

https://dspace.org/

# The construction and phenotypic characterization of mycobacterial mutants deficient in DNA glycosylases

**Goosens, Vivianne Jacoba**

**2009-04-09T08:53:06Z**
http://hdl.handle.net/10539/6883

Downloaded from a DSpace repository. https://dspace.org/

# The Construction and Phenotypic Characterization of Mycobacterial Mutants Deficient in DNA Glycosylases

Vivianne Jacoba Goosens

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, in fulfilment of the requirements for the degree of Master of Science.

Johannesburg, 2008

‘Keep on swimming, swimming, swimming’

- Dory

In Walt Disney’s Finding Nemo

# Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the degree of Masters of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

_____

(Vivianne J Goosens)

_____ day of _____ 2008

# Abstract

*Mycobacterium tuberculosis* is an exquisitely adapted intracellular pathogen that encounters hostile, host-derived reactive nitrogen and oxygen intermediates during the course of infection of its human host. These radicals cause DNA damage, which is repaired through various pathways to allow for the continued survival of the organism. Base excision repair (BER) is one such pathway, which depends on DNA glycosylases to identify and excise damaged DNA bases. Formamidopyrimidine DNA glycosylase (Fpg/ MutM/ FAPY) and Endonuclease VIII (Nei) are such enzymes, which both target oxidatively damaged DNA and together, form the Fpg family of DNA glycosylases. Bioinformatic analyses identified two copies each of Fpg and Nei-encoding genes in *M. tuberculosis* as well as in its non-pathogenic relative, *Mycobacterium smegmatis*. To understand the role of these multiple glycosylases in the maintenance of genomic integrity and survival of mycobacteria, the genes encoding the four Fpg/Nei glycosylases were individually deleted in *M. smegmatis* strain $mc^2155$ by homologous recombination. In addition to the four single mutants, double and triple Fpg and Nei glycosylase knockout mutants were generated by sequential gene knockout. When compared to the parental strain, the single and double mutants showed no variation in growth kinetics, no increased sensitivity to hydrogen peroxide and no increase in spontaneous mutation rates. However, a slight increase in frequency of spontaneous C→T transition mutations was observed in double knockout mutants compared to the wild type and single mutant strains. These results suggest that these enzymes may be part of an extensive network of enzymes which collectively work to enhance the overall survival of *M. smegmatis* through the repair of oxidatively damaged DNA.

# Acknowledgements

I am enormously grateful for the funding I received from the Molecular Mycobacterial Research Unit (MMRU), Medical Research Council (MRC), the National Research Foundation (NRF) and the University of the Witwatersrand during the course of my Masters degree.

I would like to thank my supervisor Dr. Bhavna Gordhan for her patience, guidance, encouragement and unfailing support. My co-supervisor Prof. Valerie Mizrahi for her invaluable input and the ideal environment she has created for students to grow and develop.

I would also like to thank the people of the MMRU; it is due to their constant support, reassurance, guidance, wit and friendship that I have reached this point and still love what I'm doing.

And finally, to my family and friends, thank you. To both my parents who, without question, are incredibly proud and support me in every way. To my little sister who always says the sweetest things, my friends who pick me up and make me laugh, and the Candy's who encourage me. And last, but definitely not least, I would like to thank Geoff, who holds me together and is my everything.

# Presentations

Parts of this work have been presented at the following conferences:

1. **Goosens, V. J**., Mizrahi V. and Gordhan B. G. (2007)

   The construction and phenotypic characterization of mycobacterial mutants defective in DNA glycosylases. MRC Research Day, 18th October 2007. MRC, Cape Town, South Africa (poster presentation).

2. **Goosens, V. J**., Mizrahi V. and Gordhan B. G. (2008)

   The Construction and phenotypic characterization of mycobacterial mutants defective in DNA glycosylases. The Bio-08 conference, 21-25 January 2008. 1820 Settlers' Monument, Grahamstown, South Africa. (oral presentation)

# Table of contents

Declaration ..... iii

Abstract ..... iv

Acknowledgements ..... v

Presentations ..... vi

Table of contents ..... vii

List of figures ..... xii

List of tables ..... xv

Nomenclature ..... xvi

1. Introduction ..... 1

1.1 *M. tuberculosis* infection ..... 3

1.2 Oxidative stress and the mycobacterial response ..... 4

1.3 Base excision repair (BER) ..... 6

1.4 Fpg and Nei DNA glycosylases ..... 7

1.5 Gene knockout by homologous recombination ..... 10

1.6 Aim ..... 12

2. Materials and methods ..... 13

2.1 Bioinformatic tools ..... 13
2.1.1 Artemis and ACT ..... 13
2.1.2 BLAST searches ..... 13
2.1.3 ClustalW ..... 14
2.1.4 Cluster of orthologous groups (COG) ..... 14
2.1.5 GenoList ..... 14
2.1.6 InterPro ..... 15
2.1.7 Prosite ..... 15
2.1.8 Pfam ..... 15
2.1.9 TIGR-CMR ..... 15
2.2 Bacterial strains, plasmids and maintenance of strains. ..... 16
2.2.1 Bacterial strains and culturing conditions ..... 16
2.2.2 Cloning vectors ..... 18
2.2.3 Assessment of cell viability ..... 21
2.2.4 Replica plating ..... 21
2.2.5 Spotting assays ..... 22
2.3 DNA manipulation ..... 22
2.3.1 DNA extraction ..... 22
2.3.1.1 Chromosomal DNA extraction from *M. smegmatis* ..... 22
2.3.1.1.1 Bulk chromosomal DNA extraction ..... 22
2.3.1.2 Small-scale chromosomal DNA extraction ..... 23
2.3.1.3 Plasmid extraction from *E. coli* ..... 23
2.3.2 Enzymatic modification of DNA ..... 25
2.3.2.1 Polymerase Chain Reaction (PCR) amplification ..... 25
2.3.2.2 Restriction digestion ..... 26
2.3.2.3 Ligation ..... 27
2.3.2.4 Phosphorylation ..... 27
2.3.2.5 De-phosphorylation ..... 28
2.3.3 DNA precipitation ..... 28

2.3.3.1 Isopropanol precipitation ..... 28
2.3.3.2 Ethanol precipitation ..... 28
2.3.4 The separation of DNA fragments by agarose gel electrophoresis ..... 29
2.3.5 Recovery of DNA from agarose gel ..... 29
2.3.6 Quantification of DNA ..... 30
2.3.7 Southern blot techniques ..... 30
2.3.7.1 Electro-blotting ..... 30
2.3.7.2 Radioactive labelling and hybridization ..... 31
2.3.7.3 Non-radioactive labelling and hybridization ..... 32
2.4 Transformation of bacteria ..... 33
2.4.1 Chemical transformation in *E. coli* ..... 34
2.4.1.1 Preparation of chemically competent *E. coli* ..... 34
2.4.1.2 Transformation of *E. coli* ..... 34
2.4.2 Electroporation into *M. smegmatis* ..... 35
2.4.2.1 Preparation of electro-competent *M. smegmatis* ..... 35
2.4.2.2 Electroporation ..... 35
2.5 DNA sequencing ..... 36
2.6 The construction and identification of site specific deletion mutants in *M. smegmatis* ..... 36
2.6.1.1 Construction of knockout vectors ..... 37
2.6.1.2 Identification of Single Cross Overs (SCOs) ..... 38
2.6.1.3 Identification of Double Crossovers (DCOs) ..... 39
2.7 Phenotypic characterization of mutant strains ..... 40
2.7.1 Growth kinetics ..... 40
2.7.2 Sensitivity to oxidative stress ..... 41
2.7.2.1 Hydrogen peroxide susceptibility assays ..... 41
2.7.3 Mutational assessments ..... 42
2.7.3.1 Mutation frequency ..... 42
2.7.3.2 Mutation rate ..... 43

3. Results ..... 48

3.1 Bioinformatics analysis ..... 48

3.1.1 Identification of DNA glycosylase-encoding *fpg* and *nei* genes in *M. smegmatis* ..... 48

3.1.2 Comparison of Fpg and Nei glycosylases in *M. tuberculosis* and *M. smegmatis* ..... 52

3.2 Construction of suicide vectors ..... 55

3.2.1 Construction of suicide vectors for the deletion of *fpgI* ..... 56

3.2.2 Construction of suicide vectors for the deletion of *fpgII* ..... 59

3.2.3 Construction of suicide vectors for the deletion of *neiI* ..... 61

3.2.4 Construction of suicide vectors for the deletion of *neiII* ..... 62

3.3 The generation of knockout mutants ..... 65

3.4 Phenotypic characterisation of the single and double mutant strains ..... 76

3.4.1 Growth kinetics under standard culture conditions ..... 76

3.4.2 Sensitivity to oxidative stress as generated by hydrogen peroxide ..... 78

3.4.2.1 Effects of hydrogen peroxide on cell viability ..... 78

3.4.2.2 Effects of hydrogen peroxide on outgrowth ..... 81

3.4.3 Effect of loss of *fpg* and/or *nei* gene function on mutational rates and spectra in *M. smegmatis* ..... 83

3.4.3.1 Mutation frequency ..... 83

3.4.3.2 Mutation rate analyses ..... 84

3.4.3.3 Mutation spectrum in *M. smegmatis rpoB* mutants ..... 88

4. Discussion and Conclusion ..... 91

5. Appendices ..... 99

5.1 Appendix A ..... 99

5.1.1 Culture media, supplementation and antibiotic stock solutions ..... 99

5.1.2 DNA extraction solutions ..... 100

5.1.3 Agarose gel electrophoresis solutions ..... 101
5.1.4 DNA molecular weight marker ..... 101
5.1.5 Southern blot solutions ..... 102
5.1.6 Solutions for chemical transformation of *E. coli* ..... 103
5.2 Appendix B ..... 104
5.2.1 PCR and sequencing primers used in this study ..... 104
5.3 Appendix C ..... 108
5.3.1 Bioinformatic analysis of *fpg* and *nei* DNA glycosylases ..... 108
5.3.2 Vector maps and cloning strategies used to generate suicide vectors ..... 110
5.3.3 Confirmation of suicide vectors by restriction digests ..... 116
5.3.3.1 *fpgI* based suicide vectors ..... 116
5.3.3.2 *fpgII* based suicide vectors ..... 116
5.3.3.3 *neiI* based suicide vectors ..... 117
5.3.3.4 *neiII* based suicide vectors ..... 117
5.3.3.5 Mutation identified the Δ*fpgII* downstream primer ..... 118
6. References ..... 119

# List of figures

Fig. 1. The universal base excision repair pathway. ..... 6
Fig. 2. An example of the results of a Luria-Delbrück distribution ..... 46
Fig. 3. Alignments of the amino acid sequences of Fpg and Nei proteins from various micro organisms. ..... 51
Fig. 4. Genomic context comparisons of *fpg* and *nei* DNA glycosylases in *M. tuberculosis* and *M. smegmatis.* ..... 54
Fig. 5. PCR amplification strategy for the generation of the deleted *fpgI* gene fragment. ..... 56
Fig. 6. Restriction analysis of p2NIL-based marked and unmarked *fpgI* suicide vectors digested with various restriction endonucleases. ..... 58
Fig. 7. PCR amplification strategy for generating the *fpgII* deletion gene fragment.. 59
Fig. 8. Restriction digestion of the p2NIL based unmarked *fpgII* suicide vector digested with various restriction endonucleases. ..... 60
Fig. 9. PCR amplification strategy for the generation of a deleted *neiI* gene fragment. ..... 61
Fig. 10. Restriction digestion of the p2NIL based marked and unmarked suicide vectors for *neiI* digested with various restriction endonucleases. ..... 62
Fig. 11. PCR amplification strategy for the generation of a deleted *neiII* gene fragment. ..... 63
Fig. 12. Restriction analysis of the p2NIL based marked and unmarked *neiII* suicide vectors cut *neiII* with various restriction endonucleases. ..... 64
Fig. 13. Replica-plating of possible DCOs of the unmarked Δ*fpgI* knockout mutant. 65
Fig. 14. Genotypic characterisation of the Δ*fpgI* mutant of *M. smegmatis* ..... 67
Fig. 15. A schematic representation of the sequential deletion of *fpg* and *nei* in *M. smegmatis.* ..... 68
Fig. 16. Genotypic characterization of the marked *fpgI* region of *M. smegmatis* ..... 69

Fig. 17. Genotypic characterization of the *fpgII* region of *M. smegmatis* for the single Δ*fpgII* and double Δ*fpgI*Δ*fpgII* mutants ..... 70
Fig. 18. Genotypic characterization of the marked Δ*neiI* region of *M. smegmatis* for the Δ*neiI*::*hyg* mutant ..... 71
Fig. 19: Genotypic characterization of the *neiI* region of *M. smegmatis* for the single Δ*neiI* and triple Δ*fpgI*Δ*fpgII*Δ*neiI* mutants ..... 72
Fig. 20. Genotypic characterization of the marked *neiII* region of *M. smegmatis* for the Δ*neiII* mutant ..... 73
Fig. 21. Genotypic characterization of *neiII* region of *M. smegmatis* for the double Δ*neiI*Δ*neiII* mutant ..... 74
Fig. 22. Genotypic characterization of unmarked *fpgI* region of *M. smegmatis* for the double Δ*neiI*Δ*neiII*Δ*fpgI* mutant ..... 75
Fig. 23. Comparative growth kinetics ..... 77
Fig. 24. Assessment of the wild type strain with various concentrations of hydrogen peroxide ..... 80
Fig. 25. The effect of 2.5 mM hydrogen peroxide on the cell viability of log-phase wild type, single and double knockout mutant strains grown in Sauton's media ..... 81
Fig. 26. The ability of wild type and Δ*fpgI*::*hyg* to recover from hydrogen peroxide treatment ..... 82
Fig. 27. The spontaneous mutation frequencies of wild type, single *fpgI* and *neiI* knockout mutants as measured by rifampicin resistance. ..... 84
Fig. 28. Graphical illustration of the number of cultures needed to ensure a precise mutation rate when using the MSS-likelihood algorithms. ..... 85
Fig. 29. Estimation of *Nt*-values by measuring the average CFU in individual parallel cultures. ..... 86
Fig. 30. Amino acid changes in the RRDR region of the *rpoB* gene in rifampicin resistant mutants isolated from wild type, single and double knockout mutants. ..... 90
Fig. 31. The DNA molecular weight marker λIII used in this study ..... 101

Fig. 32. The plasmid maps for pcrSMART and TOPO vector used. ..... 110
Fig. 33. The plasmid maps for pGem3Zfp, p2NIL, pGOAL17 pGOAL19 and pIJ964 ..... 111
Fig. 34. The plasmid maps of vectors generated during the construction of *fpgI* knockout suicide vectors ..... 112
Fig. 35. The plasmid maps of vectors generated during the construction of the *fpgII* knockout suicide vector. ..... 113
Fig. 36. The plasmids maps of vectors generated in this study during the construction of *neiI* knockout suicide vectors ..... 114
Fig. 37. The plasmid maps of vectors generated during the construction of *neiII* knockout suicide vectors ..... 115
Fig. 38. Restriction analysis of p2NILΔ*fpgI* with various restriction endonucleases. ..... 116
Fig. 39. Restriction analysis of p2NILΔ*fpgII* with various restriction endonucleases. ..... 116
Fig. 40. Restriction analysis of p2NILΔ*neiI* with various restriction endonucleases. ..... 117
Fig. 41. Restriction analysis of p2NILΔ*neiII* with various restriction endonucleases. ..... 117
Fig. 42. An illustration of the mutation in the Δ*fpgII* downstream primer. ..... 118

# List of tables

Table 1. Bacterial strains used in this study ..... 17
Table 2. Plasmids used in this study ..... 18
Table 3. The percent similarity between *M. tuberculosis* and *M. smegmatis* Fpg and Nei glycosylases ..... 53
Table 4. Spontaneous mutation rates resulting in rifampicin resistance of the parental, single and double knockout mutant strains as determined by fluctuation analysis. ..... 87
Table 5. The different types of DNA mutations identified within the RRDR region. 89
Table 6. Antibiotic and supplement stock solution. ..... 100
Table 7. Oligonucleotides used for PCR amplification in this study. ..... 104
Table 8. Table of primers used to sequence knockout fragments ..... 106
Table 9. Primers and conditions used to confirm DCOs by PCR ..... 107
Table 10. Identification of *fpg* and *nei* DNA glycosylase homologues in other mycobacterial genomes while using GenoList and CMR genomic comparisons. ..... 108
Table 11. BLASTp searches of *M. smegmatis* Fpg and Nei proteins to other organisms. ..... 109

# Nomenclature

| | |
|---|---|
| A | Adenine |
| a.a | Amino acid |
| ACT | Artemis comparison tool |
| AIDS | Acquired immune deficiency syndrome |
| amp | Ampicillin |
| $amp^R$ | Ampicillin resistant |
| AP | Apurinic/apyrmidinic |
| *aph* | Gene encoding kanamycin resistance |
| AphC | Alkylhydroperoxide reductase |
| BLAST | Basic local alignment search tool |
| BCG | Bacillus Calmette and Guèrin |
| BER | Base excision repair |
| Bp | Base pair |
| BSA | Bovine serum albumin |
| C | Cytosine |
| CFU | Colony forming units |
| CMR | Central microbial resources |
| COG | Cluster of orthologous groups |
| DCO | Double crossover |
| DIG | Dioxygen |
| $dH_2O$ | De-ionized water |
| DNA | Deoxyribonucleic acid |

| | |
|---|---|
| dNTPs | Deoxyribonucleotide triphosphates |
| DOTS | Directly observed therapy short-course |
| DS | Downstream |
| dTTP | Deoxythymine triphosphate |
| dUTP | Deoxyuridine triphosphate |
| EDTA | Ethylenediaminetetra acetic acid |
| EMBL-EBI | European molecular biology laboratory-European bioinformatics institute |
| ExPaSy | Expert protein analysis systems |
| Fpg | Formamidopyrimidine DNA glycosylase |
| G | Guanine |
| G+C | Guanine and cytosine |
| Gh | 5-guanidinohydantoin |
| H | Hour |
| HIV | Human immuno-deficiency virus |
| hyg | Hygromycin |
| *hyg* | Gene encoding hygromycin B resistance |
| $hyg^R$ | Hygromycin B resistant |
| IPTG | Isopropyl-beta-D-thiogalactopyranoside |
| KatG | Catalase |
| kb | Kilobase |
| km | Kanamycin |
| km - | Kanamycin negative |
| km + | Kanamycin positive |
| $km^R$ | Kanamycin resistant |

| | |
|---|---|
| *lacZ* | Gene encoding β-galactosidase activity |
| LB | Luria-Bertani broth |
| *m*-value | Number of mutational events |
| MDR-TB | Multi-drug resistant tuberculosis |
| MM | Minimal media |
| MSS | Ma-Sandri-Sarkar maximum-likelihood algorithms |
| MIC | Minimal inhibitory concentration |
| min | Minute |
| MOPS | 3-(N-morpholino) propanesulfonic acid |
| $N_0$ | Initial inoculum |
| *Nt* | Final size of the population |
| Nei | Endonuclease VIII |
| NER | Nucleotide excision repair |
| N/P | Not performed |
| N/R | Not recognised |
| Nth | Endonuclease III |
| OD | Optical density |
| $OD_{600}$ | Optical density at 600 nm |
| Ori | Origin of replication |
| *ori*E | *E. coli* origin of replication |
| PCR | Polymerase chain reaction |
| *r* | Pre-existing resistant mutants |
| *rif* | Rifampicin |
| $rif^R$ | Rifampicin resistant |

| | |
|---|---|
| RNA | Ribonucleic acid |
| RNase | Ribonuclease |
| RNI | Reactive nitrogen intermediates |
| ROI | Reactive oxygen intermediates |
| rpm | Revolutions per minute |
| *rpo*B | Codes for the β subunit of RNA polymerase |
| RT-PCR | Reverse transcriptase-polymerase chain reaction |
| *sacB* | Gene encoding levansucrase |
| SCO | Single crossover |
| SDS | Sodium dodecyl sulphate |
| SOD | Superoxide dismutase |
| Sp | Spiroiminodihydrantoin |
| T | Thymine |
| TIGR | The institute for genomic research |
| TBE | Tris, boric acid and EDTA |
| Tween | Polyoxyethylene sorbitan monooleate |
| $\mu$ | Mutation rate |
| US | Upstream |
| v/v | Volume/volume |
| WHO | World health organisation |
| XDR-TB | Extensively drug resistant tuberculosis |
| X-gal | 5-bromo-4-chloro-3-indolyl- β-galactoside |

# 1. Introduction

The bacillus shaped bacterium *Mycobacterium tuberculosis*, the causative agent of tuberculosis, was first identified in the late 19th century by Robert Koch. It has since been extensively researched, but has remained one of the most successful and arguably the most destructive human bacterial pathogen (Jacobs Jr, 2000). Despite the discovery of antibiotics, the implementation of the 'directly observed therapy short-course' (DOTS) programmes and the considerable efforts made into the development of novel drugs and vaccines, little headway leading to the eradication of this disease has been made. In 1993, the World Health Organisation (WHO) declared tuberculosis an unprecedented global heath emergency (Maartens and Wilkinson, 2007; Meya and McAdam, 2007). However, 12 years later, in 2005, 8.8 million new cases and 1.6 million deaths of tuberculosis were reported (WHO, 2007). The failure to eradicate this disease is in part due to the slow, out-dated methods of diagnosis, the occurrence of multi- (MDR-TB) and extensively drug-resistant tuberculosis (XDR-TB), the increased occurrence of HIV co-infections as well as the lack of a truly effective vaccine (Corbett *et al.*, 2003; Maartens and Wilkinson, 2007; Meya and McAdam, 2007).

Both the incidence and death rates associated with this disease are astonishing when one considers that drug-susceptible tuberculosis is a fully treatable disease. Although treatment of *M. tuberculosis* infection has been complicated by its innate resistance to a range of drugs and its ability to exist in a poorly defined persistent state (Levin and Rozen, 2006; Nguyen and Thompson, 2006), an effective antibiotic regimen is available. The regimen, or the DOTS programme, comprises of a 'short course' treatment, which involves the administration of the four first-line drugs: rifampicin, isoniazid, pyrazinamide and ethambutol, for a 6-8 month period (Maartens and Wilkinson, 2007). However, due to the long duration of treatment and the potent drug

toxicity, patient non-compliance is common. This non-compliance, in combination with the lack of new drugs since the 1970s, has led to the emergence of MDR-TB and XDR-TB (Meya and McAdam, 2007).

MDR-TB is defined as tuberculosis caused by strains of *M. tuberculosis* that are resistant to rifampicin and isoniazid and is treated with an array of second line drugs over an extended period of time. These drugs, however, have diminished efficacy and increased side effects (Meya and McAdam, 2007). The existence of strains of *M. tuberculosis* that are resistant to these two first-line drugs, combined with the difficulty in treating and diagnosing MDR-TB, has led to the emergence of XDR-TB. XDR-TB is a global phenomenon associated with strains resistant to rifampicin and isoniazid, as well as a fluoroquinolone and any one of the injectable drugs, amikacin, kanamycin (km) or capreomycin. This complicates and lengthens treatment time and results in diminished cure rates (Meya and McAdam, 2007). Recent observations of a possibly highly virulent strain/strain family of drug-resistant *M. tuberculosis,* which was responsible for an outbreak of XDR-TB at Tugela Ferry in the KwaZulu Natal province in South Africa (Gandhi *et al.*, 2006), have alarmed the international community and has further underscored the need for novel drug therapies.

Currently, the only available vaccine against tuberculosis is Bacillus Calmette and Guèrin (BCG). It was derived from an avirulent strain of *Mycobacterium bovis* in the early 20$^{th}$ century and has since been globally administered to four billion individuals (Jacobs Jr, 2000; Meya and McAdam, 2007). Although BCG shows variable efficacy against adult pulmonary disease, it is still administered due to its ability to control the disease in children (Fine, 1995; Jacobs Jr, 2000; Kaufmann, 2005; Young and Dye, 2006). However, recent findings showing an the increased occurrence of disseminated BCG disease in HIV-infected infants have further highlighted the need for better vaccine alternatives (Hesseling *et al.*, 2007; Hesseling *et al.*, 2007).

Therefore, the development of novel drug therapies and new vaccines to treat tuberculosis is an important area of research. An increased understanding of specific genes, proteins and pathways associated with survival, virulence and cellular maintenance could lead to the identification of appropriate targets and ultimately lead to the design of novel drugs and vaccines.

## 1.1 *M. tuberculosis* infection

A tuberculosis infection is contracted by the inhalation of aerosolised *M. tuberculosis* bacilli. Once in the lung, these bacilli are engulfed by alveoli macrophage`s and are able to survive by preventing macrophage maturation. At this point the infection is considered a primary infection and, under normal conditions, the host's immune response localises the infection by surrounding the bacilli with necrotic tissue and macrophages to form what is referred to as a granuloma (Glickman and Jacobs, 2001; Kaufmann, 2005; Kaufmann *et al.*, 2005).

The granuloma encompasses the bacilli and creates a hostile environment by maintaining a low pH, limiting nutrients and oxygen as well as becoming the site of constant immune attack (Russell, 2007). This effectively contains the bacilli indefinitely, thus preventing systemic spread as well as any further symptoms of disease. If the host becomes immuno-compromised, however, the granuloma undergoes changes that result in resuscitation of quiescent bacteria and re-infection of new tissues by *M. tuberculosis* - this is referred to as a secondary infection. It is at this point that the usual symptoms of tuberculosis present and, if untreated, results in host fatality (Glickman and Jacobs, 2001; Kaufmann, 2005).

The granuloma is the site of consistent immune attack. This attack is waged by macrophages and damage is caused by the action of reactive oxygen intermediates (ROI) and reactive nitrogen intermediates (RNI) (Nathan and Shiloh, 2000; Zahrt and Deretic, 2002). These reactive species attack cellular components damaging lipids, proteins and nucleic acids which ultimately result in irreparable damage and cellular death (Storz and Imlay, 1999).

The bacterium can remain within the granuloma indefinitely with only 5%-10% lifetime risk of reactivation in healthy individuals (Toossi and Ellner, 1998). However, the likelihood of reactivation of latent TB infection is increased drastically by weakened immune systems, such as in the case of HIV positive individuals, where the risk approaches 10% per annum (Corbett *et al.*, 2003). In order for *M. tuberculosis* to cause secondary infection, the bacterium must overcome the hostile environment under which it is suppressed. Hence, the response to oxidative stress and nitrosative stress must be important, not only to the survival of *M. tuberculosis*, but also to its re-emergence, re-growth and spread.

## 1.2 Oxidative stress and the mycobacterial response

ROI are normal by-products of aerobic growth; however, cells have designed mechanisms to effectively cope with a degree of oxidative stress. Apart from the normal cellular activities, oxidative stress is also generated by the immune system to limit infection (Storz and Imlay, 1999; Termini, 2000), and has been shown to be a major secondary mechanism of cellular death brought about by some antibiotics (Dwyer *et al.*, 2007; Kohanski *et al.*, 2007). Infectious agents have therefore developed additional mechanisms for dealing with oxidative stress in order to ensure their survival. These common response mechanisms operate at various levels and include: interfering with the synthesis of reactive species in macrophages, breaking

down of the ROI and repair of macromolecules that are damaged by such species, particularly DNA (Storz and Imlay, 1999; Zahrt and Deretic, 2002).

The general oxidative stress response has been extensively characterised in *Escherichia coli* and the key regulators recognized as OxyR and SoxRS (Storz and Imlay, 1999). In mycobacteria, the overall regulation of this response is ambiguous, because not only are SoxRS-like proteins absent in all mycobacteria but the OxyR gene is often non-functional, as in *M. tuberculosis* and *M. smegmatis* (Zahrt and Deretic, 2002). Although the transcriptional regulation is poorly understood, enzymes involved in the degradation of ROI are present in mycobacteria and have been the subject of intense investigation. Catalase (KatG), superoxide dismutase (Sod) and alkylhydroperoxide reductase (AphC) have been generally conserved and shown to be involved in the mycobacterial oxidative stress response (Dussurget and Smith, 1998; Manca *et al.*, 1999; Edwards *et al.*, 2001; Master *et al.*, 2002).

In addition to this, the mechanisms by which DNA is repaired are also unusual in mycobacteria. Although two classical pathways for DNA repair, namely nucleotide excision repair (NER) and base excision repair (BER), are present in mycobacteria (Mizrahi *et al.*, 2000), the third universal mechanism of DNA repair, mismatch repair, is absent (Mizrahi and Andersen, 1998; Mizrahi *et al.*, 2000). In spite of this heritable DNA repair deficiency, mycobacteria, however, show no signs of hypermutability when cultured *in vitro* under standard conditions (David and Newman, 1971). This study focused on one component of the BER system in mycobacteria.

## 1.3 Base excision repair (BER)

BER is known to be the fundamental mechanism for the repair of oxidatively damaged DNA (Demple and Harrison, 1994). It is a multi-protein, multi-step repair pathway that involves the identification and excision of the damaged base by DNA glycosylases (Fig. 1a), the removal of the phosphate backbone by AP (apurinic/apyrmidinic) endonucleases (Fig. 1b) and the conversion of the 3′ termini to a hydroxyl group by a diesterase (dRpase) (Fig. 1c). These steps provide the substrate for DNA polymerases and ligases to re-insert new nucleotides (Fig. 1d), thereby repairing the damaged DNA (Demple and Harrison, 1994; Friedberg *et al.*, 1995).

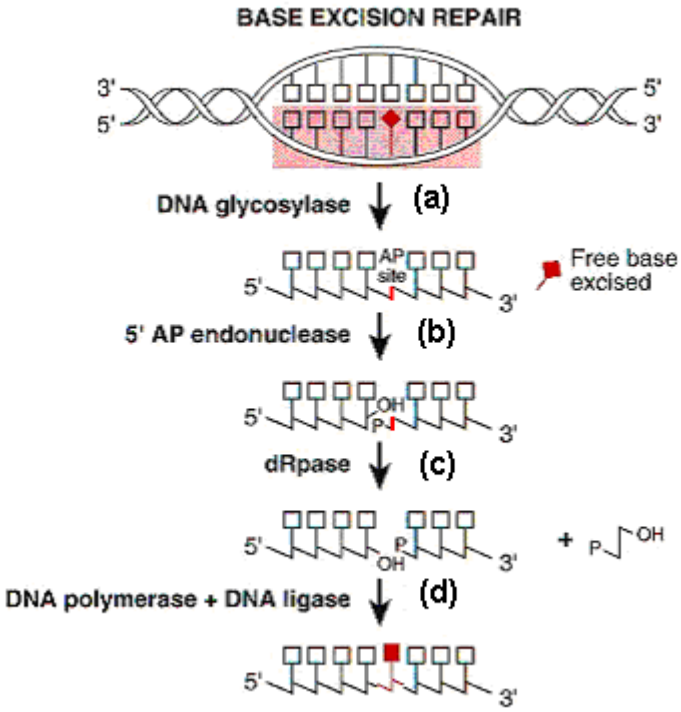



**Fig. 1. The universal base excision repair pathway.**

Figure modified from Friedberg *et al*, 1993.

The first step of BER utilizes the activity of specialised DNA glycosylases. There are three DNA glycosylases which are directly involved in the recognition and removal of oxidatively damaged DNA lesions: endonuclease III (Nth), endonuclease VIII (Nei) and formamidopyrimidine DNA glycosylase (Fpg/MutM/Fapy)(Krokan *et al.*, 1997). These three DNA glycosylases act on a plethora of oxidized lesions which, if not repaired, can be mutagenic, fatal or silent (Krokan *et al.*, 1997). Oxidized thymine and adenine lesions are seldom pro-mutagenic; however, some can block replication and hence, are lethal. Examples of such lesions include the thymine derivatives thymine glycol and 5-formyluracil (Wallace, 2002). Oxidized cytosine lesions are predominantly pro-mutagenic and result most frequently in C→T transitions. Oxidized guanine lesions are known to cause both pro-mutagenic and lethal lesions (Wallace, 2002), where 2,6-diamino-5-formamidopyrimidine (FapyG) is an example of a lethal lesion while 8-oxoguanine (8oxoG) is a pro-mutagenic lesion. 8oxoG is the most well studied lesion caused by oxidative stress, and has been shown to cause G → T transversions (Krokan *et al.*, 1997; Wallace, 2002). 8oxoG can, however, be further oxidized to form other important pro-mutagenic lesions such as spiroiminodihydrantoin (Sp) and 5-guanidinohydantoin (Gh). Sp and Gh are both known to cause G → T and G → C transversions (Hazra *et al.*, 2001; David *et al.*, 2007).

The focus of this project is on the Fpg and Nei DNA glycosylases which make up the Fpg family of DNA glycosylases (Krokan *et al.*, 1997; Zharkov *et al.*, 2003).

## 1.4 Fpg and Nei DNA glycosylases

The Fpg family of DNA glycosylases is a family of structurally related DNA glycosylases made up of formamidopyrimidine DNA glycosylase (Fpg/MutM/Fapy) and endonuclease VIII (Nei) (Krokan *et al.*, 1997; Wallace *et al.*, 2003; Zharkov *et*

*al.*, 2003; Fromme *et al.*, 2004). Both types of glycosylases identify and excise oxidatively damaged DNA lesions. Homologues of this family have been identified in plants, bacteria, viruses and vertebrates. Fpg is generally found in bacteria and plants, with the lone example in the eukaryotes being in *Candida albicans* (Wallace *et al.*, 2003; Zharkov *et al.*, 2003; Murphy and George, 2005; Bandaru *et al.*, 2007; Scortecci *et al.*, 2007). Nei homologues, though occasionally found in bacteria, have been identified in numerous vertebrates including humans (Wallace *et al.*, 2003), and homologues have also been identified in the Mimivirus (Bandaru *et al.*, 2007). Three human Nei homologous have been characterised and mutations in these genes have been implicated in numerous cancers as well as metabolic disorders (Wallace *et al.*, 2003; David *et al.*, 2007). Since these DNA glycosylases show conservation throughout the domains of life it suggests that they are important in general cellular survival.

The Fpg and Nei glycosylases are bi-functional enzymes that display both glycosylase and AP endonuclease activity (Krokan *et al.*, 1997; Fromme *et al.*, 2004) and excise the damaged base in a two-step process referred to as δ- β elimination (Wallace *et al.*, 2003; Zharkov *et al.*, 2003). Before the damaged base is excised, the DNA glycosylase binds the minor grove of the DNA and pushes the damaged base into its active site (Wallace *et al.*, 2003; Zharkov *et al.*, 2003). The damaged base is then removed by the cleavage of the N-glycosyl bond (δ elimination), after which the remaining deoxyribose group is nicked and the phosphodiester backbone removed (β elimination), resulting in both a 3′ and 5′ phosphate group. The 3′ phosphate group is replaced with a hydroxyl group by a 3′ diesterase to generate the substrate for a polymerase and ligase. The continuation of the BER process is discussed in section 1.3 (Friedberg *et al.*, 1995; Cunningham, 1997; Wilson III *et al.*, 1998).

Though structurally related, Fpg and Nei DNA glycosylases do not share the same primary substrate specificity (Krokan *et al.*, 1997; Wallace *et al.*, 2003). The main

substrates of Fpg are oxidized purines bases (guanine and adenine) while oxidized pyrimidines (thymine and cytosine) are recognised by Nei (Krokan *et al.*, 1997; Blaisdell *et al.*, 1999; Hazra *et al.*, 2001; Wallace, 2002; Wiederholt *et al.*, 2005; David *et al.*, 2007). Substantial substrate overlap does, however, exist as over and above the primary targets, Fpg has been shown to repair 5-formyluracil (Zhang *et al.*, 2000) and other thymine and cytosine lesions (D'Ham *et al.*, 1999), while Nei repairs Sp and Gh in addition to 8oxoG (Hazra *et al.*, 2000; Hazra *et al.*, 2001; Matsumoto *et al.*, 2001; Wiederholt *et al.*, 2005; David *et al.*, 2007). It is important to note that Nei shares the same substrate specificity with another DNA glycosylase present in mycobacteria, namely endonuclease III (Nth) (Krokan *et al.*, 1997; Mizrahi and Andersen, 1998; Wallace *et al.*, 2003). This substrate overlap between the various glycosylases suggests a possible adaptation which allows for an increased survival of organisms during oxidative stress conditions.

