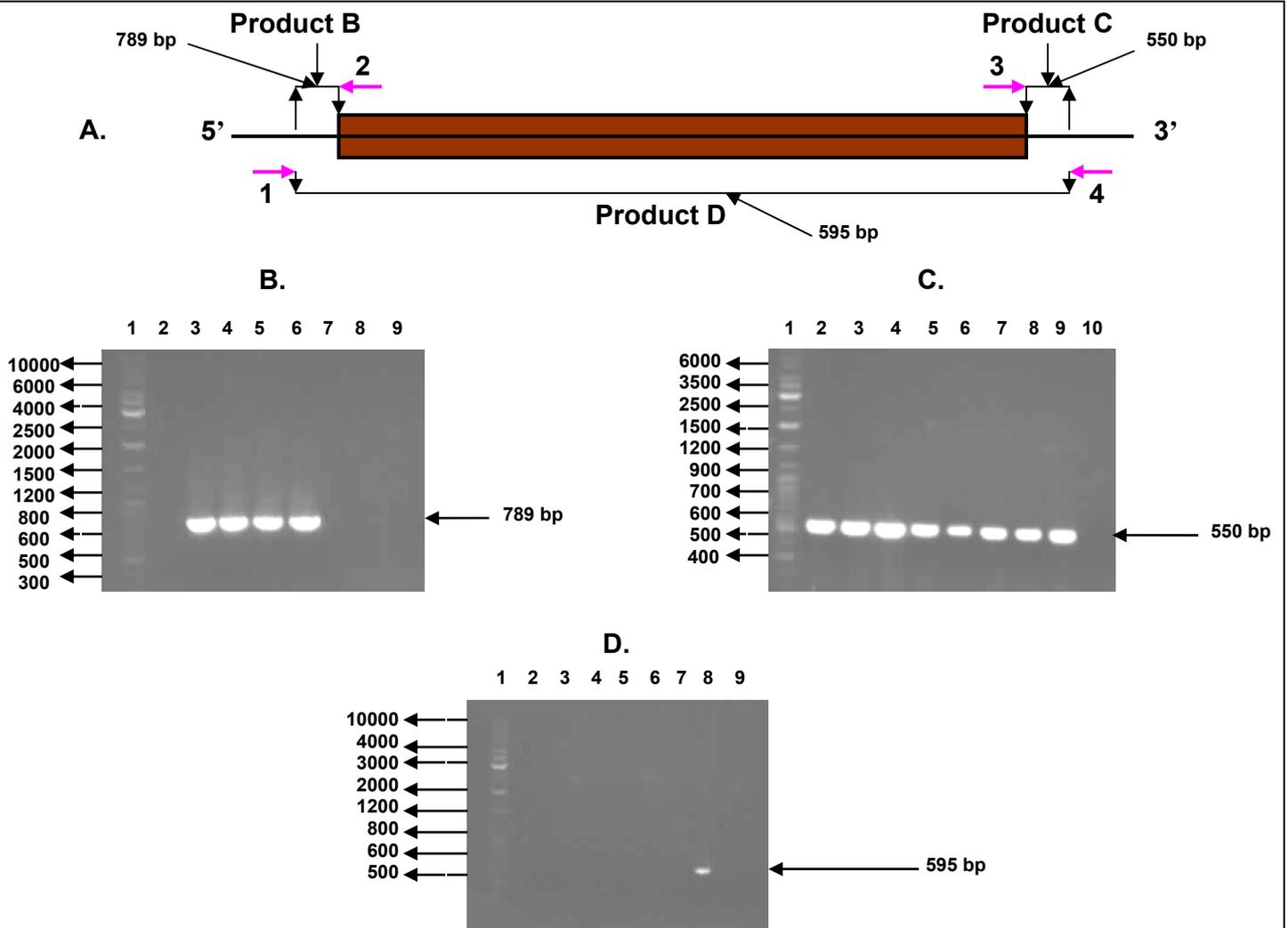


Figure 3.10. Screening of region 3 knockout mutants. (A) Diagram of the upstream (5'-end) and downstream (3'-end) of region 3 indicating primer positions and sizes of the amplification products (B-D). PCR amplicons of region 3 deletion allele were run on a 1% agarose gel with ethidium bromide staining, 10 μ l of the PCR reactions were loaded in each well. (B) Lane 1: Gene ruler molecular weight marker. Lane 2-8: PCR screening of DCO mutant alleles using primers upstream (5'-end, MSMEG_0605f and MSMEG_0605r, primers set 1 and 2) of Region 3, generating a fragment size of 789 bp), with no amplification in Lane 2, 7 and 8. Lane 9: Water used as negative control (C) Lane 1: Gene ruler molecular weight marker. Lane 2 to 9: PCR screening of DCO mutant alleles using primers downstream (3'-end, MSMEG_0616f and MSMEG_0616r, primers set 3 and 4) of Region 3, generating a fragment size of 550 bp. Lane 10: Water used as negative control. (D) Lane 1: Gene ruler molecular weight marker. Lane 2-8: PCR screening of DCO mutant allele, using primers flanking the deletion allele, i.e. forward primer (5'-end MSMEG_0605f, primer 1) and reverse primer (3'-end MSMEG_0616r, primer 4), generating a fragment of size, 595 bp in Lane 8 and no amplification in Lane 2-7. Lane 9: Water was used as negative control. The arrows indicate the fragment size in kilobases (kb)

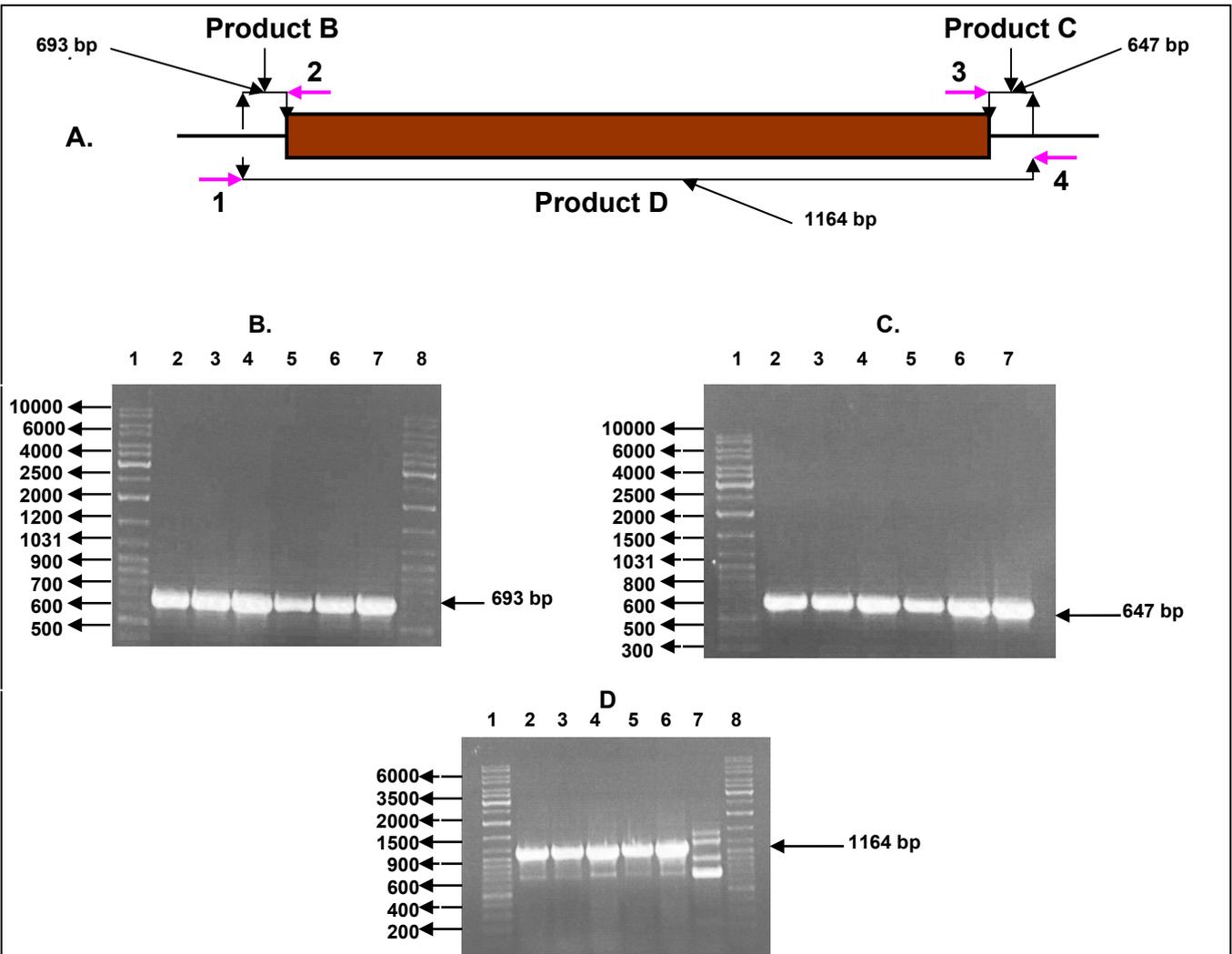


Figure 3.11. Screening of region 4 knockout mutants. (A) Diagram of the primers upstream (5'-end) and downstream (3'-end) of region 4 indicating primer positions and sizes of the amplification products (B-D). PCR amplicons of region 4 deletion allele were run on a 1% agarose gel with ethidium bromide staining, 10 μ l of the PCR reactions were loaded in each well. (B) Lane 1 and 8: Gene ruler molecular weight marker. Lane 2-6: PCR screening of the 5 DCO mutant alleles using primers upstream (5'-end, MSMEG_1532f and MSMEG_1532r, primer set 1 and 2) of Region 4, generating a fragment size of 693 bp. Lane 7: *M. smegmatis* mc²155 was used as positive control. (C) Lane 1: Gene ruler molecular weight marker. Lane 2-6: PCR screening of the 5 DCO mutant alleles using primers downstream (3'-end, MSMEG_1538f and MSMEG) 1538r, primer set 3 and 4) of Region 4 (fragment size of 647 bp). Lane 7: *M. smegmatis* mc²155 was used as positive control. (D) Lane 1 and 8: Gene ruler molecular weight marker. Lane 2-6: PCR screening of the 5 DCO mutant alleles, using primers flanking the deletion allele, i.e. forward primer (5'-end MSMEG_1532f, primer 1) and reverse primer (3'-end MSMEG_1538r, primer 4), generating a fragment size of 1164 bp. Lane 7: *M. smegmatis* mc²155 was used as positive control. The arrows indicate the fragment size in kilobases (kb).