The pro-mutagenic guanine lesion 8oxoG is the target of a specifically evolved repair system, the 8oxoG or "GO" system. This system removes and repairs this pro-mutagenic lesion at three levels (Krokan *et al.*, 1997). The first level, performed by the enzyme MutT, a Nudix hydrolase, hydrolyses 8oxo-dGTP and 8oxo-GTP thereby ensuring that the dGTP in the nucleotide pool remains undamaged. The second level of protection is provided by Fpg as it repairs the damaged base once it has been incorporated into the DNA strand, and the third level is provided by MutY which removes the mis-incorporated adenine across from 8oxoG if replication has already occurred (Krokan *et al.*, 1997; Horst *et al.*, 1999; Fowler *et al.*, 2003). All the components of the GO system, originally characterised in *E. coli*, have been identified in *M. tuberculosis* (Mizrahi and Andersen, 1998). Four copies of MutT have been identified in *M. tuberculosis* and *M. smegmatis,* of which two have been implicated as anti-mutators (Dos Vultos *et al.*, 2006). The amplification of GO system genes and the substrate overlap of the glycosylases involved underscore the

potential importance of these genes in the overall maintenance of genome integrity in mycobacteria.

The objective of this study was to gain further insights into the role(s) of the Fpg and Nei DNA glycosylases in the survival of mycobacteria. In order to achieve this, the *fpg* and *nei* genes in *M. smegmatis* were sequentially knocked-out by homologous recombination and mutant strains assessed under various *in vitro* conditions.

## 1.5 Gene knockout by homologous recombination

Phenotypic characterization performed on strains defective in specific genes allows for the elucidation of the gene's importance, its function and its cellular associations. Targeted mutagenesis is achieved by allelic exchange which takes advantage of the cell's intrinsic homologous recombination machinery.

Homologous recombination is a biological process used in the repair of double stranded breaks during replication. This natural phenomenon occurs due to the ability of homologous regions of DNA to cross over one another and takes advantage of the occurrence of two copies of the genome directly after replication. If one of these genome copies is damaged, the second can act as a template for the re-synthesis of the affected strand, thereby removing the damage (Voet and Voet, 1995).

This natural repair process has been successfully manipulated in the laboratory, where the template for re-synthesis (previously the second copy of the genome) is replaced with a suicide vector carrying an inactivated copy of the targeted gene and its flanking regions. This suicide vector lacks a mycobacterial origin of replication (*ori*) and therefore, is unable to replicate episomally or express genes within the

target organism unless it incorporated into the mycobacterial chromosome. The suicide vector does, however, contain an *E. coli ori* to enable it to replicate and be manipulated in an *E. coli* host (Gordhan and Parish, 2001). Crossing over of the DNA from the vector onto the target chromosome occurs within the flanking regions of homology. A second cross-over event between the flanking sequence on the other side of the mutation and the chromosome results in expulsion of the vector sequence and replacement of the target gene on the chromosome by the inactivated copy to create a knockout mutant strain, which can be identified by a two-step selection process that relies on selectable and counter-selectable genetic markers carried on the suicide vector, as outlined below (Gordhan and Parish, 2001).

Cross-over events between the suicide vector and the chromosome could occur on either one or on both sides of the disrupted gene. A single crossover event (SCO) is what occurs when the entire suicide vector is inserted on one side, next to the targeted gene on the chromosome, while a double crossover (DCO), occurs when the targeted gene is completely replaced with the inactivated allele. In the first step of this process, antibiotic markers are used to select for SCOs. In the second step, a SCO recombinant is grown to allow for the second recombination event to occur. This is then identified using a counter-selectable marker which selects against the SCO that carries the suicide vector backbone by conferring conditional lethality, thereby allowing rare products of a second crossover event – DCOs – to be identified (Gordhan and Parish, 2001).

With the increased availability of sequenced genomes, targeted gene knockout based on homologous recombination has become a very widely applied tool for reverse genetic studies in both eukaryotes and prokaryotes (Griffiths *et al.*, 2000; Gordhan and Parish, 2001; Strachan and Read, 2003).

## 1.6 Aim

*Mycobacterium smegmatis* was used as a model organism to study the *fpg* and *nei* genes in mycobacteria as it has a relatively fast rate of replication and is non-pathogenic (Jacobs Jr, 2000). The aim of this project was to inactivate the *fpg* and *nei* genes in *M. smegmatis* by homologous recombination, individually or in selected combinations, in order to assess their individual and collective roles in the maintenance of genome integrity and survival of *M. smegmatis* under conditions of oxidative stress. The findings in *M. smegmatis* could be useful for informing studies of the *fpg* and *nei* homologues in *M. tuberculosis*, which ultimately could lead to an increased understanding of mechanisms of pathogenesis, survival and mutagenesis of this pathogen in its intracellular habitat in a human host.

# 2. Materials and methods

## 2.1 Bioinformatic tools

Various bioinformatic tools were used to identify and analyze the glycosylases of the Fpg DNA glycosylase family in *M. smegmatis* and *M. tuberculosis*. These tools are listed below.

### 2.1.1 Artemis and ACT

(http://www.sanger.ac.uk/Software/Artemis/ )

Artemis and the Artemis comparison tool (ACT) are genome viewers created by the Wellcome Trust Sanger Institute. The DNA, amino acid and G+C content of the annotated sequences can be viewed and direct genome comparisons made between two or three genomes.

### 2.1.2 BLAST searches

(http://www.ncbi.nlm.nih.gov/BLAST/)

Basic Local Alignment Search Tool (BLAST) is a program that compares DNA or protein sequences to genome sequences listed within the database, which allows for the identification of similar regions or proteins within other organisms (Altschul *et al.*, 1990). A query sequence is inserted into the algorithm of the BLAST server, which results in a list of sequences that pair-up to the query sequence and an associated bit score and E-value. The bit score indicates the strength of the alignment, while the E-value indicates the statistical chance of an accidental alignment (Madden, 2003).

There are different types of BLAST searches that allow for various combinations of DNA and Protein alignments. These include nucleotide to nucleotide (BLASTn), nucleotide to protein (BLASTx) and protein to protein BLAST searches (BLASTp).

### 2.1.3 ClustalW

(http://www.ebi.ac.uk/Tools/clustalw/)

ClustalW is a program designed by the European Bioinformatics Institute that aligns multiple DNA or protein sequences. Multiple sequences are aligned according to similarities which allow for the identification of relationships and differences both visually and by pairwise score calculations. The pairwise score is a percentage identity score that values the strength alignments.

### 2.1.4 Cluster of orthologous groups (COG)

(http://www.ncbi.nhi.gov/COG/ )

Cluster of Orthologous Groups is a database of sequences (gene or protein) which have been grouped according to their evolutionary relatedness. It allows for comparisons between sequences from different species, suggesting possible common ancestral lineages and hence insinuating similar functions for the query sequence.

### 2.1.5 GenoList

(http://GenoList.pasteur.fr/ )

GenoList is a server hosted by the Institut Pasteur that maintains various sequenced and annotated genomes, such as: *Mycobacterium tuberculosis, Mycobacterium bovis, Mycobacterium leprae, Mycobacterium ulcerans, Mycobacterium bovis BCG* and *Mycobacterium marinum* as well as other non-mycobacterial organisms such as *Escherichia coli* and *Bacillus subtilis*. It allows the user to easily identify genes, their corresponding DNA and amino acid sequences as well as their functional assignments.

### 2.1.6 InterPro

(http://www.ebi.ac.uk/interpro/ )

InterPro is a database offered by the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI). It combines the information from protein related databases; Pfam, Prints, SMART, Prosite and PRODOM. This database retrieves information on protein families, domains and sequence repeats and thus allows one to predict biological function.

### 2.1.7 Prosite

(http://www.expasy.ch/prosite/ )

Prosite is a database of protein families and domains run by the Expert Protein Analysis Systems (ExPaSy). Prosite groups conserved protein sequences and the conservation of families and domains allows for predictions of protein structure as well as biological function.

### 2.1.8 Pfam

(http://www.sanger.ac.uk/Saftware/Pfam )

Pfam is a protein family database run by the Wellcome Trust Sanger Institute. It is a large collection of protein family domains i.e. functional regions previously identified in other proteins. Pfam allows the user to identify family domains in a given protein, thereby providing possible functions of the protein. Pfam provides two levels of search: A and B. Pfam-A gives high-quality, manually-checked results. While Pfam-B, which is also referred to as PRODOM, gives automated results which are useful when no Pfam-A is identified.

### 2.1.9 TIGR-CMR

(http://cmr.TIGR.org/TIGR-scripts/CMR/CmrHomePage.cgi)

The Institute for Genomic Research (TIGR) is a non-profit, genomic-focused organization that recently merged with a number of similar organizations to form the

new J. Craig Venter Institute. This institute hosts a number of databases that allow users to look in detail at various, freely-available genomes, including the Central Microbial Resources (CMR). The CMR allows users to navigate between all the completed and available prokaryotic genomes. A number of analytical tools such as BLAST searches, protein motif searches and G+C content analyzers allow the user to identify specific genes within a genome and graphically display the genomic regions. It also allows for the comparisons of genomes based on sequence homology, gene attributions and genomic context.

## 2.2 Bacterial strains, plasmids and maintenance of strains.

For all solutions and reagents used in this section refer to Appendix A, section 5.1.1.

### 2.2.1 Bacterial strains and culturing conditions

Single *E. coli* colonies were grown in Luria-Bertani broth (LB) with shaking overnight at 37ºC (Labcon shaking incubator), while colonies transformed with large plasmids (>8kb) were grown with shaking at 30°C (New Brunswick Scientific Innova 400 incubator shaker). Where appropriate, the media was supplemented with the following antibiotics: ampicillin (amp) (100 µg/ml), kanamycin (50 µg/ml) and/or hygromycin (hyg) (100 µg/ml).

*E. coli* liquid cultures were plated onto Luria-Bertani agar plates (LA) and grown either at 37°C (Incotherm Labotec Incubator) or 30°C (Heraeus Instrument Incubator). Where appropriate, LA plates were supplemented with antibiotics (100 µg/ml amp, 50 µg/ml km, 100 µg/ml hyg), 50 µl of 5-bromo-4-chloro-3-indolyl- β-galactoside (X-gal), 50 µl of isopropyl-β-D-thiogalactopyranoside (IPTG), and/or 5% sucrose.

Single *M. smegmatis* colonies were grown shaking at 37ºC in LB, Sauton's or minimal media (MM). To prevent clumping, 0.05-0.1% Tween 80 was added to liquid cultures and where appropriate, km (25 μg/ml) and/or hyg (50 μg/ml) was included in the culture medium.

Liquid cultures were plated onto solid MM and incubated at 37ºC for 3 days. Where appropriate, agar media was supplemented with antibiotics km (25 μg/ml km), hyg (50 μg/ml), and rif (200 μg/ml), 50 μl X-gal and/or 2-5% sucrose.

All bacterial strains used in this study are listed in Table 1 and were stored at -70ºC in 30% glycerol (v/v).

**Table 1. Bacterial strains used in this study**

| Strain | Characteristics | Origin |
|---|---|---|
| *Escherichia coli* | | |
| DH5α | *sup*E44 Δ*lacU169* (Φ80 *1acZ*ΔM15 *hsd*R17 recA1 *end*A1 *gyr*A96 *thi*-1 *rel*A1 | (Hanahan, 1983) |
| *Mycobacterium smegmatis* | | |
| $mc^2155$ | *ept*-1, efficient plasmid transformation mutant of $mc^26$ | (Snapper *et al.*, 1990) |
| Δ*fpgI* | Mutant of $mc^2155$ carrying an internal 679bp deletion in MSMEG_2419 | This study |
| Δ*fpgII* | Mutant of $mc^2155$ carrying an internal 301bp deletion in MSMEG_5545 | This study |
| Δ*fpgI*Δ*fpgII* | Mutant of $mc^2155$ carrying a internal 679bp and 301bp deletions in MSMEG_2419 and MSMEG_5545 respectively | This study |
| Δ*neiI* | Mutant of $mc^2155$ carrying an internal 486bp deletion in MSMEG_1756 | This study |
| Δ*neiII* | Mutant of $mc^2155$ carrying an internal 637bp a deletion in MSMEG_4683 | This study |
| Δ*neiI*Δ*neiII* | Mutant of $mc^2155$ carrying internal 486bp and 637bp deletions in MSMEG_1756 and MSMEG_4683 respectively | This study |

| | | |
|---|---|---|
| Δ*fpgI*Δ*fpgII*Δ*neiI* | Mutant of $mc^2155$ carrying internal 679bp, 301bp and 486bp deletions in MSMEG_2419, MSMEG_5545 and MSMEG_1756 respectively | This study |
| Δ*neiI*Δ*neiII*Δ*fpgI* | Mutant of $mc^2155$ carrying internal 486bp, 637bp and 679bp deletions in MSMEG_1756, MSMEG_4683 and MSMEG_2419 respectively | This study |

### 2.2.2 Cloning vectors

All plasmids used and generated in this study are listed in Table 2. All corresponding maps are shown in Appendix C, Fig. 32-40.

**Table 2. Plasmids used in this study**

| Plasmids | Characteristics | Size | Source/reference |
|---|---|---|---|
| pGEM-3Zf(+) | *E. coli* cloning vector, $amp^R$, *lacZ*-alpha, *ori*E | 3 199bp | Promega |
| pcDNA3.1/V5-His-Topo | *E. coli* cloning vector, Topoisomerase Activity, *ori*E, $amp^R$ | 5 533bp | Invitrogen |
| P2NIL | *E. coli* cloning vector and mycobacterial suicide plasmid; $km^R$, *ori*E | 4 753bp | (Parish and Stoker, 2000) |
| pcrSMART™-HC Km | *E. coli* cloning vector, $km^R$, *ori*E | 1 788bp | Lucigen |
| pIJ963 | Plasmid carrying *hyg* as a *Bgl*II cassette; $amp^R$ $hyg^R$, *ori*E | 4 403bp | (Blondelet-Rouault *et al.*, 1997) |
| pGOAL17 | Plasmid carrying *lacZ* and *sacB* genes as a *Pac*I cassette; $amp^R$, *ori*E | 8 855bp | (Parish and Stoker, 2000) |
| pGOAL19 | Plasmid carrying *lacZ, hyg* and *sacB* genes as a *Pac*I cassette; $amp^R$, *ori*E | 10 435bp | (Parish and Stoker, 2000) |
| pOLYG | Mycobacterial and *E. coli* shuttle vector; $hyg^R$, *ori*E | 5 315bp | (Ó Gaora *et al.*, 1997) |

| | | | |
|---|---|---|---|
| pTOPO Δ*fpgI*us | TOPO replication vector carrying deleted *fpgI* upstream region; *ori*E, amp$^R$ | 6 581bp | This study |
| pTOPO Δ*fpgI*ds | TOPO replication vector carrying deleted *fpgI* downstream region; *ori*E, amp$^R$ | 6 581bp | This study |
| pcrSMART Δ*fpgI*ds | pcrSMART replicating vector carrying deleted *fpgI* downstream region; km$^R$, *ori*E | 2 836bp | This study |
| pcrSMART Δ*fpgIIu*s | pcrSMART replicating vector carrying deleted *fpgII* upstream region; km$^R$, *ori*E | 2 707bp | This study |
| pcrSMART Δ*fpgII*ds | pcrSMART replicating vector carrying deleted *fpgII* downstream region; km$^R$, *ori*E | 2 770bp | This study |
| pcrSMART Δ*neiIIu*s | pcrSMART replicating vector carrying deleted *neiII* upstream region; km$^R$, *ori*E | 2 741bp | This study |
| pcrSMART Δ*neiII*ds | pcrSMART replicating vector carrying deleted *neiII* downstream region; km$^R$, *ori*E | 2 768bp | This study |
| pGEM Δ*fpgI*us+ | pGEM cloning vector carrying deleted *fpgI* upstream region and part of TOPO vector; amp$^R$, *lacZ*-alpha, *ori*E | 5 436bp | This study |
| pGEM Δ*fpgI* | pGEM cloning vector carrying deleted *fpgI* ; amp$^R$, *lacZ*-alpha, *ori*E | 5 244bp | This study |
| pGEM Δ*neiII*ds | pGEM cloning vector carrying deleted *neiII* downstream region; amp$^R$, *lacZ*-alpha, *ori*E | 5 006bp | This study |
| pGEM Δ*neiII* | pGEM cloning vector carrying deleted *neiII*; amp$^R$, *lacZ*-alpha, *ori*E | 4 104bp | This study |
| p2NIL Δ*fpgI* | p2NIL knockout vector carrying an unmarked deletion mutation in *fpgI;* km$^R$, *ori*E | 6 520bp | This study |
| p2NIL Δ*fpgII* | p2NIL knockout vector carrying an unmarked deletion mutation in *fpgII;* km$^R$, *ori*E | 6 309bp | This study |
| p2NIL Δ*neiI* | p2NIL knockout vector carrying an unmarked deletion mutation in *neiI;* km$^R$, *ori*E | 6 644bp | This study |
| p2NIL Δ*neiII* | p2NIL knockout vector carrying an unmarked deletion mutation in *neiII*; km$^R$, *ori*E | 6 282bp | This study |

| | | | |
|---|---|---|---|
| p2NIL Δ*fpgI*::*hyg* | p2NIL knockout vector carrying a *hyg* marked deletion mutation in *fpgI;* hyg$^R$; km$^R$, *ori*E | 8 278bp | This study |
| p2NIL Δ*neiI*::*hyg* | p2NIL knockout vector carrying a *hyg* marked deletion mutation in *neiI*; hyg$^R$; km$^R$, *ori*E | 8 402bp | This study |
| p2NIL Δ*neiII*::*hyg* | p2NIL knockout vector carrying a *hyg* marked deletion mutation in *neiII;* hyg$^R$; km$^R$, *ori*E | 8 034bp | This study |
| p2NIL Δ*fpgI*::*hyg*::*p17* | p2NIL Δ*fpgI*::*hyg* knockout vector carrying selectable and counter selectable markers; hyg$^R$ km$^R$, *lac*Z, *sacB*, *ori*E | 14 637bp | This study |
| p2NIL Δ*fpgI*::*p19* | p2NIL Δ*fpgI* knockout vector carrying selectable and counter-selectable markers; km$^R$, hyg$^R$, *lac*Z, *sacB*, *ori*E | 14 459bp | This study |
| p2NIL Δ*neiI*::*hyg*::*p17* | p2NIL Δ*neiI*::*hyg* knockout vector carrying selectable and counter-selectable markers; hyg$^R$, km$^R$, *lac*Z, *sacB*, *ori*E | 14 761bp | This study |
| p2NIL Δ*neiI*::*p19* | p2NIL Δ*neiI* knockout vector carrying selectable and counter-selectable markers; hyg$^R$, km$^R$, *lac*Z, *sacB*, *ori*E | 14 583bp | This study |
| p2NIL Δ*neiII*::*hyg*::*p17* | p2NIL Δ*neiII*::*hyg* knockout vector carrying selectable and counter-selectable markers; hyg$^R$, km$^R$, *lac*Z, *sacB*, *ori*E | 14 393bp | This study |
| p2NIL Δ*neiII*::*p19* | p2NIL Δ*neiII* knockout vector carrying selectable and counter-selectable markers; hyg$^R$, km$^R$, *lac*Z, *sacB*, *ori*E | 14 221bp | This study |
| p2NIL Δ*fpgII*::*p19* | p2NIL Δ*fpgII* knockout vector carrying selectable and counter-selectable markers; km$^R$, hyg$^R$, *lac*Z, *sacB*, *ori*E | 16 744bp | This study |

### 2.2.3 Assessment of cell viability

The growth of cultures in liquid media was assessed by measuring the optical density (OD) at 600 nm ($OD_{600}$) using either a Shimadzu UV-1601 UV-visible spectrophotometer or a WPA Biowave C0800 cell density meter. Cultures with $OD_{600}$ above 0.8 were diluted 1:10 and the $OD_{600}$ re-measured for accuracy.

Viable cell counts were assessed by enumerating colony forming units (CFUs). The cultures were serially diluted ten-fold ($10^{-1}$ - $10^{-8}$), vortexed vigorously and 100 µl of the appropriate dilution was spread in duplicate onto solid media. The colonies on the plates were enumerated and viable cells per ml of culture were calculated as follows:

$$\text{CFU/ml} = \frac{\text{number of colonies on plate} \times \text{dilution factor}}{\text{volume plated}}$$

### 2.2.4 Replica plating

A single colony was resuspended in 30 µl sterile $dH_2O$ containing 0.05% Tween 80 and 5 µl-10 µl of the suspension was spotted onto parallel plates with and without the appropriate supplementation. Once the liquid was absorbed into the media, the plates were incubated for 1-3 days at 37°C before scoring for growth or no growth.

### 2.2.5 Spotting assays

Spotting assays were performed to compare growth between differentially treated cultures. Differentially treated parallel cultures were serially diluted ($10^{-1}$-$10^{-8}$) and 5 µl -10 µl of each dilution was spotted consecutively onto plates. Once the liquid was absorbed into the media, the plates were incubated for 1-3 days at 37°C and then assessed for growth.

## 2.3 DNA manipulation

Various enzymatic reactions were used to amplify, clone and quantify the DNA. For all solutions and reagents used in this section refer to Appendix A, sections 5.1.2 - 5.1.5.

### 2.3.1 DNA extraction

#### 2.3.1.1 Chromosomal DNA extraction from *M. smegmatis*

##### 2.3.1.1.1 Bulk chromosomal DNA extraction

Chromosomal DNA was isolated using a modified cetyltrimethylammonium bromide (CTAB) extraction method (Larsen, 2000). Cells were collected, either by scraping a loop-full of cells from a plate or by harvesting 1ml of a log-phase culture on a bench top centrifuge (Eppendorf Centrifuge 5415D) and resuspended in 500 µl TE. The cell

suspension was heated at 95ºC for 5 min or 65ºC for 20 min before adding 50 µl of lysozyme (10 mg/ml) and incubating overnight at 37ºC. Seventy µl of 10% SDS and 6 µl proteinase K (10 mg/ml) were added to the suspension, which was incubated at 65ºC for 30 min to 2 h. The solution was then treated with 100 µl of 5 mM NaCl and mixed. Eighty µl of CTAB solution was then added and the mixture incubated at 65ºC for 10 min. The DNA was purified by adding 400 µl chloroform: isoamyl alcohol (24:1 v/v) and centrifuging the solution for 10 min at 13 000 rpm. The aqueous phase containing the DNA was removed and precipitated with 2.5 volumes of 100% ethanol at 13 000 rpm for 20 min. The DNA pellet was washed with 70% ethanol, dried in a vacuum centrifuge (SpeedVac, Savant, Farmingdale NY, USA) and resuspended in 100-200 µl of sterile $dH_2O$.

#### 2.3.1.2 Small-scale chromosomal DNA extraction

Single *M. smegmatis* colonies were resuspended in 50 µl of water and boiled for 20 min. Forty-five µl of chloroform: isoamyl alcohol (24:1 v/v) was added and the solution left to stand at room temperature for 20 min before centrifugation at 13 000 rpm for 5 min. The aqueous phase containing the DNA was removed and used for further DNA manipulations such as PCR amplification.

#### 2.3.1.3 Plasmid extraction from *E. coli*

##### 2.3.1.3.1 Small-scale preparation of plasmids

Two methods were used for small-scale isolation of plasmids, namely the TENS and the Mini-prep methods, respectively (Sambrook *et al.*, 1989; Zhou *et al.*, 1990).

TENS method

One ml of a log phase culture was harvested at 13 000 rpm for 5 min. The pellet was resuspended in 50-100 µl of the remaining supernatant and 300 µl of fresh TENS buffer was added. The solution was mixed by inversion before adding 1.5 µl RNase A (10 mg/ml) and incubated at 42° for 10 min. 150 µl 3 M sodium acetate (pH 5.5) was then added and the solution kept on ice for 5-30 min before centrifugation for 5 min at 13 000 rpm. The DNA was ethanol precipitated, dried in a vacuum centrifuge and resuspended in 50 µl sterile $dH_2O$.

Mini-prep method

One ml of a log-phase culture was centrifuged at 13 000 rpm for 5 min and the cells resuspended in 100 µl solution I (solutions are detailed in Appendix A). Two hundred µl of solution II was added and the suspension mixed by inversion before adding 150 µl of solution III. The mixture was incubated on ice for 5 min and centrifuged at 13 000 rpm for 5 min. The supernatant was collected and treated with 1.5 µl RNase A (10 mg/ml) for 15 min at 42ºC. The DNA was precipitated with 350 ml isopropanol, the pellet rinsed with 70% ethanol and vacuum dried before resuspension in 20 µl sterile $dH_2O$.

#### 2.3.1.3.2 Maxi-prep or large-scale preparation

One hundred ml of a log-phase culture was harvested by centrifugation at 5000 rpm for 5 min (Beckman Coulter Allegra X-22R Centrifuge). The DNA extraction procedure was similar to the mini-prep method, except 10 times as much solution I, II and III was used. After addition of solution III, the suspension was centrifuged for 10 min at 8 000 rpm. Seven hundred and twenty µl aliquots of supernatant were placed

in 1.5 µl Eppendorf tubes and incubated with 2 µl RNase A (10 mg/ml) for 15 min at 42ºC. The DNA was precipitated with 800 µl isopropanol and vacuum dried before resuspension in 100 µl sterile $dH_2O$. For certain procedures that required ultra pure DNA, the sample underwent an additional purification step. One tenth of the volume of 3 M sodium acetate (pH 5.5) and 350 µl phenol: chloroform (1:1 v/v) was added to the resuspended DNA and the solution centrifuged at 13 000 rpm for 10 min. The aqueous phase was removed, treated with 350 µl of chloroform: isoamyl alcohol (24:1 v/v) and centrifuged at 13 000 rpm for 10 min. The aqueous phase was once again removed and the DNA precipitated with 2.5 volumes of 100% ethanol. The DNA pellet was rinsed with 70% ethanol, vacuum dried and resuspended in 100-200 µl of sterile $dH_2O$.

## 2.3.2 Enzymatic modification of DNA

### 2.3.2.1 Polymerase Chain Reaction (PCR) amplification

Oligonucleotides used in this study were designed using two bioinformatic programs. Primer3 (http://frodo.wi.mit.edu/) identified possible primers within the specified region, whilst Generunner 3.05 (http://www.genelink.com/tools/gl-downloads.asp) was used to confirm the stability of the primers and the absence of secondary structures. Primers were obtained from Inqaba Biotech Ltd. All primer sequences used in this study are detailed in Appendix B, Table 7 - Table 9. Taq DNA polymerase (Roche Biochemicals) was used to optimize PCR amplification before high fidelity polymerases, such as Expand High Fidelity Polymerase (Roche Biochemicals) or Phusion Polymerase (Finnzymes) were used to amplify genomic fragments.

The PCR reactions were set up in 20 µl final volume as follows: 10% - 20% PCR buffer as per the manufacturer recommendations, 10 mM dNTP (1.25 mM stock solutions), 10% G+C rich solution (Roche Biochemicals), 1 µM forward and reverse primers, 5-100 µg template DNA and 0.2 µl of the appropriate DNA polymerase. Three control reactions were included in all PCR experiments to rule out non-specific amplification, namely, a reaction lacking DNA, a reaction lacking the forward primer and a reaction lacking the reverse primer.

PCR reactions with Taq and Expand DNA polymerases were carried out using the following conditions; 5 min denaturation (94°C), 25-30 cycles of 15 s of denaturation (94°C), 30 s of annealing (56-65°C) and 45 – 90 s elongation (72°C), with a final elongation step for 5-7 min. Standard conditions for the Phusion polymerase were as follows; 1 min denaturation (98°C), 25-30 cycles of denaturation (98°C) for 10 s , 15 s of annealing (56-65°C) and 15-30 s elongation (72°C), with a final elongation cycle of 5-7 min.

#### 2.3.2.2 Restriction digestion

Restriction enzymes were purchased from Roche Biochemicals, AEC Amersham Bioscience, and New England Biolabs. Restriction enzyme digestions were performed as per the manufacturer's instructions using the specified buffers and when necessary bovine serum albumin (BSA) and/or triton X-100 was added. Approximately 1 µg plasmid DNA was digested in a total volume of 10-20 µl for 1 hour at 37°C (unless otherwise instructed), while 5-10 µg of chromosomal DNA was digested overnight at 37°C in a total volumes of 30-100 µl. Double digests were performed in a single step if both the enzymes had identical buffer requirements. If the two enzymes had different buffer requirements, the enzyme with the lower salt buffer was used first to digest the DNA. After digestion the enzyme was heat

denatured at 65°C for 20 min, before addition of the second enzyme with the appropriate buffer.

#### 2.3.2.3 Ligation

Fast-link™ ligation kit (Epicentre Technologies) was used as per the manufacturer's instructions. Briefly, 1 µl ligase, 1.5 µl 10 mM ATP and 1.5 µl of the supplied ligation buffer was added to the vector and insert reaction in a total volume of 15 µl. The reaction was incubated at room temperature for 10-15 min and then stopped by heat denaturation at 65°C for 20 min. In order to optimize cloning events the vector to insert ratios of 1:1, 1:2 or 1:3 were calculated as follows.

$$\text{Concentration of insert DNA (ng)} = \frac{\text{the size of the insert (bp)} \times \text{concentration of vector (ng)}}{\text{total size of the vector (bp)}}$$

#### 2.3.2.4 Phosphorylation

A phosphate group at the 5′ end of the insert DNA is needed for ligation of the DNA into the vector. PCR products do not contain 5′ phosphate groups hence primers were phosphorylated prior to amplification. T4 polynucleotide kinase supplied in the pcrSMART kit (Lucigen) was used as per the manufacturer's instructions. One µl of T4 polynucleotide kinase was added to the PCR primers, incubated at 37°C for 15 min and then used directly in the PCR amplification reactions. The PCR cycling conditions do not completely denature the T4 polynucleotide kinase and therefore the PCR products were purified by gel electrophoresis before further manipulations.

#### 2.3.2.5 De-phosphorylation

To prevent self re-ligation of digested vectors, the 5′ phosphate group was removed by de-phosphorylation using Calf Alkaline phosphatase (Roche Biochemicals), Shrimp Alkaline phosphatase (Roche Biochemicals) or Antarctic Alkaline phosphatase (New England Biolabs) as per the manufacturer's instructions. One µl phosphatase and its appropriate buffer was added to linearised vector DNA and incubated for 1 hr at 37°C. The phosphatase was inactivated by either heat inactivation at 65°C for 20 min (Shrimp Alkaline phosphatase and Antarctic Alkaline phosphatase) or the vector DNA was purified by gel electrophoresis (Calf Alkaline phosphatase) before further cloning manipulations were carried out.

### 2.3.3 DNA precipitation

#### 2.3.3.1 Isopropanol precipitation

0.8 times volume of isopropanol was added to the DNA solution and incubated for 10 min at room temperature. The DNA was pelleted at 13 000 rpm for 10-20 min, rinsed with 70% ethanol, vacuum dried and resuspended in an appropriate volume of sterile water.

#### 2.3.3.2 Ethanol precipitation

0.1 times volume of 3 M sodium acetate (pH 5.5) and 2.5 volumes of chilled 100% ethanol were added to the DNA solution. The DNA was allowed to precipitate at -

20°C for 5-30 min before pelleting at 13 000 rpm for 15-20 min. The pellet was washed with 70% ethanol, vacuum-dried and resuspended in sterile water.

### 2.3.4 The separation of DNA fragments by agarose gel electrophoresis

DNA was separated and analysed using general electrophoretic techniques (Sambrook *et al.*, 1989). Agarose gels (0.8-1%) were prepared in TAE buffer to which 0.5 µg/ml of ethidium bromide was added. The DNA fragments were separated in TAE buffer at 80-100 volts and lambda DNA molecular weight markers (Roche Biochemicals) were used to assess fragment sizes of samples (Appendix A, Fig. 31). Gels were visualised under UV-light using either the GelDoc system (Bio-Rad) or G Box (Synergene).

### 2.3.5 Recovery of DNA from agarose gel

DNA was excised from agarose gels and purified using the Nucleospin kit (Macherey-Nagel) as per the manufacturer's instructions. Briefly, gel fragments containing DNA were melted at 40°C and the suspension loaded onto a column which bound the DNA. The column was then washed before the DNA was eluted with 30-40 µl of sterile $dH_2O$.

## 2.3.6 Quantification of DNA

DNA was quantified on the Nanodrop ND-1000 Spectrophotometer and on agarose gels using lambda DNA molecular weight markers (Roche Biochemicals). The DNA bands of the lambda molecular weight markers have known concentrations (Appendix A, Fig. 31) therefore the band intensities of the DNA samples can be compared to the molecular marker concentration and approximate concentrations of the sample DNA can be estimated.

## 2.3.7 Southern blot techniques

Southern blotting allows for the identification of a DNA sequence of interest within a complex mixture of fragments. It involves the separation of digested DNA by agarose gel electrophoresis, the transfer of DNA onto a nitrocellulose membrane followed by hybridization of specifically-labelled probes to the DNA on the membrane. The probes are labelled using either radioactive isotopes or by the chemiluminescence techniques. Both techniques allow for the hybridized fragments to be visualized on X-ray film.

### 2.3.7.1 Electro-blotting

Between 4 and 8 μg of chromosomal DNA was digested overnight and separated by electrophoresis (section 2.3.4). The agarose gel was treated with depurination solution for 10-15 min, rinsed with $dH_2O$, and then incubated in denaturation solution for 10-15 min. The agarose was sandwiched together with nitrocellulose paper (Hybond$^{TM}$-N nylon membrane, Amersham Biosciences) between two layers of 3 mm Whatman

filter paper and sponges in a TE 22 Transfer cassette. The cassette was placed in a TE 22 Mini Transphor unit (Hoefer) and the DNA transferred onto the nitrocellulose paper in TBE buffer at 0.5A for 2hr at 4°C. After the transfer the DNA was cross-linked onto the nylon membrane at 1200 mJ/cm$^2$ (UV Stratalinker 1800, Stratagene).

### 2.3.7.2 Radioactive labelling and hybridization

#### 2.3.7.2.1 $^{32}$P labelling of probes

Specific DNA fragments were labelled with the radioactive isotope $^{32}$P as per the manufacturer's instructions using the Random Primer Labelling Kit (Roche Biochemicals Bioscience). The reaction consisted of 100 µg template DNA, 2 µl random primer mix, 1 µl Klenow enzyme and 2 mM of each nucleotide, including radioactively labelled dCTP ([$\alpha^{32}$P]-CTP). The reaction was carried out for 30 min at 37°C and stopped with 70 µl TE buffer (pH 8). Non-incorporated radioactive nucleotides were removed by eluting the labelled probe twice through G-25 Sephadex columns at 4400g for 3 min. The probe was used immediately or stored at -20°C for a short period of time till required.

#### 2.3.7.2.2 Hybridization

The hybridization procedure with radioactive probes was followed as described by (Sambrook *et al.*, 1989)). The nitrocellulose membrane with bound DNA was pre-treated with hybridization solution containing 30 µl of heat-denatured fish sperm DNA (10mg/ml, Roche Biochemicals) in roller bottles (Techne Part No FHB II) at 42°C for 1.5-2 h in a hybridization oven (Techne Hybridiser HB-1). The labelled

probe was heat-denatured at 95°C for 5 min, rapidly cooled on ice and added to the pre-hybridized membrane and hybridization carried out at 42°C overnight. Following hybridization, the nitrocellulose membrane was washed twice at 42°C for 15 min with solution I followed by washes with solution II and solution III at 42°C for 15 min each. A final stringent wash was performed with solution IV at 65°C for 30 min. The blot was then placed in a cassette (3M Trimax T16 or Okamoto) and exposed to X-ray film (AGFA CP-G Plus Medical X-ray film) for 1 – 7 days at -70°C.