CHAPTER 4: DISCUSSION AND CONCLUSIONS

This study aimed to investigate and compare the subcellular localization of the ancestral, PE35 (Rv3872), and PPE68 (Rv3873) from *M. smegmatis*; as well as the most recently duplicated PE and PPE (PGRS and MPTR) proteins from *M. tuberculosis* in the model organism *M. smegmatis*. We further aimed to investigate whether the ESAT-6 secretion apparatus (or parts thereof) are involved in the trafficking of both the ancestral and recent PE and PPE proteins in the model organism *M. smegmatis*. The genome of *M. tuberculosis* H37Rv contains two large glycine-rich gene families which code for the PE and PPE proteins. The PE protein family consists of 99 members and the PPE protein family consists of 68 members, while analysis of the genome sequence of *M. smegmatis* revealed only two pairs each of the PE and PPE gene families. None of the other members of the PE or PPE gene families, including any of the PE_PGRS or PPE-MPTR genes, could be detected within the *M. smegmatis* genome. The first pair corresponds to the Rv3872/3 orthologues (MSMEG0062 and MSMEG0063) from ESAT-6 (*esx*) gene cluster region 1 (70% and 55% similarity to the *M. tuberculosis* H37Rv proteins, respectively), while the second pair corresponds to the Rv0285/6 orthologues (MSMEG0608 and MSMEG0609) from ESAT-6 (*esx*) gene cluster region 3 (87% and 64% similarity to the *M. tuberculosis* H37Rv proteins, respectively) (Gey van Pittius *et al*, 2006). These two gene pairs have been shown to be required for *in vivo*, and *in vitro* growth, respectively, in *M. tuberculosis* H37Rv (Pym *et al*, 2002; Sasseti and Rubin, 2003). Thus, the only PE and PPE genes present within the *M. smegmatis* genome are found within two ESAT-6 (*esx*) gene cluster regions, (region 1 and region 3) (Gey van Pittius *et al*, 2006). The sub-cellular localization of these proteins remains to be determined, a few of the PE_PGRS proteins have been considered as possible virulence factors in *M. marinum* (Ramakrishnan *et al*, 2000), and some are cell surface constituents, involved in interaction between mycobacteria and the macrophage (Brennan *et al*, 2001; Chaithra *et al*, 2005).

In order to achieve the set aims for this study, we selected the PE_PGRS (Rv2615c) and PPE-MPTR (Rv0442c, and Rv0878c) proteins from *M. tuberculosis* H37Rv based on their smaller sizes as compared to other PE and PPE proteins and also with the hope that they will amplify better. It has been shown that they are difficult to amplify due to their large/bigger sizes and also their GC rich regions. Western blot analysis of cell fractions showed that these proteins were not secreted in *M. smegmatis*. The sub-cellular localization of ancestral PE35 and PPE68 proteins, using Western blotting for analysis of cell fractions also showed no secretion of these proteins. The absence of secretion of the ancestral PE and

PPE proteins was not expected and was in contrast to other unpublished studies which has identified the presence of these proteins in culture filtrates in *M. tuberculosis*. In order to generate the tools to investigate whether the ESAT-6 secretion apparatus (or parts thereof) are involved in the trafficking of both the ancestral and recent PE and PPE proteins, we knocked out two of the three ESAT-6 gene cluster regions, region 3 and region 4 in *M. smegmatis*. This provides a source for future studies to investigate the effect of region 1 on secretion of the PE and PPE genes and other genes within these region.

The reason for the maintenance of the gene pairing of the ancestral PE and PPE genes is still unclear, although these genes may be functionally related and co-transcribed. There is some early evidence for the latter from gene expression data obtained during adaptation to nutrient starvation (the gene pairs Rv0285/86 (PE5/PPE4), Rv1195/96 (PE13/PPE18), Rv1386/87 (PE15/PPE20) and Rv2431c/30c (PE25/PPE41) are downregulated and the pair Rv1169c/68c (PE11/PPE17) is up-regulated (Betts *et al.* 2002; Gey van Pittius *et al.*, 2006). For secretion to occur, the recombinant PE35 and PPE68 proteins must have partners (i.e. native proteins to allow duplex formation and allow co-expression to occur) within the model organism, as shown in the studies by Tundup *et al.*, (2006), who recently demonstrated that the genes from at least one of these PE-PPE gene pairs, Rv2430c/31c, are co-transcribed and that the gene products interact with each other to form a hetero-tetramer. Strong *et al.* (2006), expanded these findings by determining the structure of the Rv2430c/31c protein interaction, and demonstrated that the PE/PPE protein pair forms a 1:1 complex. The absence of secretion of the ancestral PE/PPE proteins in this study was unexpected, because these proteins had native proteins to allow complex formation and coexpression to occur in *M. smegmatis*. The reason why there was no secretion may be due to the fact that the HA tag might interfere with the dimerization of the mutant PE and PPE with the wild type PE and PPE produced by *M. smegmatis*, preventing secretion. The other reason could be misfolding of the recombinant proteins as well as cleavage of the HA tag/proteolytic degradation and also that the native proteins could have interacted with each other and not with the recombinant proteins, to prevent secretion. Secretion observed in other studies could also be due to cell lysis, and/or different growth conditions as compared to the conditions used in this study. We hypothesise that the development of antibodies specific for PE and PPE proteins will give better results in tracking the secretion of these proteins, as the native proteins could be investigated without episomal expression and HA tag interference. This will be the subject for future studies.

In a collaborative study with Dr Wilbert Bitter and Abdallah Abdallah from Free University of Amsterdam, it was shown that the expression of the recombinant HA-tagged PPE-MPTR and PE_PGERS

proteins Rv0442c and Rv2615c was very good in *M. marinum* (much better than Rv2430/31c). They showed that although both proteins are not secreted in high amounts (only about 10-20% was found in the culture supernatant), the secretion of these proteins is completely dependent on ESX-5 (ESAT-6 gene cluster region 5) (data not shown). Gey van Pittius *et al.*, (2006) have shown that the PGRS and MPTR subgroups are the last duplications of the PE and PPE families respectively, and that they have each originated from the ESAT-6 cluster region 5 duplicates. They showed that region 2 and 5 were not duplicated in the genome of *M. smegmatis*, and hypothesized that any genes which were duplicated from these regions would not be present in the genome of *M. smegmatis*. Thus, PE_PGRS and PPE-MPTR genes are absent in this organism and this, together with the fact that there is no region 5 present, could explain the absence of secretion of these proteins in *M. smegmatis* in this study. This remains to be proven in future studies. Future studies to investigate the secretion of the PE_PGRS and PPE-MPTR will consist of transforming the HA tagged constructs into *M. bovis* BCG and monitoring secretion using SDS-PAGE and Western blotting. This will enable us to determine if the HA tag interferes with secretion, since *M. bovis* contains the EAST-6 gene cluster region 5.

In conclusion, we aimed to investigate and compare the subcellular localization of the ancestral, PE35 (Rv3872), and PPE68 (Rv3873) from *M. smegmatis*; as well as the most recently duplicated PE and PPE (PGRS and MPTR) proteins from *M. tuberculosis* in the model organism *M. smegmatis*. Western blot analysis showed that the PE35 and PPE68 proteins were not secreted in *M. smegmatis*, and the absence of secretion of the recombinant PE35 and PPE68 proteins may be explained by the presence of the HA tag which may interfere with the binding of these proteins with the non-recombinant PE35 and PPE68 proteins, thereby inhibiting their secretion. The absence of secretion of the PE_PGRS and PPE-MPTR recombinant proteins of *M. tuberculosis* in the model organism *M. smegmatis* may also be explained by the HA-tag or by the absence of the ESAT-6 gene cluster region 5 (ESX-5) in *M. smegmatis*. This remains to be proven in future studies. In order to generate the tools to investigate whether the ESAT-6 secretion apparatus (or parts thereof) are involved in the trafficking of both the ancestral and recent PE and PPE proteins, we knocked out two of the three ESAT-6 gene cluster regions, region 3 and region 4 in *M. smegmatis*. This provides a source to investigate the effect of region 1 on secretion of the PE and PPE genes. Single genes will be knocked out within the remaining region 1 and the secretion of the PE and PPE genes will be monitored. The findings of this study provide relevant tools and an important starting point to study the secretion of the PE and PPE proteins and their relationship with the ESAT-6 gene cluster regions. This could lead to the development of strategies to terminate or enhance secretion of

these antigens, thereby influencing the immunogenicity of the pathogens, which may ultimately have an impact on the design and development of vaccines.

CHAPTER 5: FUTURE DIRECTIONS

Future studies to investigate the secretion of the PE and PPE proteins and their relationship with the ESAT-6 gene clusters will be aimed at the following:

- ❖ Double knockout studies of ESAT-6 gene cluster region 3 and region 4, from the genome of *M. smegmatis*, needs to be done, in order to generate an *M. smegmatis* mutant in which only ESAT-6 gene cluster region 1 is functional.
- ❖ Single genes from the remaining ESAT-6 gene cluster region 1 can be knocked out sequentially, and the PE and PPE proteins and other proteins within the region be monitored to determine whether they are secreted, and how they function.
- ❖ In order to further investigate the secretion of the PE_PGRS and PPE-MPTR, the HA tagged constructs must be transformed into *M. bovis* BCG and secretion monitored using SDS-PAGE and Western blotting.
- ❖ To confirm if the ESAT-6 gene cluster region 5 regulates secretion of the PGRS and MPTR, this region will be cloned/transformed into *M. smegmatis*, the HA-tagged proteins will be co-transformed and allowed to express, after which secretion will be monitored using SDS-PAGE and Western blotting. This will confirm whether the PE_PGRS and PPE-MPTR genes are influenced by/involved with/secreted by the ESAT-6 gene clusters, specifically ESAT-6 gene cluster region 5.

CHAPTER 6: ADDENDUM A TO E

6.1. ADDENDUM A

6.1.1. Photographic representations of the agarose gels showing the PCR amplification results:

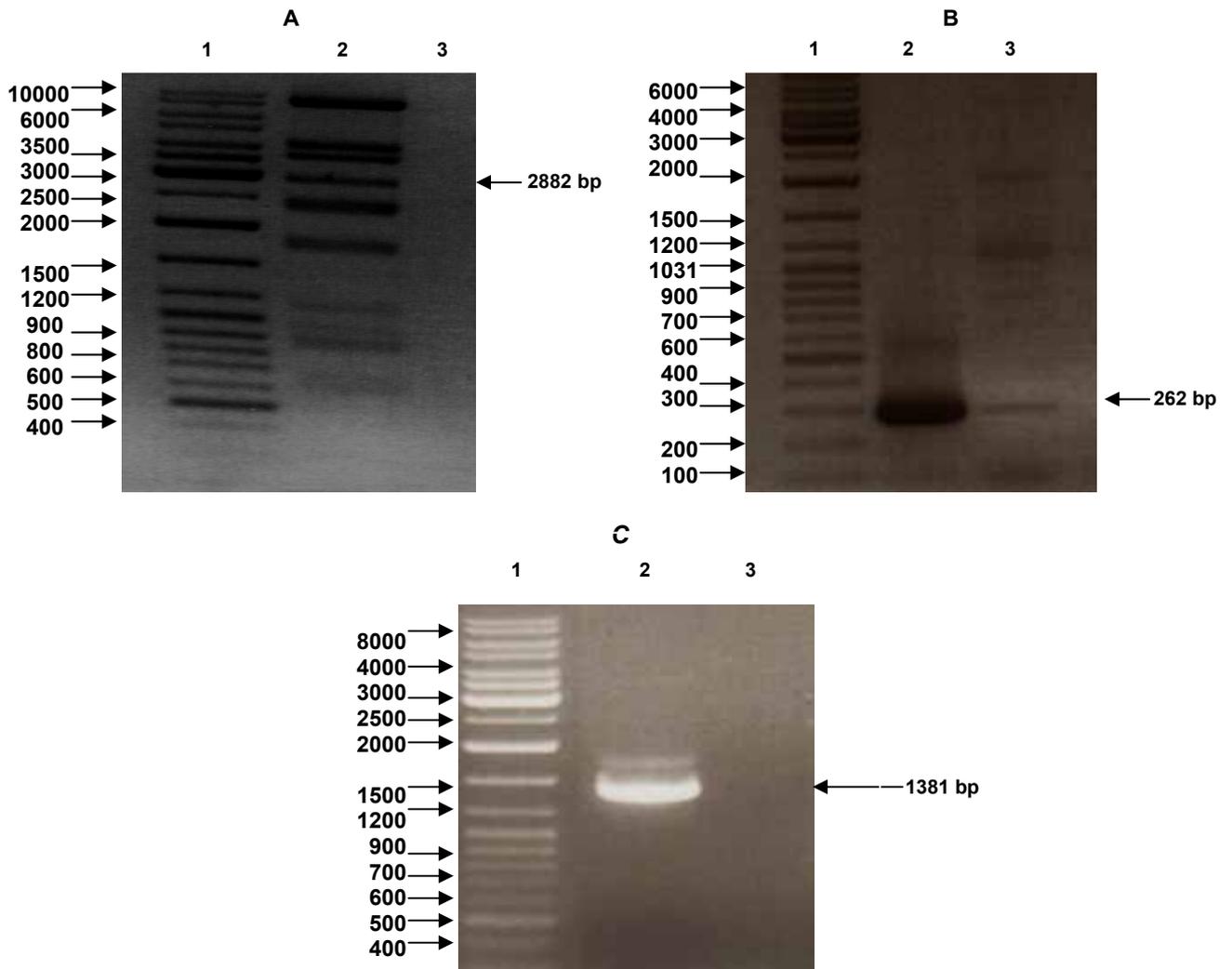
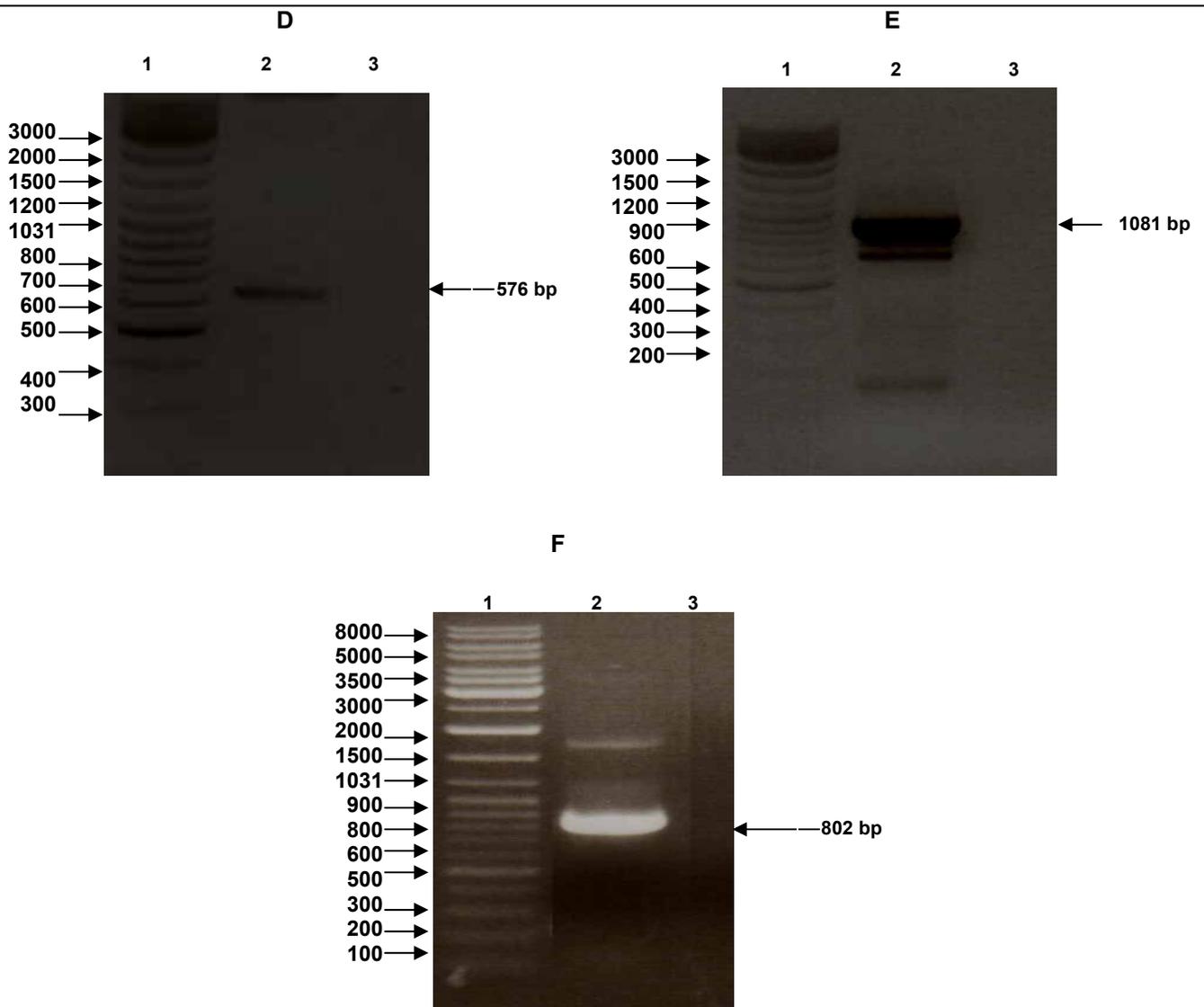
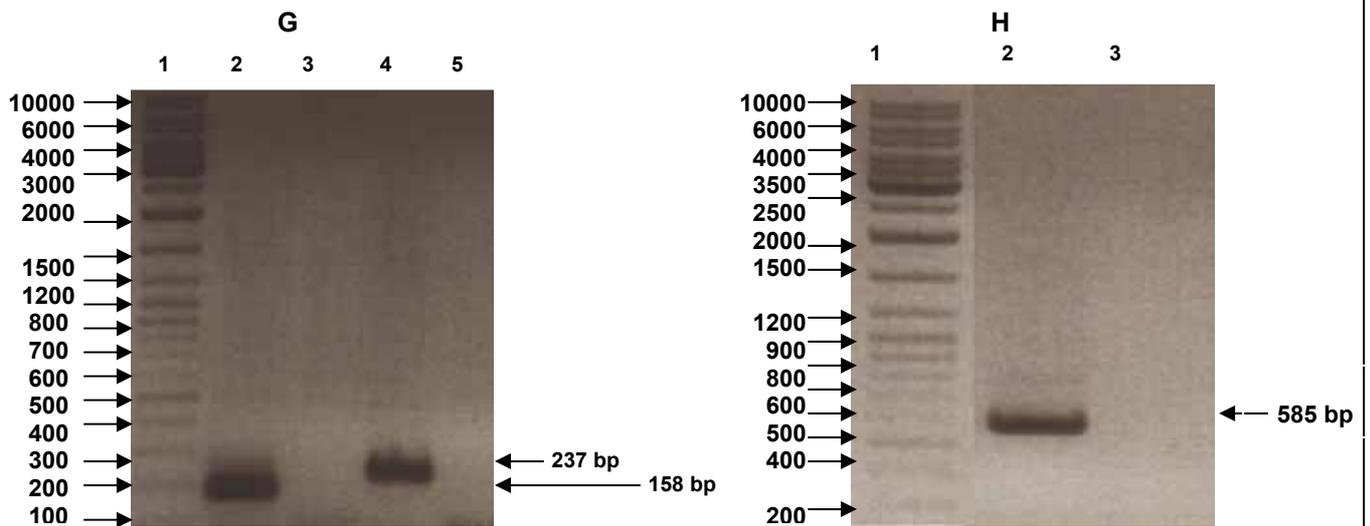


Figure A.1. PCR amplification of the *PE35* and *PPE68* genes using the primers in Table 2.1, and the digestion of the cosmid from A to H. This figure shows the photographic representations of a 1% agarose gel with ethidium bromide staining, where 10 μ l of the PCR amplicon reactions of both genes were loaded in each well. **(A)** Lane 1: Gene ruler molecular weight marker. Lane 2: Cosmid digested with the restriction enzymes *Bgl*I and *Sma*I, generating a fragment size of 2882 bp. **(B)** Lane 1: Gene ruler molecular weight marker. Lane 2: *PPE68*N-terminal HA tagged megaprimer, using the primers *PPE68f* and *Bcl*I generating a product of 262 bp. Lane 3: H_2O as control. **(C)** Lane 1: Gene ruler molecular weight marker. Lane 2: *PPE68*N-terminal HA tagged megaprimer, using the primers *PPE68f* and *Bst*XI generating a product of 1381 bp. Lane 3: H_2O control loaded. The fragment sizes (bp) are indicated with arrows.