#### 2.3.7.3 Non-radioactive labelling and hybridization

##### 2.3.7.3.1 Chemiluminescence labelling of probes

Chemiluminescence probes were generated using the PCR DIG Probe Synthesis kit (Roche Biochemicals) as per manufacturer's instructions. The labelled probe was generated by replacing dTTP with digoxigenin-labelled dUTP (DIG- dUTP) in a PCR amplification reaction. To confirm DIG-dUTP incorporation both the DIG labelled PCR product (probe) and a non-labelled PCR product were analyzed on a 0.8% agarose gel. The DIG-dUTP labelled product (probe) migrates at a slower rate as it had a higher molecular weight than its non-labelled counterpart.

##### 2.3.7.3.2 Hybridization

Membrane-probe hybridization was performed using the DNA High Prime DNA labelling and Detection Starter Kit II (Roche Biochemicals) as per manufacturer's instructions. Pre-hybridization was performed in roller bottles (Hybaid HB-OV-BM) at 42°C for 30 min in a hybridization oven (Hybaid Micro-4) by incubating the membrane in 10 ml of DIG Easy Hyb solution. The probe was heat-denatured at 95°C

for 10 min, rapidly cooled on ice, added to the pre-hybridized membrane and incubated overnight at 42°C. After hybridization the membrane was washed twice at room temperature for 5 min with solution I and once at 65°C for 30 min in solution II.

#### 2.3.7.3.3 Chemiluminescence detection

DIG-labelled DNA probes are immuno-detected by antibodies attached to an alkaline phosphatase. The substrate CSPD is then added and de-phosphorylated by the alkaline phosphatase. The de-phosphorylation of CSPD causes chemiluminescence at maximum wavelength of 477 nm which can be detected on X-ray film.

Immunological detection of hybridized membranes was carried out by immersing the blot in 20 ml washing buffer for 5 min, followed by incubation for 30 min in 120 ml blocking solution at room temperature. The membrane was then incubated in 20 ml antibody solution for 30 min and excess antibody was removed with washing buffer for 30 min. The blot was then equilibrated in 20 ml detection buffer and immediately placed in a hybridization bag (Roche Biochemicals) with 1ml CSPD before incubation at 37°C for 15 min. Positively hybridized bands were detected by exposure of the blot to X-ray film (AGFA CP-G Plus Medical X-ray film) at room temperature for 30 min - 2 h.

## 2.4 Transformation of bacteria

For all reagents and solutions used in this section refer to Appendix A, section 5.1.6.

## 2.4.1 Chemical transformation in *E. coli*

### 2.4.1.1 Preparation of chemically competent *E. coli*

This procedure was obtained from Dr. P. Stolt (personal communication). One hundred ml of *E. coli* culture was grown in LB to mid-log phase and harvested at 4°C at 4000 rpm (Beckman Coulter Allegra™ X-22R centrifuge). The pellet was resuspended at 4°C in 33 ml ice-cold RF1 solution before centrifugation at 3000 rpm for 15 min at 4°C. The cells were resuspended in 8 ml ice-cold RF2 solution, incubated at 4°C for 30 min - 2h and snap-frozen in 500 µl aliquots at -70°C until required.

### 2.4.1.2 Transformation of *E. coli*

The transformation procedure was performed as previously described (Sambrook *et al.*, 1989). Plasmid DNA was added to 200 µl chemically competent *E. coli* cells and incubated on ice for 15 min. The cells were heat shocked for 90 s at 42°C and allowed to recover in 1ml LB for 1 h at 37°C. Transformants were selected on media containing appropriate antibiotics.

## 2.4.2 Electroporation into *M. smegmatis*

### 2.4.2.1 Preparation of electro-competent *M. smegmatis*

Electroporations were performed as described previously (Larsen, 2000; Gordhan and Parish, 2001).

One hundred ml of culture was grown to mid-log phase and cooled on ice. All further manipulations were performed at 4°C. The culture was pelleted twice at 4000 rpm for 10 min (Beckman Coulter Allegra™ X-22R centrifuge) and resuspended in 10ml chilled 10% glycerol. The washed cells were re-pelleted at 4000 rpm for 10 min and resuspended in 2ml chilled 10% glycerol. Five hundred µl aliquots of cells were centrifuged for 30 s at 13 000 rpm, resuspended in 400 µl glycerol and stored on ice until needed.

### 2.4.2.2 Electroporation

Plasmid DNA (2-4 µg) was added to electro-competent cells and transferred to pre-cooled 0.2cm electroporation cuvettes (Bio-Rad). Cells were pulsed in a GenePulser™ (Bio-Rad) with the following specifications; 2.5kV, 1000Ω and a capacitance of 25µF. After pulsing the cells were rescued in 1ml 2×TY media and incubated at 37°C for 2-3 h. The cells were then plated onto appropriately supplemented media and incubated at 37°C for 5 days.

Electroporation efficiency was assessed by electroporating 1ng of a replicating plasmid into electrocompetent cells. The efficiency was calculated as shown below

and cells were considered competent at a frequency of $10^4$ transformants per nanogram of DNA.

$$\text{transformants/ng} = \frac{\text{number of transformed cells}}{\text{concentration of DNA inserted}}$$

## 2.5 DNA sequencing

Plasmid DNA was isolated by the maxi-prep method (section 2.3.1.3.2), whilst chromosomal DNA from *M. smegmatis* was isolated by the small scale extraction procedure (section 2.3.1.2).

All sequencing was outsourced to either Inqaba Biotech or to the Molecular Biology Department of the University of Cape Town. Sequencing data was analysed using the SeqManager module of the Lasergene suite of programs (DNAstar). Sequencing primers were designed using the bioinformatic programs Primer3 and Generunner. All sequencing primers used in this study are listed in Appendix B, Table 8.

## 2.6 The construction and identification of site specific deletion mutants in *M. smegmatis*

The knockout vectors and mutant strains of *M. smegmatis* were constructed as previously described (Gordhan and Parish, 2001). Maps of vectors created in this study are illustrated in Appendix C, Fig. 32 and Fig. 33.

#### 2.6.1.1 Construction of knockout vectors

The construction of deletion mutant alleles was initiated by the PCR amplification of the upstream (5′) and downstream (3′) regions flanking the targeted gene, as described in section 2.3.2.1. These fragments are approximately 1kb in size as this length of DNA flanking the deleted gene as sufficient for homologous recombination (Springer *et al.*, 2004). The upstream and downstream homologous sequences were included to enable site-specific recombination between the suicide plasmid and the corresponding sequence on the chromosome. The PCR amplified products were then cloned separately into *E. coli* replicating vectors (TOPO or pcrSMART vectors). The PCR primers used to amplify the upstream and downstream fragments contained unique restriction enzyme sites engineered at the 3′ end to facilitate cloning of these fragments into the various vectors used to construct the knockout vectors. During PCR amplification there is a possibility that the DNA polymerase may introduce mutations in the amplicons. Mutations in the flanking DNA regions could have polar effects on genes downstream of the target with potential phenotypic consequences. To rule out this possibility, all amplicons were sequenced prior to use in allelic mutagenesis.

Once the accuracy of the sequence was confirmed, the upstream and the downstream fragments were excised from their respective replicating vectors and ligated into the suicide vector p2NIL either as a single-step, three-way cloning or by consecutive ligation reactions to create the inactivated allele of the target gene. When consecutive ligations were carried out, the upstream and downstream fragments were first ligated into the pGEM vector before cloning the deleted allele into p2NIL.

The gene knockout methodology employs a two-step strategy to select mutants using *hyg*, *aph* and *lacZ* selectable markers and the *sacB* counter-selectable marker. The

*lac*Z gene expresses ß-galactosidase, which hydrolyses the histological marker, X-gal, to form a blue colour, thus facilitating the identification of SCO recombinants. The counter-selectable *sacB* gene expresses the enzyme levansucrase which converts sucrose to the toxin, levan, which causes cell death, and thus provides a means of counter-selecting against SCOs.

These markers are carried as gene cassettes on plasmids pGOAL17, pGOAL19 and pIJ963 (Blondelet-Rouault *et al.*, 1997; Parish and Stoker, 2000). The plasmid pGOAL17 contains the *sacB* and *lac*Z cassettes, whilst pGOAL19 contains *sacB*, *hyg* and *lac*Z. These pGOAL cassettes were excised with the *Pac*I restriction enzyme and inserted into the unique *Pac*I site in p2NIL in a one step cloning process (Parish and Stoker, 2000). The *hyg* cassette was excised as a *Bgl*II fragment from pIJ963 which was used to mark the knockout construct by cloning it at the junction of the ligated upstream and the downstream fragments of the disrupted gene.

The marked and unmarked suicide vectors were verified by restriction analysis and tested in *E. coli* for levansucrase activity (sucrose sensitivity), ß-galactosidase activity (blue/white colour selection) as well as km and hyg resistance prior to use for targeted gene knockout in *M. smegmatis.*

#### 2.6.1.2 Identification of Single Cross Overs (SCOs)

The suicide vector (2-4 µg) was electroporated into electro-competent *M. smegmatis* cells (section 2.4.2). A negative control containing no plasmid DNA and a positive control containing 1ng of a replicating vector (pOLYG) were used to assess electroporation efficiency (section 2.4.2.2). Electroporated cells were spread on plates containing km, hyg and X-gal and SCOs were identified as blue colonies after 5-7

days. Chromosomal DNA from several SCOs was extracted (section 2.3.1.1.1) and genotypically analysed by Southern blot analysis (section 2.3.7) to confirm the site-specificity of recombination and to distinguish between products of upstream *vs*. downstream crossover events.

#### 2.6.1.3 Identification of Double Crossovers (DCOs)

Once SCOs were genotypically confirmed they were passaged without selection (*i.e.* in the absence of km) to allow second crossover events to occur, which would result in vector loss from the chromosome. For the isolation of *hyg*-marked DCOs, hyg was maintained during all subsequent selection procedures. The log-phase culture was serially diluted ($10^0$-$10^{-6}$) and dilutions $10^0$ – $10^{-4}$ were plated onto LA plates containing X-gal and 5% sucrose whilst dilutions $10^{-4}$– $10^{-6}$ were plated on LA plates containing X-gal alone. In the presence of sucrose, *sacB* expression generates levansucrase resulting in *ca*. $10^4$-fold killing. Hence, white colonies in the presence of sucrose indicated that these colonies had lost the vector backbone containing the *sacB* gene since they were unable to produce the toxin in the presence of sucrose. These white, sucrose-resistant colonies were replica plated (section 2.2.4) onto LA plates with and without km to distinguish between DCO events (i.e. loss of the *aph*-containing vector backbone, km-sensitive) and sucrose-resistance through the generation of spontaneous *sacB* mutants (km-resistant). km-sensitive clones were confirmed as DCOs by PCR (section 2.3.2.1) and Southern blot (section 2.3.7) analysis. In the case of unmarked DCOs, the second crossover event can either restore the wild type allele when it occurs on the same side of the mutation as the first, or generate the mutant allele when it occurs on the opposite side of the mutation. Therefore, an initial PCR screen to distinguish between these possibilities was included to reduce the number of clones to be confirmed by Southern blot analysis. PCR primers flanking the deletion were used to identify smaller amplicons for the

deletion mutation allele compared to the wild type allele. For *hyg*-marked mutations, all white, hygromycin-resistant, km-sensitive colonies were expected to carry the mutated allele. Once possible DCOs were identified by PCR, Southern blot analysis was carried out to confirm the mutations and to ensure their site-specificity. Details of primers and amplification conditions used for PCR confirmation are shown in Appendix B, Table 9.

## 2.7 Phenotypic characterization of mutant strains

The parental strain together with the mutant strains generated in this study were phenotypically characterised by comparing their growth kinetics under normal culture conditions, sensitivity to oxidative stress (induced by hydrogen peroxide, $H_2O_2$) and their ability to repair DNA by measuring the frequency and spectrum of mutations leading to rifampicin resistance.

### 2.7.1 Growth kinetics

Single colonies of the mutant and wild type strains were inoculated in 10 ml of MM media and the cultures grown to log phase. One ml of this pre-culture was used to inoculate 100ml of MM to give a starting inoculum of *ca*. 2 x $10^6$ CFU/ml ($OD_{600}$ = 0.01). The growth of each strain was monitored at regular intervals over a 36 h period both by optical density measurements at $OD_{600}$ and by viable cell count assessments (section 2.2.3). Bacterial growth was represented graphically either as CFU/ml or $OD_{600}$ *vs.* time.

## 2.7.2 Sensitivity to oxidative stress

The sensitivity of the mutant strains to oxidative stress generated by $H_2O_2$ was assessed. $H_2O_2$ was stored in the dark at 4°C.

### 2.7.2.1 Hydrogen peroxide susceptibility assays

#### 2.7.2.1.1 Assessment of cell viability post hydrogen peroxide treatment

The growth phase of the cultures and the optimal $H_2O_2$ concentration at which this agent conferred toxicity on *M. smegmatis* was ascertained in the parental strain. Different concentrations of $H_2O_2$ (2.5 mM, 5 mM, 10 mM, 20 mM) was added to cultures in lag, log and stationary phase and growth monitored over a period of 6 h by spotting 5 µl-10 µl of serially diluted cultures ($10^0$ -$10^{-8}$) onto plates. The plates were incubated at 37°C and growth measured after 3 days.

Once conditions were optimized, the parental and mutant strains were grown in approximately 50 ml of LB, Sauton's or MM medium to mid-log phase. Forty ml of this culture was removed and placed into a new flask to which 2.5 mM $H_2O_2$ added. The flasks were maintained shaking at 37°C for the remainder of the experiment and growth was recorded at regular intervals by plating 1ml and $10^0$-$10^{-6}$ dilutions onto MM plates. Cell viability post $H_2O_2$ treatment was represented graphically as CFU/ml v*s.* time. Paired Student's *t*-tests were performed to assess the statistical significance between the parental and knockout mutant strains using GraphPad Prism Software (http://www.graphpad.com/quickcalcs/ttest2.cfm).

#### 2.7.2.1.2 Assessment of cell outgrowth post hydrogen peroxide treatment

Cell outgrowth was assessed in the manner described by Jain *et al*. (2007). Single colonies of the mutant strains and parental wild type were picked from MM plates and grown for 48 h. Five hundred µl of this culture was used to inoculate 50 ml of LB and allowed to grow for 6 h with shaking at 37°C. Forty five ml of this culture was removed, placed in a new flask and treated with 3 mM $H_2O_2$. The growth was measured by $OD_{600}$ measurements at regular interval for 8 days. Cell outgrowth post $H_2O_2$ treatment was represented graphically as $OD_{600}$ *vs*. time.

## 2.7.3 Mutational assessments

### 2.7.3.1 Mutation frequency

Mutation frequencies are a measure of the fraction of mutants in a given culture. A single colony was inoculated into 10 ml of LB media and grown to an $OD_{600}$ of 1 - 1.5. One ml of this culture was spread neat onto an MM plate supplemented with 200 µg/ml rifampicin to select for rifampicin-resistant mutants whereas aliquots from a dilution series of the culture ($10^{-4}$-$10^{-6}$) were plated onto drug-free control plates to obtain the total number of cells per ml of culture. Plates were incubated at 37°C and CFUs were scored after 3 days on plates without supplementation. Colonies in the presence of rifampicin were scored after 5-7 days and the mutation frequency was calculated as shown below.

$$\text{Mutation frequency} = \frac{\text{number of rifampicin resistant colonies (CFU/ml)}}{\text{total number of cells (CFU/ml)}}$$

### 2.7.3.2 Mutation rate

The mutation rate of a cell is a measurement of the likelihood of it gaining a mutation during its lifetime. Luria-Delbrück fluctuation analysis is the most efficient method of measuring mutation rates (Rosche and Foster, 2000). Fluctuation assays compare the average mutation frequencies between parallel cultures inoculated from a common source. This analysis takes into account the variations between parallel cultures, considers the possibility of mutations arising early or late during growth and allows for the estimation of the number of mutational events or the *m*-value.

All cultures are started from a common, low inoculum ($N_0$) and are checked for any pre-existing mutants (*r*). The cultures are then aliquoted into several parallel tubes and allowed to grow for 5-7 days, thus giving each culture an equal chance of producing mutants. The estimation of the number of mutational events (*m*-value) and the approximate final size of the population (*Nt*) is then used to calculate the final mutation rate ($\mu$) as discussed in section 2.7.3.2.2 (Rosche and Foster, 2000).

#### 2.7.3.2.1 Experimental procedure

The experimental procedure for determining mutation rates in *M. smegmatis* was performed as previously described (Rosche and Foster, 2000; Machowski *et al.*, 2007). Single colonies were grown in MM to an $OD_{600}$ of 1-1.5 (late log phase) which represents approximately $10^8$ cells/ml. A dilution series ($10^{-5}$-$10^{-8}$) of this pre-culture was plated onto MM to measure $N_0$ and a 1 ml aliquot was plated onto MM containing rifampicin to confirm the absence of pre-existing mutants (*r*). Following this a dilution series of the pre-culture was prepared, which reduced the cell density from ~$10^8$ to ~$10^2$ cells/ml. To prevent clumping, the cultures were vigorously mixed

at each dilution step. The final diluted culture (200 ml of a ~$10^2$ cells/ml) was stirred continuously with a magnetic stirrer (FMH Instruments Heat stirrer, Stuart heat stir SB162) to ensure homogeneity whilst 2.5 ml aliquots were dispensed in 25-30 cultures tubes. The parallel culture tubes were grown shaking at 37°C for 5-7 days in covered beakers containing moist tissue to prevent excess evaporation. After incubation, 100 µl from each culture tube was sampled and a dilution series prepared to determine the *Nt* value. The entire remaining culture from each tube was plated onto an MM agar plate supplemented with 200 µg/ml rifampicin. The plates were incubated at 37°C and scored after 3 days for total cell count (*Nt*) or after 5-7 days for rifampicin resistant ($rif^R$) mutants. Once consistent CFUs in parallel cultures were observed, *Nt*-values were determined from only 5 dedicated tubes so that sampling from the experimental tubes did not bias the mutation assessment (Rosche and Foster, 2000).

#### 2.7.3.2.2 Statistical analysis

The mutation rate was calculated by dividing the likely number of mutations per culture (*m*-value) by the number of cells in the culture (*Nt* value), as shown in Equation 5 (Rosche and Foster, 2000). The number of cells in the culture was estimated by the CFUs in the parallel cultures. The *m*-value is a measure of the probability of a cell sustaining a mutation in its lifetime and this is obtained by a combination of calculations resulting in a Luria-Delbrück distribution (Rosche and Foster, 2000). This distribution allows one to calculate the most probable *m*-value. Equation 1 and Equation 2 were used to estimate the *m*-value. There are various calculations used to estimate the *m*-value, but the $P_o$ method was used in this study as the number of $rif^R$ colonies arising from one or more of the parallel cultures equalled zero (Rosche and Foster, 2000). A Luria-Delbrück distribution was formulated using the two Ma-Sandri-Sarkar (MSS) Maximum-Likelihood algorithms, shown as

Equation 3 and Equation 4. The extrapolated results from Equation 3 correspond to all possible *m*-values, taking into account the estimated *m*-value. These *m*-values are plotted on the *x*-axis and the extrapolated results from Equation 4 correspond to all possible *y*-axis values. These probable *x* and *y* values are then plotted to give a Luria-Delbrück distribution, an example of which is shown in Fig. 2. The peak of this distribution is the most probable *m*-value which is used to determine the mutation rate (Equation 5).

| | Equation 3 (m-value estimates) | Equation 4 |
|---|---|---|
| m1 | -0.8 | -3.0E-06 |
| m2 | -0.6 | -1.1E-09 |
| m3 | -0.4 | -4.9E-14 |
| m4 | -0.2 | -1.5E-20 |
| m5 | 0.0 | 0.0E+00 |
| m6 | 0.2 | 1.5E-22 |
| m7 | 0.4 | 4.6E-20 |
| m8 | 0.6 | 2.7E-19 |
| initial m | 0.8 | 3.0E-19 |
| m9 | 1.0 | 1.3E-19 |
| m10 | 1.2 | 2.9E-20 |
| m11 | 1.4 | 4.2E-21 |
| m12 | 1.6 | 4.3E-22 |
| m13 | 1.8 | 3.4E-23 |
| m14 | 2.0 | 2.2E-24 |
| m15 | 2.2 | 1.2E-25 |
| m16 | 2.4 | 5.6E-27 |

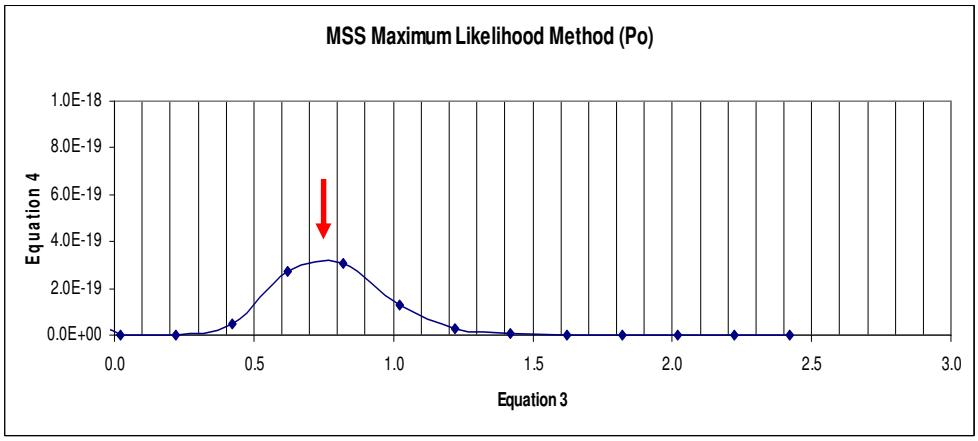



**Fig. 2. An example of the results of a Luria-Delbrück distribution**

The table lists the results of Equation 3 and Equation 4, which when plotted onto a graph allows for the identification of the most probable m-value (the highest point indicated by the arrow).

These calculations were performed using Microsoft Office Excel spreadsheets developed by Dr. E. Machowski (MMRU).

The calculations were as follows:

Equation 1 - $P_o$ value estimation

$$P_o = \frac{\text{number tubes with zero}}{\text{total number of tubes}}$$

Equation 2 - $P_o$ $m$-value equation

$$p_o = e^{-m}$$

$$m = -ln\ p_o$$

Equation 3

$$p_o = e^{-m};\ p_r = \frac{m}{r} \sum_{i=0}^{r-1} \frac{p_i}{r - 1 + 1}$$

Equation 4

$$f(r|m) = \Pi f(r_i|m)$$

Where $f(r|m) = p_r$ from equation 3

Equation 5

$$\mu = \frac{m\ x\ ln2}{Nt}$$

# 3. Results

## 3.1 Bioinformatics analysis

### 3.1.1 Identification of DNA glycosylase-encoding *fpg* and *nei* genes in *M. smegmatis*

A single *fpg* gene had been identified in the *M. tuberculosis* H37Rv genome (Mizrahi and Andersen, 1998). Since genome sequence data for both *M. smegmatis* (http://cmr.tigr.org) and *M. tuberculosis* (Cole *et al.*, 1998) strains were available at the time that this study was initiated, it was important to scan these genomes for *fpg/nei* homologues using several bioinformatics tools. Using the *M. tuberculosis* genome database Tuberculist, the previously reported *fpg* gene was identified as Rv2924c and upon an extended analysis, a second *fpg* (Rv0944) and two *nei* glycosylases (Rv3297 and Rv2464c) were identified. Using the programmes ACT, BLAST and CMR, homologues of these genes were also identified in *M. smegmatis,* two of which were annotated as *fpg* glycosylases (MSMEG_2419 and MSMEG_5545) and two as *nei* glycosylases (MSMEG_1756 and MSMEG_4683).

The proteins encoded by the *M. tuberculosis* and *M. smegmatis fpg* and *nei* homologues were classified within the same Clusters of Orthologous Groups (COG), namely: the formamidopyrimidine DNA glycosylase cluster, COG0266. Members of this COG were identified in 37 other genomes covering 5 phyla including Proteobacteria, Firmicutes, Deinococcus-Thermus, Cyanobacteria and Actinomycetes. In the majority of these genomes, only one copy of either *fpg* or *nei* was present and rarely both. However, within the Actinomycetes, several copies of both genes were identified. Additional searches using BLASTn (nucleotide-

nucleotide), GenoList and CMR databases revealed similar *fpg* and *nei* glycosylase-encoding genes in other mycobacterial species (*M. ulcerans, M. avium, M. avium paratuberculosis, M. bovis* and *M. leprae)*. This extensive genome sequence analysis also uncovered a fifth possible *fpg* and *nei* glycosylase-encoding gene in *M. avium* and *M. avium paratuberculosis* (Appendix B, Table 10). In addition to this fifth putative *fpg/nei* gene, an extra gene encoding the DNA glycosylase Endonuclease III (Nth) was identified in *M. avium.*

The general Fpg DNA glycosylase family structure consists of two domains separated by a flexible hinge (Wallace *et al.*, 2003; Zharkov *et al.*, 2003). The N-terminal domain contains the active site, while the C-terminal contains a four-cysteine zinc finger domain as well as a helix-two-turn-helix (H2TH), both of which are involved in binding the DNA (Gilboa *et al.*, 2002; Zharkov *et al.*, 2002; Wallace *et al.*, 2003; Zharkov *et al.*, 2003). Pfam, Prosite and InterPro scans were used to identify these domains within the *M. smegmatis* glycosylases. Differences in the results of these scans are caused by the individual algorithms of the programs used, cut-offs and whether the databases were based on curated (Pfam-A) or automatically generated results (Prosite and InterPro).

The Pfam scans identified Pfam-A criteria for the N-terminal domain and the H2TH portion of the C-terminal domain in all four of the *M. smegmatis* glycosylases. However, Pfam-A criteria for the zinc finger domain was only identified in the proteins encoded by the *fpg* gene, MSMEG_2419, and the *nei* gene, MSMEG_1756. Therefore Prosite and InterPro scans were performed. Using Prosite and InterPro the predicted zinc finger domain was identified in both the proteins encoded by the second *fpg* gene (MSMEG_5545) and the second *nei* gene (MSMEG_4683). Therefore, bioinformatics analysis identified the necessary protein domains required for the classification of these predicted proteins in the Fpg family of DNA glycosylases.

BLASTp (protein-protein) searches were performed to compare *M. smegmatis* Fpg and Nei glycosylases to other Actinobacteria and structurally characterised Fpg and Nei proteins found in *E. coli, Thermus thermophilus, Geobacillus stearothermophilus (*previously named *Bacillus stearothermophilus*), and *Lactococcous lactis* (Gilboa *et al.*, 2002; Zharkov *et al.*, 2003; Golan *et al.*, 2005). The results indicated significant similarity (low E-values) between the *M. smegmatis* Fpg and Nei glycosylases and to those previously characterised (Appendix C, Table 11). The *M. smegmatis* Fpg and Nei proteins were further compared to other Fpg and Nei proteins by ClustalW alignments (Fig. 3) (Wallace *et al.*, 2003; Zharkov *et al.*, 2003). These alignments highlighted the strong resemblance between the Fpg and Nei proteins and also allowed one to distinguish between the Fpg and Nei glycosylases. Unique amino acid residues at positions 4, 155 and 172 distinguished the protein as either Fpg or Nei (Wallace *et al.*, 2003; Zharkov *et al.*, 2003). At position 4, Fpg proteins have a conserved leucine residue whereas Nei proteins have a conserved glycine, while at position 155 a lysine was conserved in Fpg and either an alanine, glycine or serine residue was present in Nei. Greater variations were observed at position 172, where Fpg sequences had maintained a neutral alanine or valine residue, with the exception of MSMEG_5545 and Rv0944 where a serine is found, while the polar residues cysteine, arginine and serine were found in the Nei sequence as indicated in the third box in Fig. 3. These alignments clearly highlight similarities and differences between the Fpg and Nei glycosylases and confirmed the predicted annotations of these proteins listed in the CMR database.

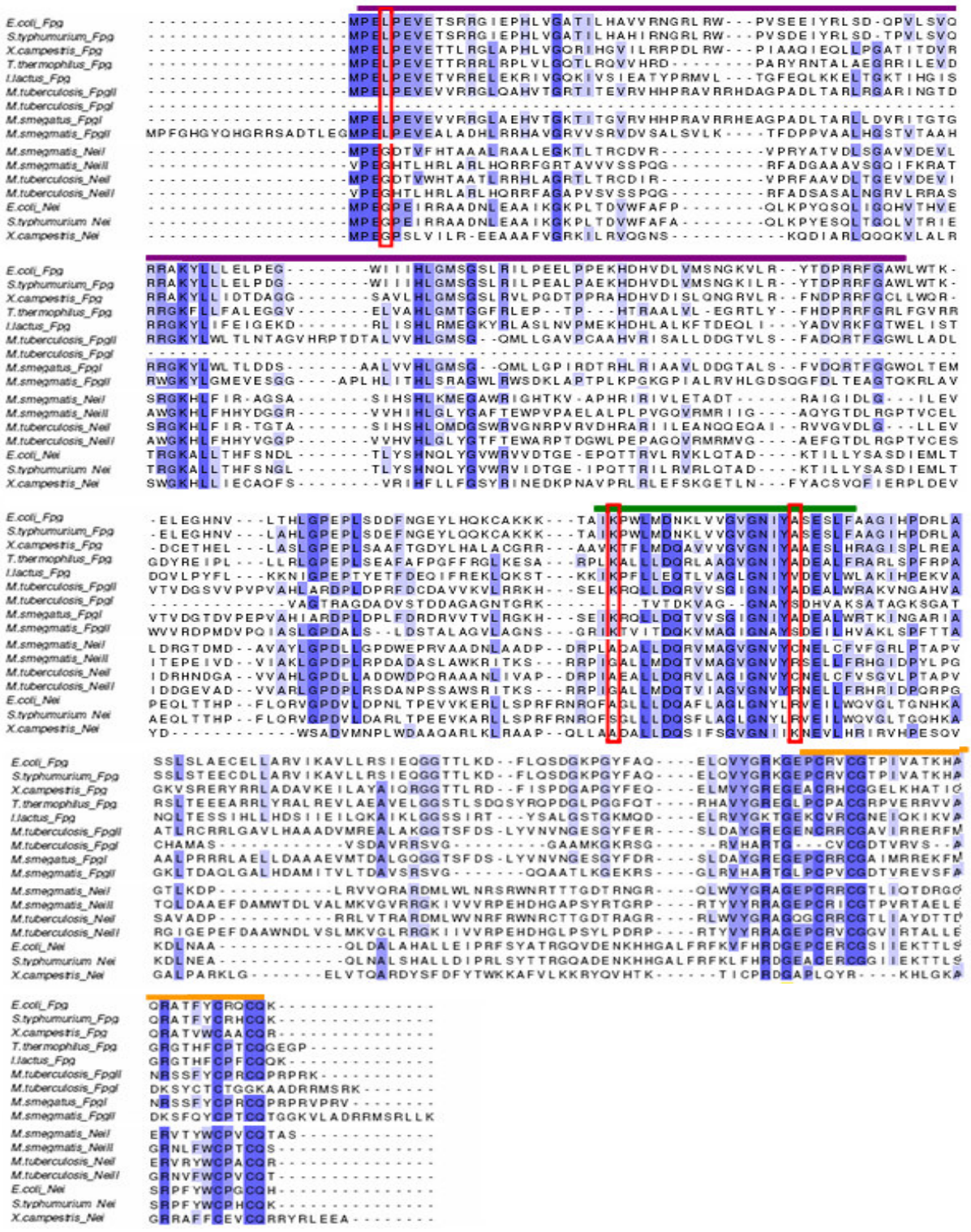



**Fig. 3. Alignments of the amino acid sequences of Fpg and Nei proteins from various micro organisms.**

Sequences were downloaded from CMR for *E. coli, S. typhimurium, Xanthomonas campestris, T. thermophilus, L. lactis, M. tuberculosis* and *M. smegmatis.* Differences between Fpg and Nei residues are highlighted in the red boxes, while the N-terminal domain is indicated in purple. The H2TH domain is indicated in green and the hairpin loop of the zinc finger domain indicated in orange. This figure is based on similar alignments performed by Wallace *et al.* (2003).

Interestingly, ClustalW protein alignments showed that the *M. tuberculosis* FpgII glycosylase is missing the entire N-terminal active site (Fig. 3) suggesting that this truncation is likely to abrogate FpgII glycosylase function in *M. tuberculosis*.

In summary, the bioinformatic analysis identified two putative Fpg and two putative Nei DNA glycosylases in *M. smegmatis* and *M. tuberculosis*. These glycosylases had maintained the necessary domains for Fpg and Nei protein function and showed strong resemblances to each other as well as to previously characterised Fpg family glycosylases. These data support the annotation of the predicted *M. smegmatis* proteins as Fpg and Nei DNA glycosylases, respectively.

### 3.1.2 Comparison of Fpg and Nei glycosylases in *M. tuberculosis* and *M. smegmatis*

Preliminary bioinformatic analysis showed that *M. smegmatis* and *M. tuberculosis* have similar Fpg and Nei DNA glycosylase homologues. The homologues from both organisms were examined further to identify differences and similarities.

*M. tuberculosis* and *M. smegmatis* Fpg and Nei glycosylase protein alignments were generated using ClustalW and their percentage pairwise scores, denoting protein resemblances, were calculated (Table 3). Scores above 30% were considered significant.

**Table 3. The percent similarity between *M. tuberculosis* and *M. smegmatis* Fpg and Nei glycosylases**

| | Rv2924c | Rv0944 | Rv3297 | Rv2464c | MSMEG_2419 | MSMEG_5545 | MSMEG_1756 | MSMEG_4683 |
|---|---|---|---|---|---|---|---|---|
| Rv2924c | | 17.5 | 28.67 | 25.09 | 76.12 | 29.15 | 25 | 23.67 |
| Rv0944 | 17.5 | | 25.81 | 24.49 | 24.1 | 58 | 20.34 | 15.22 |
| Rv3297 | 28.67 | 25.81 | | 32.35 | 27.4 | 25.81 | 37.5 | 33.7 |
| Rv2464c | 25.09 | 24.49 | 32.35 | | 24.28 | 15.22 | 28.3 | 73.51 |
| MSMEG_2419 | 76.12 | 24.1 | 27.4 | 24.28 | | 29.15 | 20 | 24.28 |
| MSMEG_5545 | 29.15 | 58 | 25.81 | 15.22 | 29.15 | | 66.73 | 29.75 |
| MSMEG_1756 | 25 | 20.34 | 37.5 | 28.3 | 20 | 66.67 | | 24.53 |
| MSMEG_4683 | 23.67 | 15.22 | 33.7 | 73.51 | 24.28 | 29.75 | 24.53 | |

Percentages highlighted in red indicate the highest pairwise score. The homologous pairing of the *M. smegmatis* Nei-encoding protein MSMEG_1756, and the *M. tuberculosis* Nei-encoding protein Rv3297, is indicated in yellow and is considered significant.

Related genes from different organisms are found in similar genomic contexts suggesting that these genes are evolutionarily related. Genetic context comparisons for the *fpg* and *nei* genes of *M. tuberculosis* and *M. smegmatis* were carried out using CMR (Fig. 4). Analysis of genetic homology (Table 3) and gene synteny (Fig. 4) allowed for identification and pairing of the *fpg/nei* homologues in *M. tuberculosis* and *M. smegmatis,* thus the four othologues were annotated as: *fpgI*, *fpgII*, *neiI* and *neiII* (illustrated in Fig. 4).