(D) Lane 1: Gene ruler molecular weight marker. Lane 2: PE35N-terminal HA tagged megaprimer, amplified using the primers PE35f and BclI_r, generating a product of 576 bp. Lane 3: H₂O as control. **(E)** Lane 1: Gene ruler molecular weight marker. Lane 2: PE35N-terminal HA tagged amplified using the megaprimer obtained in **(D)** and Sall I in order to engineer the flanking restriction site to allow directional cloning, generating a product of 1081 bp loaded. Lane 3: H₂O as control. **(F)** Lane 1: Gene ruler molecular weight marker. Lane 2: PE35C-terminal HA tagged megaprimer, amplified using the primers PE35r and Sall generating a product of 802 bp. Lane 3: H₂O as control 3. The fragment sizes (bp) are indicated with arrows.



(G) The amplification of PE35N- and C-terminal HA tagged megaprimers. Lane 1: Gene ruler molecular weight marker. Lane 2 and 4: PE35-N and PE35-C megaprimers were amplified using the primers PE35f and Pstlr, and PE35r and Pstlf; generating products of 158 bp and 237 bp, respectively. Lane 3 and 5: H₂O as control. **(H)** Lane 1: Gene ruler molecular weight marker. Lane 2: PPE68C-terminal HA tagged megaprimer, amplified using the primers PPE68r and SphI generating a product of 585 bp. Lane 3: H₂O as control. The fragment sizes (bp) are indicated with arrows.

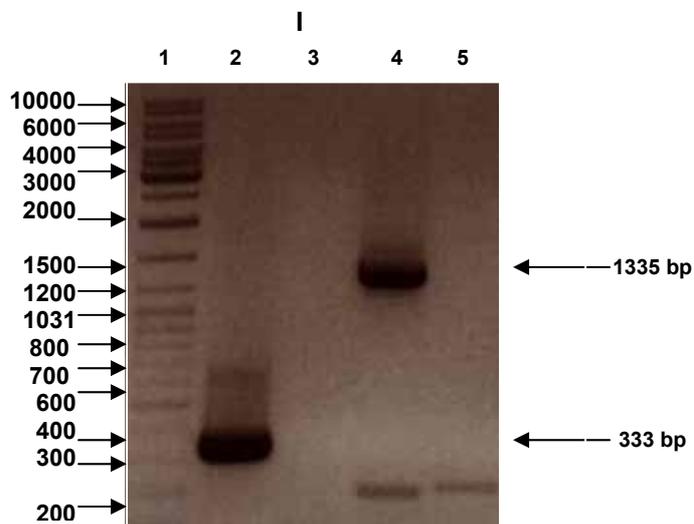


Figure A.2. PCR amplification of the PE (Rv3872) and PPE (Rv3873). This figure shows the photographic representations of a 1% agarose gel with ethidium bromide staining, where 10 μ l of the PCR amplicon reactions of both genes were loaded in each well. Lane 1: Gene ruler molecular weight marker. Lane 2: PE was amplified using the primers PEF and PER, generating a product of size, 333 bp. Lane 3: H₂O as control. Lane 4: PPE was amplified using the primers PPEf and PPEr, generating a product of size, 1335 bp. Lane 5: H₂O as control. The fragment sizes (bp) are indicated with arrows.

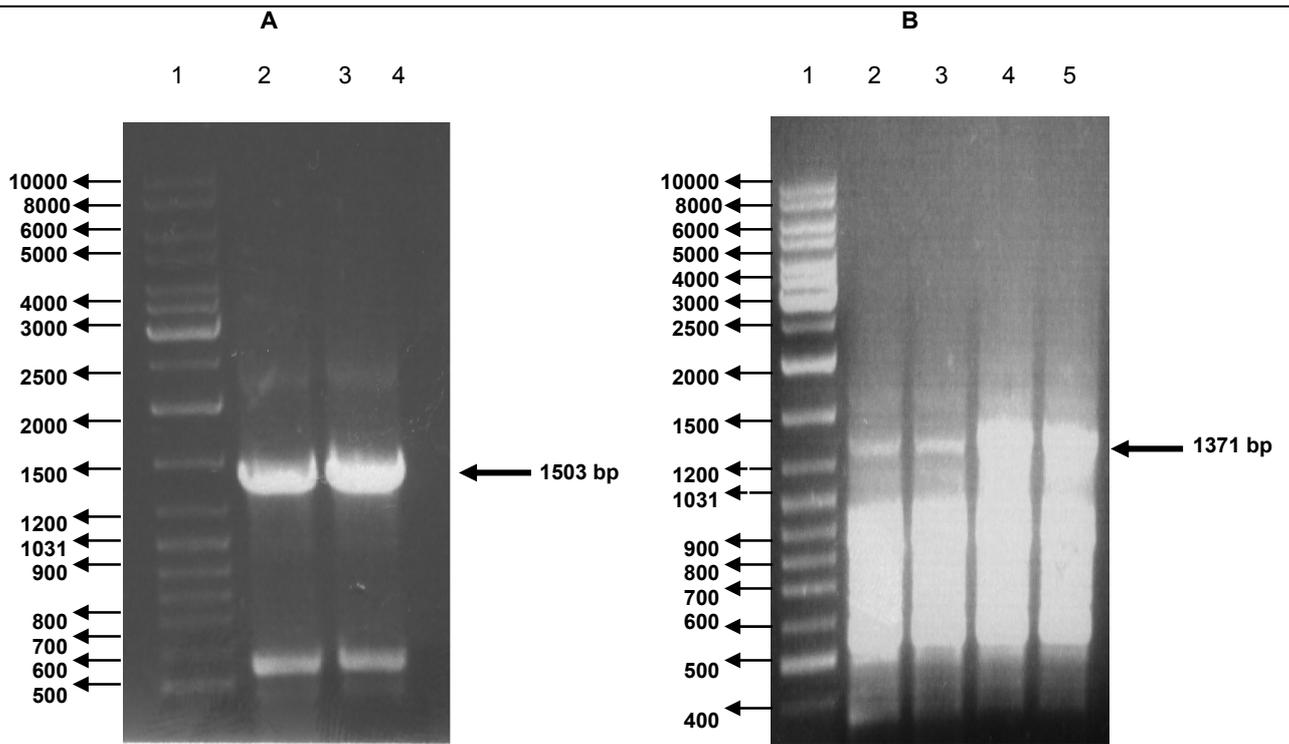


Figure A.3. PCR amplification of the PPE-MPTR genes (*Rv0442c* and *Rv0878c*, respectively). This figure shows the photographic representations of a 1% agarose gel with ethidium bromide staining, where 10 μ l of the PCR amplicon reactions of both genes were loaded in each well. (A) Lane 1: Gene Ruler molecular weight marker. Lane 2 and 3: PPE-MPTR, *Rv0442c* amplified using the inner primers, generating a fragment size of 1503 bp. Lane 4: H₂O control. (B) Lane 1: Gene ruler molecular weight marker. Lane 2,3,4 and lane 5: PPE-MPTR, *Rv0878c* amplified using the inner primers, generating a fragment size of 1371 bp. The fragment sizes (bp) are indicated with arrows.