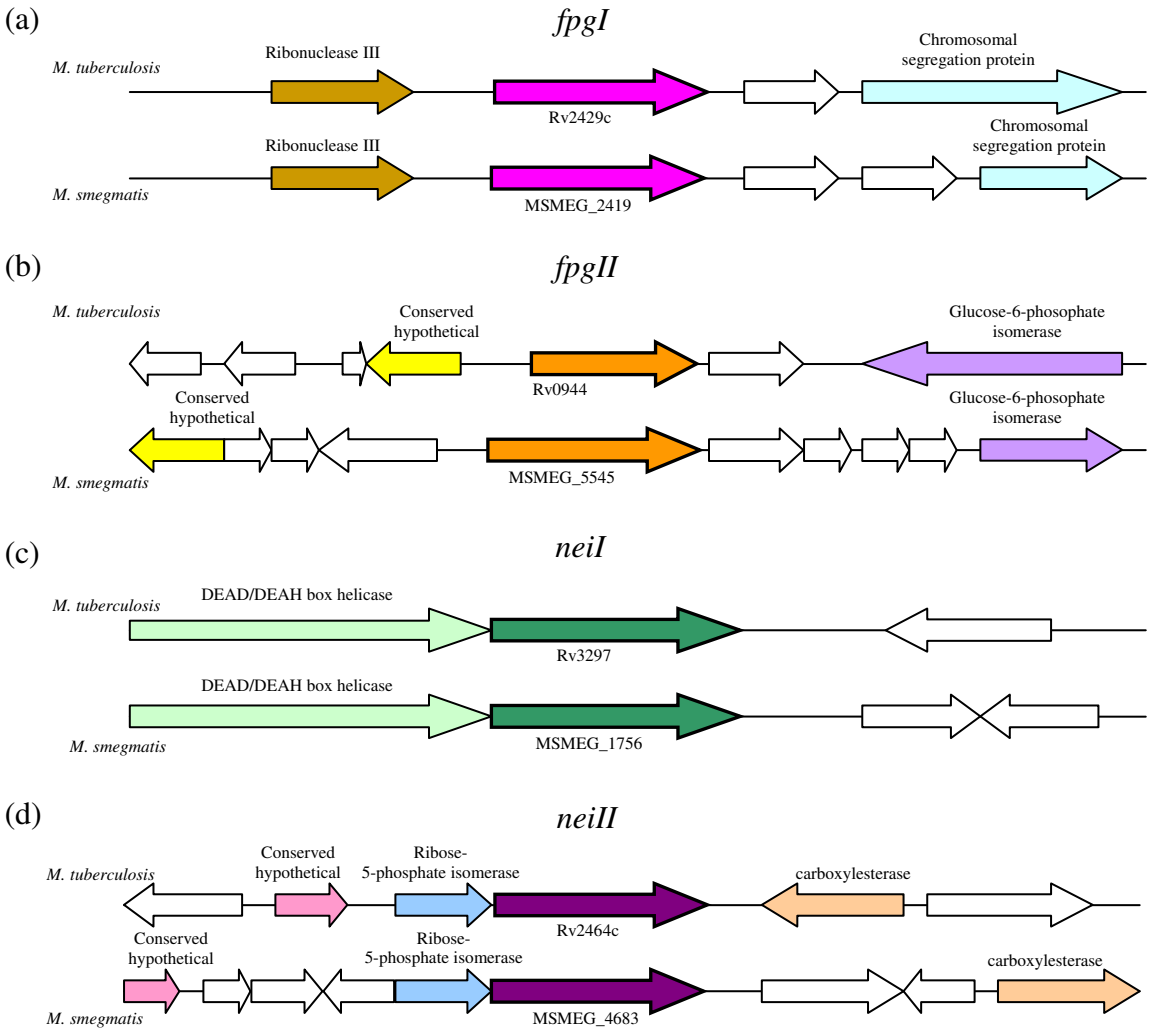



**Fig. 4. Genomic context comparisons of *fpg* and *nei* DNA glycosylases in *M. tuberculosis* and *M. smegmatis*.**
Block arrows in the same colour (other than white) denote homologous genes.

The *nei* gene MSMEG_1756 showed a strong similarity to the *fpg* gene MSMEG_5545 (Table 3). Although this similarity could possibly indicate an increased functional overlap of these *fpg* and *nei* genes, they were not considered to be orthologues. The pairwise score for the *nei* glycosylases Rv3297 and MSMEG_1756 was significant at 37.5% (Table 3) and these genes shared similar genomic contexts (Fig. 4c) and were therefore considered orthologous. Interestingly,

*neiI* appeared to be operonic with a gene encoding a DNA helicase (Fig. 4c). This operonic relationship was confirmed by operonic predictions in *M. tuberculosis* (Roback *et al.*, 2007). Since these genes are both involved in DNA repair, their coordinate expression suggests a possible interaction between Nei and the helicase during the repair of damaged DNA.

Genomic comparisons of the genes encoding Fpg and Nei glycosylases in other mycobacteria showed that *M. ulcerans, M. leprae, M. avium, M. avium paratuberculosis* and *M. bovis* maintained all four *fpg* and *nei* genes in the same genomic context as *M. tuberculosis* and *M. smegmatis*. In contrast, in *M. leprae*, an organism with a highly degenerate genome (Young and Robertson, 2001), only one gene, *fpgI,* was functionally conserved. Interestingly, an additional glycosylase belonging to the Fpg family was identified in a conserved genomic context in both *M. avium* and *M. avium paratuberculosis*. *M. bovis* displays a truncation in the *fpgII* similar to the one observed in *M. tuberculosis*. This truncation of the *fpgII* as well as the presence of the fifth glycosylase in some mycobacteria is consistent with the evolutionary history and relatedness of mycobacterial species, as predicted by phylogenetic analyses (Gey van Pittius *et al.*, 2006).

## 3.2 Construction of suicide vectors

Once the putative glycosylase–encoding genes were identified, suicide vectors carrying inactivated copies of these genes were constructed. The plasmid maps and cloning strategies are listed in Appendix C, Fig. 34 to Fig. 37.

### 3.2.1 Construction of suicide vectors for the deletion of *fpgI*

hyg-marked (p2NILΔ*fpgI*::*hyg*::p17) and unmarked (p2NILΔ*fpgI*::p19) vectors containing a deleted allele of *fpgI* was constructed by PCR amplification. Availability of the complete genome sequence of *M. smegmatis* mc$^2$155 (http://cmr.tigr.org) allowed for the design of PCR primers which amplified both upstream and downstream flanking fragments, as shown in Fig. 5. The primer sequences, PCR reaction conditions and the resultant amplicons are detailed in Appendix B, Table 7.

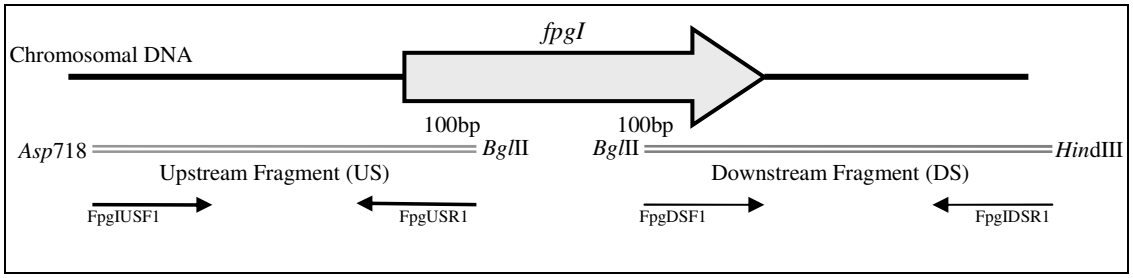



**Fig. 5. PCR amplification strategy for the generation of the deleted *fpgI* gene fragment.**
Primers are annotated by arrows Appendix B, Table 7.

Upstream and downstream fragments were amplified using Expand high fidelity polymerase (Roche Biochemicals) and cloned separately into the TOPO vector (Invitrogen) to the generate pTopoΔ*fpgI*us and pTopoΔ*fpgI*ds replicating vectors. These plasmids were transformed into competent *E. coli* (DH5α) cells and transformants were selected on media containing ampicillin. The vectors containing the deletion fragments from the resultant transformants were confirmed by restriction enzyme analysis, and plasmids from the correct clones were sequenced to ensure no mutations were introduced during the PCR amplification.

Sequence analysis identified a mutation in pTopoΔ*fpgI*ds. Therefore, the PCR amplification of the downstream *fpgI* fragment was repeated using a different polymerase and cloning vector, namely, Phusion polymerase (Finnzymes) and pcrSMART (Lucigen) respectively. The primers were phosphorylated prior to amplification with Phusion polymerase, and amplification generated a blunt-ended PCR product that was cloned in the pcrSMART vector to produce the pcrSMARTΔ*fpgI*ds plasmid. This plasmid was transformed into *E.coli* DH5α and the resultant transformants were screened by restriction analysis. The fragment was sequenced and no mutations were identified.

The upstream fragment was then excised from pTopoΔ*fpgI*us as an *Asp*718/*Sma*I fragment and cloned into *Asp*718/*Sma*I-digested pGEM-3Zf(+) to generate pGEMΔ*fpgI*us. The downstream fragment was excised as a *Hin*dIII/*Bgl*II fragment from pcrSMARTΔ*fpgI*ds and cloned into *Hin*dIII/*Bgl*II-digested pGEMΔ*fpgI*us+ to generate pGEMΔ*fpgI*. The pGEMΔ*fpgI* plasmid therefore contained the mutant *fpgI* allele (Δ*fpgI*) and its flanking regions (Appendix C, Fig. 34). The Δ*fpgI* region from pGEMΔ*fpgI* was excised as an *Asp*718/*Hin*dIII fragment and cloned in *Asp*718/ *Hin*dIII-digested p2NIL to generate p2NILΔ*fpgI*. The integrity of this vector was confirmed by restriction enzyme analysis (Appendix C, Fig. 38).

To generate a marked suicide vector, the *hyg* cassette from pIJ963 (Blondelet-Rouault *et al.*, 1997) was inserted into p2NILΔ*fpgI* at the unique *Bgl*II site, after which the *lacZ-sacB* from pGOAL17 (Parish and Stoker, 2000) was cloned as a *Pac*I cassette into the unique *Pac*I site on the p2NIL vector backbone to generate p2NILΔ*fpgI*::*hyg*::p17. The unmarked suicide vector, p2NILΔ*fpgI*::p19, was constructed by inserting the *lac*Z-*hyg-sacB* cassette from pGOAL19 (Parish and Stoker, 2000) into the *Pac*I restriction enzyme site of p2NILΔ*fpgI*.

Since insertion of these marker cassettes substantially increases the plasmid size, all further culturing was carried out at 30°C to minimise DNA rearrangements. All final constructs were thoroughly checked by restriction analysis to ensure that no rearrangements occurred during the cloning process (Fig. 6). The knockout vectors were confirmed to maintain km and hyg resistance as well as *lac*Z and *sacB* expression in *E. coli*. In the presence of β-galactosidase, the correct clones were blue and conferred $10^4$-fold killing of *E. coli* when plated on media containing 5% sucrose.

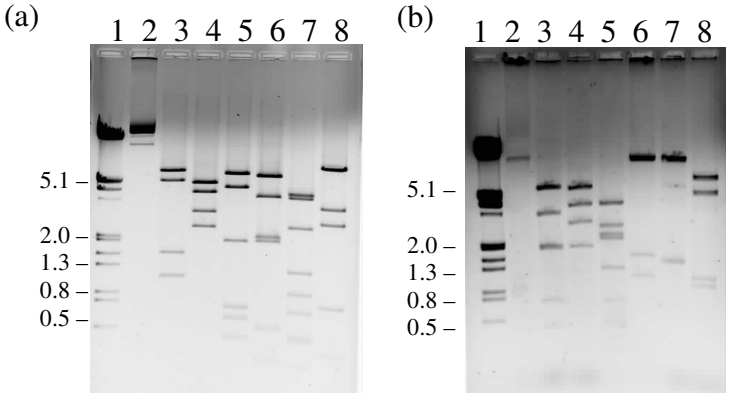



**Fig. 6. Restriction analysis of p2NIL-based marked and unmarked *fpgI* suicide vectors digested with various restriction endonucleases.**
Marker sizes in bp are shown adjacent to the fragment. (a) Digestion confirming plasmid p2NILΔ*fpgI*::19; Lane 1, molecular weight marker λIII; lane 2, uncut p2NIL Δ*fpgI*::19; lane 3, *Asp*718 digest (6520bp, 5197bp, 1580bp, 1132bp); lane 4, *Nru*I (4954bp, 4179bp, 2961p, 2365bp); lane 5, *Eco*RI digest (6122p, 4539bp, 1863bp, 766bp, 635bp, 534bp); lane 6, *Pst*I digest (5696p, 3847bp, 2031bp, 1979bp, 539bp); lane 7, *Pvu*I (4057bp, 3809bp, 1303bp, 981bp, 762bp, 480bp, 453bp, 266bp); lane 8, *Pvu*II digest (7300bp, 3276bp, 2557bp, 819bp). (b) Digestion confirming plasmid p2NILΔ*fpgI*::*hyg*::p17; lane 1, molecular weight marker λIII; lane 2, uncut p2NILΔ*fpgI*::hyg::p17; lane 3, *Eco*RI digest (6083bp, 3594bp, 1863bp, 1842bp, 766bp, 543bp); lane 4, *Nru*I digest (5739bp, 4046bp, 2961bp, 1891bp); lane 5, *Pvu*I digest (4057bp, 2739bp, 2384bp, 2229bp, 1303bp, 766bp, 480bp, 453bp); lane 6, *Sma*I digest (11556bp, 1620bp, 1227bp); lane 7, *Sal*I digest (11473bp, 1520bp, 1463bp); lane 8, *Asp*718 digest (7225bp, 5197bp, 1132bp, 1053bp).

### 3.2.2 Construction of suicide vectors for the deletion of *fpgII*

The unmarked suicide vector containing the deleted *fpgII* gene was constructed using a similar approach as for the deletion construct of *fpgI*. The upstream and downstream fragments were PCR amplified using Phusion polymerase, as detailed in Fig. 7 and in Appendix B, Table 7. The fragments were then cloned separately into the pcrSMART vector to generate pcrSMARTΔ*fpgII*us and pcrSMARTΔ*fpgII*ds. The resultant *E. coli* transformants were selected on km, the plasmids from positive clones were analysed by restriction endonucleases and the correct clones sequenced to exclude mutations during amplification of the deletion fragments.

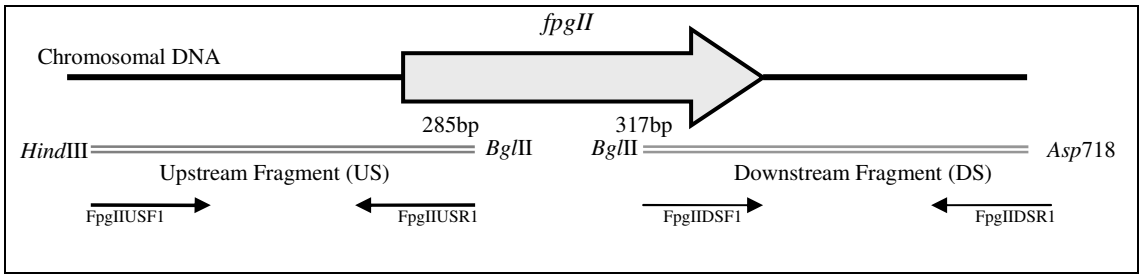



**Fig. 7. PCR amplification strategy for generating the *fpgII* deletion gene fragment.**
Primers are annotated by arrows.

Sequencing revealed no mutations in the upstream fragment; however, a single site mutation was present in the *Asp*718 restriction enzyme site (Appendix C, Fig. 42). The mutation, which abrogated the restriction site, occurred beyond the downstream homologous region and would therefore not affect the homologous recombination. Hence, it was not necessary to re-amplify the fragment and subsequent cloning of this fragment was performed using the unique *Eco*RI site in the pcrSMART vector together with the other restrictions sites incorporated into the primer sequences.

The *fpgII* deletion upstream and downstream fragments were excised from pcrSMARTΔ*fpgII*us and pcrSMARTΔ*fpgII*ds as *Hin*dIII/*Bgl*II and *Eco*RI/*Bgl*II fragments as illustrated in Appendix C, Fig. 35. In a single three-way cloning step, the upstream and downstream Δ*fpgII* fragments were cloned into p2NIL at the *Eco*RI and *Hin*dIII sites, to generate the p2NILΔ*fpgII* vector. The integrity of the construct was confirmed by restriction enzyme digestion (Appendix C, Fig. 39).

Finally, the *hyg-lacZ-sacB* cassette from pGOAL19 was cloned as a blunt-ended cassette into the *Xmn*I site of p2NIL since the unique *Pac*I site was lost during the cloning process. The stability of the final unmarked suicide vector, p2NILΔ*fpgII*::p19, was confirmed by restriction enzyme digests (Fig. 8) and positive clones were tested for sucrose sensitivity and β-galactosidase activity in *E. coli*, as described above.

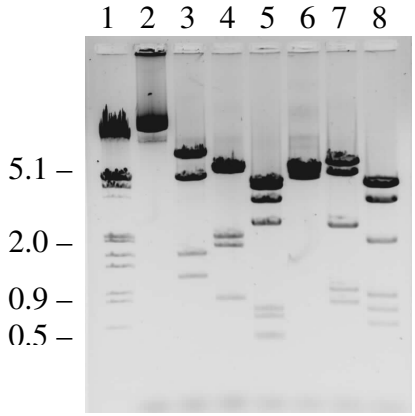



**Fig. 8. Restriction digestion of the p2NIL based unmarked *fpgII* suicide vector digested with various restriction endonucleases.**
Marker sizes in bp are shown adjacent to the fragment. The digestion of p2NILΔ*fpgII*::p19: lane 1; molecular weight marker λIII, lane 2; uncut p2NILΔ*fpgII*::p19; lane 3, *Asp*718 digest (8835bp, 5197bp, 1580bp, 1132bp); lane 4, *Pst*I digest (6309bp, 5696bp, 2031bp, 1800bp, 908bp); lane 5, *Pvu*I digest (4630bp, 4146bp, 3277bp, 2384bp, 726bp, 648bp, 480bp); lane 6, *Sma*I digest (6629bp, 5534bp, 4981bp); lane 7, *Hin*dIII/ *Bgl*II digest (7048bp, 5559bp, 2336bp, 976bp, 739bp); lane 8, *Bgl*II/*Eco*RI digest (4539bp, 4529bp, 3343bp, 1863bp, 919bp, 766bp, 607bp, 150bp).

### 3.2.3 Construction of suicide vectors for the deletion of *neiI*

The first cloning step in the construction of a vector containing a deleted allele of *neiI* was carried out by Mr. S. Nathoo, a visiting student in the MMRU. The upstream and downstream regions flanking the *neiI* gene were amplified using primers engineered with unique restriction sites, as detailed in Fig. 9 and in Appendix B, Table 7. The resultant fragments were cloned into the TOPO vector and, in a single three-way cloning step, the upstream and downstream fragments were inserted into *Hin*dIII/*Asp*718 digested p2NIL to generate p2NILΔ*neiI*. My involvement was to verify the integrity of the p2NILΔ*neiI* vector by restriction enzyme digestion (Appendix C, Fig. 40) and sequence the constructs to ensure that the amplified fragments did not contain any mutations.

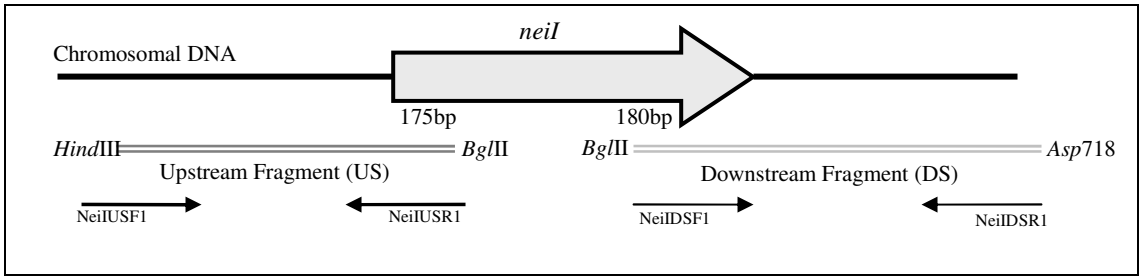



**Fig. 9. PCR amplification strategy for the generation of a deleted *neiI* gene fragment.**
Primers are annotated by arrows

Once the p2NILΔ*neiI* vector was confirmed to be correct, the p2NILΔ*neiI*::*hyg*::p17 vector was generated by cloning the *hyg* cassette (from pIJ963) and the *lacZ-sacB* cassette (from pGOAL17) into the unique *Bgl*II and *Pac*I sites in p2NIL respectively. For the unmarked suicide vector, the *hyg-lacZ-sacB* cassette from pGOAL19 was inserted at the *Pac*I site in p2NILΔ*neiI* to generate p2NILΔ*neiI*::p19 (Appendix C, Fig. 36).

The integrity of both the marked and unmarked *neiI* deletion constructs was confirmed by restriction enzyme digestions (Fig. 10), and *E. coli* clones carrying these vectors were shown to maintain sucrose sensitivity and β galactosidase activity.

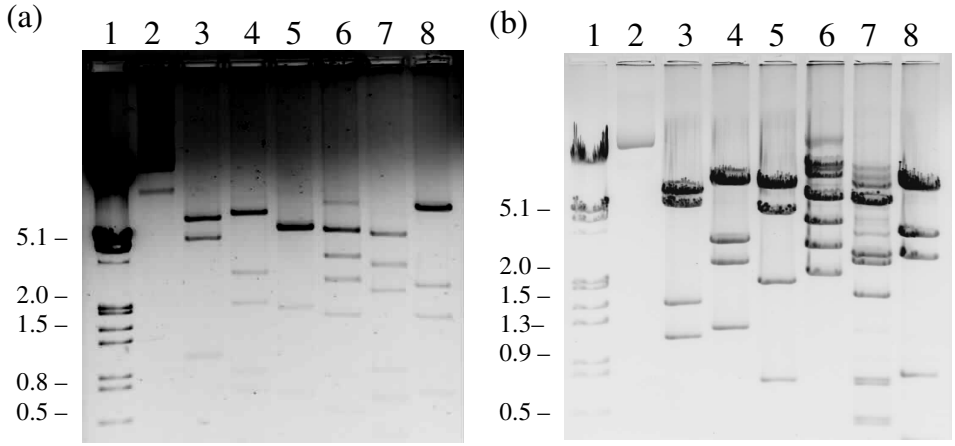



**Fig. 10. Restriction digestion of the p2NIL based marked and unmarked suicide vectors for *neiI* digested with various restriction endonucleases.**
Marker sizes in bp are shown adjacent to the fragment. (a) Digestion of p2NILΔ*neiI*::hyg::p17; lane 1, molecular weight marker λIII; lane 2, uncut p2NILΔ*neiI*::hyg::p17; lane 3, *Asp*718 digest (7281bp, 5197bp, 1132bp, 1121bp); lane 4, *Bam*HI digest (7911bp, 3072bp, 2044bp, 948bp, 786bp); lane 5, *Eco*RI digest (6094bp, 5888bp, 1863bp, 766bp, 750bp); lane 6, *Pst*I digest (5696bp, 3933bp, 2761bp, 1753bp); lane 7, *Pvu*I digest (5523bp, 3437bp, 2384bp, 1003bp, 755bp); lane 8, *Pvu*II digest (9088bp, 2557bp, 1790bp, 819bp, 363bp) (b) Digestion of p2NILΔ*neiI*::p19; lane 1, molecular weight marker λIII; lane 2, uncut p2NILΔ*neiI*::p19; lane 3, *Asp*718 digest (6644bp, 5197bp, 1580bp, 1132bp); lane 4, *Bam*HI digest (7911bp, 3072bp, 2366bp, 1234bp); lane 5, *Eco*RI digest (7265bp, 4539bp, 1863bp, 766bp); lane 6, *Pst*I digest (5695bp, 3933bp, 2756bp, 2031bp); lane 7, *Pvu*I digest (5523bp, 2583bp, 2384bp, 1679bp, 755bp, 726bp, 480bp, 453bp); lane 8, *Pvu*II digest (7330bp, 3370bp, 2557bp, 819bp, 363bp).

### 3.2.4 Construction of suicide vectors for the deletion of *neiII*

The final deletion construct that was made targeted the *neiII* gene. The upstream and downstream region containing flanking homologous sequence were PCR amplified, as detailed in Fig. 11 and in Appendix B, Table 7. The resultant PCR products were cloned separately into the pcrSMART cloning vectors to generate

pcrSMARTΔ*neiII*us and pcrSMARTΔ*neiII*ds. The insertion was confirmed by restriction digest and checked for mutations by sequence analysis. Primers for the amplification of the downstream fragment had unique restriction endonuclease sites incorporated, whilst only one restriction site had to be engineered into the reverse primer amplifying the upstream fragment since a *Bgl*II site was present about 58bp into the primer as shown in Fig. 11.

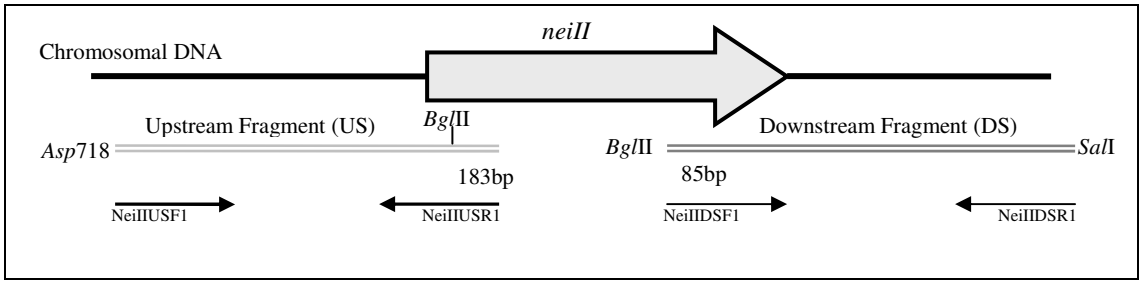



**Fig. 11. PCR amplification strategy for the generation of a deleted *neiII* gene fragment.**
Primers are annotated by arrows

To generate the deleted Δ*neiII* fragment, the downstream fragment was excised from pcrSMARTΔ*neiII*ds by digestion with *Sal*I and *Eco*RV resulting in one blunt (*Eco*RV) and one sticky end (*Sal*I). This fragment was cloned into *Sal*I/*Sma*I-digested pGEM-3Zf(+) to yield pGEMΔ*neiII*ds. The upstream fragment was cloned into pcrSMART to create pcrSMARTΔ*neiII*us. This fragment was then excised as an *Asp*718/*Bgl*II fragment and cloned into the *Asp*718/*Bgl*II-digested pGEMΔ*neiII*ds to generate pGEMΔ*neiII*. The deleted *neiII* fragment was excised from pGEMΔ*neiII* with *Asp*718 and *Hin*dIII and cloned into the same sites of p2NIL to generate p2NILΔ*neiII*. The constructs were checked by restriction enzyme digests for integrity (Appendix C, Fig. 41).

The p2NILΔ*neiII* vector was marked with the *hyg* cassette, which was cloned as a *Bgl*II fragment from pIJ963 into the unique *Bgl*II site of the vector. Finally, the *lacZ*-

*sacB* cassette from pGOAL17 was cloned into the unique *Pac*I site of p2NILΔ*neiII* to generate p2NILΔ*neiII*::*hyg*::p17. The unmarked suicide vector, p2NILΔ*fpgI*::p19, was constructed by inserting the *hyg-lac*Z-*sacB* cassette from pGOAL19 into the *Pac*I site of p2NILΔ*fpgI*. Both vectors were confirmed by restriction analysis (Fig. 12) and *E. coli* strains harbouring these vectors were shown to maintain sucrose-sensivity and β-galactosidase activity.

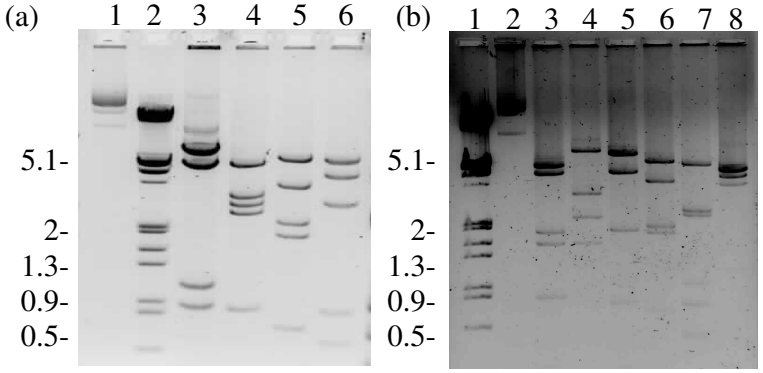



**Fig. 12. Restriction analysis of the p2NIL based marked and unmarked *neiII* suicide vectors cut *neiII* with various restriction endonucleases.**
Marker sizes in bp are shown adjacent to the fragment. (a) p2NILΔ*neiII*::hyg::p17 digestion; Lane 1, uncut p2NILΔ*neiII*::hyg; lane 2, molecular weight marker λIII; lane 3, *Asp*718 digest (7098bp, 5197bp, 1132bp, 936bp); lane 4, *Bam*HI digest (7325bp, 3072bp, 2757bp, 2447bp, 901bp); lane 5, *Eco*RI (5911bp, 3594bp, 2259bp, 1863bp, 766bp); lane 6, *Pst*I digest (5696bp, 4298bp, 2829bp, 897bp, 673bp) (b) p2NILΔ*neiII*::p19 digestion; lane 1, molecular weight marker λIII; lane 2, uncut p2NILΔ*neiII*::p19; lane 3, *Asp*718/*Hin*dIII digest (5197bp, 4436bp, 1846bp, 1580bp, 739bp); lane 4, *Bam*HI digest (7331bp, 3072bp, 2228bp, 1590bp); lane 5, *Eco*RI digest (6903bp, 4539bp, 1863bp, 766bp); lane 6, *Pst*I digest (5696bp, 3933bp, 2021bp, 1882bp, 649bp); lane 7, *Pvu*I digest (5523bp, 2503bp, 2384bp, 972bp, 726bp, 724bp, 480bp, 455bp); lane 8, *Nru*I digest (4954bp, 4353bp, 3741bp, 1170bp)

## 3.3 The generation of knockout mutants

All knockout mutants were generated as discussed in section 2.6. Briefly, the relevant suicide vectors were electroporated into electrocompetent $mc^2155$ and SCOs were identified as blue colonies on plates containing X-gal, km and hyg. SCO recombinants were passaged in the absence of antibiotics and putative DCOs were identified as white, sucrose-resistant colonies. These colonies were further confirmed to have lost the suicide vector backbone and its *lac*Z-*sacB* cassette by replica-plating the white, sucrose-resistant colonies in the presence and absence of km, an example of which is shown in Fig. 13. In this example, seven colonies showed no growth in the presence of km, suggesting that these colonies lost the vector backbone and were therefore considered as DCOs. However, a large majority of the colonies in this example were km resistant suggesting that these colonies must have maintained the p2NIL vector backbone, and that the sucrose resistance was acquired as a result of a mutation in the *sacB* gene.

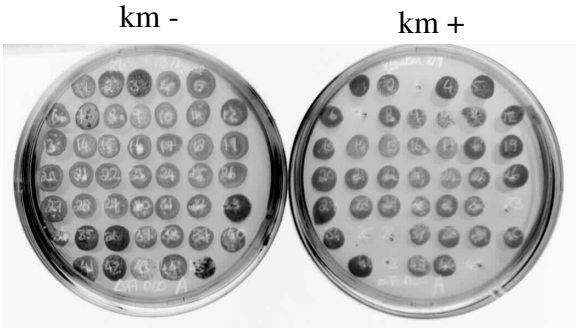



**Fig. 13. Replica-plating of possible DCOs of the unmarked Δ*fpgI* knockout mutant.**
White sucrose resistant colonies were resuspended in 30 µl of sterile distilled water and 5 µl spotted on 7H10 plates with and without km (20 µg/ml)

The white, sucrose-resistant, km-sensitive DCOs could either be a knockout mutant (if the second crossover event occurred on the opposite side of the deletion mutation compared to the first crossover event), or regenerated wild type (if the second crossover event occurred on the same side as the first). As an initial screen, clones

were examined by PCR using primers that amplified a smaller amplicon for the deleted allele compared to the wild type allele. PCR amplification was a useful means of screening for unmarked mutants since, in the absence of antibiotic selective pressure, the frequency of wild type revertants was generally high and as a result, many clones had to be screened to identify the DCO. As a final confirmation, the positive clones identified by PCR were genotyped by Southern blot analysis to ensure that the gene replacement was targeted and the upstream and downstream flanking regions did not undergo rearrangements.

Fig. 14 shows an example of the PCR and Southern blot analysis for the Δ*fpgI* deletion mutant. PCR amplification of the *fpgI* deletion region resulted in a 2.4kb amplicon for the wild type allele (lane 2) and a 1.4kb fragment for the mutant allele (lane 9). In lane 6, both the wild type and the mutant alleles were amplified implying that the clone was a SCO. This SCO must have acquired mutations in the *aph*, *sacB* and *lacZ* genes to result in km sensitivity, sucrose resistance and white coloration in the presence of X-gal. The correct DCO was analyzed further by Southern blot analysis (Fig. 14 c).

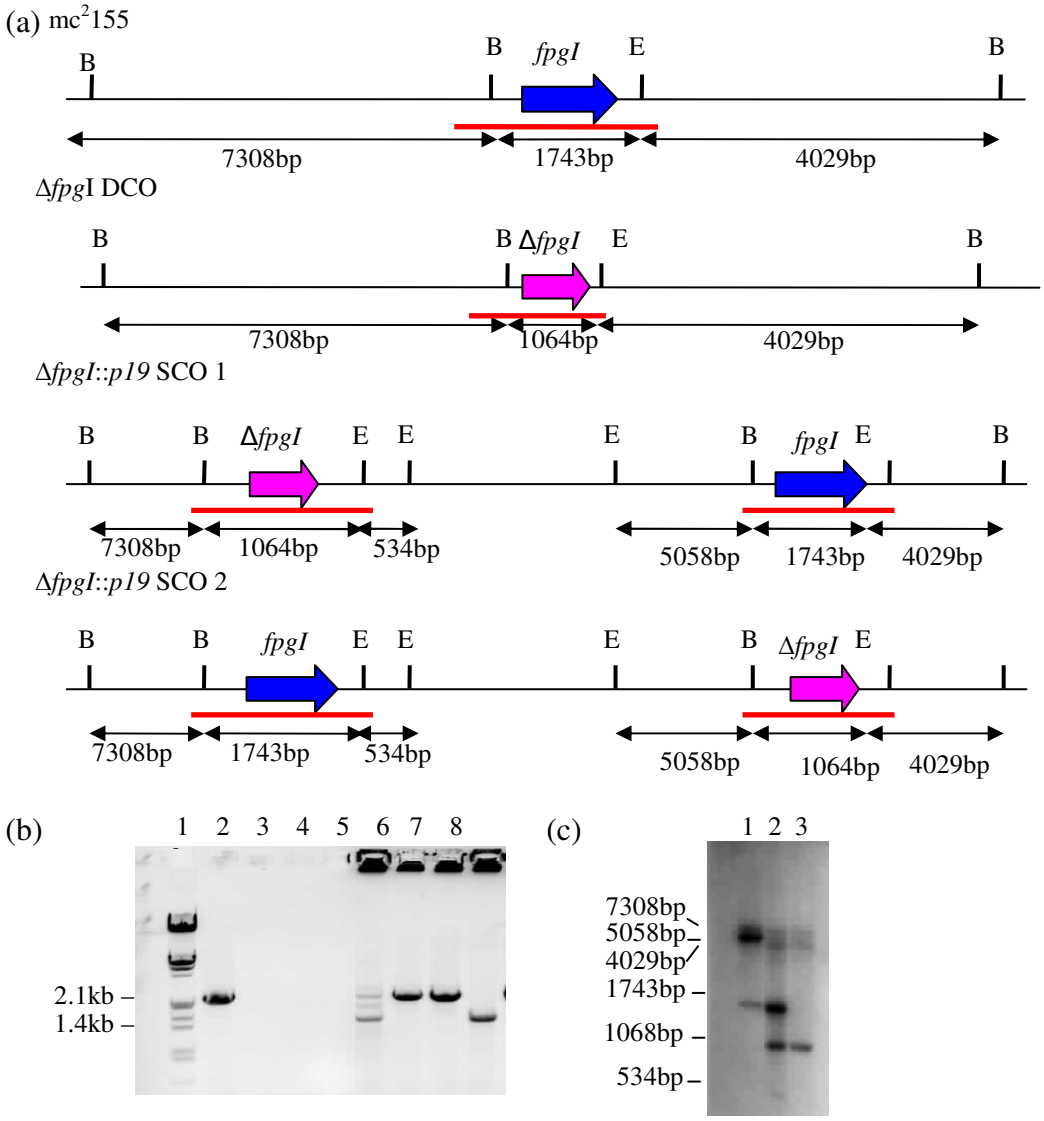



**Fig. 14. Genotypic characterisation of the Δ*fpgI* mutant of *M. smegmatis***
(a) Schematic representation of allelic replacement by homologous recombination. Single recombination events between homologous regions shared between the homologous regions in the inactivated construct and the chromosome on either side of the inactivated gene will generate SCO1 or SCO2 strain and a second crossover event will result in a DCO recombinant. (b) PCR amplification of chromosomal DNA from the parental strain and possible Δ*fpgI* knockout mutants using primers detailed in Appendix B, Table 9. Lane 1, molecular weight marker λIII; lane 2, mc$^2$155 (2.1kb); lane 3, forward primer only (negative control); lane 4, reverse primer only (negative control); lane 5, no DNA control; lanes 6-9, possible DCOs (1.4kb). (c) Southern blot analysis of the Δ*fpgI* knockout mutant. Chromosomal DNA from the parental strain, the SCO and the DCO was digested with *Eco*RI and *Bam*HI and the fragments separated on an agarose gel. The digested DNA was probed with a 2kb fragment (shown as a red line in panel (a)). Lane 1, wild type (7308bp, 4029bp, 1743bp); lane 2, Δ*fpgI* SCO (7308bp, 5058bp, 4029bp, 1743bp, 1064bp, 534bp); lane 3: Δ*fpgI* DCOs (7308bp, 4029bp, 1064bp).