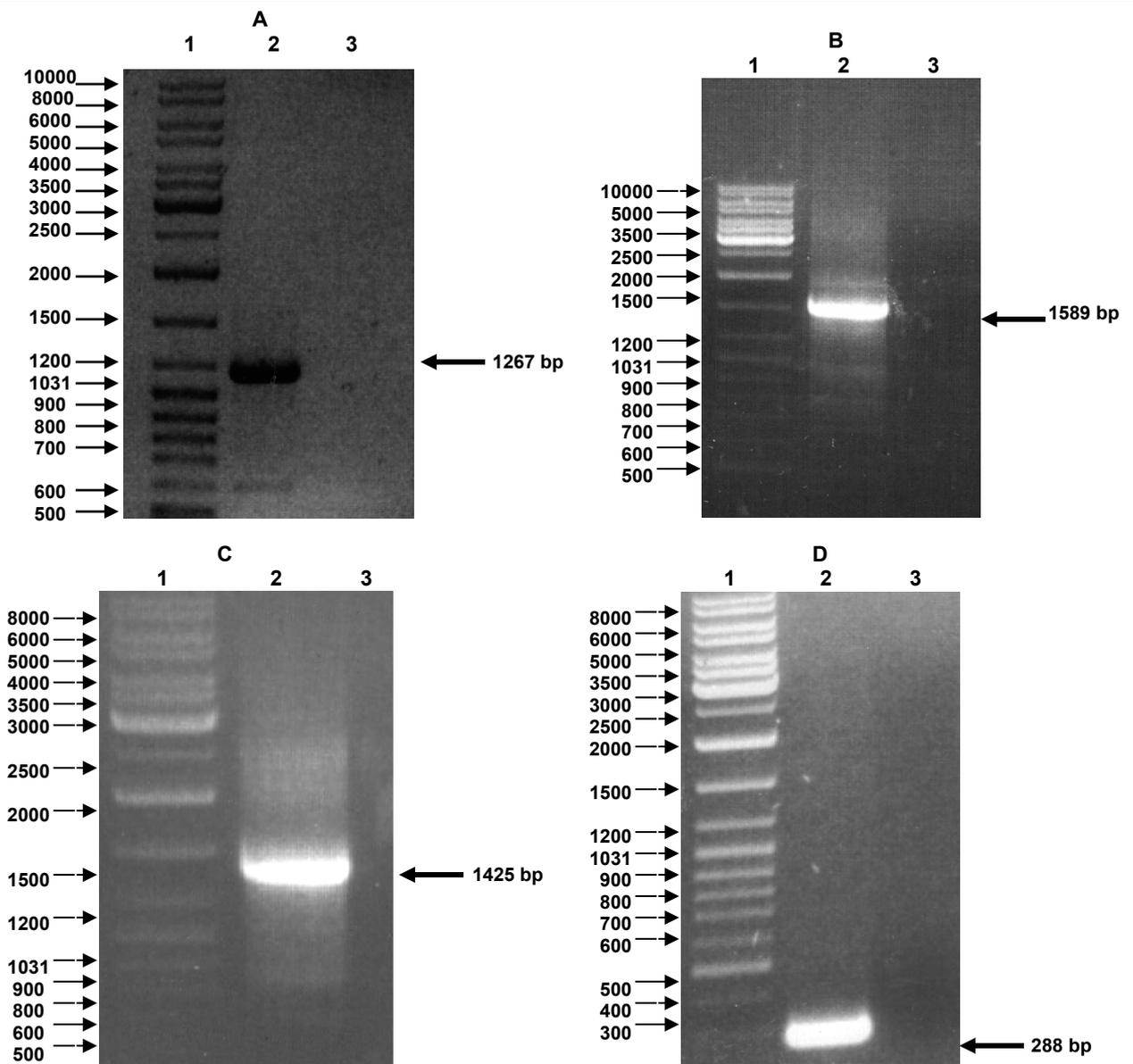


Figure A.4. PCR amplification of the PE_PGRS genes (*Rv0978c* and *Rv2615c*, using the outer and inner primers, Nested PCR) and *SmegESAT-6* (*Rv3875*) gene. This figure shows the photographic representations of a 1% agarose gel with ethidium bromide staining, where 10 μ l of the PCR amplicon reactions of both genes were loaded in each well. **(A)** Lane 1: Gene ruler molecular weight marker. Lane 2: PE_PGRS, *Rv0978c* amplified using the outer primers, generating a fragment size of 1267 bp. Lane 3: H₂O control. **(B)** Lane 1: Gene ruler molecular weight marker. Lane 2: PE_PGRS, *Rv2615c* amplified using the outer primers, generating a fragment size of 1589 bp. Lane 3: H₂O as control. **(C)** Lane 1: Gene ruler molecular weight marker. Lane 2: PE_PGRS, *Rv2615c* amplified using nested PCR; i.e. using product in **(B)** as a template and the inner primers generating a fragment size of 1425 bp, in Lane 3. **(D)** Lane 1: Gene ruler molecular weight marker. Lane 2: ESAT-6 (*Rv3875*, MSMEG_0066) amplified using the primers *SmegESAT-6f* and *SmegESAT-6r*, generating a fragment size of 288 bp, in Lane 3. The fragment sizes (bp) are indicated with arrows.

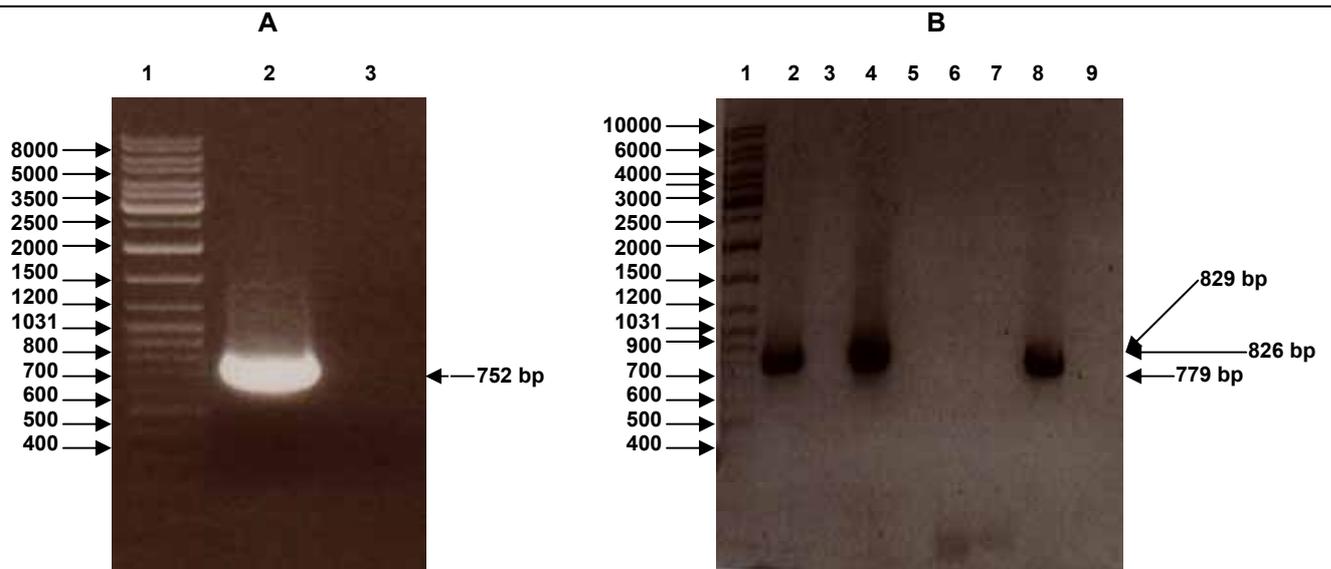


Figure A.5. PCR amplification of the regions upstream and downstream of the *M. smegmatis* ESAT-6 gene cluster region 1 and region 4. This figure shows the photographic representations of a 1% agarose gel with ethidium bromide staining, where 10 μ l of the PCR amplicons of both genomic regions upstream and downstream of region 3 and 4 were loaded in each well. **(A)** Lane 1: Gene ruler molecular weight marker. Lane 2: Genomic region DownMSMEG_1538 (DownstreamRv3444) of region 4 amplified using the primers downstreamMSMEG_1538 f and downstreamMSMEG_1538 r, generating a product of size, 752 bp. Lane 3: H₂O as control. **(B)** Lane 1: Gene ruler molecular weight marker. Lane 2: Genomic region UpstreamMSMEG_0605 (UpstreamRv0282) of region 3 amplified using the primers upstreamMSMEG_0605 f and upstreamMSMEG_0605 r, generating a product of size, 826 bp. Lane 3: H₂O as control. Lane 4: Genomic region DownstreamMSMEG_0616 (DownstreamRv0292) of region 3 amplified using the primers downstreamMSMEG_0616 f and downstreamMSMEG_0616 r, generating a product of size, 829 bp. Lane 5: H₂O as control. Lane 8: Genomic region UpstreamMSMEG_1532 (UpstreamRv3450c) of region 4 amplified using the primers upstreamMSMEG_1532 f and upstreamMSMEG_1532 r, generating a product of size, 779 bp. Lane 3: H₂O as control. Lane 9: H₂O as control. The fragment sizes (bp) are indicated with arrows.

6.2. ADDENDUM B.

6.2.1. Radioactive Labeling of Probe

A Sephadex spin column was used in the purification of the probes. The column was constructed by putting glass wool in the bottom of a syringe. The syringe was filled with Sephadex G50M in STE buffer (pH 8.0). The syringe was then flicked to remove air bubbles and to allow Sephadex to settle to the bottom of the syringe. Two hundred nanograms (200 ng) of DNA probe was added in 10 ml distilled water using a Random Primed Labelling Kit tube. Thirty two microlitres (32 μ l) of sterile water was added to the tube, and incubated at 100 °C for 5 minutes, with vortexing in between. The mixture was put on ice and 5 μ l [α -³²P]-dCTP was added. Thereafter 2,5 μ l Random Primed Labeling Enzyme (stored at room temperature) was added giving a total of 50 μ l. The mixture was quick-spun in a micro-centrifuge and gently mixed. The mixture was incubated at 37 °C for 60 minutes and thereafter 2 μ l "STOP" solution was added to the mixture, gently mixed and quick-spun. For removal of salts, the tip of the prepared column was put in an Eppendorf tube and both were put in a larger tube. Fifty microlitres (50 μ l) STE buffer was added to the column and spun in a desktop centrifuge at Nr 2 speed for 3 minutes, after which the washing step was repeated. The spin-through was discarded and a clean Eppendorf tube was put into the bottom of the syringe. The sample was added to the column and spun at Nr 2 speed in a bench centrifuge for 3 minutes, after which the sample was collected in a clean Eppendorf tube. Fifty microliters (50 μ l) STE buffer was added to the sample, after which it was centrifuged. Radio-actively labeled probe was collected at the bottom of the Eppendorf tube and aspirated and put into a clean labeled Eppendorf tube.

6.2.2. Hybridization and Detection

The baked filters were floated on the surface of a tray of 2x SSC solution until they have become thoroughly wetted from beneath. The filters were then submerged for 5 minutes, transferred into Pre-washing solution, and incubated for 30 minutes at 50°C with shaking. The filters were not allowed to dry out during any of the following stages at this point. Kimwipes soaked in Pre-washing solution were used to scrape the surfaces of the membranes to get rid of the bacterial debris (this reduces the background). The filters were sealed in plastic bags and Pre-hybridization solution, i.e. 30 ml for 18x19 cm membrane, was added. These were then incubated overnight at 42°C. ³²P-labeled probe was denatured by heating for 5 minutes to 100°C, and was immediately put on ice for 5 minutes. Fifty microliters (50 μ l) of the probe was added to the Hybridization solution. The Hybridization solution

containing the probe was added to the membranes, which were then sealed in plastic wrap. These were allowed to hybridize overnight at 42°C, whereafter the blots were washed in Washing buffer at room temperature for 30 minutes. The washing step was repeated. Thereafter the blots were washed again in Washing buffer at 50°C for 30 minutes. This step was repeated. The membranes were sealed in plastic wrap for autoradiography.

6.2.3. Solutions for Colony Blotting

A. 10 % SDS

Dissolve 10 g of SDS in 90 ml distilled water and heat to dissolve the crystals. Adjust the pH to 7.2 with HCl, adjust the volume to 100 ml. Store at room temperature.

B. Denaturing solutions

1.5 M NaCl

Dissolve 87.66 g of sodium chloride (NaCl, MW= 58.44 g/mol) in 1 liter of distilled water and sterilize by autoclaving. Store at room temperature.

0.5 M NaOH

Dissolve 20 g of sodium hydroxide (NaOH, MW=40.00 g/mol) in 1 liter of distilled water and sterilise by autoclaving. Store at room temperature.