The marked and unmarked vectors, p2NILΔ*fpgI*::*hyg*, p2NILΔ*fpgI,* p2NILΔ*neiI*::*hyg,* p2NILΔ*neiI,* p2NILΔ*neiII*::*hyg* and p2NILΔ*fpgII* were electroporated into the parental mc$^2$155 strain and the resultant transformants were screened by a two-step selection process to identify mutants lacking each of the four glycosylase-encoding genes. Deletion of the *fpg* and/or *nei* genes individually or in combination may have adverse effects or could be lethal in *M. smegmatis* thereby making selection of double crossover mutants difficult. The use of marked knockout constructs would facilate selection of mutants due to the presence of antibiotic.

Single marked and unmarked *fpg* and *nei* knockout mutants were therefore generated, and the unmarked mutants used as hosts for the sequential inactivation of the remaining glycosylase encoding genes. In this manner, the double knockout mutants, Δ*fpgI*Δ*fpgII* and Δ*neiI*Δ*neiII*, were generated. These were then used as background strains for the knockout of a third DNA glycosylase-encoding gene to generate the triple mutants, Δ*fpgI*Δ*fpgII*Δ*neiI* and Δ*neiI*Δ*neiII*Δ*fpgI*. The sequential deletion of the *fpg* and *nei* genes is illustrated in Fig. 15, and the confirmation by PCR and Southern blotting of the single, double and triple knockout mutants is illustrated in Fig. 16 - Fig. 22.

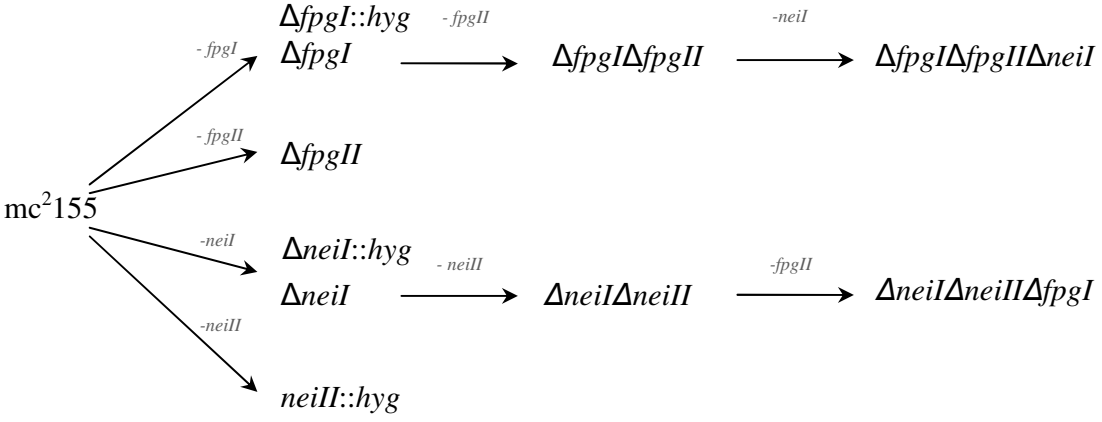



**Fig. 15. A schematic representation of the sequential deletion of *fpg* and *nei* in *M. smegmatis.***

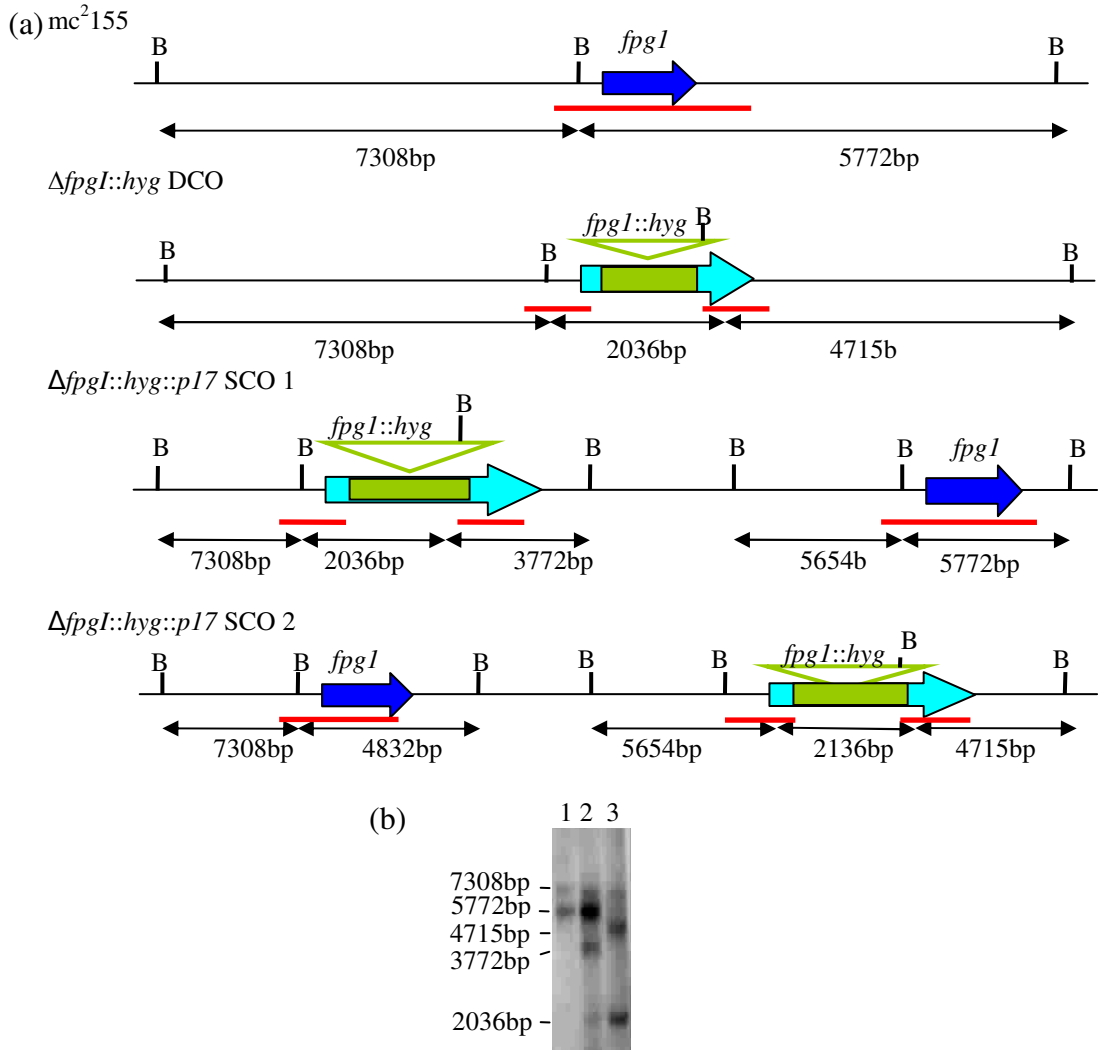



**Fig. 16. Genotypic characterization of the marked *fpgI* region of *M. smegmatis***
(a) Schematic representation of allelic replacement by homologous recombination. (b) The Southern blot analysis of the Δ*fpgI* marked-knockout mutant. Chromosomal DNA from the parental strain, the SCO and the DCO was digested with *Bam*HI and the fragments separated on an agarose gel. The digested DNA was probed with a 2kb fragment (shown as a red line in a). The expected sizes are as follows: lane 1, the wild type mc$^2$155 (73008bp, 5772bp); lane 2, Δ*fpgI*::*hyg* SCO 1 (7308bp, 5772bp, 5654bp, 3772bp, 2036bp); lane 3, Δ*fpgI*::*hyg* DCO (7308bp, 4715bp, 2036bp).

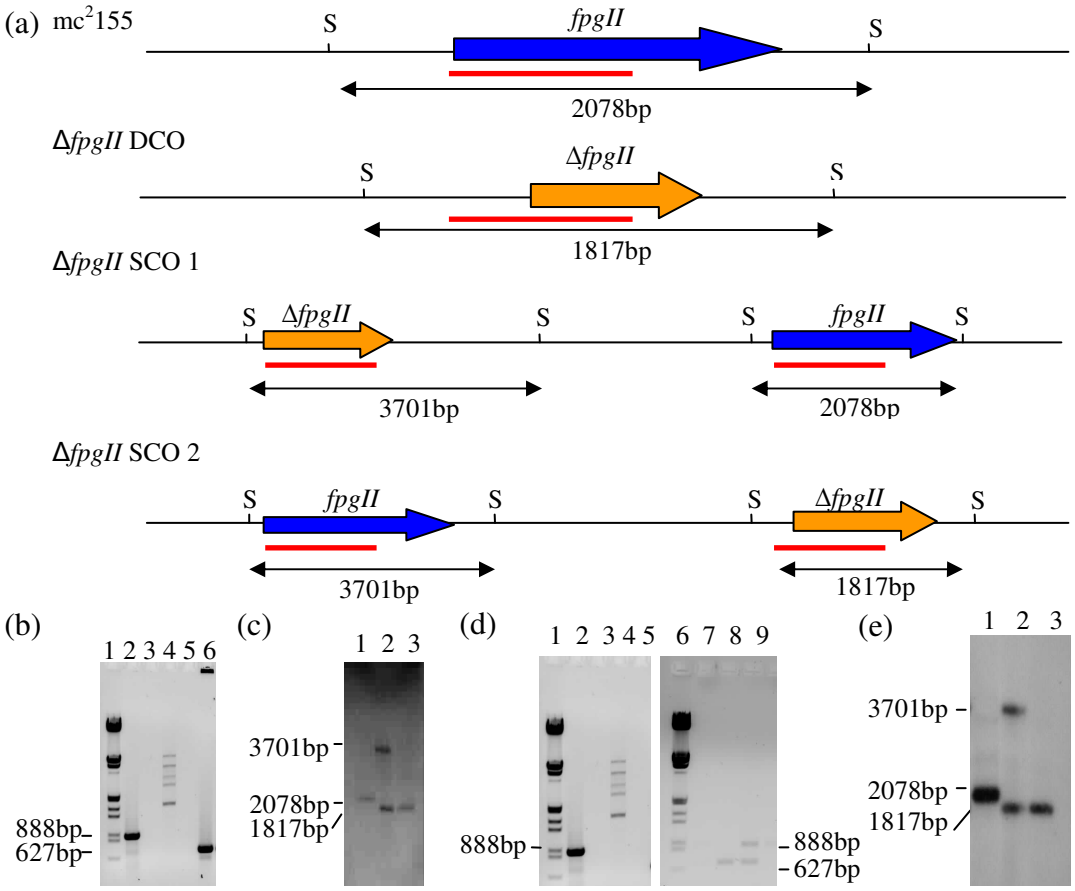



**Fig. 17. Genotypic characterization of the *fpgII* region of *M. smegmatis* for the single Δ*fpgII* and double Δ*fpgI*Δ*fpgII* mutants**

(a) Schematic representation of allelic replacement by homologous recombination. (b) PCR amplification of chromosomal DNA from the parental strain and possible Δ*fpgII* knockout mutants using primers detailed in Appendix B, Table 9. Lanes represent the following: lane 1, molecular weight marker λIII; lane 2, mc$^2$155 (888bp); lane 3, forward primer only (negative control); lane 4, reverse primer only (negative control); lane 5, no DNA control; lane 6, possible DCO (627bp). (c) The Southern blot analysis of the Δ*fpgII* mutant. Chromosomal DNA from the parental strain, the SCO and the DCO was digested with *Sma*I (S) and the fragments separated on an agarose gel. The digested DNA was probed with a 0.9kb fragment (shown as a red line in a). The expected sizes are as follows: lane 1, wild type mc$^2$155 (2078bp); lane 2, Δ*fpgII* SCO 1 (3701bp, 1817bp); lane 3, Δ*fpgII* DCO (1817bp). (d) PCR amplification of chromosomal DNA from the parental strain and putative Δ*fpgI*Δ*fpgII* mutants, the primers used are detailed in Appendix B, Table 9. Lanes represent the following: lanes 1 and 6; molecular weight marker λIII; lane 2, mc$^2$155 (888bp); lane 3, forward primer only (negative control); lane 4, reverse primer only (negative control); lane 5, no DNA control; lane 8, DCO (627bp); and lane 9, SCO with mutant and wild type alleles (e) The Southern blot analysis of Δ*fpgI*Δ*fpgII* mutants. The Chromosomal DNA was digested with *Sma*I and the DNA probed with a 0.9kb fragment. Lanes represent the following: lane 1, mc$^2$155 (2078bp); lane 2, Δ*fpgI*Δ*fpgII* SCO 2 (3701bp, 1817bp); lane 3, Δ*fpgI*Δ*fpgII* (1817bp).

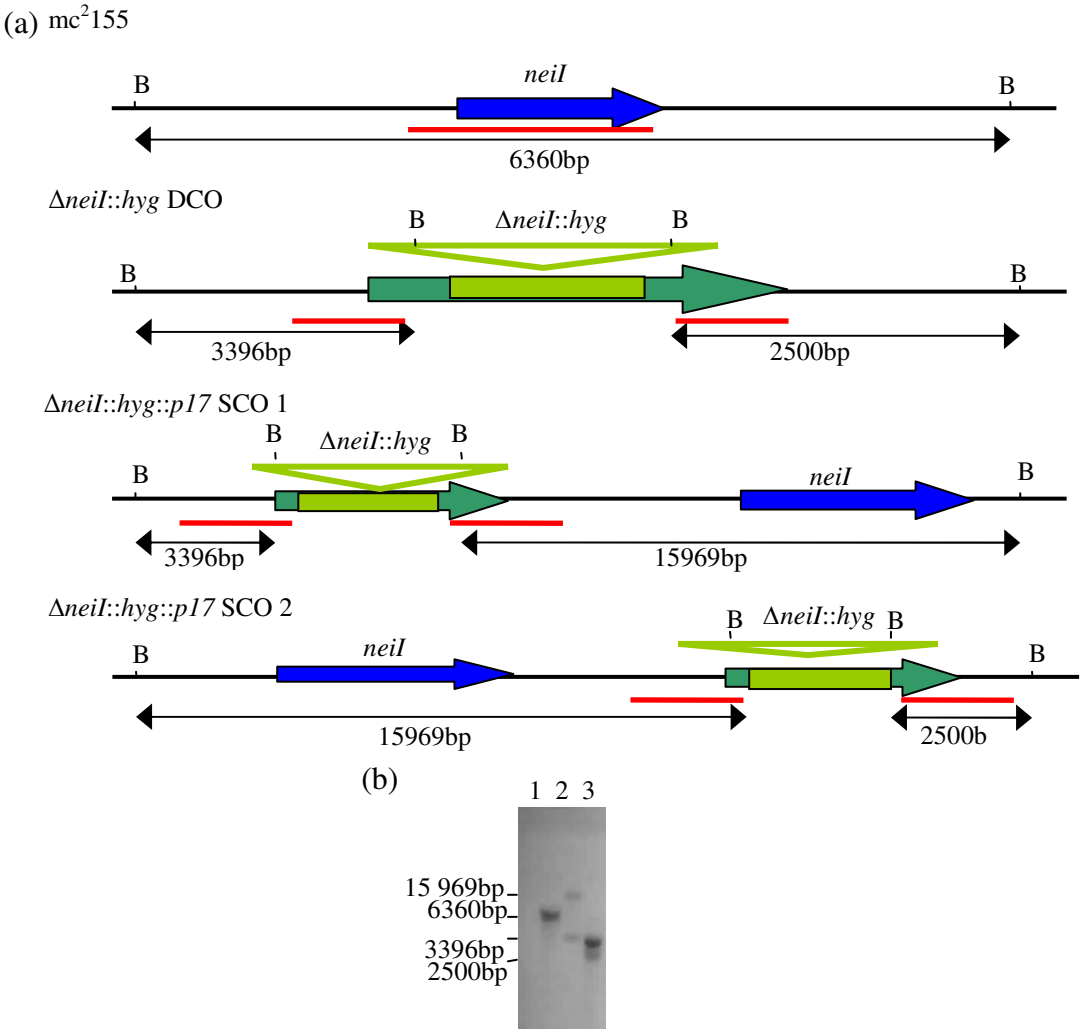



**Fig. 18. Genotypic characterization of the marked Δ*neiI* region of *M. smegmatis* for the Δ*neiI*::*hyg* mutant**

(a) Schematic representation of allelic replacement by homologous recombination. (b) The Southern blot analysis of the Δ*neiI*::*hyg* mutant. Chromosomal DNA from the parental strain, the SCO and the DCO was digested with *Bgl*II (B) and the fragments separated on an agarose gel. The digested DNA was probed with a 1.1kb fragment (shown as a red line in panel (a)). The expected sizes are as follows: lane 1, wild type mc²155 (6360bp); lane 2, Δ*neiI*::*hyg* SCO 1 (15 969bp, 3396bp); lane 3, Δ*neiI*::*hyg* DCO (3396bp, 2500bp).

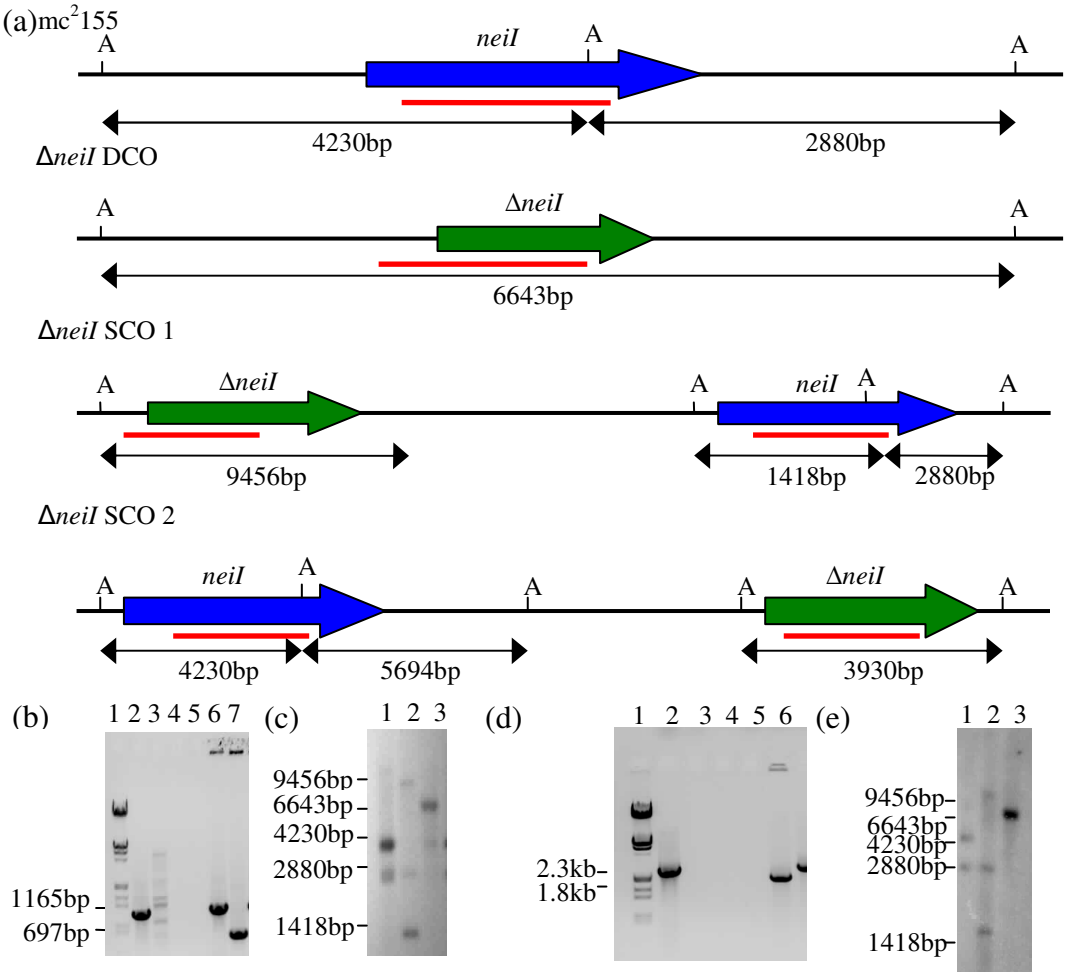



**Fig. 19: Genotypic characterization of the *neiI* region of *M. smegmatis* for the single Δ*neiI* and triple Δ*fpgI*Δ*fpgII*Δ*neiI* mutants**

(a) Schematic representation of allelic replacement by homologous recombination. (b) PCR amplification of chromosomal DNA from the parental strain and possible Δ*neiI* knockouts mutants using primers detailed in Appendix B, Table 9. Lanes represent the following: lane 1, molecular weight marker λIII; lane 2, mc²155 (1165bp); lane 3; forward primer only (negative control); lane 4, reverse primer only (negative control); lane 5, no DNA control; lane 6, DCO with wild type allele (1165bp); and lane 7, DCO with mutant allele (697bp). (c) The Southern blot analysis of the Δ*neiI* mutant. Chromosomal DNA from the parental strain, the SCO and the DCO was digested with *Asp*718 (A) and the fragments separated on an agarose gel. The digested DNA was probed with a 1.1kb fragment (shown as a red line in panel (a)). The expected sizes are as follows: lane 1, wild type mc²155 (4230bp, 2880bp); lane 2, Δ*neiI* SCO (9456bp, 2880bp, 1418bp); lane 3, Δ*neiI* DCOs (6643bp).(d) PCR amplification of chromosomal DNA from the parental strain and putative Δ*fpgI*Δ*fpgII*Δ*neiI* mutants, the primers used are detailed in Appendix B, Table 9. Lanes represent the following: lane 1: molecular weight marker λIII; lane 2, mc²155 (2.3kb); lane 3, forward primer only (negative control); lane 4, reverse primer only (negative control); lane 5, no DNA control; lane 6, DCO with mutant allele (1.8kb). (e) The Southern blot analysis of putative Δ*fpgI*Δ*fpgII*Δ*neiI* mutants, Southern blot analysis were performed as detailed in (c). Lane 1, wild type mc²155; lane 2, Δ*fpgI*Δ*fpgII*Δ*neiI* SCO 1; lane 3, Δ*fpgI*Δ*fpgII*Δ*neiI* DCOs

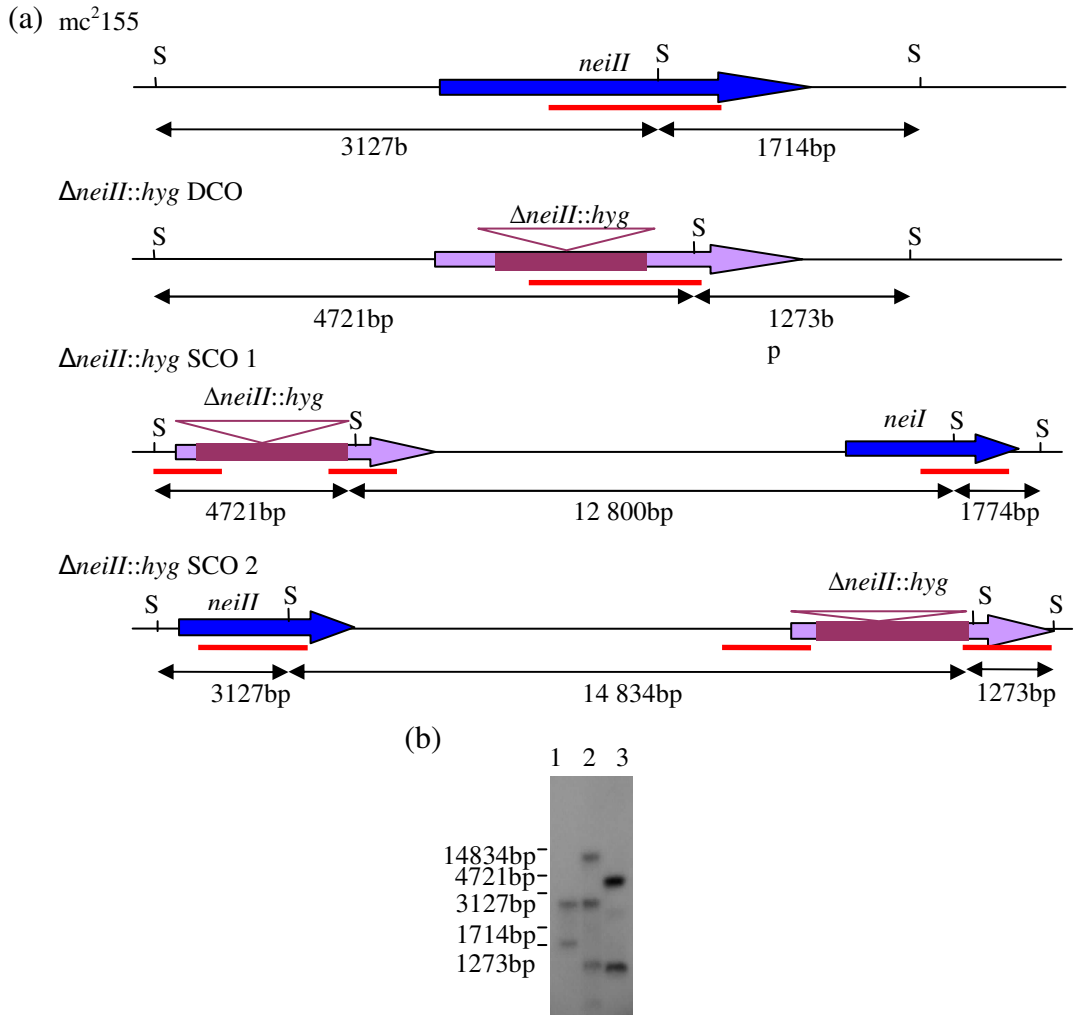



**Fig. 20. Genotypic characterization of the marked *neiII* region of *M. smegmatis* for the Δ*neiII* mutant**

(a) Schematic representation of allelic replacement by homologous recombination. (b) The Southern blot analysis of Δ*neiII*::*hyg* mutant. Chromosomal DNA from the parental strain, the SCO and the DCO was digested with *Sph*1 (S) and the fragments separated on an agarose gel. The digested DNA was probed with a 1.8kb fragment (shown as a red line in a). The expected sizes are as follows: lane 1, wild type mc²155 (3127bp, 1714bp); lane 2, Δ*neiII*::*hyg* SCO 2 (14831bp, 3127bp, 1273bp); lane 3, Δ*neiII*::*hyg* DCO (4721bp, 1273bp).

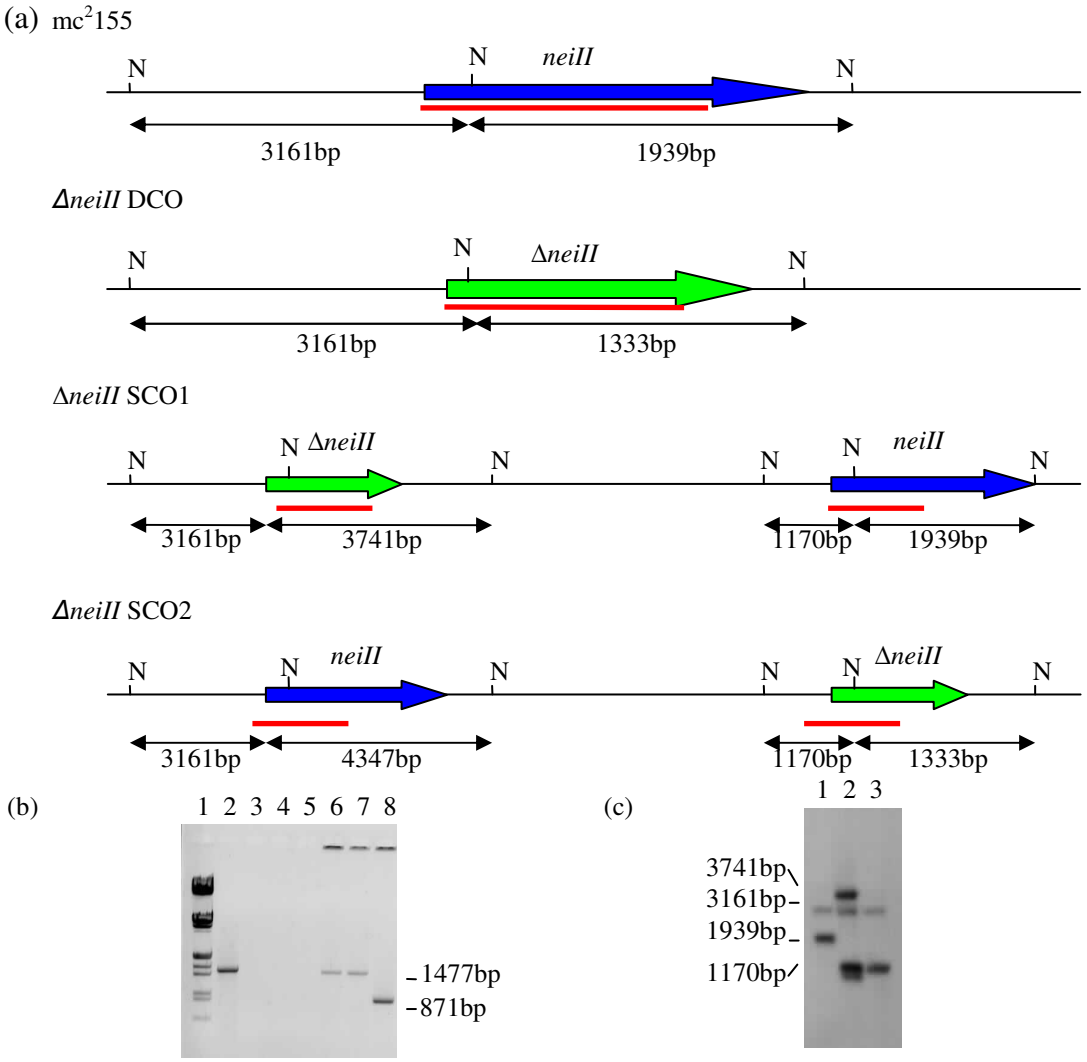



**Fig. 21. Genotypic characterization of *neiII* region of *M. smegmatis* for the double Δ*neiI*Δ*neiII* mutant**

(a) Schematic representation of allelic replacement by homologous recombination. (b) PCR amplification of chromosomal DNA from the parental strain and putative DCOs using primers detailed in Appendix B, Table 9. Lane 1, molecular weight marker λIII; lane 2, mc²155 (1477bp); lane 3, forward primer only (negative control); lane 4, reverse primer only (negative control); lane 5, no DNA control; lanes 6 and 7, wild type DCO (1477bp); lane 8, mutant DCO (871bp). (c) Southern blot analysis of the Δ*neiI*Δ*neiII* mutant. Chromosomal DNA from the parental, the SCO and the DCO was digested with *Nru*I (N) and the fragments separated on an agarose gel. The digested DNA was probed with a 1.8kb fragment (shown as a red line in a). Lane 1, wild type mc²155 (3161bp, 1939bp); lane 2, Δ*neiI* Δ*neiII* SCO 2 (4347bp, 3161bp, 1333bp 1170bp); lane 3, Δ*neiI* Δ*neiII* (3161bp, 1333bp)

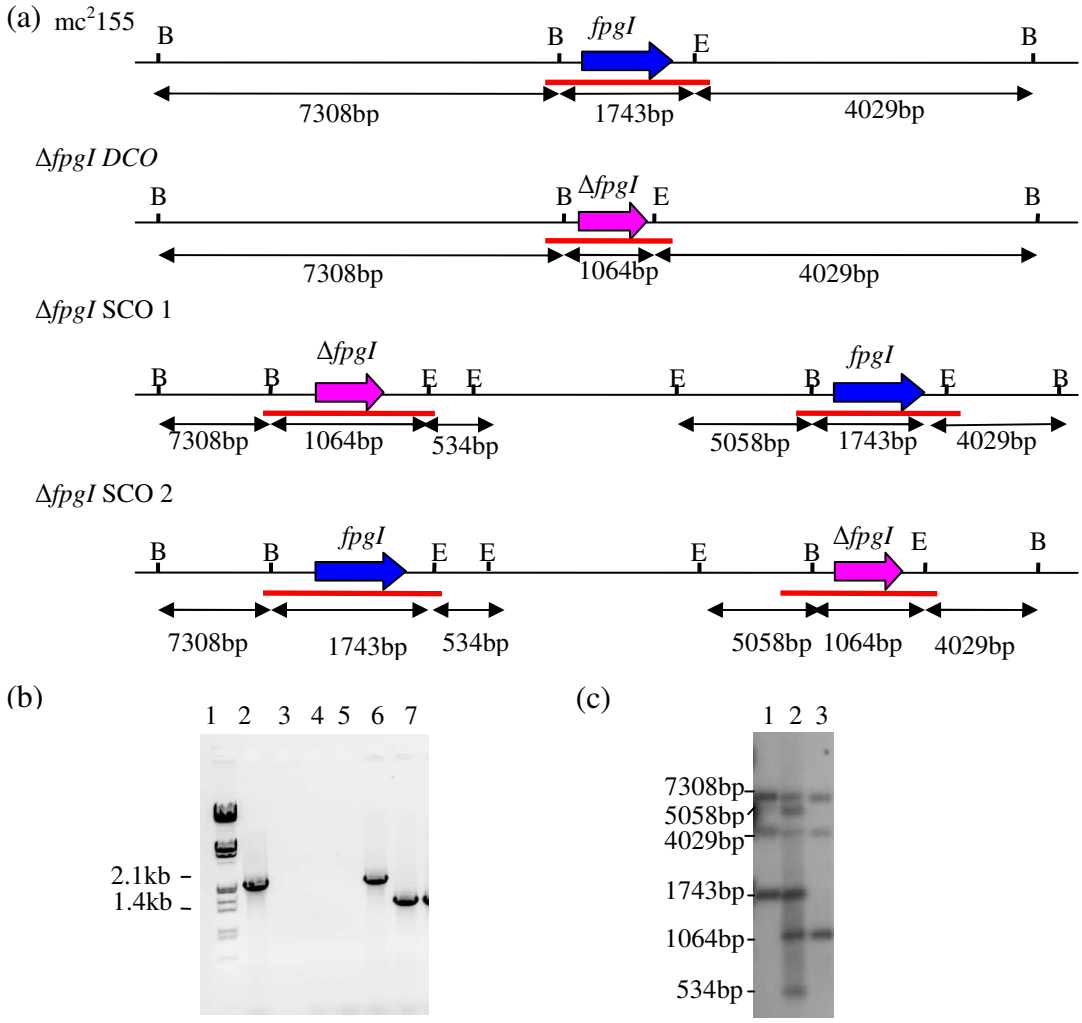



**Fig. 22. Genotypic characterization of unmarked *fpgI* region of *M. smegmatis* for the double Δ*neiI*Δ*neiII*Δ*fpgI* mutant**

(a) Schematic representation of allelic replacement by homologous recombination. (b) PCR amplification of chromosomal DNA from the parental strain and Δ*neiI*Δ*neiII*Δ*fpgI* mutants using primers detailed in Appendix B, Table 9. Lane 1: molecular weight Marker λIII, lane 2: mc²155 (2.1kb), lane 3: forward primer only (negative control), lane 4: reverse primer only (negative control), lane 5: no DNA control, lane 6: wild type revertant DCO (2.1kb), lane 7: mutant DCO (1.4kb). (c) The Southern blot analysis of the Δ*neiI*Δ*neiII*Δ*fpgI* mutant. Chromosomal DNA from the parental, SCO and DCO strains was digested with *Eco*RI and *Bam*HI, and probed with a 2kb fragment (shown as a red line in a). Expected sizes: lane 1, wild type mc²155 (7308bp, 4029bp, 1743bp); lane 2, Δ*neiI*Δ*neiII*Δ*fpgI* SCO (7308bp, 5058bp, 4029bp, 1743bp, 1064bp, 534bp); lane 3, Δ*neiI*Δ*neiII* Δ*fpgI* DCOs (7308bp, 4029bp, 1064bp).

## 3.4 Phenotypic characterisation of the single and double mutant strains

The single and double mutants were phenotypically characterised alongside the parental strain by comparing the following: (1) growth kinetics under normal culture conditions, (2) sensitivity to oxidative stress generated by $H_2O_2$, and (3) ability to repair DNA by measuring the rate and spectrum of spontaneous mutation to rifampicin resistance.

### 3.4.1 Growth kinetics under standard culture conditions

*M. smegmatis* strains were inoculated in MM media and the growth of the cultures measured at regular intervals by optical density measurements and viable cell counts (Fig. 23) over a 36h period, as described in section 2.7.1. No differences in growth kinetics were observed between the parental, single and double knockout mutant strains.

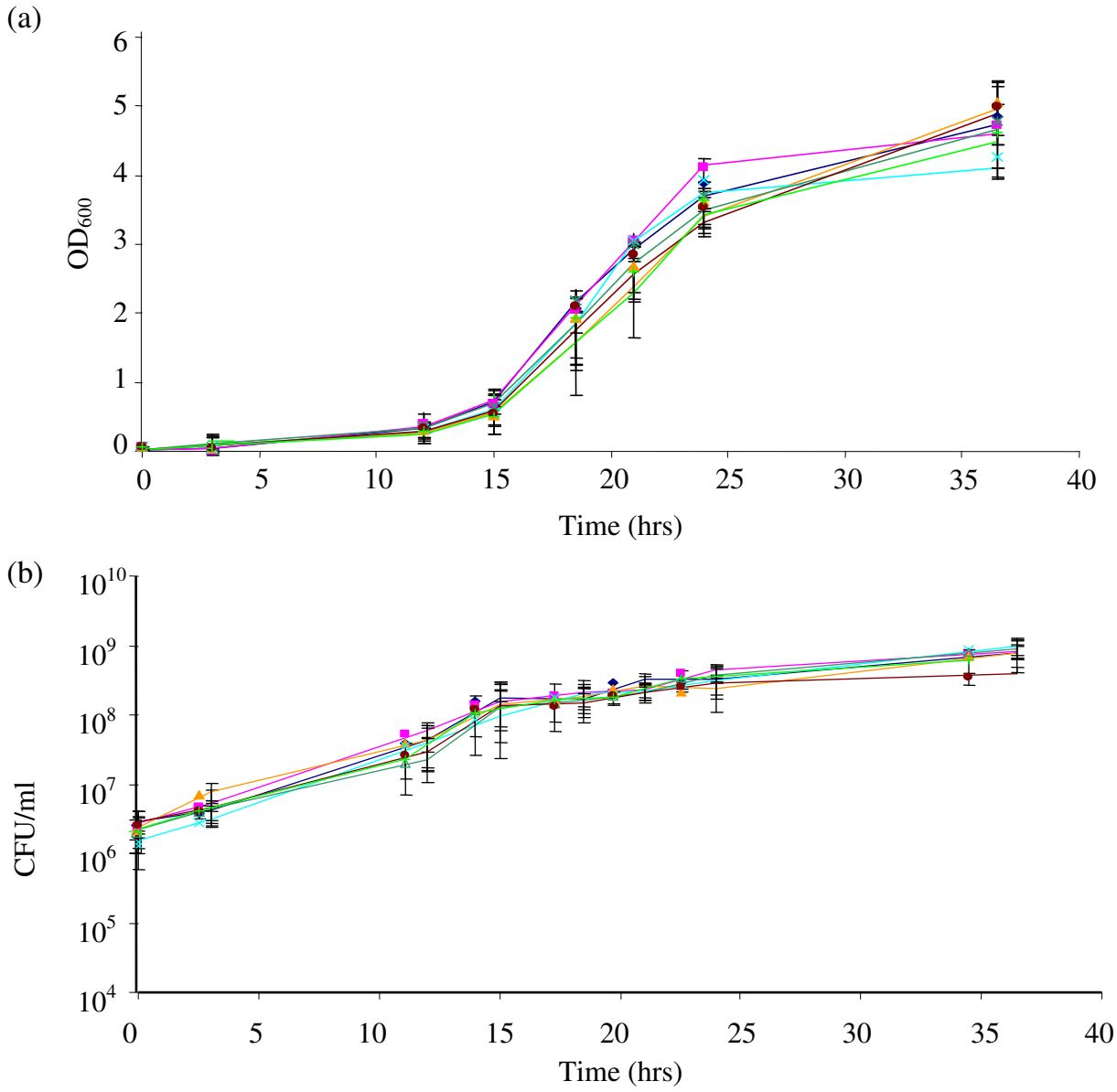



**Fig. 23. Comparative growth kinetics**

(a) Measurements by $OD_{600}$ and (b) viable cells counts. Data represent averages and standard deviations from three independent experiments. The strains were represented as follows: $mc^2155$ (—◆—); Δ*fpgI*::*hyg* (—■—); Δ*fpgII* (—▲—); Δ*neiI*::*hyg* (—×—); Δ*neiII*::*hyg* (—△—), Δ*fpgI*Δ*fpgII* (—●—); Δ*neiI*Δ*neiII*(—+—).

### 3.4.2 Sensitivity to oxidative stress as generated by hydrogen peroxide

In order to investigate the individual and combined roles of Fpg and Nei DNA glycosylases in the repair of oxidatively damaged DNA, the single and double mutants were assessed for cell viability and outgrowth under conditions of oxidative stress as generated by the addition of exogenous $H_2O_2$ to the various cultures.

#### 3.4.2.1 Effects of hydrogen peroxide on cell viability

In order to observe whether the removal of one or more *fpg* and *nei* affected the strains ability to withstand oxidative stress generated by $H_2O_2$, the optimal concentration and growth point at which to add $H_2O_2$ was determined. In previous studies in mycobacteria, the concentration of $H_2O_2$ used and the growth stage at which $H_2O_2$ was added varied considerably (Rosner and Storz, 1994; Sherman *et al.*, 1995; Gordhan *et al.*, 1996; Newton *et al.*, 2005). In these studies, 1-10 mM $H_2O_2$ was used to generate oxidative stress (Rosner and Storz, 1994; Sherman *et al.*, 1995; Gordhan *et al.*, 1996; Newton *et al.*, 2005). It was also observed that the susceptibility of *M. smegmatis* to $H_2O_2$ varied in a growth phase dependent manner, (Sherman *et al.*, 1995) suggesting that the optimal point at which to add $H_2O_2$ needed to be established. A concentration of 10 mM $H_2O_2$ was thus selected as a starting point for the present study.

Cultures of the parental strain in lag, log or stationary phase were treated with 10 mM $H_2O_2$ and allowed to grow for a further 6 h. At various time intervals over the 6 hour period, samples were removed from each culture and a dilution series was spotted onto MM to assess the killing caused by $H_2O_2$. This revealed that the lag-phase culture showed approximately 1 $\log_{10}$ killing after 10 min of treatment and after 1.5 h, complete killing was observed. The log-phase cultures showed a 3 $\log_{10}$ reduction in CFUs after 1.5 h, while the stationary phase cultures showed little killing over the 6 h period. Hence, for all further experiments, $H_2O_2$ was added to log phase cultures as during this phase of growth, the cells displayed sensitivity to $H_2O_2$ without rapid killing.

Further experiments included establishing the optimal concentration of $H_2O_2$ for use in the study. Log-phase wild type cultures were treated with various concentrations of $H_2O_2$ and the cell viability measured over 2.5 h. The results showed that concentrations of 5 mM and higher caused 2.5 log killing within 2 h of treatment (Fig. 24) while cultures treated with 2.5 mM $H_2O_2$ resulted in CFUs which remained consistent for approximately 1.5 h before declining. Therefore, for the purpose of this study, $H_2O_2$ was used at a concentration of 2.5mM for all further experiments.

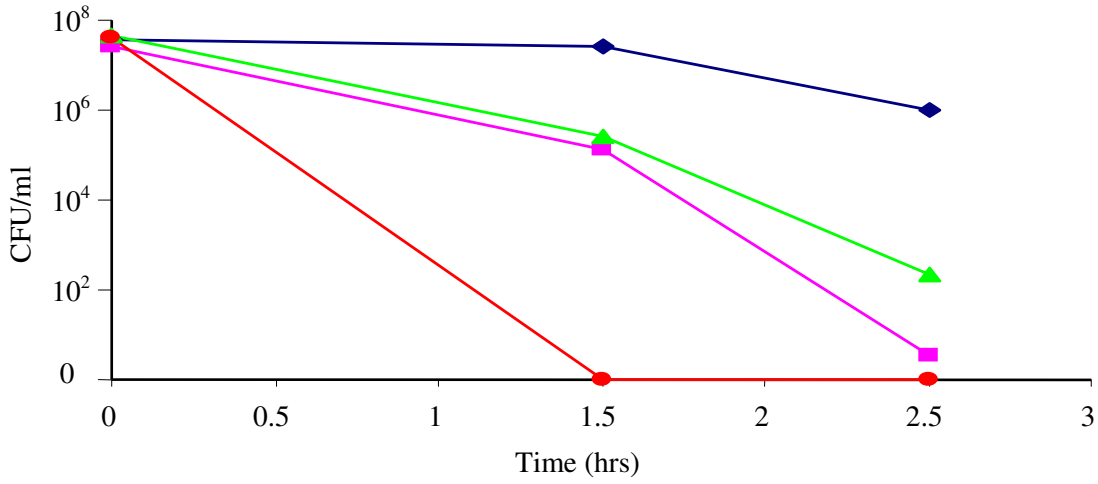



**Fig. 24. Assessment of the wild type strain with various concentrations of hydrogen peroxide.**
Log phase cultures treated with 2.5 mM (—●—), 5 mM (—▲—), 10 mM (—■—) and 20 mM(—●—) $H_2O_2$ and viable cell counts assessed over a 2.5h period.

The wild type, single and double mutants were tested under the optimised $H_2O_2$ treatment conditions established as described above. The results from several experiments showed that there were no significant differences in growth or killing between the parental, single or double knockout mutants when exposed to oxidative stress as generated by treatment with $H_2O_2$ at a concentration of 2.5 mM (Fig. 25). However, there were variations in the overall killing rates between separate experiments performed on different days. To test whether this variation was due to the type of media employed or the batch of $H_2O_2$ used, the experiments were repeated using different types of growth media (LB, Sauton's and MM) with different batches of $H_2O_2$ and all precautions were taken to keep conditions the same between experiments (temperature, initial concentrations, media, volumes, flasks). These experiments showed that the variation was independent of the type of media or the batch of $H_2O_2$ used (data not shown). Hence under conditions tested no increased sensitivity to $H_2O_2$ was observed in the *fpg* and *nei* knockout mutant strains when compared to the parental strain.

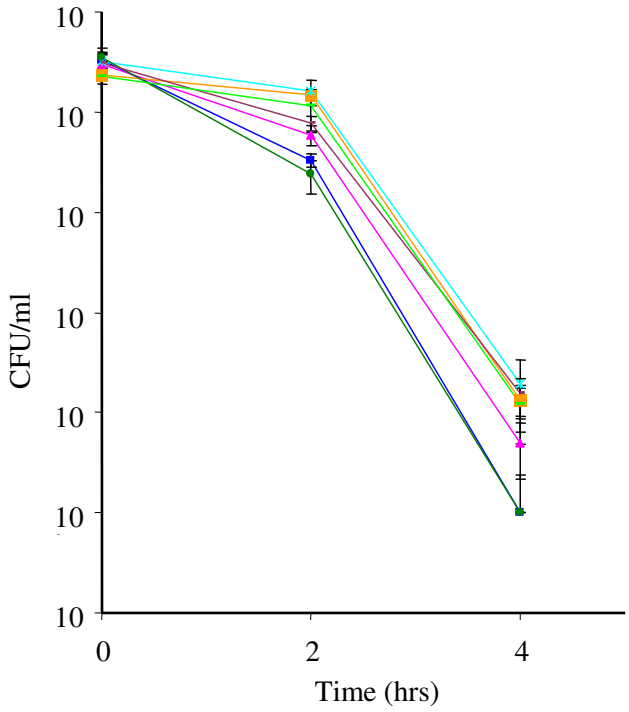



**Fig. 25. The effect of 2.5 mM hydrogen peroxide on the cell viability of log-phase wild type, single and double knockout mutant strains grown in Sauton's media**
The data represents a single experiment and error bars denote standard deviations between technical replicates. Students *t*-test analysis showed no significant differences between strains (*P*-values >0.1 for all points). Strains are represented as: $mc^2155$ (—◆—), Δ*fpgI*::*hyg* (—■—), Δ*fpgI* (—▲—), Δ*neiI*::*hyg* (—×—), Δ*neiII*::*hyg* (—△—), Δ*fpgI*Δ*fpgII* (—●—), Δ*neiI*Δ*neiII*(—+—).

### 3.4.2.2 Effects of hydrogen peroxide on outgrowth

During the course this study, another group independently reported the construction and characterisation of a hyg-marked Δ*fpgI* knockout mutant strain (Jain *et al.*, 2007). This group reported that their *fpgI* mutant was impaired for recovery after exposure to $H_2O_2$, as evidenced by the kinetics of outgrowth of a lag-phase culture of the mutant strain after treatment with 3 mM $H_2O_2$ (Jain *et al.*, 2007). This outgrowth following $H_2O_2$ treatment assay was therefore used to assess the *fpgI* mutant generated in this study alongside the wild type control. Lag-phase cultures of the parental and Δ*fpgI* mutant strain were treated with 3 mM $H_2O_2$ and monitored for growth as detailed in section 2.7.2.1.2. This assay, which was performed according to the published

procedure (Jain *et al.*, 2007), differed from the $H_2O_2$ assay described above (section 3.4.2.1) as the $H_2O_2$ was added to the culture during the lag phase, a slightly higher concentration of $H_2O_2$ (3 mM) was used, and the growth was determined only by optical density.

The results of this experiment (Fig. 31) showed that there was no recovery of either the wild type or the Δ*fpgI*::*hyg* mutant even after 8 days, suggesting that both cultures were completely killed by the 3 mM $H_2O_2$ treatment.

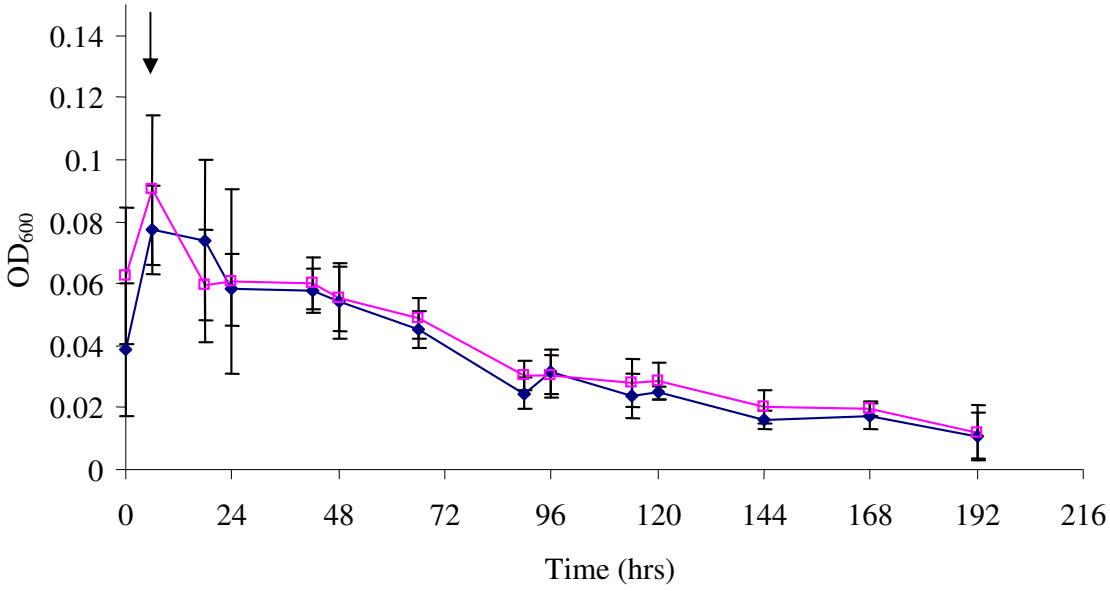



**Fig. 26. The ability of wild type and Δ*fpgI*::*hyg* to recover from hydrogen peroxide treatment**

Early log phase cultures of mc$^2$155 (—◆—) and Δ*fpgI*::*hyg* (—■—) were treated with $H_2O_2$ at the time point indicated by the arrow. The data are shown as averages and standard deviations from three independent experiments.

### 3.4.3 Effect of loss of *fpg* and/or *nei* gene function on mutational rates and spectra in *M. smegmatis*

Fpg and Nei glycosylases are DNA repair enzymes with predicted anti-mutator roles (Horst *et al.*, 1999). Therefore the effect of the loss of these glycosylases on spontaneous mutagenesis in *M. smegmatis* was assessed. The mutational target used for this study was resistance to the drug, rifampicin. A high level of rifampicin resistance in *M. smegmatis* results from mutations in a defined region of the *rpoB* gene, which encodes the β subunit of RNA polymerase – the rifampicin resistance-determining region (RRDR) (Musser, 1995; Telenti *et al.*, 1997; Ramaswamy and Musser, 1998; Karunakaran and Davies, 2000; Morlock *et al.*, 2000). Investigations analyzing the rate at which $rif^R$ mutants arose spontaneous during growth and the spectra of *rpoB* mutations found within the resistant mutants allowed comparisons to be made between the wild type and Fpg/ Nei glycosylase-deficient mutants.

#### 3.4.3.1 Mutation frequency

Mutation frequencies are a measure of the average number of mutants generated in a single culture. In this study, mutation frequencies were calculated as the ratio of the number of rifampicin mutants to the total number of cells, as described in section 2.7.3.1 (Martinez and Baquero, 2000; Rosche and Foster, 2000). Two separate experiments indicated that the Δ*fpgI*::*hyg* mutant had a higher mutation frequency compared to the wild type or the Δ*neiI*::*hyg* mutant (Fig. 27). The mutation frequencies showed variability between the two experiments (Fig. 27), but this was expected because mutations can occur randomly at any point during the growth of the cell. Hence, if a mutation occurs early during cell growth, after subsequent replications an increased number of that mutant will be represented on the plate, giving an apparently higher mutation frequency. Therefore, an apparently high mutation frequency might not, necessarily, be due to an increased number of

mutational events but, rather the result of an early mutational event. To accurately identify if a mutator phenotype is associated with the *fpgI* and *neiI* single and double knockout mutants, a more accurate mutation rate assessment was performed on the single and double knockout mutants.

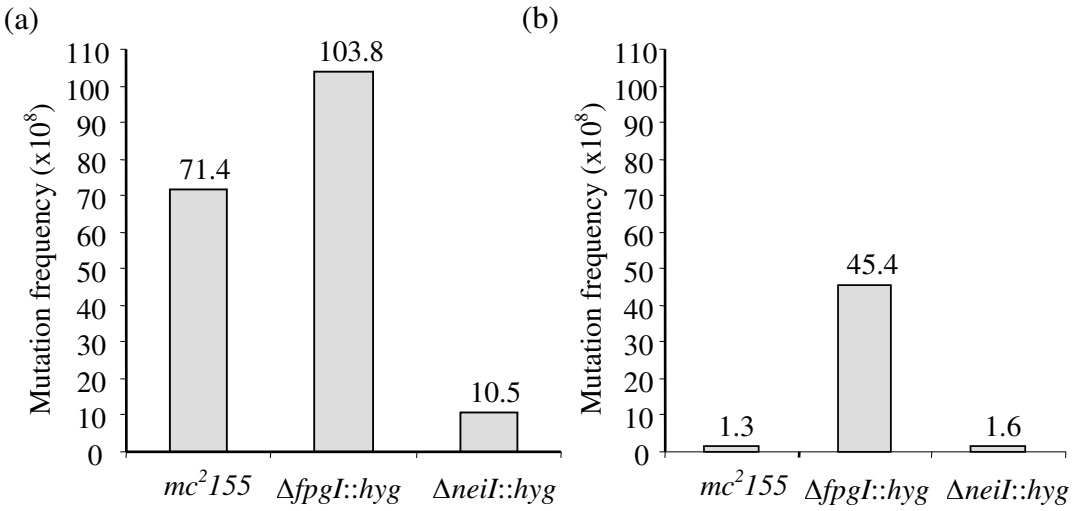



**Fig. 27. The spontaneous mutation frequencies of wild type, single *fpgI* and *neiI* knockout mutants as measured by rifampicin resistance.**
The data represents the results of two separate experiments shown in (a) and (b).

### 3.4.3.2 Mutation rate analyses

Mutation rate is a measure of the probability that a mutational event would occur in the lifetime of a cell. This is accurately measured by fluctuation assays. Fluctuation assays identify mutation rates by analyzing the range of mutation frequencies in a series of parallel cultures, and the final mutation rates are calculated by taking into account the number of cells (*Nt*-value) and the number of mutational events (*m*-value) occurring in these parallel cultures (Rosche and Foster, 2000), as described in section 2.7.3.2.2.

The precision and reproducibility of the mutation rate between separate assays is related to the size of the *m*-value as well as the total number of parallel tubes used within the assay. This relationship can be mathematically correlated and is illustrated in Fig. 28 which shows that the lower the probability of a mutational event occurring (i.e. *m*-value) the larger the number of parallel tubes needed for statistical analysis (Rosche and Foster, 2000). Therefore, the sample size (number of parallel cultures) necessary to generate precise and reproducible results needed to be ascertained. In the first set of assays for the parental, single and double knockout mutant strains, 25 parallel tubes were used to evaluate the *m*-values. The *m*-value ranged between 0.5 and 1.1, therefore in further experiments the number of parallel tubes was increased to 30.

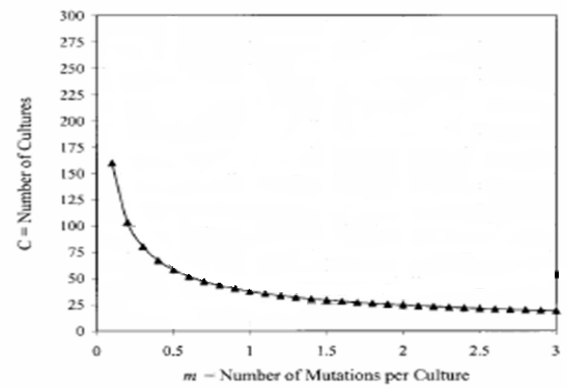



**Fig. 28. Graphical illustration of the number of cultures needed to ensure a precise mutation rate when using the MSS-likelihood algorithms.**
The *m*-value is represented on the *x*-axis with the corresponding number of parallel cultures needed on the *y*-axis (figure taken and adapted from Rosche and Foster, 2000).

The final calculation used to estimate the mutation rate is dependent on an accurate *Nt*-value. In the first set of fluctuation assays the *Nt*–value was calculated by measuring CFUs from every parallel sample of the parental and mutant strains (Fig. 29). Since the observed variation between the parallel samples was shown to be minimal (Fig. 29), for all further experiments, the *Nt*-values were calculated from only 5 cultures grown alongside the experimental cultures that were sampled for the presence of rifampicin-resistant mutants.

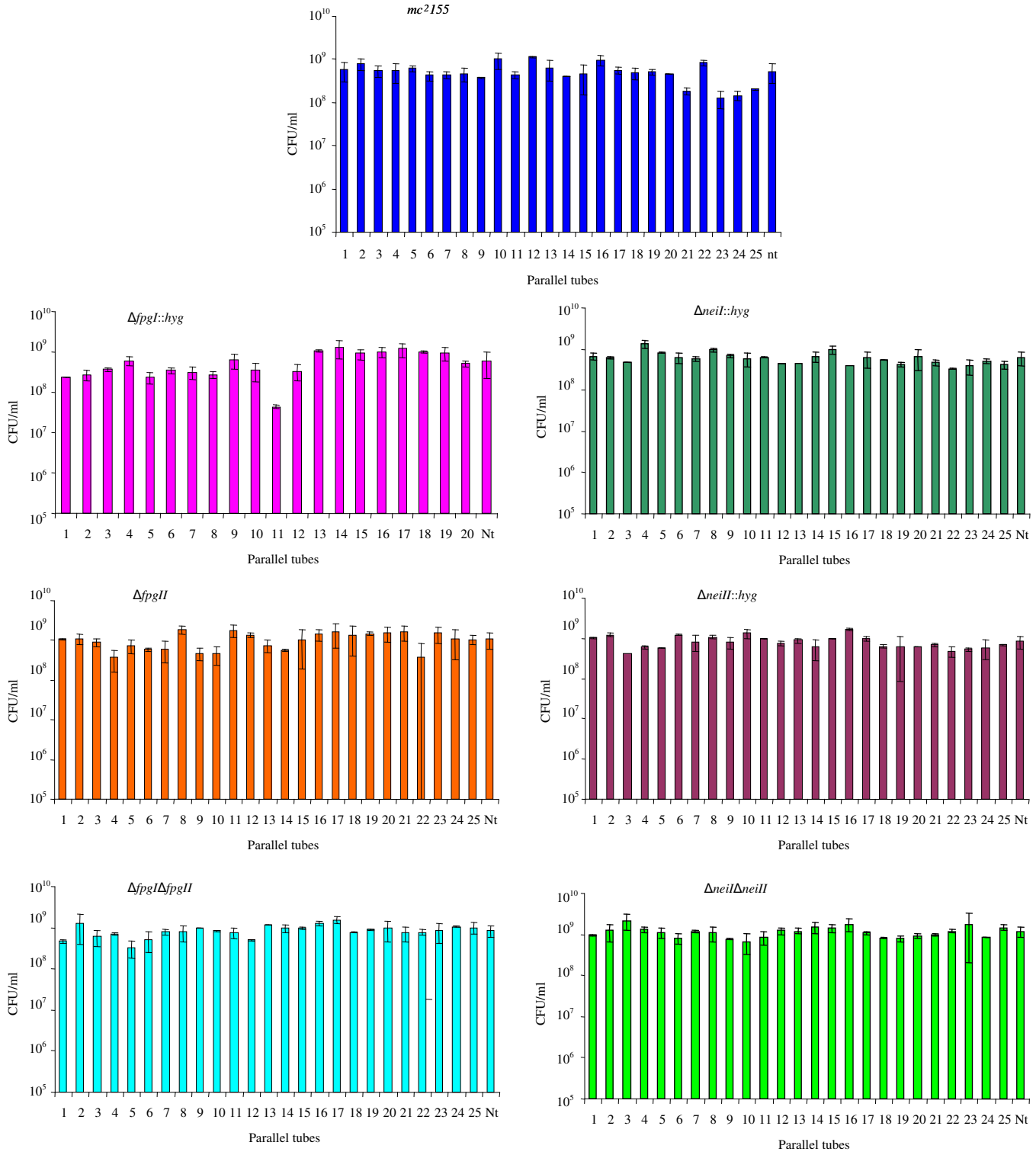



**Fig. 29. Estimation of *Nt*-values by measuring the average CFU in individual parallel cultures.**

Statistics allow for comparisons to be made between different strains and biological replicates, thereby giving a degree of confidence to the comparisons (*e.g*. standard deviations). However, due to the intrinsic sensitivity of the mutation rate to the *Nt*-value and the extrapolations used to calculate the mutation rate, it is not possible to calculate the differences between mutation rates with a degree of statistical confidence (Rosche and Foster, 2000). Nonetheless, if mutation rates are drastically dissimilar (>10-fold) one could conclude that a difference in mutation rates between the strains examined exists. Hence, fluctuation assays measuring the mutation rates for the parental, single and double knockout mutant strains were performed, the results of which are shown in Table 4.

**Table 4. Spontaneous mutation rates resulting in rifampicin resistance of the parental, single and double knockout mutant strains as determined by fluctuation analysis.**

| Strain | Mutation rates calculated after each fluctuation assay ($\times 10^{-10}$) | | | | | Average Mutation Rate ($\times 10^{-10}$) | Fold change relative to mc²155 |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | |
| mc²155 | 9.8 | 5.9 | 1.3 | 1.2 | N/P | 4.5 | 1.00 |
| Δ*fpgI* | 23 | 3.2 | 0.8 | 10 | 7.2 | 8.9 | 2.0 |
| Δ*fpgII* | 1.0 | 13 | 0.6 | N/P | N/P | 5.0 | 1.1 |
| Δ*fpgI*Δ*fpgII* | 4.8 | 0.7 | 0.9 | 25 | N/P | 7.8 | 1.7 |
| Δ*neiI* | 5.1 | 0.8 | 5.0 | N/P | N/P | 3.6 | 0.8 |
| Δ*neiII* | 4.3 | 2.3 | 6.0 | 7.5 | N/P | 5.1 | 1.1 |
| Δ*neiI*Δ*neiII* | 13 | 11 | 3.0 | N/P | N/P | 8.8 | 2.0 |

* N/P Not performed

In Table 4**,** the average mutation rates were calculated from three to five independent fluctuation assays for the parental and mutant strains. However, this analysis suggested that no significant differences in mutation rates were observed.

### 3.4.3.3 Mutation spectrum in *M. smegmatis rpoB* mutants

The parallel cultures used in fluctuation assays minimize the probability of $rif^R$ colonies originating from a common source (Rosche and Foster, 2000). Therefore, genomic DNA from 13-14 $rif^R$ colonies isolated off several plates generated from a number of fluctuation assays conducted on the parental and mutants strains was PCR-amplified across the RRDR region and the amplicon sequenced to identify the types of lesions that resulted in rifampicin resistance. To minimise the introduction of errors during PCR amplification, the high fidelity polymerase Expand (Roche Biochemicals) was used.

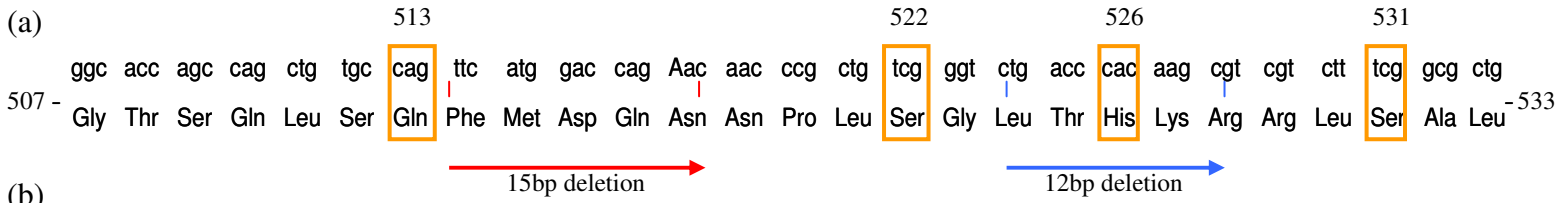
Detailed sequence analysis of the specific types of DNA mutations in the RRDR region is shown in Table 5, while the resultant amino acid changes and deletions which occurred in the RRDR region is shown in Fig. 30. Table 5 shows that a 16% increase in C→T (or G→A) transitions in the Δ*fpgI* mutant was observed when compared to the parental strain. This increase was absent in the single Δ*fpgII* mutant, but the double *fpg* knockout mutant showed a 27% increase. Similarly a 22% increase in C→T transitions (or G→A) was observed in Δ*neiI*, which was further amplified to 38% in the double *nei* knockout mutant. An increase of 17% in C→A transversions was observed for the Δ*neiII* mutant compared to the parental strain, however no additive increase in mutations was observed for the double *nei* knockout mutant.

**Table 5. The different types of DNA mutations identified within the RRDR region**

| Mutation | Percentage (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | $mc^2155$ (12/13)[a] | Δ*fpgI* (12/13)[a] | Δ*fpgII* (13/14)[a] | Δ*fpgI* Δ*fpgII* (14/14)[a] | Δ*neiI* (14/14)[a] | Δ*neiII* (14/14)[a] | Δ*neiI* Δ*neiII* (11/14)[a] |
| C→G/A→T | 25 | 25 | 31 | 23 | 14 | 43 | 20 |
| C→G/T→A | 42 | 58 | 46 | 69 | 64 | 28.5 | 80 |
| C→G/G→C | 8 | 8 | 15 | 0 | 7 | 0 | 0 |
| A→T/G→C | 25 | 0 | 0 | 8 | 14 | 28.5 | 0 |
| A→T/C→G | 0 | 8 | 8 | 0 | 0 | 0 | 0 |

[a]Proportion of samples that had mutations in the RRDR region.

(a)




(b)

| Mutation | Alteration | mc$^2$155 (13) | Δ*fpgI* (13) | Δ*fpgII* (14) | Δ*fpgI* Δ*fpgII* (14) | Δ*neiI* (14) | Δ*neiII* (14) | Δ*neiI* Δ*neiII* (14) |
|---|---|---|---|---|---|---|---|---|
| cag→aag | Gln513 →Lys | 23.1 | 23.1 | 28.6 | 21.4 | 14.3 | 42.7 | 14.4 |
| tcg→ ttg | Ser522 →Leu | - | 7.4 | 7.1 | 7.1 | - | - | - |
| cac→cgc | His526→Arg | 23.1 | - | - | 7.1 | 14.3 | 28.5 | 21.4 |
| cac→tac | His526→Tyr | 23.1 | 30.8 | 7.1 | 7.1 | 28.6 | 7.1 | 7.1 |
| cac→ccc | His526→Pro | - | 7.7 | 7.1 | - | - | - | - |
| cac→gac | His526→Asp | - | - | 7.1 | - | - | - | - |
| tcg →ttg | Ser531→Leu | 15.4 | 15.4 | 28.6 | 50 | 35.7 | 21.4 | 28.6 |
| tcg →tgg | Ser531→Trp | 7.7 | 7.4 | 7.1 | - | 7.1 | - | - |
| 15bp deletion | Phe514 - Asn518 | - | - | - | 7.1 | - | - | - |
| 12bp deletion | Leu524 - Arg528 | - | - | - | - | - | - | 7.1 |
| Unknown | | 7.7 | 7.4 | 7.1 | - | - | - | 21.4 |

**Fig. 30. Amino acid changes in the RRDR region of the *rpoB* gene in rifampicin resistant mutants isolated from wild type, single and double knockout mutants.**

(a) The codon and corresponding amino acid sequence of the RRDR region. Areas where mutations occurred are highlighted in the boxes and areas where deletions occurred are indicated by arrows and lines. (b) Table indicating the specific mutations, alterations and percentage frequency of the mutations for the single and double mutants compared to the parental strain. The numbers of mutants isolated from each of the mutant and parental strains are indicated in brackets and those mutations which did not map to the RRDR region are labelled as unknown.