C. Neutralizing solution

Dissolve 87.66 g NaCl (1.5 M), 0.37 g Na₂EDTA (0.001 M), and 60.5 g Tris-HCl (0.5 M) in 1 liter of distilled water, and adjust pH to 7.2. Store at room temperature.

D. 20x SSC

Dissolve 176.4 g Sodium citrate (0.3 M), hydrous or 154.8g anhydrous and 350.6 g NaCl (3 M) in 1.8 liters of distilled water and adjust the pH to 7.0 with a few drops of NaOH. Adjust the volume to 2 liters with distilled water, and sterilize by autoclaving. Use as 2x SSC solution (200 ml of 20x SSC in 2 liters of distilled water).

6.2.4. Solutions for Hybridization and Detection

A. Prewashing solution

Add 250 ml of a 20x SSC, 50 ml of a 10% SDS solution and 2 ml of a 0.5 M EDTA solution to a final volume of 1 liter with distilled water.

B. Pre-hybridization and Hybridization solution

For the preparation of Pre-hybridization solution 250 ml 20x SSPE, 50 ml 10% SDS, 20 ml 10 mg/ml Herring Sperm DNA and 50 ml of 100x Denhardt's solution are combined. To make Hybridization solution, a radioactively labelled probe is added to the Pre-hybridization solution.

C. Washing solution

Combine 200 ml 20x SSC and 20 ml 10% SDS and make up to a final volume of 2 liters with distilled water.

D. 20x SSPE

Dissolve 175.3 g NaCl, 27.6 g $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$, 7.4 g EDTA in 800 ml of distilled water and adjust the pH to 7.4 with NaOH and adjust the volume to 1 liter with distilled water. Sterilize by autoclaving and store at room temperature.

6.3. ADDENDUM C.

6.3.1. Protein concentration determination

To determine the protein concentration of the recombinant proteins, the Bio-Rad Protein assay was used. The protein concentrations of the samples were as indicated in Table C.1 and C.2, and in Figure C1 (A and B).

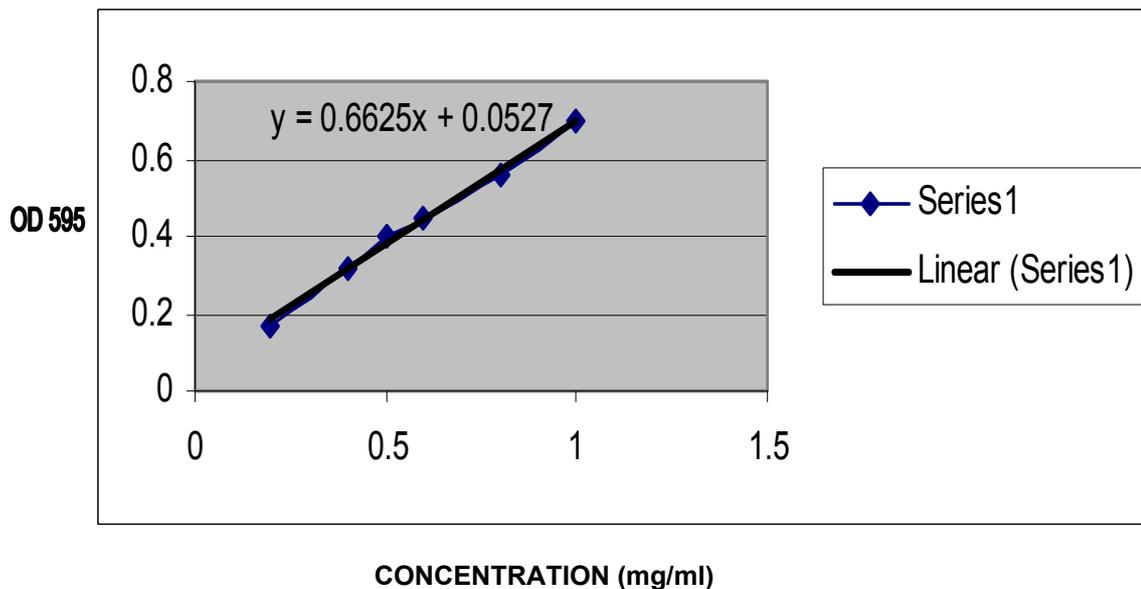
Table C.1. Absorbance of the standards and that of the samples and their concentrations.

BSA Standard Concentration Series:				
BSA Dilutions (mg/ml)		Reading		
0.2		0.165		
0.4		0.312		
0.5		0.404		
0.6		0.444		
0.8		0.559		
1		0.695		
Samples:				
Sample name	Sample reading	Concentration of dilution (mg/ml), 10X Dilution	Concentration of sample (mg/ml) multiplied by 10 (X10)	For 3 micrograms (microliters)
Rv0442c -CF	0.422	0.55743396	5.57433962	0.538180341
Rv0442c -WCL	0.305	0.38083019	3.80830189	0.787752675
Rv0442c -CWL	0.233	0.27215094	2.72150943	1.102329451
Rv0442c -MEM	0.703	0.98158491	9.81584906	0.305628172
Rv0442c -CYT	0.341	0.43516981	4.35169811	0.689386056
Rv0878c -CF	0.406	0.53328302	5.33283019	0.562553071
Rv0878c -WCL	0.234	0.27366038	2.73660377	1.096249311
Rv0878c -CW	0.239	0.28120755	2.81207547	1.066827697
Rv0878c -MEM	0.271	0.32950943	3.29509434	0.910444343
Rv0878c -CYT	0.425	0.56196226	5.61962264	0.533843674
Rv2615c -CF	0.584	0.89819537	8.98195371	0.334003057
Rv2615c -WCL	0.26	0.26265202	2.6265202	1.142195668
Rv2615c -CW	0.128	0.00372695	0.03726952	80.49473684
Rv2615c -MEM	0.248	0.23911338	2.39113378	1.254634947
Rv2615c -CYT	0.302	0.34503727	3.4503727	0.869471291

Table C.2. Absorbance of the standards and that of the samples and their concentrations.

BSA Standard Concentration Series:				
BSA Dilutions (mg/ml)			Reading	
0.2			0.171	
0.4			0.32	
0.5			0.417	
0.6			0.496	
0.8			0.546	
1			0.647	
Samples:				
Sample name	Sample reading	Concentration (mg/ml), 10X Dilution	Concentration of Samples, multiplied 10 (X10)	For 3 micrograms (microliters)
Rv3872-CF	0.132	0.012	0.116256158	258.051
Rv3872-WCL	0.292	0.327	3.268965517	9.177
Rv3872-CWL	0.184	0.114	1.1408867	26.295
Rv3872-MEM	0.212	0.169	1.692610837	17.724
Rv3872-CYT	0.56	0.855	8.549753695	3.509
Rv3873-CF	0.274	0.291	2.914285714	10.294
Rv3873-WCL	0.231	0.207	2.066995074	14.514
Rv3873-CWL	0.161	0.069	0.687684729	43.625
Rv3873-MEM	0.154	0.055	0.549753695	54.570
Rv3873-CYT	0.47	0.678	6.77635468	4.427

A.



B.

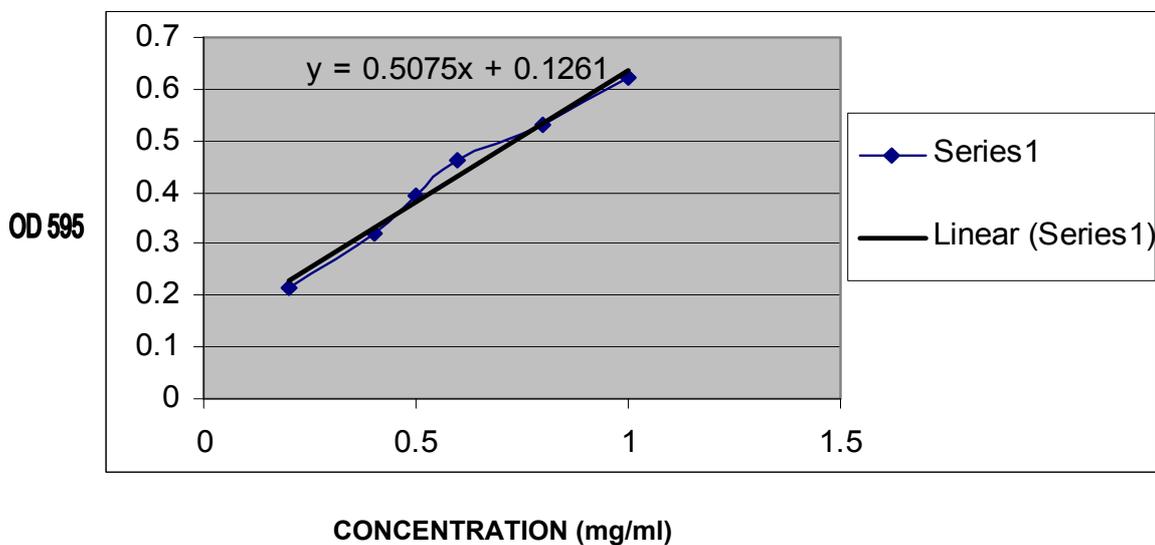


Figure C.1. Standard curves for the protein assay, absorbance against protein concentration. (A) A graph curve for *M. tuberculosis* PPE-MPTR, Rv0442c, Rv0878c and PE_PGRS, (Rv2615c) expressed in *M. smegmatis*. (B) A graph curve for the PE35 (Rv3872) and PPE68 (Rv3873) of *M. smegmatis*.

6.4. ADDENDUM D

6.4.1. SDS PAGE Buffers and solutions preparations.

A. Acrylamide-bisacrylamide (30%)

Dissolve 30 g Acrylamide and 0.8 g Bis-acrylamide in 100 ml distilled water. Filter through Whatman filter paper no.1. Store at 4°C in the dark.

B. 1.5 M Tris-HCl; pH 8.8

Dissolve 36.3 g Tris in approximately 150 ml distilled water. Titrate to pH 8.8 with HCl and make up to 200 ml with distilled water.

C. 1 M Tris-HCl; pH 6.8

Dissolve 121.1 g Tris in 800 ml of distilled water. Adjust the pH to 6.8 with HCl. Adjust the volume of the solution to 1 liter with distilled water.

D. Staining gel buffer.

Dissolve 3 g to make 500 mM Tris in approximately 40 ml distilled water. Titrate with HCl to pH 6.8. Make up to 50 ml with dH₂O. Filter through the Whatman filter paper no.1. Store at 4°C .