# 4. Discussion and Conclusion

*M. tuberculosis*, the causative agent of tuberculosis, is exposed to considerable DNA-damaging, oxidative stress during its infection cycle. However, it is still able to, not only cause primary infections but, is also, after indefinite periods of latency, able to re-emerge and cause secondary disease (Glickman and Jacobs, 2001; Kaufmann, 2005; Russell, 2007). Hence, *M. tuberculosis* must have mechanisms that can repair oxidatively damaged DNA to allow for its survival and further infection.

Fpg and Nei glycosylases are members of a structurally related family of DNA glycosylases involved in the identification, excision and repair of oxidatively-damaged DNA. Genome analysis performed in this study allowed for the identification of four putative *fpg* and *nei* DNA glycosylase-encoding genes in *M. tuberculosis* and *M. smegmatis.* Bioinformatic analysis confirmed that these *M. tuberculosis* and *M. smegmatis* genes were homologues. Further, it was observed that the *M. smegmatis fpg* and *nei* encoded proteins had all maintained the domains necessary for protein function, and showed a strong resemblance to previously characterised Fpg and Nei DNA glycosylases in *E. coli*, *T. thermophilus*, *G. stearothermophilus* and *L. lactis* (Gilboa *et al.*, 2002; Zharkov *et al.*, 2003; Golan *et al.*, 2005). Therefore, these glycosylases were considered to have a functional role in their host organism.

The presence of homologues of functional glycosylases in mycobacteria, as well as the high level of substrate overlap between these *fpg* and *nei* encoding genes, indicates possible redundancy of these genes. Most organisms have maintained one member of this family but, rarely both. It has, therefore, been suggested that the retention and the redundancy of these BER enzymes in intracellular organisms, such as *M. tuberculosis* and *S. typhimurium*, might be a deliberate adaptation that has

allowed increased survival of these pathogens when exposed to oxidative stress (Suvarnapunya *et al.*, 2003).

In order to expand this hypothesis and to identify the individual and/or combined role(s) of the Fpg and Nei glycosylases in mycobacterial DNA repair, single, double and triple knockout mutant strains were generated in *M. smegmatis* by homologous recombination and phenotypically assessed. No difference in growth kinetics was observed between the parental, single and double knockout mutant strains. These results concur with previously published data, where a single *fpgI* mutant of *M. smegmatis*, and *fpg* and *nei* knockout mutants of *S. typhimurium* showed no growth defects under normal culture conditions (Suvarnapunya *et al.*, 2003; Suvarnapunya and Stein, 2005; Jain *et al.*, 2007).

When the parental, single and double knock-out mutants were treated with 2.5 mM $H_2O_2$ no difference in killing rates was observed, however, variations in killing rates between biological replicates was noted. This variation could be due to the decomposition of $H_2O_2$, since $H_2O_2$ spontaneously decomposes into water and oxygen, in a manner dependent on temperature, initial concentration and pH (Brown *et al.*, 2000). Therefore, although attempts were made to ensure conditions between experiments remained the same, there might have been differences in the effective concentration of $H_2O_2$ that the cells were exposed to between biological replicates. However, since the single and double knockout mutants consistently behaved as the parental strain, the results reported in this study suggested that the loss of one or two *fpg* and *nei* genes did not affect the ability of *M. smegmatis* to recover from oxidative stress generated by 2.5 mM $H_2O_2$. Similarly, in *E. coli* and *S. typhimurium,* no increased sensitivity to $H_2O_2$ was observed in *fpg* and *nei* mutants (Asad *et al.*, 1995; Saito *et al.*, 1997; Alhama *et al.*, 1998; Suvarnapunya *et al.*, 2003) and a slight increase in sensitivity to $H_2O_2$ was observed, in *E. coli* only with the double *nei nth* mutant (Saito *et al.*, 1997; Suvarnapunya *et al.*, 2003).

The single *fpgI* mutant was tested for its ability to recover from oxidative stress as reported by Jain *et al.* (Jain *et al.*, 2007). Their study showed that lag-phase wild type cultures took 12 h to recover after exposure to 3mM $H_2O_2$ before entering log phase, whereas the *fpgI* mutant entered log phase after 48 h (Jain *et al.*, 2007). However, the Δ*fpgI*::*hyg* mutant strain generated in this study together with the parental strain showed no recovery even after 8 days, suggesting that both cultures were unable to tolerate the oxidative stress generated by the $H_2O_2$. The reason for the differences between the findings reported in this study and those described by Jain *et al.* are currently unknown and warrants further investigation. Additional studies could include phenotypic analysis of the mutants under different sources of oxidative stress, for example ionizing radiation or the superoxide generator plumbagin. Ionizing radiation induces similar DNA damage to oxidative stress and the mechanisms used in DNA repair are the same as those used in oxidatively damaged DNA (Blaisdell and Wallace, 2001; Suvarnapunya *et al.*, 2003; Suvarnapunya and Stein, 2005). The free-radicals produced in this manner are of a consistent amount and do not spontaneously decompose (Henle and Linn, 1997). Plumbagin, on the other hand, releases a superoxide radical via a specific and controlled enzymatic interaction in a non-spontaneous, reliable manner. Plumbagin has also been shown to cause oxidative stress capable of killing *M. smegmatis* (Farr *et al.*, 1985; Walkup and Kogoma, 1989; Rawat *et al.*, 2004). The resulting damage caused by ionizing radiation and plumbagin could be of a more consistent concentration as opposed to those generated by $H_2O_2$ and therefore may not show the variable killing rate as observed for $H_2O_2$.

Two different methods were used in this study to assess spontaneous mutagenesis in the glycosylase-depleted strains: mutation frequencies and mutation rates. Limited mutation frequency analysis suggested an elevated mutation frequency for the *fpgI* mutant, which is concordant with the 3.8-fold increase observed by Jain *et al.* for an independently generated *M. smegmatis* mutant lacking the *fpgI* gene (Jain *et al.*,

2007). However, mutation frequencies do not accurately reflect the number of mutations incurred by a specific cell. Therefore, fluctuation assays were performed to accurately ascertain changes in mutation rates between the parental and mutant strains. The fluctuations assays showed that there was no significant variation in mutation rates between the parental, single or double knock-out mutants.

The mutation rates data reported in this study can be broadly compared to the mutation frequency data reported in several other studies. *E. coli fpg* and *nei* mutants and *S. typhimurium nei* knock-out mutants did not display any increase in the frequencies of rifampicin mutations (Jiang *et al.*, 1997; Blaisdell *et al.*, 1999; Fowler *et al.*, 2003), and mutator phenotypes only became apparent in *E. coli* but not in *S. typhimurium*, when the *nth* and/or *mutY* encoding genes were inactivated in addition to the *fpg* and *nei* encoding genes (Jiang *et al.*, 1997; Blaisdell *et al.*, 1999; Suvarnapunya *et al.*, 2003; Suvarnapunya and Stein, 2005).

Although the glycosylase-deficient mutants analysed in this study did not show obvious mutator phenotypes, sequence analysis of the RRDR region of $rif^R$ colonies obtained from the parental, single and double mutants did suggest a trend towards an increased frequency of C→T transitions in the *fpg* and *nei* mutants. However, the sample size for this analysis was small suggesting that the data could be biased. Hence, any future study would need to include a larger sample set of $rif^R$ colonies before any significant conclusions can be made. The first sign of a weakened DNA repair network is an increase in C→T (G→A) transitions as these mutations are the most abundantly occurring mutations under normal conditions (Wang *et al.*, 1998; Wallace, 2002). These transition mutations originate from lesions in cytosine on the sense, or lesions in guanine on the anti-sense strand (Wang *et al.*, 1998; Wallace, 2002). Since the primary pro-mutagenic substrates for Fpg and Nei are guanine and cytosine lesions respectively, the trend of an increased frequency of C→T transition mutations observed in the *fpg* and *nei* deletion mutants from the limited sample size

analysed in this study, was as predicted (Jiang *et al.*, 1997; Krokan *et al.*, 1997; Wang *et al.*, 1998; Wallace, 2002; Wallace *et al.*, 2003).

In *E. coli* studies uniquely designed *lac* genes have been used as neutral mutational targets for recognising the types of mutations that occur in the absence of a wide range of DNA repair genes, including *fpg* and *nei*. The *lac* genes were altered at a unique glutamic acid codon, so that only a precise transversion or transition allowed for the reversion of a $lac^-$ phenotype to a $lac^+$ phenotype (Cupples and Miller, 1989). Using this method *E. coli fpg* deletion mutants showed an increase in $lac^+$ reversions when the erroneous base needed to be altered from G→T, it however, did not increase G→C transversions (Cabrera *et al.*, 1988; Michaels *et al.*, 1991; Michaels *et al.*, 1992; Matsumoto *et al.*, 2001; Fowler *et al.*, 2003). The *nei* deletion mutants showed no increase in reversion of C→T, nor G→C (Saito *et al.*, 1997; Matsumoto *et al.*, 2001). Therefore, the deletion of *fpg* and *nei* genes in *M. smegmatis* did not give comparable mutations as seen in *E. coli fpg* and *nei* deletion mutants. These differences could be due to the dissimilar systems used to assess mutations in *E. coli* and *M. smegmatis*, as mutations in the *lac* reporter system does not effect the survival of the organism while mutations in *rpoB* gene could. Further there might be slightly different functions between the DNA glycosylases in the two organisms, or possibly due to the different complement of back-up DNA repair enzymes in *E. coli* and *M. smegmatis* which could influence the recognition and repair of damaged bases.

The sequencing data was extended to analyse the codon and amino acid changes in the RRDR region. Although the DNA from a few colonies showed deleted regions, mutations were predominantly single polymorphisms resulting in changes at His526 (32.5%), Ser531 (32%) and Gln513 (23.9%), which corresponded to previously reported mutations in *M. smegmatis* and *M. tuberculosis* (Musser, 1995; Karunakaran and Davies, 2000; Warner, 2005; Jain *et al.*, 2007). Ninety four percent of the total $rif^R$ colonies sequenced in this study showed mutations in the RRDR

region, which was in agreement with previous studies (Musser, 1995; Ramaswamy and Musser, 1998; Warner, 2005). In contrast the *M. smegmatis fpgI* knockout mutant of (Jain *et al.*, 2007) showed that only 30-50% of the $rif^R$ colonies had mutations within the RRDR region (Jain *et al.*, 2007). The large differences in the percentages of $rif^R$ colonies could be due to the different concentrations of rifampicin used. The level of antibiotic resistance ranges from weakly to strongly resistant between individual resistors, and resistance can be caused not only by genotypic change but by phenotypic adaptations as well. The rifampicin MIC of *M. smegmatis* ranges between 4-32 μg/ml (Hetherington *et al.*, 1995; Alexander *et al.*, 2003) therefore, by selecting on a concentration close to the MIC the chances of selecting non-genotypic resistors is increased. Jain *et al.* used 50 μg/ml rifampicin compared to 200 μg/ml used in this and other studies (Karunakaran and Davies, 2000; Boshoff *et al.*, 2003; Warner, 2005). Therefore, by selecting on 200μg/ml, even though the possibility of excluding relatively low-level resistant mutants is increased, and those selected are more likely to have mutations in the RRDR region. A low level of rifampicin resistance in *M. smegmatis* is associated with ADP-ribosyl transferase activity where rifampicin is inactivated by ribosylation (Quan *et al.*, 1997). Although this was not established in the Jain *et al.* study, there was a chance that the $rif^R$ colonies that did not have mutations in the RRDR region could have their resistance associated with ADP-ribosyl transferase activity. Furthermore, the 42 $rif^R$ colonies sequenced in the RRDR by Jain *et al.* showed that rifampicin resistance in 26% of the colonies was caused by C→T transitions as compared to 36% in the parental strain and that an increase (from 24%-46%) in A→G (or T→C) transversion had occurred when *fpgI* was absent (Jain *et al.*, 2007). The lack of correlation between the two studies could be attributed to the different concentrations of rifampicin used or possibly due to deviations in the methods of isolating mutants. In this study the colonies were picked after fluctuation assays whilst the colonies tested by Jain *et al*. were from rifampicin frequency assays.

Guanine lesions resulting in G→T transversion are known to occur at a high frequency, are highly mutagenic and the pro-mutagens generated are substrates recognized by both Fpg and Nei DNA glycosylases (Michaels and Miller, 1992; Hazra *et al.*, 2001; David *et al.*, 2007). Hence, one would expect an increase in this type of transversion in the *M. smegmatis fpg* and *nei* deletion mutants. However, this was not observed under the conditions tested, which is possibly due to the presence of other DNA repair enzymes able to prevent these transversions, such as Nth and components of the GO system (Krokan *et al.*, 1997), or that these mutations in the RRDR were not well tolerated and therefore the resultant mutants were non viable.

The absence of phenotypic variation in mutants missing one or two DNA glycosylases could be due to the functional redundancy among the Fpg and Nei homologues, as well as the presence of damage avoidance mechanisms, detoxification pathways and/or a number of other DNA repair mechanisms present in *M. smegmatis*. These repair enzymes share a large potential overlap in substrate specificities and hence have the possible ability to compensate for the loss of the glycosylases (Blaisdell *et al.*, 1999; Hazra *et al.*, 2000; Hazra *et al.*, 2001; Matsumoto *et al.*, 2001; Wallace *et al.*, 2003; Wiederholt *et al.*, 2005). It also suggests that an *in vitro* phenotype might only become apparent once all four *fpg* and *nei* glycosylases are inactivated and possibly only after removal of other DNA repair enzymes such as *nth* and/or *mutY*. Hence, this study will need to be extended by constructing a quadruple mutant lacking all four *fpg/nei* genes and phenotypically characterizing the triple and quadruple knockout mutants under identical conditions as for the single and double mutants.

Expression data on each of the DNA repair genes could also be informative. The Jain *et al*. study suggested that *fpgI* was expressed under normal growth conditions (Jain *et al.*, 2007), but not much is known about the expression of *fpgII*, the *nei* glycosylases, and the *nth* glycosylase in *M. smegmatis*. Therefore, the panel of

mutants generated in this study is a rich resource for assessing the expression levels of the remaining glycosylases by RT-PCR. Such studies could provide insight into the possible compensatory and or redundant role(s) of these genes thus broadening our understanding of the function of these glycosylases in the overall DNA repair network in *M. smegmatis*. Since *M. tuberculosis* has similar Fpg/Nei homologues, the observations and conclusions made in the non-pathogenic fast growing *M. smegmatis,* can ultimately be useful to inform and guide the investigation of these glycosylases in *M. tuberculosis*.

# 5. Appendices

## 5.1 Appendix A

### 5.1.1 Culture media, supplementation and antibiotic stock solutions

All media was made in 1 litre of deionised water and sterilized by autoclaving at 121℃ for 10-20 min.

| | |
|---|---|
| Luria-Bertani broth (LB) | 10 g tryptone, 10 g NaCl, 5 g yeast |
| Luria-Bertani agar plates (LA) | 10 g tryptone, 10 g NaCl, 5 g yeast extract, 15 g agar |
| 2xTY media | 16 g tryptone, 10 g yeast extract, 5 g NaCl |
| Sauton's media | 4 g asparagines, 0.5 g magnesium sulphate, 2 g citric acid, 0.5 g potassium dihydrogen orthophosphate, 0.05g ammonium ferric citrate, 48 ml glycerol, pH 7.2 |
| Middlebrook 7H10 minimal media plates (MM) | 19 g 7H10 powder, 5 ml glycerol, 1.1 M glucose, 1.46 M salt |
| Middlebrook 7H9 minimal media (MM) | 4.7 g 7H9 powder, 2 ml glycerol, 1.1 M glucose, 1.46 M salt |

**Table 6. Antibiotic and supplement stock solution.**

| | Stock solution | Solvent |
|---|---|---|
| Antibiotic | | |
| Ampicillin | 100 mg/ml | 50% $dH_2O$<br>50% ethanol |
| Kanamycin | 50 mg/ml | $dH_2O$ (filter sterilized) |
| Hygromycin | 50 mg/ml | $dH_2O$ (filter sterilized) |
| Rifampicin | 100 mg/ml | dimethyl formamide |
| Supplement | | |
| X-gal | 20 mg/ml | dimethyl formamide |
| IPTG | 200 mg/ml | $dH_2O$ (filter sterilized) |
| Sucrose | 75% | $dH_2O$ (autoclaved) |
| Tween 80 | 25% | $dH_2O$ (filter sterilized) |
| $H_2O_2$ | 8 mM | |

### 5.1.2 DNA extraction solutions

| | |
|---|---|
| TE buffer | 10 mM tris-HCl , 1 mM EDTA, pH 8 |
| CTAB solution | 10% CTAB, 0.7 M NaCl |
| TENS buffer | 10 mM tris-HCl pH 7.5, 1 mM EDTA, 0.1 M NaOH, 0.5% SDS |
| Solution I | 0.5 M glucose, 0.5 M EDTA, tris-HCl pH 8 |
| Solution II | 10 M NaOH, 10% SDS |
| Solution III | 5 M potassium acetate, glacial acetic acid |

### 5.1.3 Agarose gel electrophoresis solutions

| | |
|---|---|
| TAE buffer | 1 mM EDTA, 40 mM tris-acetic acid pH 8.5 |

### 5.1.4 DNA molecular weight marker

Marker λIII

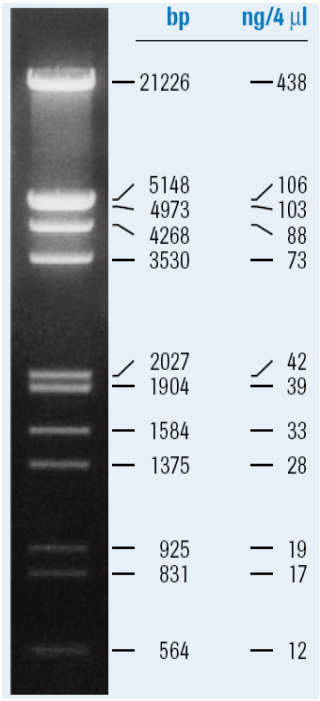



**Fig. 31. The DNA molecular weight marker λIII used in this study**
Supplied by Roche Biochemicals.

### 5.1.5 **Southern blot solutions**

Electroblotting

| | |
|---|---|
| Depurination solution | 0.25 M HCl |
| Denaturation solution | 0.5 M NaOH; 1.5 M NaCl |
| TBE buffer | Tris-borate-EDTA pH 8.0, Sigma |

Hybridization and radioactive labelling solutions

| | |
|---|---|
| Pre-hybridization solution | 0.5% SDS; 6 × SSC; 5 × Denhardt's solution; 50% de-ionized formamide |
| G-25 Sephadex columns | 100 g of Sephadex in 500ml TAE buffer |

Radioactive wash solutions

| | |
|---|---|
| Solution I | 2 × SSC, 0.1% SDS |
| Solution II | 0.5 × SSC, 0.1% SDS |
| Solution III | 0.1 × SSC, 0.1% SDS |
| Solution IV | 0.1x SSC, 1% SDS |

Non-radioactive washing solutions

| | |
|---|---|
| Solution I | 2x SSC, 0.1% SDS |
| Solution II | 0.5x SSC, 0.1% SDS |

Immuno-detection and chemiluminescence solutions

| | |
|---|---|
| CSPD | Disodium 2-chloro-5-(4-methoxyspiro (2-dioxetane-3,2 (2-dioxetane-3,2'-(5'-chloro)-tricyclo[3.3.1.1. 3, 7. ]decan)-. 4-yl)-1-phenyl phosphate |
| Washing buffer | 0.01 M maleic acid, 0.015 M NaCl, 0.3% triton X-100, pH 7.5 |
| Blocking solution | 0.01 M maleic acid, 0.015 M NaCl, 10 × blocking solution supplied, pH 7.5 |
| Antibody solution | 25 ml blocking solution, 1 µl of anti-DIG-AP per plot |
| Detection buffer | 50 mM $MgCl_2$, 0.1 M Tris·HCl pH 9, 1 M NaCl |

### 5.1.6 Solutions for chemical transformation of *E. coli*

| | |
|---|---|
| RF1 solution | 100 mM $RbCl_2$, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$, 15% (v/v) glycerol pH 5.8 |
| RF2 solution | 10 mM MOPS, 10 mM $RbCl_2$, 75 mM $CaCl_2$, 15% (v/v) glycerol pH 6.8 |

## 5.2 Appendix B

### 5.2.1 PCR and sequencing primers used in this study

**Table 7. Oligonucleotides used for PCR amplification in this study.**

| Name of fragment | Names of primers | Oligonucleotide sequence [a] | Annealing Temperature and DNA polymerase | Properties of amplicons |
|---|---|---|---|---|
| *fpgI* Upstream | FpgIUSF1 | 5′*GGTACC* ACACCGTCGACCTGGAAC 3′ | Taq/Expand =58°C | 1048bp amplicon generated from wild type *fpgI* allele, retaining only 100bp of 5′ end of the *fpg1* gene using FpgIUSF1and FpgIUSR1 primers |
| | FpgIUSR1 | 5′*AGATCT* CACACCGGTGATCGTCTT 3′ | | |
| *fpgI* Downstream | FpgIDSF1 | 5′*AGATCT* ACCGCTCGTCGTTCTACT 3′ | Taq/Expand =58°C | 1048bp amplicon generated from wild type *fpgI* allele, retaining only 100bp of 3′ end of the *fpg1* gene and using FpgIDSF1and FpgIDSR1 primers |
| | FpgIDSR1 | 5′*AAGCTT* GTACGCTCAACCCAGAGA 3′ | | |
| *fpgII* Upstream | FpgIIUSF1 | 5′ *AGATCT* ACAGGTGCGTG ATGAGGT 3′ | Taq/Expand =55°C Phusion = 56°C | 982bp amplicon generated from wild type *fpgII* allele, retaining only 317bp of 5′ end of the *fpgII* gene and using FpgIIUSF1and FpgIIUSR1 primers |
| | FpgIIUSR1 | 5′ *AAGCTT* CTGATCGGGTTTCGGTTG 3′ | | |
| *fpgII* Downstream | FpgIIDSF1 | 5′ *GGTACC* GGCCTTGCTCGCACAGTA 3′ | Taq/Expand =55°C Phusion = 56°C | 913bp amplicon generated from wild type *fpgII* allele, retaining only 317bp of 3′ end of the *fpgII* gene and using FpgIIDSF1and FpgIIDSR1 primers |
| | FpgIIDSR1 | 5′ A*GATCT* TGATCACCGACCAGAAGG 3′ | | |
| *neiI* Upstream | NeiIUSF1 | 5′ *AAGCTT* GTGGCGCGTATTTCTTCC 3′ | Taq/Expand =54°C | 1098bp amplicon generated from wild type *neiI* allele, retaining only 175bp of ’5 end of the *neiI* gene and using NeiIUSF1and NeiIUSR1 primers |
| | NeiIUSR1 | 5′ *AGATCT* ATGAACAGATGTTTGCCG 3′ | | |

| | | | | |
|---|---|---|---|---|
| *neiI* Downstream | NeiIDSF1 | 5′ *AGATCT* GCCAGCTGTGGGTGTACG 3′ | Taq/Expand =58°C | 1110bp amplicon generated from wild type *neiI* allele, retaining only 180bp of ’3 end of the *neiI* gene and using NeiIDSF1and NeiIDSR1 primers |
| | NeiDSR1 | 5′ *GGTACC* GCATCCTGTGCGGTGTTG 3′ | | |
| *neiII* Upstream | NeiIIUSF1 | 5′GGTACC ATCACCAGATCGTCGGGATA 3′ | Taq/Expand =55°C<br>Phusion = 56°C | 974bp amplicon generated from wild type *neiII* allele, retaining only 183bp of 5′ end of the *neiII* gene and using NeiIIUSF1and NeiIIUSR1 primers |
| | NeiIIUSFI | 5′ CCGTCGTAGTGGTGGAACAG GTG 3′ | | |
| *neiII* Downstream | NeiIIDSR1 | 5′ *GTCGAC* GTTCCACACGTAGCGTTCCT 3′ | Taq/Expand =55°C<br>Phusion = 56°C | 992bp amplicon generated from wild type *neiII* allele, retaining only 85 bp of 3′ end of the *neiII* gene and using NeIDSF1and NeiIDSR1 primers |
| | NeiIIDSF1 | 5′ *AGATCT* GTGGTACACCCGTACGAACC 3′ | | |
| RRDR region of *rpoB* | MsmrpoBF1 | 5′ GCAGACCCTGATCAACATCC 3′ | Expand = 60°C | 184bp amplicon generated from wild type *rpoB* allele using MsmrpoBF1 and MsmrpoBR1primers |
| | MsmrpoBR1 | 5′ ACGTCGCGGACCTCGAGG 3′ | | |

[a] Restriction enzymes sites are indicated in italics

**Table 8. Table of primers used to sequence knockout fragments**

| Template | Name of primer | Sequence |
|---|---|---|
| *fpgI* Upstream | FpgISeqUF1 | 5′ ACACCGTCGACCTGGAAC 3′ |
| | FpgISeqUF2 | 5′ CCACCGCAGCTATTCGTA 3′ |
| | FpgISeqUF3 | 5′ AATCTATCTCGAACACGG 3′ |
| | FpgISeqUR1 | 5′ CACACCGGTGATCGTCTT 3′ |
| | FpgISeqUR2 | 5′ GCCGAACAGGCGCAGGAT 3′ |
| | FpgISeqUR3 | 5′ CGCCGTTCTCATACGAAT 3′ |
| *fpgI* Downstream | FpgISeqDF1 | 5′ ACCGCTCGTCGTTCTACT 3′ |
| | FpgISeqDF2 | 5′ CGACATGTCGAACAGGGC 3′ |
| | FpgISeqDF3 | 5′ CTGCCCGTATTGCGGTTG 3′ |
| | FpgISeqDR1 | 5′ GTACGCTCAACCCAGAGA 3′ |
| | FpgISeqDR2 | 5′ CGCAGCCTGCTTACGGGG 3′ |
| | FpgISeqDR3 | 5′ TACTTCTGTGCGGGCGAT 3′ |
| *fpgII* Upstream | FpgIIseqUR1 | 5′ GGGATTACCGACGGGTTC 3′ |
| | FpgIISeqUF2 | 5′ GGAACCCGTCGGTAATCC 3′ |
| | FpgIISeqUR2 | 5′ ATCGCGATCACGGTCACT 3′ |
| | FpgIIseqUF3 | 5′ CCACAGTGACCGTGATCG 3′ |
| *fpgII* Downstream | FpgIISeqDR1 | 5′ ACGGATCGTCTGGAGGAG 3′ |
| | FpgIISeqDF2 | 5′ GGCCACCTTGATGTACGG 3′ |
| | FpgIISeqDR2 | 5′ TTCGCCGACAAGTCCTTC 3′ |
| | FpgIISeqDF3 | 5′ CGGGCAGTACTGGAAGGA 3′ |
| *neiI* Upstream | NeiISeqUF1 | 5′ TCGCCACGCTGTACAAGG 3′ |
| | NeiISeqUR1 | 5′ CCTTGTACAGCGTGGCGA 3′ |
| | NeiISeqUF2 | 5′ GGATCCCGTCACTGCTCG 3′ |
| | NeiISeqUR2 | 5′ CGAGCAGTGACGGGATCC 3′ |
| | NeiIUExtraF | 5′ CACTTCCAGGCCGAGTTG 3′ |
| | NeiIIUExtraR | 5′ CCAGCGACTCCACGAAGTA 3′ |
| *neiI* Downstream | NeiISeqDF3 | 5′ GAGGATCACCACAACCCG 3′ |
| | NeiISeqDR3 | 5′ CGGGTTGTGGTGATCCTC 3′ |
| | NeiISeqDF4 | 5′ AGTCCGATCCGTAATGCG 3′ |
| | NeiISeqDR4 | 5′ CGCATTACGGATCGGACT 3′ |
| | NeiISeqDF5 | 5′ ACCGTGCGCGCCCACTTC 3′ |

| | | |
|---|---|---|
| *neiII* Upstream | NeiIISeqUF | 5′ ATCACCAGATCGTCGGGATA 3′ |
| | NeiIISeqUF2 | 5′ TTAGGGTTAGCGCCATGC 3′ |
| | NeiIISeqUF3 | 5′ CACAACAACGCGAACCTG 3′ |
| | NeiIISeqUR2 | 5′ TCGTACTCGTAAGCCCCG 3′ |
| | NeiIISeqUR3 | 5′ ACCACGGAGCAGACAGGA 3′ |
| *neiII* Downstream | NeiIISeqDF | 5′ GTGGTACACCCGTACGAACC 3′ |
| | NeiIISeqDF4 | 5′ CTCCCCGATTTCGAACCT 3′ |
| | NeiIISeqDF5 | 5′ ATGATGGCGATCCTGACC 3′ |
| | NeiIISeqDR1 | 5′ GTTCCACACGTAGCGTTCCT 3′ |
| | NeiIISeqDR5 | 5′ GGTCAGGATCGCCATCAT 3′ |
| | NeiIISeqDR6 | 5′ GAACGAGAAGTTCAGTGC 3′ |

**Table 9. Primers and conditions used to confirm DCOs by PCR**

| Template | Name of primers | Annealing temp | Amplicon sizes |
|---|---|---|---|
| *fpgI* | FpgISeqUF2 and FpgISeqDR2 | 56°C-57°C. | $mc^2155$ = 2142bp<br>mutant = 1463bp |
| *fpgII* | FpgIISeqDF3 and FpgIISeqUR1 | 58°C | $mc^2155$ = 888bp<br>mutant = 627bp |
| *neiI* | NeiISeqDR3 and NeiISeqUF2<br>0r<br>NeiISeqDF4 and NeiIUSF1 | 56°C-57°C.<br>55°C | $mc^2155$ = 1165bp<br>mutant = 697bp<br>or<br>$mc^2155$ = 2312bp<br>mutant = 1845bp |
| *neiII* | NeiIISeqDR5 and NeiIISeqUF3 | 58°C. | $mc^2155$ = 1477bp<br>mutant = 871bp |

## 5.3 Appendix C

### 5.3.1 Bioinformatic analysis of *fpg* and *nei* DNA glycosylases

**Table 10. Identification of *fpg* and *nei* DNA glycosylase homologues in other mycobacterial genomes while using GenoList and CMR genomic comparisons.**

| Organism | Annotation | Homologue identified by genomic context analysis |
|---|---|---|
| *M. leprae* | ML1658 | *fpgI* |
| | ML0148 | *fpgII* pseudogene |
| | ML1483c | *neiII* pseudogene |
| *M. bovis* | Mb2949c | *fpgI* |
| | Mb0969 | *fpgII* similarly truncated as *Rv0944* |
| | Mb3325 | *neiI* |
| | Mb2491c | *neiII* |
| *M. ulcerans* | MUL_2031 | *fpgI* |
| | MUL_4418 | *fpgII* |
| | MUL_2650 | *neiI* |
| | MUL_3737 | Possible *neiII* |
| *M. avium* | MAV_3782 | *fpgI* |
| | MAV_1066 | *fpgII* |
| | MAV_4269 | *neiI* |
| | MAV_1708 | *neiII* |
| | MAV_3149 | Same as MAP_1328 |
| *M. avium paratuberculosis* | MAP_2994c | *fpgI* |
| | MAP_0889 | *fpgII* |
| | MAP_3416 | *neiI* |
| | MAP_2284c | *neiII* |
| | MAP_1328 | same as MAV_3149 |

**Table 11. BLASTp searches of *M. smegmatis* Fpg and Nei proteins to other organisms.**

| | *M. tuberculosis* (H37Rv) | Actinobacteria | *E. coli* | *T. thermophilus* | *G. stearothermopilus* | *L. lactis* |
|---|---|---|---|---|---|---|
| MSMEG_2419 (FpgI) | $4e^{-95}$ | $1e^{-122} – 1e^{-51}$ | $6e^{-29}$ | $9e^{-24}$ | $7e^{-23}$ | $1e^{-23}$ |
| MSMEG_5545 (FpgII) | $7e^{-97}$ | $1e^{-132} – 1e^{-73}$ | $6e^{-29}$ | $6e^{-24}$ | $7e^{-23}$ | $1e^{-23}$ |
| MSMEG_1756 (NeiI) | $1e^{-91}$ | $6e^{-100} – 1e^{-51}$ | $2e^{-18}$ | N/R* | N/R* | N/R* |
| MSMEG_4683 (NeiII) | $8e^{-109}$ | $3e^{-119} – 1e^{-51}$ | $6e^{-16}$ | $1e^{-17}$ | $1e^{-21} – 1e^{-19}$ | $2e^{-13} –2e^{-11}$ |