E. 10% and 20% SDS (Sodium dodecyl sulphate).

Dissolve 10 g (for 10% SDS) and 20 g (for 20%SDS), respectively in 90 ml distilled water. Heat to solute the crystals. Adjust the pH to 7.2 with HCl. Adjust the volume to 100 ml. Store at room temperature.

F. Running buffer (SDS-PAGE)

Dissolve 6g Tris to make 25 mM Tris, 28.8 g to make 192 mM Glycine and 20 ml SDS (10%) and make up to 2 liters with distilled water. Store at room temperature or 4°C.

G. EDTA (0.5 M)

Dissolve 186.12 g of disodium ethylene diaminetetra-acetate. H₂O in 800 ml distilled water, stir vigorously on a magnetic stirrer. Adjust the pH to 8.0 with NaOH and sterilize by autoclaving. Store at room temperature.

H. 2X Reducing SDS Sample Buffer

Add 3.4 ml 1 M Tris-HCl (pH 6.8), 1 ml β -mercaptoethanol, 3 ml SDS (20%), 2 ml Glycerol, 500 μ l Bromophenol blue (0.75 %), and 200 μ l EDTA (0.5 M). Store at -20°C or 4°C.

I. Coomassie blue

Dissolve 2.5 g Coomassie blue R250 in a solution of 450 ml Methanol, 450 ml dH₂O, and 100 ml Acetic acid. Filter through Whatman paper. Store at room temperature.

J. Destain solution.

Add 375 ml Glacial Acetic acid, 250 ml Methanol, 100 ml Glycerol, and 1275 ml distilled water. Make up to 2000 ml.

K. Ammonium persulphate (APS) (10% w/v)

Dissolve 1 g Ammonium per sulphate in 10 ml distilled water. Store at 4°C.

6.5. ADDENDUM E

6.5.1. Western blotting buffers

A. Blotting/Transfer buffer

Dissolve 3.03 g Tris (25 mM), 14.4 g Glycine (192 mM) and 200 ml Methanol (20%) in 800 ml of distilled water. Store at 4°C.

B. TBS-T Wash buffer (1%)

Add 50 ml 20 mM Tris, (1 M, pH 7.6), 137 ml/40.03 g 137 mM NaCl (5 M), and 5 ml Tween 20.

Make up to a final volume of 5 liters with distilled water.

C. Blocking buffer

Dissolve 10 g Fat free milk powder in 200 ml of TBS-T wash buffer. Add 4 ml of 5-10 mM Sodium Azide (NaN₃) to prevent microbial growth. Store at 4°C.

CHAPTER 7: LIST OF REFERENCES

- Adindla, S. and Guruprasad, L. (2003). Sequence analysis corresponding to the PPE and PE proteins in *Mycobacterium tuberculosis* and other genomes. *J Biosci*, **28**:169-179.
- Aldovini, A., Husson, R. N. & Young, R. A. (1993). The *uraA* locus and homologous recombination in *Mycobacterium bovis* BCG. *J Bacteriol* **175**, 7282-7289.
- Andersen, P., Andersen, A.B., Sorensen, A.L., and Nagai, S. (1995). Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice. *J. Immunol.* **154**:3359-3372.
- Azad, A. K., Sirakova, T. D., Fernandes, N. D. & Kolattukudy, P. E. (1997). Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J Biol Chem.* **272**, 16741-16745.
- Azad, A. K., Sirakova, T. D., Rogers, L. M. & Kolattukudy, P. E. (1996). Targeted replacement of the mycocerosic acid synthase gene in *Mycobacterium bovis* BCG produces a mutant that lacks mycosides. *Proc Natl Acad Sci USA* **93**, 4787-4792.
- Balasubramanian, V., Pavelka, M. S., Jr, Bardarov, S. S., Martin, J., Weisbrod, T. R., McAdam, R. A., Bloom, B. R. & Jacobs, W. R., Jr (1996). Allelic exchange in *Mycobacterium tuberculosis* with long linear recombination substrates. *J Bacteriol* **178**, 273-279.
- Banu, S., Honore, N.B., Saint-Joanis, B., Philpott, D., Prevost, M.C. and Cole, S.T. (2002). Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol. Microbiol.* **44**:9-19.
- Berthet, F.X., Rasmussen, P.B., Rosenkrands, I., Andersen, P., Gicquel, B. A. (1998). *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular mass culture filtrate protein (CFP-10). *Microbiology* **144**:3195-3203.

Betts J.C., Lukey, P.T., Robb, L.C., McAdam, R.A., Duncan, K. (2002). Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol* **43**:717-731.

Brennan, M.J., Delogu, G., Chen, Y., Bardarov, S., Kriakov, J., Alavi, M. and Jacobs Jr., W.R. (2001). Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect. Immun.* **69**:7326–7333.

Brennan, M.J., Gey van Pittius, N., Espitia, C (2005). The PE and PPE multigene families of Mycobacteria. *Tuberculosis and the Tubercle Bacillus*, Chapter 33:513-25.

Brodin, P., Eiglmeier, K., Marmiesse, M., Billault, A., Garnier, T., Niemann, S., *et al.* (2002). Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect Immun* **70**:5568–5578.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., and Cole, S.T. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA.* **99**:3684-3689.

Brusasca, P.N., Colangeli, R., Lyashchenko, K.P., Zhao, X., Vogelstein, M., Spencer, J.C., McMurray, D.N., Gennaro, M.L. (2001). Immunological characterization of antigens encoded by the RD1 region of the *Mycobacterium tuberculosis* genome. *Scand. J. Immunol.* **54**:448-452.

Calmette, A and Guerin, C. (1920). *Ann. Inst. Pasteur* **34**:553.

Chaitra, M.G., Hariharaputran, S., Chandra, N.R., Shaila, M.S., Nayaka, R (2005). Defining putative T cell epitopes from PE and PPE families of proteins of *Mycobacterium tuberculosis* with vaccine potential. *Vaccine* **23**:1265-1272.

Chakhaiyar, P., and Hasnain, S.E. (2004). Defining the mandate of tuberculosis research in a postgenomic era. *Med. Princ. Pract.* **13**: 177–184.

Colangeli, R., Spencer, J.S., Bifani, P., Williams, A., Lyashchenko, K., Keen, M.A., Hill, P.J., Belisle, J., Gennaro, M.L. (2000). MTSa-10, the product of the Rv3874 gene of *Mycobacterium tuberculosis*, elicits tuberculosis-specific, delayed-type hypersensitivity in guinea pigs. *Infect. Immun.* **68**:990-993.

Cole S.T. (1998). Comparative mycobacterial genomics. *Curr Opin Microbiol* **1**:567–571.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry III CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream, M.A, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.

Cole, S.T. (2002). Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiology* **148**:2919-2928.

Cole, S.T., (1999). Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett.* **452**:7-10.

Cole, S.T., and Barrell, B.G. (1998). Analysis of the genome of *Mycobacterium tuberculosis* H37Rv. *Novartis. Found. Symp.* **217**:160-172.

Cousins, D. V., Bastida, R., Cataldi, A., Quse, V., Redrobe, S., Dow, S., Duignan, P., Murray, A., Dupont, C., Ahmed, N., Collins, D.M., R. Butler, W., Dawson, D., Rodriguez, D., Loureiro, J., Romano, M.I., Alito, A., Zumarraga, M., and Bernardelli, A. (2003). Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int. J. Syst. Evol. Microbiol.* **53**:1305–1314.

Crubezy, E., Ludes, B., Poveda, J., Clayton, J., Crouau-Roy, B., and Montagnon, D. (1998). Identification of *Mycobacterium* DNA in an Egyptian Pott's disease of 5,400 years old, C.R. *Acad. Sci III.* **321**: 941-951.

Daugelat, S., Kowall, J., Mattow, J., Bumann, D., Winter, R., Hurwitz, R., Kaufmann, S.H.E. (2003). The RD1 proteins of *Mycobacterium tuberculosis*: expression in *Mycobacterium smegmatis* and biochemical characterization. *Microbes and Infection* **5**:1082-1092.

Delogu, G., and Brennan, M.J. (2001). Comparative immune response to PE and PE-PGRS antigens of *Mycobacterium tuberculosis*. *Infect. Immunol.* **69**:5606-11.

Demangel, C., Brodin, P., Cockle, P.J., Brosch, R., Majlessi, L., Leclerc, C., and Cole, S.T. (2004). Cell Envelope Protein PPE68 Contributes to *Mycobacterium tuberculosis* RD1 Immunogenicity Independently of a 10-Kilodalton Culture Filtrate Protein and ESAT-6. *Infection and Immunity*: 2170-2176.

Dillon, D.C., Alderson, M.R., Day, C.H., Bement, T., Campos-Neto, A., Skeiky, Y.A., Vedvick, T., Badaro, R., Reed, S.G., Houghton, R (2000). Molecular and immunological characterization of *Mycobacterium tuberculosis* CFP-10, an immunodiagnostic antigen missing in *Mycobacterium bovis* BCG. *J. Clin. Microbiol.* **38**:3285-3290.

Dillon, D.C., Alderson, M.R., Day, C.H., Lewinsohn, D.M., Coler, R., Bement, T., *et. al.* (1999). Molecular characterization and human T cell responses to a member of a novel *Mycobacterium tuberculosis* mbt39 gene family. *Infect. Immunol.* **67**:2941-50.

Doran, T.J., Hodgson, A.L., Davies, J.K., and Radford, A.J. (1992). Characterisation of a novel repetitive DNA sequence from *Mycobacterium bovis*. *FEMS Microbiol. Lett.* **75**: 179-185.

Ehlers, S. (1993). The wolf at the door. Some thoughts on the biochemistry of the tubercle bacillus. *S. Afri. Med. J.* **83**:900-903.

Elhers, S (1999). Immunity to tuberculosis: a delicate balance between protection and pathology. *FEMS Immunol. Med. Microbiol.* **23**:149-158.

Espitia, C., Laclete, J.P., Modragon-Palomimno, M., Amador, A., Campuzano, J., Martens, A., *et. al.* (1999). The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis* : a new family of fibronectin-binding proteins. *Microbiology* **145**:387-95.

Gey van Pittius, N., Gamielien, J., Hide, W., Gordon, D.B., Siezen, R.J., and Beyers, A.D. (2001). The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C gram-positive bacteria. *Genome Biology* 2(10): research 0044.1-0044.18.

Gey van Pittius, N.C., Sampson, S.L., Lee, H., Kim, Y., Helden, P.D., and Warren, R.M. (2006). Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (*esx*) gene cluster regions. *BMC Evolutionary Biology* 6(95):1-31.

Gey van Pittius, N.C., Sampson, S.L., Lee, H., Kim, Y., Van Helden, P.D., and Warren, R.M. (2003). The evolutionary history of the expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and its association with the duplication of the ESAT-6 gene clusters. *Keystone Symp*:471.

Gordon, S.V., Eiglmeier, K., Brosch, R., Gramier, T., Honore, N., Barrel, B.G, and Cole, S.T. (1999). Genomics of *Mycobacterium tuberculosis* and *Mycobacterium leprae*. In 'Mycobacteria: Molecular biology and virulence.' (Eds. C. Ratledge and J. Dale): 93-109 (Blackwell Science Ltd:Oxford).

Greth, A., Flamand J. R., and Delhomme A. (1994). An outbreak of tuberculosis in a captive herd of Arabian oryx (*Oryx leucoryx*): management. *Vet. Rec.* 134:165–167.

Harboe, M., A. S. Malin, H. S. Dockrell, H. G. Wiker, G. Ulvund, A. Holm, M. C. Jorgensen, and P. Andersen. (1998). B-cell epitopes and quantification of the ESAT-6 protein of *Mycobacterium tuberculosis*. *Infect. Immun.* 66: 717-723.

Hermans, P.W., van Soolingen, D., and van Embden, J.D. (1992). Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. *J Bacteriol* 174:4157-4165.

<http://www.umdnl.edu/ntbcweb/cd/tbhistory.htm>

<http://genolist.pasteur.fr/TubercuList/>

<http://cmr.tigr.org/tigr-scripts/CMR/GenomePage.cgi?org=gms>

Huard, R.C., Fabre, M., de Haas, P., Lazzarini, L.C.O., van Soolingen, D., Cousins, D., and Ho, J.L. (2006). Novel Genetic Polymorphisms That Further Delineate the Phylogeny of the *Mycobacterium tuberculosis* Complex. *J. Bacteriol.* **188**:4271-4278.

Husson, R. N., James, B. E. & Young, R. A. (1990). Gene replacement and expression of foreign DNA in mycobacteria. *J Bacteriol.* **172**:519-524.

Jacobs, W.R., Jr., Kalpana, G.V., Cirollo, J.D., Pascopella, L., Snapper, S.B., Udani, R.A., Jones, W., Barletta, R.G., and Bloom, B.R. (1991). Genetic systems for mycobacteria. *Methods Enzymol* **204**:537-555.

Kalpana, G. V., Bloom, B. R. & Jacobs, W. R. (1991). Insertional mutagenesis and illegitimate recombination in mycobacteria. *Proc Natl Acad Sci USA* **88**, 5433-5437.

Kana, B.D. and Mizrahi, V (2004). Molecular genetics of *Mycobacterium tuberculosis* in relation to the discovery of novel drugs and vaccines. *Tuberculosis* **84**:63-75.

Ke, S.H. and Madison EL, (1997). Rapid and efficient site-directed mutagenesis by single-tube 'megaprimer' PCR method. *Nucl. Acad. Res.* **25**:3371-3372.

Kock, R. (1882). Die Aetiologie der Tuberkulose. *Ber. Klin. Wochenschrift* **15**.

Lewis, K.N., Liao, R.L., Guinn, K.M., Hicky, M.J., Smith, S., Behr, M.A., and Sherman, D.R. (2003). Deletion of RD from *Mycobacterium tuberculosis* mimics bacillus Calmette-Guerin attenuation. *J. Infect. Dis.* **187**:117-123.

Machowski, E.E., Dawes, S., and Mizrahi, V. (2005). TB tools to tell the tale-molecular genetic methods for mycobacterial research. *The Int. J.Biochem. and Cell Biol.* **37**:54-68.

Mckinney, J.D., Jacobs, W.R., Jr., and Bloom, B.R. (1998). Persisting problems in tuberculosis. *In Emerging infections.* 51-146. (Academic Press).

Norman, E., Dellagostin, O. A., McFadden, J. & Dale, J. W. (1995). Gene replacement by homologous recombination in *Mycobacterium bovis* BCG. *Mol Microbiol* **16**, 755-760.

Okkels, L. M., I. Brock, F. Follmann, E. M. Agger, S. M. Arend, T. H. M. Ottenhoff, F. Oftung, I. Rosenkrands, and P. Andersen. (2003). PPE protein (Rv3873) from DNA segment RD1 of *Mycobacterium tuberculosis*: strong recognition of both specific T-cell epitopes and epitopes conserved within the PPE family. *Infect. Immun.* **71**:6116–6123.

Okkels, L.M., and Andersen, P. (2004). Protein-protein interactions of proteins from the ESAT-6 family of *Mycobacterium tuberculosis*. *J. Bacteriol.* 2487-2491.

Pallen, M.J. (2002) The ESAT-6/WXG100 super family and a new gram positive secretion system? *Trends Microbiol.* **10**:209–212.

Parish, T., and Stoker, N.G., (2000). Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis* *tlpA plcABC* mutant by gene replacement. *Microbiol.* **146**:1969-1975.

Proding, W. M., Eigentler, A., Allerberger, F., Schonbauer, M., and Glawschnig, W. (2002). Infection of red deer, cattle, and humans with *Mycobacterium bovis* subsp. *caprae* in western Austria. *J. Clin. Microbiol.* **40**:2270–2272.

Pym, A.S., Brodin, P., Brosch, R., Huerre, M., and Cole, S.T. (2002). Loss of RD1 contributed to the attenuation of the live tuberculosis vaccine *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol. Microbiol.* **46**:709-717.

Pym, A.S., Brodin, P., Majlessi, L., Brosch, R., Demangel, C., Williams, A., Griffiths, K.E., Marchal, G., Leclerc, C. and Cole, S.T. (2003). Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat. Med.* **9**:533–539.

Ramakrishnan, L., Federspiel, N.A., Falkow, S. (2000). Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. *Science* **288**:1436–1439.

Renshaw, P.S., Panagiotidou, P., Whelan, A., Gordon, S.V., Hewinson, G., Williamson, R.A., and Carr, M.D. (2002). Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6.CFP-10 complex. *J. Biol. Chem.* **277**: 21598-21603.

Reyrat, J.M., Berthet, F.X., and Gicquel, B. (1995). The urease locus of *Mycobacterium tuberculosis* and its utilization for the demonstration of allelic exchange in *Mycobacterium bovis* bacillus Calmette-Guerin. *Proc. Natl. Acad. Sci. USA.* **92**:8768-8772.

Salo, W.L., Aufderheide, A.C., Buckstra, J., and Holcomb, T.A. (1994). Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc. Natl. Acad. Sci. U.S.A.*: 2091-2094.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sampson, S.L., Lukey, P., Warren, R.M., van Helden, P.D., Richardson, M. and Everett, M.J. (2001). Expression, characterization and sub cellular localization of the *Mycobacterium tuberculosis* PPE gene Rv1917c. *Tuberculosis* **81**:305–317.

Sassetti, C.M., and Rubin, E.J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**:12989-12994.

Singh, K.K., Zhang, X., Patibandla, A.S., Chien, P. Jr., Laal, S. (2001). Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. *Infec Immunol.* **69**:4185-91.

Skeiky, Y.A., Owendale, P.J., Jen, S., Alderson, M.R., Dillon, D.C., Smith, S., *et. al.* (2000). T-cell expression cloning of a *Mycobacterium tuberculosis* gene encoding a putative antigen associated with the early control of infection. *J. Immunol.* **165**:7140-9.

Smith, A. M., and Klugman, K. P. (1997). "Megaprimer" method of PCR-based mutagenesis: the concentration of megaprimer is a critical factor. *BioTechniques* **22**:438-442.

Smith, I. (2003). *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. *Clinical Microbiology Reviews* **16** (3):463–496.

Smith, N. (1960). The dassie bacillus. *Tubercle* (London) **41**:203–212.

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

Stead, W.W., Eisenach, K.D., Cave, M.D., Beggs, M.L., Templeton, G.L., Thoen, C.O., Bates, J.H. (1995). When did *Mycobacterium tuberculosis* infection first occur in the New World? An important question with public health implications. *Am J Respir. Crit. Care Med.* **151**:1267–1268.

Stenger, S., and Modlin, R. T cell mediated immunity to *Mycobacterium tuberculosis*. (1999). *Curr. Opin. Microbiol.* **2**: 89-93.

Strong, M., Mallick, P., Wang, S., and Eisenberg, D. (2003). Computational and biochemical analysis of the *Mycobacterium tuberculosis* genome. *Keystone Symp.*:453.

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006). Toward the structural genomics of complexes: Crystal structure of the PE/PPE protein complex from *Mycobacterium tuberculosis*. *PNAS* **103** (21):8080-8065.

Tekaia, F., Gordon, S.V., Garnier, T., Brosch, R., Barrell B.G., Cole, S.T. (1999). Analysis of the proteome of *Mycobacterium tuberculosis in silico*. *Tuberc Lung Dis.* **79**:329-42.

Tundup, S., Akhter, Y., Thiagarajan, D., and Hasnain, S.E. (2006) Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. *FEBS Lett.* **580**:1285–1293.

van Soolingen, D., de Haas P. E., Haagsma J., Eger T., Hermans P. W., Ritacco V., Alito A., and van Embden J. D. (1994). Use of various genetic markers in differentiation of *Mycobacterium bovis* strains

from animals and humans and for studying epidemiology of bovine tuberculosis. *J. Clin. Microbiol.* **32**:2425–2433.

Wagner, J. C., and Bokkenheuser V. (1961). The *Mycobacterium* isolated from the dassie *Procavia capensis* (Pallas). *Tubercle* **42**:47–56.

Wagner, J. C., Buchanan G., V Bokkenheuser., and Leviser S. (1958). An acid-fast bacillus isolated from the lungs of the Cape hyrax, *Procavia capensis* (Pallas). *Nature* **181**:284–285.

Young, D.B. Blueprint for the white plague. (1998). *Nature* **393**:515-516.