*not recognised

### 5.3.2 Vector maps and cloning strategies used to generate suicide vectors

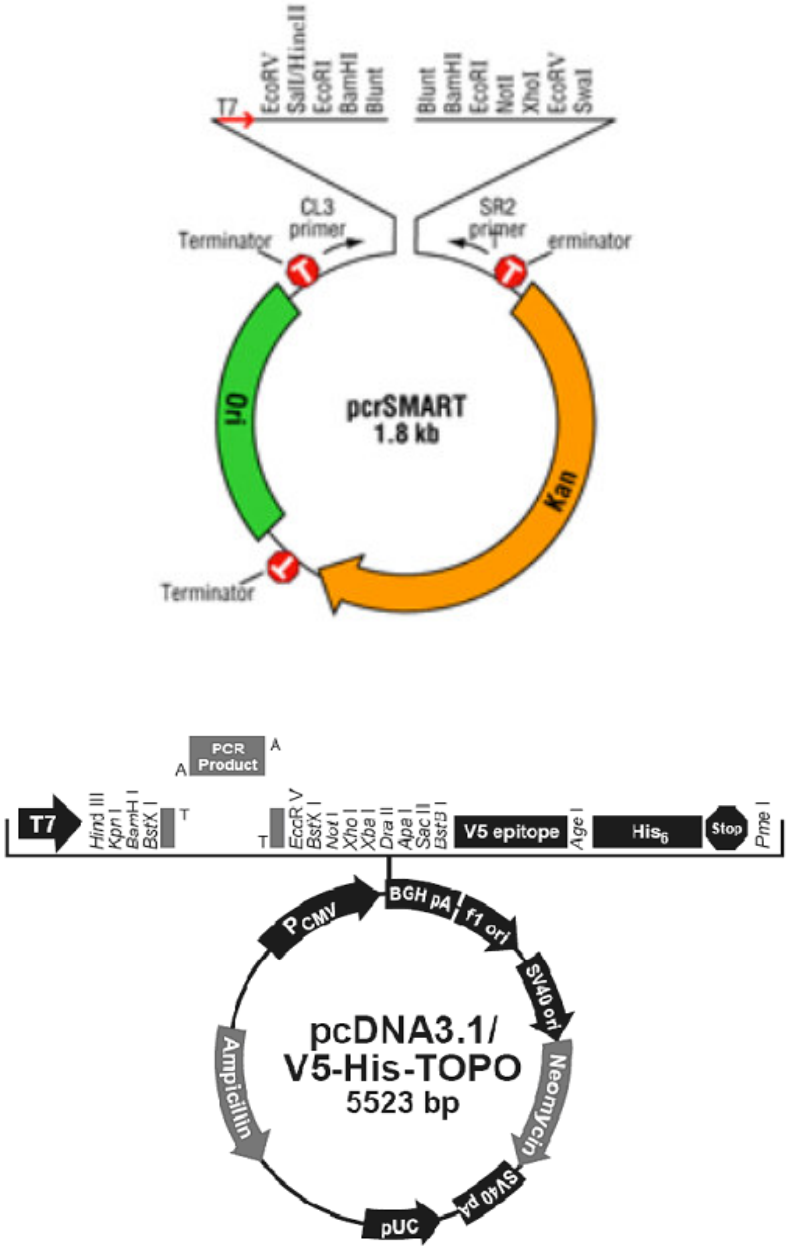



**Fig. 32. The plasmid maps for pcrSMART and TOPO vector used.**

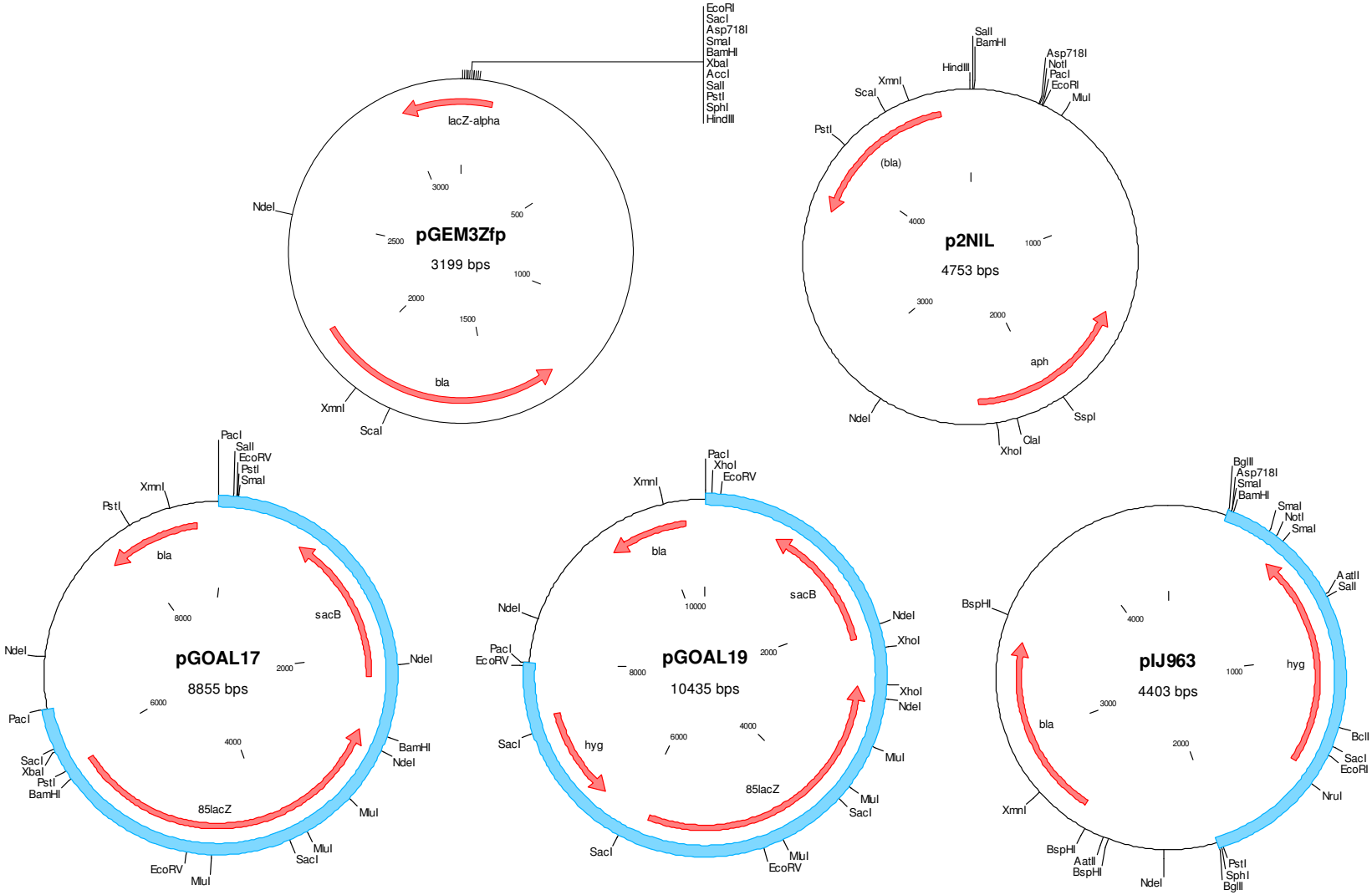



**Fig. 33. The plasmid maps for pGem3Zfp, p2NIL, pGOAL17 pGOAL19 and pIJ964**
The marker cassettes are highlighted on pGOAL17, pGOAL19 and pIJ963

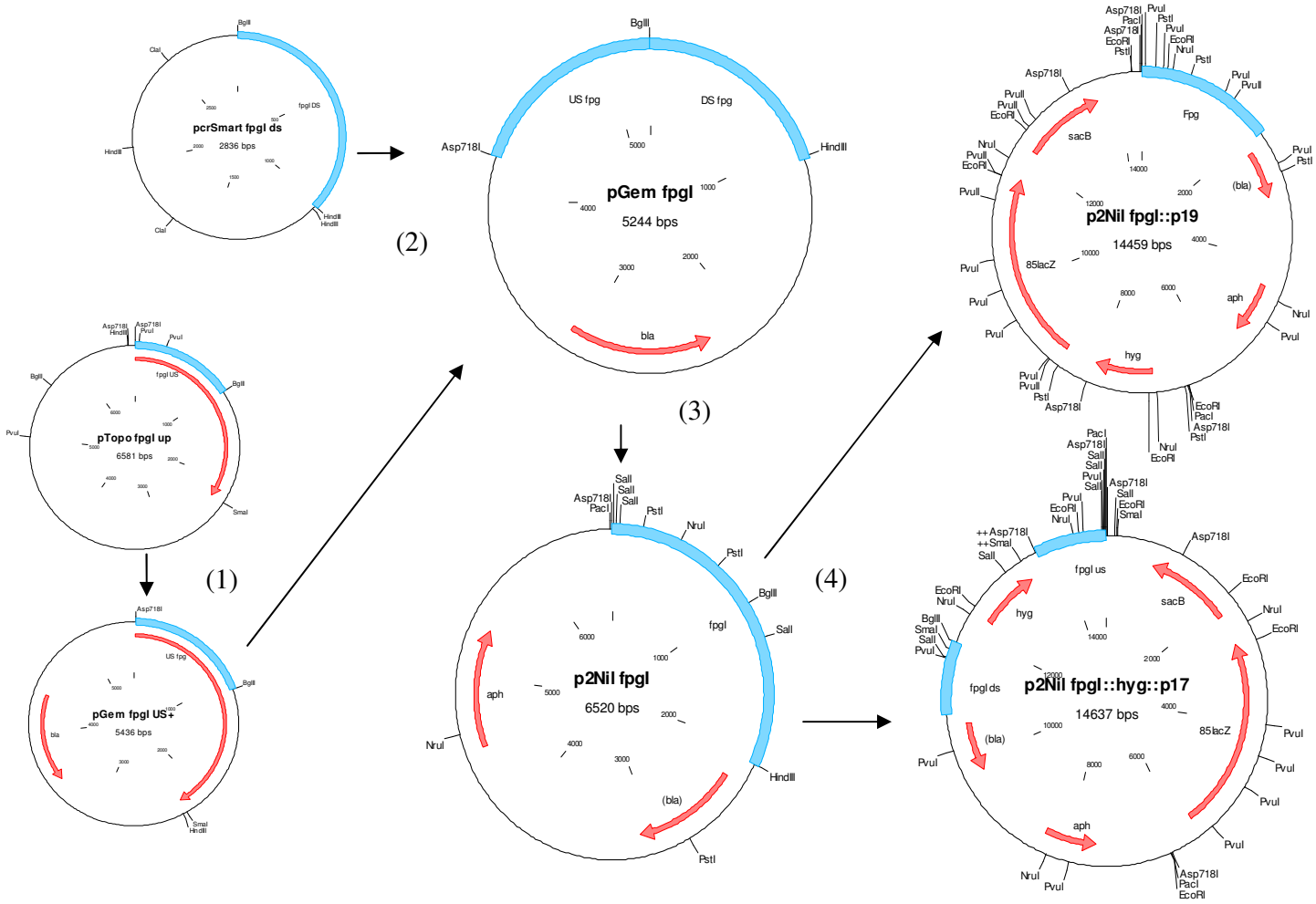



**Fig. 34. The plasmid maps of vectors generated during the construction of *fpgI* knockout suicide vectors**
The US and DS fragments are highlighted in the respective plasmid. (1) the insertion of the US fragment with part of the TOPO vector (shown in red) into pGEM (2) insertion of the DS fragment into pGEMΔ*fpgI*US (3) the insertion of the Δ*fpgI* region into p2NIL and (4) the insertion of marker cassettes.

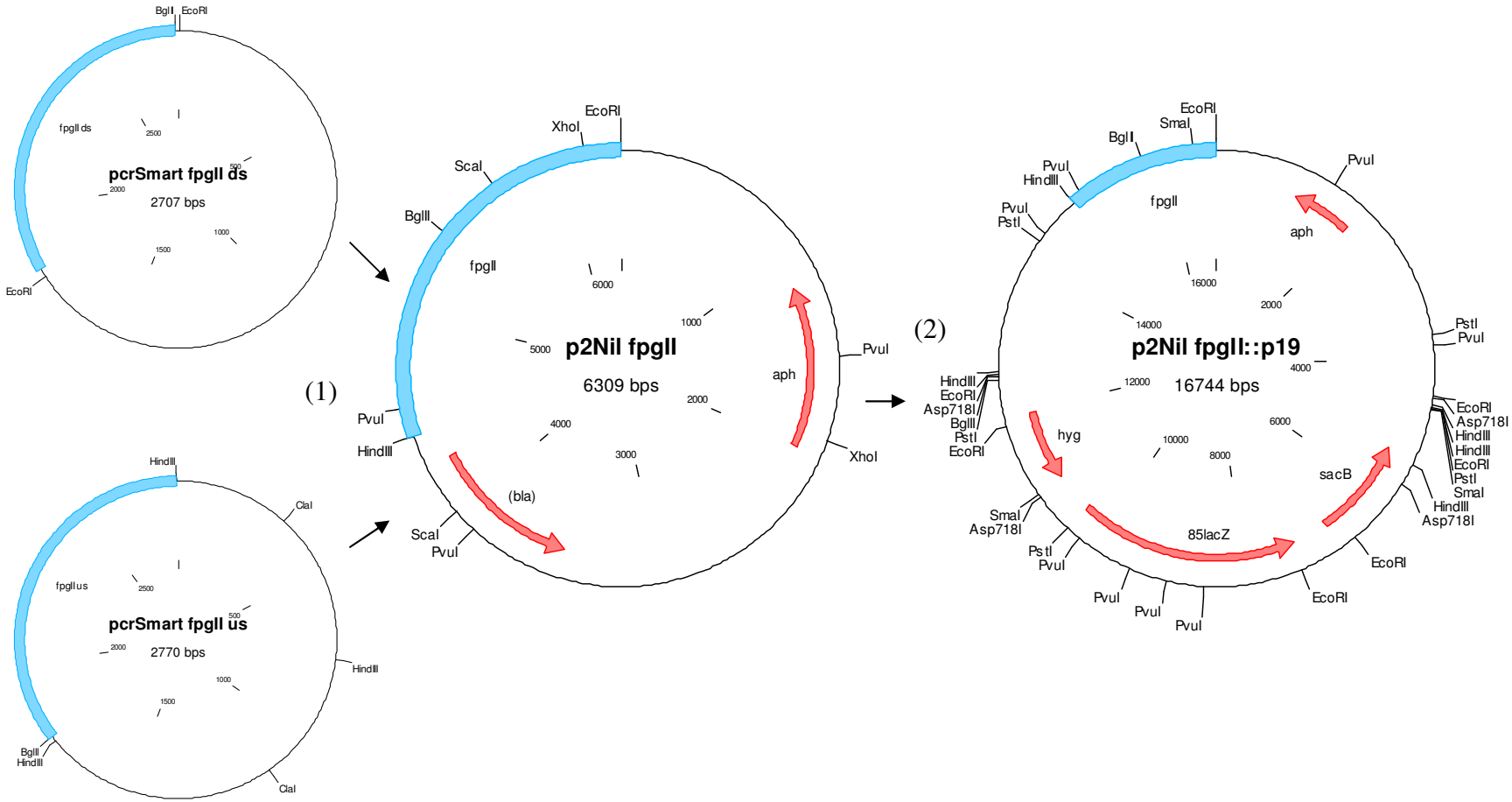



**Fig. 35. The plasmid maps of vectors generated during the construction of the *fpgII* knockout suicide vector.**
The US and DS fragments are highlighted on the respective plasmid (1) The insertion of the DS and US fragments into p2NIL. (2) The insertion of marker cassettes.

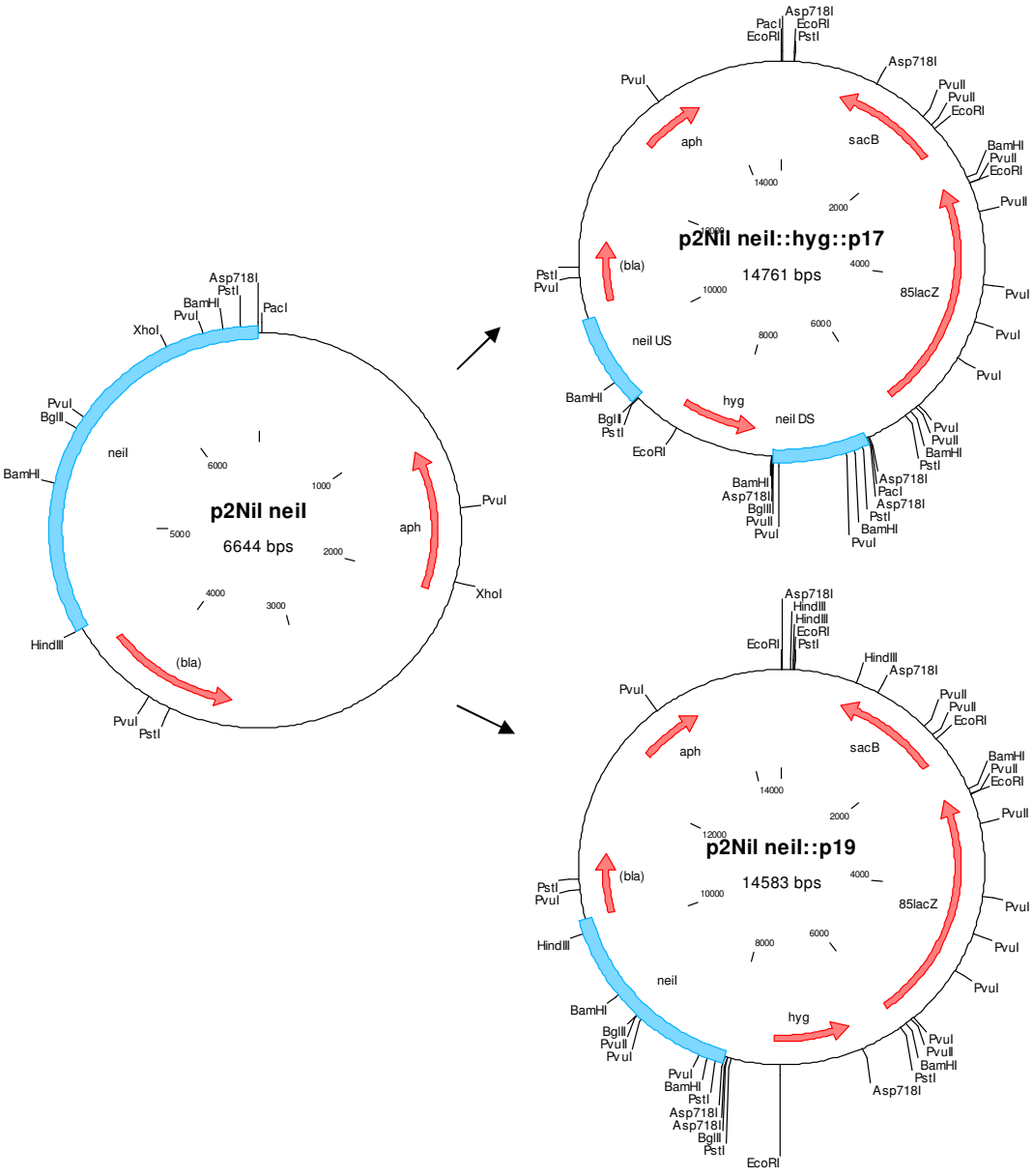



**Fig. 36. The plasmids maps of vectors generated in this study during the construction of *neiI* knockout suicide vectors**
The US and DS fragments are highlighted the respective plasmid.

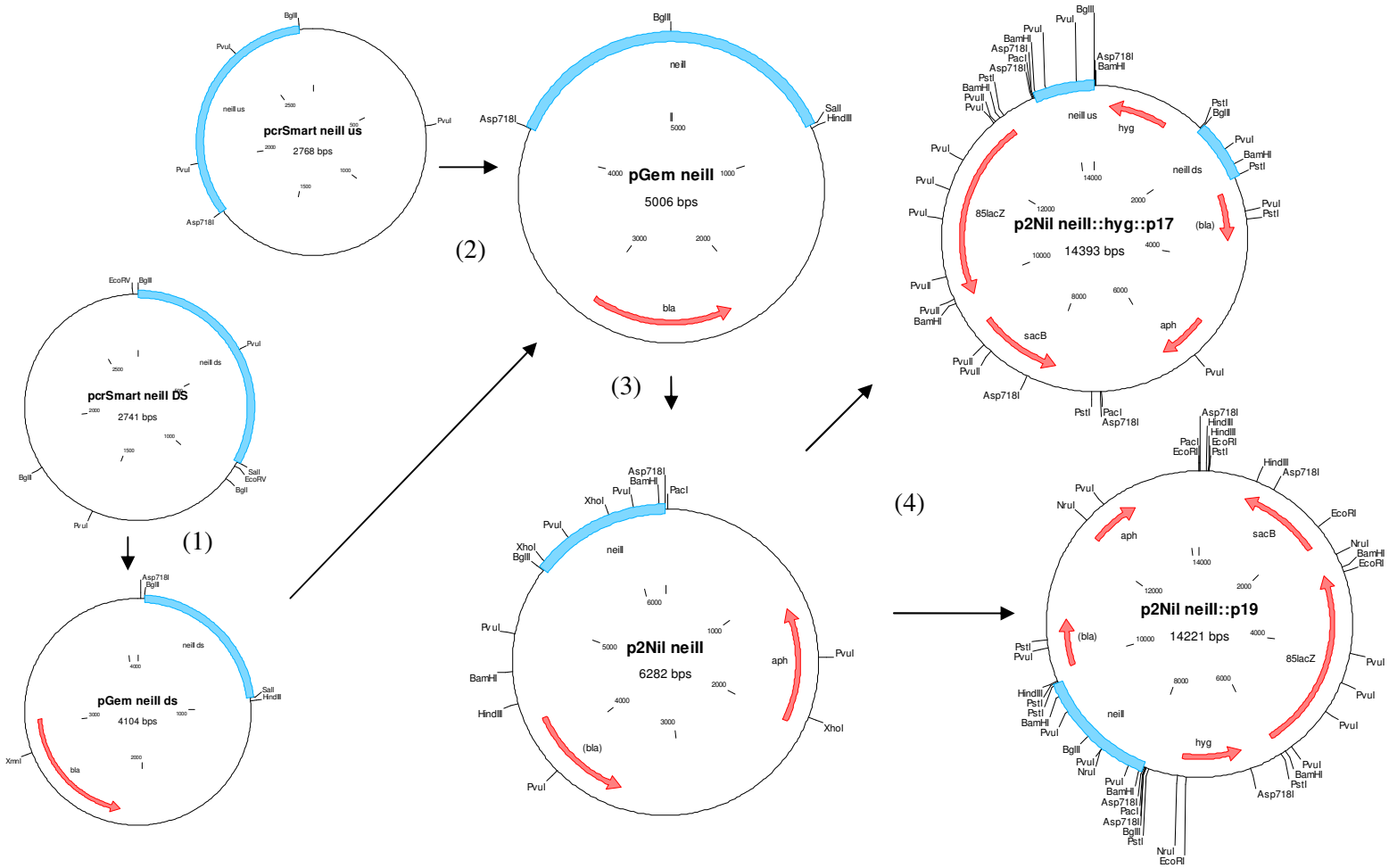



**Fig. 37. The plasmid maps of vectors generated during the construction of *neiII* knockout suicide vectors**
The upstream and downstream fragments are highlighted the respective plasmid (1) The insertion of the DS fragment into pGEM (2) insertion of the US fragment into pGEMΔ*neiII*DS (3) The insertion of the Δ*neiII* region into p2NIL and (4) the insertion of marker cassettes

## 5.3.3 Confirmation of suicide vectors by restriction digests

### 5.3.3.1 *fpgI* based suicide vectors

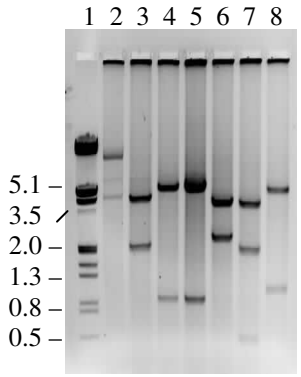



**Fig. 38. Restriction analysis of p2NILΔ*fpgI* with various restriction endonucleases.**
Marker sizes in bp are shown adjacent to the fragment. Lane 1, molecular weight marker λIII; lane 2, uncut p2NILΔ*fpgI*; lane 3, *Asp*718/*Hin*dIII digest (4436bp, 2084bp); lane 4, *Asp*718/ *Bgl*II digest (5478bp, 1042bp); lane 5, *Hin*dIII/*Bgl*II digest (5478bp, 1042bp); lane 6, *Nru*I digest (4179bp, 2341bp); lane 7, *Pst*I digest (4002bp, 1979bp, 539p); lane 8, *Sal*I digest (5256bp, 1225p).

### 5.3.3.2 *fpgII* based suicide vectors

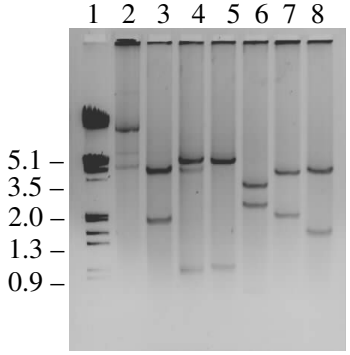



**Fig. 39. Restriction analysis of p2NILΔ*fpgII* with various restriction endonucleases.**
Marker sizes in bp are shown adjacent to the fragment. The restriction digest of p2NILΔ*fpgII*; Lane 1, molecular weight marker λIII; lane 2, uncut p2NILΔ*fpgII*; lane 3, *Eco*RI/*Hin*dIII digest (4414bp, 1895bp); lane 4, *Eco*RI/ *Bgl*II digest (5390bp, 919bp); lane 5, *Hin*dIII/*Bgl*II digest (5333bp, 919bp); lane 6, *Pvu*I digest (3277bp, 2384bp, 648bp); lane 7, *Xho*I digest (4220bp, 2089bp); lane 8, *Sca*I digest (4631bp, 1682bp)

### 5.3.3.3 *neiI* based suicide vectors

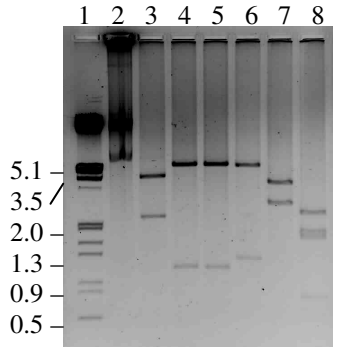



**Fig. 40. Restriction analysis of p2NILΔ*neiI* with various restriction endonucleases.**
Marker sizes in bp are shown adjacent to the fragment. Lane 1, molecular weight marker λIII; lane 2, uncut p2NILΔ*neiI*; lane 3, *Asp*718/*Hin*dIII digest (4436bp, 2208bp); lane 4, *Asp*718/ *Bgl*II digest (5534bp, 1110bp); lane 5, *Hin*dIII/*Bgl*II digest (5536bp, 1098bp); lane 6, *Bam*HI digest (5410bp, 1234bp); lane 7, *Pst*I digest (3888bp, 2756bp); lane 8, *Pvu*I digest (2384bp, 1826bp, 1679bp, 755bp)

### 5.3.3.4 *neiII* based suicide vectors

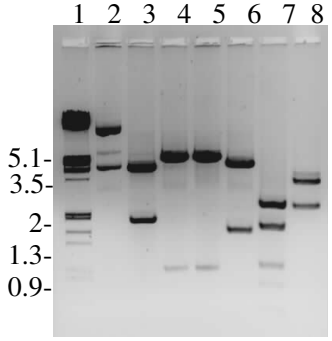



**Fig. 41. Restriction analysis of p2NILΔ*neiII* with various restriction endonucleases.**
Marker sizes in bp are shown adjacent to the fragment. lane 1, molecular weight marker λIII; lane 2, uncut p2NILΔ*neiI*; lane 3, *Asp*718/*Hin*dIII digest (4436bp, 1846bp); lane 4, *Asp*718/ *Bgl*II digest (5357bp, 925bp); lane 5, *Hin*dIII/*Bgl*II digest (5361bp, 921bp); lane 6, *Bam*HI digest (4692bp, 1590bp); lane 7, *Pvu*I digest (2384bp, 1747bp, 972bp, 724bp, 455bp); lane 8, *Xho*I digest (3467bp, 2309bp, 506bp).

### 5.3.3.5 Mutation identified the Δ*fpgII* downstream primer

(a)

```
Molecule:       fpgII downstream region,  913 bps DNA
Description:    the predicted fpgII downstream region
File Name:      DS FpgII.cm5,  dated 06 Jun 2006
Printed:        1-913 bps (Full),  format Single Strand

     1   ggggggtaccg gccttgctcg cacagtaggc ggccttgccg ccggggacac ccgtgttgcc
    61   cagcaccgag gagatcagca ccaggtgtcc accgcctgcc ttcttgaaca tctccagtgc
   121   ggcctcgatc tgcaccagac cggcgaccag gttggtctcg agcgtggcct tgttggccca
   181   cggcttgccc tcgcccagcg gccagccctt gccgatgccc gcgttgacga tgacgcggtc
   241   gataccgccg agttcctcgg acagttcgtt gaacacccgg ggcacggcct cgtgatcgtt
   301   gacgtccagc gcggcgatgg ccaccttgat gtacgggtgc cggtcggcga tctccgcctt
   361   gagctcctcc agacgatccg tgcggcgcgc gcacagcgcc aggttgcggc ctctggctgc
   421   gaattgacgg gccataccgg cgcccaggcc tgagctggca ccggtgatga ggatgttctg
   481   acgggtcacc gcagcaggat aaccgggtta gacaactgtc aagaatccgc ccgctatttg
   541   agcaggcgcg acatccgccg gtcggccagc accttaccgc cggtctggca ggtcgggcag
   601   tactggaagg acttgtcggc gaacgagacc tcccgcacgg tgtcgccgca caccgggcac
   661   ggcaggccgg tgcgcgcgtg gacccgtagg cccgagcgtt tctcgccctt caacgtcgcg
   721   gcctgctgcc ccaccgagcg gctcacggca tcggtgagca ccgtgatcat ggcgtcgtgc
   781   agcgcaccga gttgcgcatc ggtgagcttg cccgcggtcg tgaacggcga cagcttggcc
   841   acgtgcagga tctcgtcact gtaggcgttg ccgatgcccg ccatcacctt ctggtcggtg
   901   atcaagatct ccc
```

(b)

```
Molecule:       fpgII downstream region sequencing results,  913 bps DNA
Description:    the sequenced fpgII downstream region
File Name:      DS FII after sequencing.cm5,  dated 04 Jan 2007
Printed:        1-913 bps (Full),  format Single Strand

     1   ggggctaccg gccttgctcg cacagtaggc ggccttgccg ccggggacac ccgtgttgcc
    61   cagcaccgag gagatcagca ccaggtgtcc accgcctgcc ttcttgaaca tctccagtgc
   121   ggcctcgatc tgcaccagac cggcgaccag gttggtctcg agcgtggcct tgttggccca
   181   cggcttgccc tcgcccagcg gccagccctt gccgatgccc gcgttgacga tgacgcggtc
   241   gataccgccg agttcctcgg acagttcgtt gaacacccgg ggcacggcct cgtgatcgtt
   301   gacgtccagc gcggcgatgg ccaccttgat gtacgggtgc cggtcggcga tctccgcctt
   361   gagctcctcc agacgatccg tgcggcgcgc gcacagcgcc aggttgcggc ctctggctgc
   421   gaattgacgg gccataccgg cgcccaggcc tgagctggca ccggtgatga ggatgttctg
   481   acgggtcacc gcagcaggat aaccgggtta gacaactgtc aagaatccgc ccgctatttg
   541   agcaggcgcg acatccgccg gtcggccagc accttaccgc cggtctggca ggtcgggcag
   601   tactggaagg acttgtcggc gaacgagacc tcccgcacgg tgtcgccgca caccgggcac
   661   ggcaggccgg tgcgcgcgtg gacccgtagg cccgagcgtt tctcgccctt caacgtcgcg
   721   gcctgctgcc ccaccgagcg gctcacggca tcggtgagca ccgtgatcat ggcgtcgtgc
   781   agcgcaccga gttgcgcatc ggtgagcttg cccgcggtcg tgaacggcga cagcttggcc
   841   acgtgcagga tctcgtcact gtaggcgttg ccgatgcccg ccatcacctt ctggtcggtg
   901   atcaagatct ccc
```

**Fig. 42. An illustration of the mutation in the Δ*fpgII* downstream primer.**
The *Asp*718 restriction enzyme site is indicated by the line and the mutation which abrogates the restriction enzyme site is indicated in the box. (a) shows the expected sequence and (b) the actual sequence of the fragment.

# 6. References


Alexander, D. C., Jones, J. R. and Liu, J. (2003) A rifampin-hypersensitive mutant reveals differences between strains of *Mycobacterium smegmatis* and presence of a novel transposon, IS1623. *Antimicrob Agents Chemother* 47(10): 3208-13.

Alhama, J., Ruiz-Laguna, J., Rodriguez-Ariza, A., Toribio, F., Lopez-Barea, J. and Pueyo, C. (1998) Formation of 8-oxoguanine in cellular DNA of *Escherichia coli* strains defective in different antioxidant defences. *Mutagenesis* 13(6): 589-94.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol* 215(3): 403-10.

Asad, N. R., de Almeida, C. E., Asad, L. M., Felzenszwalb, I. and Leitao, A. C. (1995) Fpg and UvrA proteins participate in the repair of DNA lesions induced by hydrogen peroxide in low iron level in *Escherichia coli*. *Biochimie* 77(4): 262-4.

Bandaru, V., Zhao, X., Newton, M. R., Burrows, C. J. and Wallace, S. S. (2007) Human endonuclease VIII-like (NEIL) proteins in the giant DNA Mimivirus. *DNA Repair (Amst)* 6(11): 1629-41.

Blaisdell, J. O., Hatahet, Z. and Wallace, S. S. (1999) A novel role for *Escherichia coli* endonuclease VIII in prevention of spontaneous G-->T transversions. *J Bacteriol* 181(20): 6396-402.

Blaisdell, J. O. and Wallace, S. S. (2001) Abortive base-excision repair of radiation-induced clustered DNA lesions in *Escherichia coli*. *Proc Natl Acad Sci USA* 98(13): 7426-30.

Blondelet-Rouault, M. H., Weiser, J., Lebrihi, A., Branny, P. and Pernodet, J. L. (1997) Antibiotic resistance gene cassettes derived from the omega interposon for use in *E. coli* and Streptomyces. *Gene* 190(2): 315-7.

Boshoff, H. I., Reed, M. B., Barry, C. E., 3rd and Mizrahi, V. (2003) DnaE2 polymerase contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell* 113(2): 183-93.

Brown, L. T., LeMay, H. E. and Bursten, B. E. (2000) Chemistry The central Science. Upper Saddle River, New Yersey, Prentice Hall.

Cabrera, M., Nghiem, Y. and Miller, J. H. (1988) *mutM,* a second mutator locus in *Escherichia coli* that generates G.C----T.A transversions. *J Bacteriol* 170(11): 5405-7.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393(6685): 537-44.

Corbett, E. L., Watt, C. J., Walker, N., Maher, D., Williams, B. G., Raviglione, M. C. and Dye, C. (2003) The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med* 163(9): 1009-21.

Cunningham, R. P. (1997) DNA glycosylases. *Mutat Res* 383(3): 189-96.

Cupples, C. G. and Miller, J. H. (1989) A set of *lac*Z mutations in *Escherichia coli* that allow rapid detection of each of the six base substitutions. *Proc Natl Acad Sci USA* 86(14): 5345-9.

D'Ham, C., Romieu, A., Jaquinod, M., Gasparutto, D. and Cadet, J. (1999) Excision of 5,6-dihydroxy-5,6-dihydrothymine, 5,6-dihydrothymine, and 5-hydroxycytosine from defined sequence oligonucleotides by *Escherichia coli* Endonuclease III and Fpg proteins: kinetic and mechanistic aspects. *Biochemistry* 38(11): 3335-44.

David, H. L. and Newman, C. M. (1971) Some observations on the genetics of isoniazid resistance in the tubercle bacilli. *Am Rev Respir Dis* 104(4): 508-15.

David, S. S., O'Shea, V. L. and Kundu, S. (2007) Base-excision repair of oxidative DNA damage. *Nature* 447(7147): 941-50.

Demple, B. and Harrison, L. (1994) Repair of oxidative damage to DNA: enzymology and biology. *Annu Rev Biochem* 63: 915-48.

Dos Vultos, T., Blazquez, J., Rauzier, J., Matic, I. and Gicquel, B. (2006) Identification of Nudix hydrolase family members with an antimutator role in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *J Bacteriol* 188(8): 3159-61.

Dussurget, O. and Smith, I. (1998) Interdependence of mycobacterial iron regulation, oxidative-stress response and isoniazid resistance. *Trends Microbiol* 6(9): 354-8.

Dwyer, D. J., Kohanski, M. A., Hayete, B. and Collins, J. J. (2007) Gyrase inhibitors induce an oxidative damage cellular death pathway in *Escherichia coli*. *Mol Syst Biol* 3: 91.

Edwards, K. M., Cynamon, M. H., Voladri, R. K., Hager, C. C., DeStefano, M. S., Tham, K. T., Lakey, D. L., Bochan, M. R. and Kernodle, D. S. (2001) Iron-cofactored superoxide dismutase inhibits host responses to *Mycobacterium tuberculosis*. *Am J Respir Crit Care Med* 164(12): 2213-9.

Farr, S. B., Natvig, D. O. and Kogoma, T. (1985) Toxicity and mutagenicity of plumbagin and the induction of a possible new DNA repair pathway in *Escherichia coli*. *J Bacteriol* 164(3): 1309-16.

Fine, P. E. (1995) Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* 346(8986): 1339-45.

Fowler, R. G., White, S. J., Koyama, C., Moore, S. C., Dunn, R. L. and Schaaper, R. M. (2003) Interactions among the *Escherichia coli* MutT, MutM, and MutY damage prevention pathways. *DNA Repair (Amst)* 2(2): 159-73.

Friedberg, E. C., Walker, G. C. and Siede, W. (1995) DNA Repair and Mutagenesis. Washington D.C, ASM Press.

Fromme, J. C., Banerjee, A. and Verdine, G. L. (2004) DNA glycosylase recognition and catalysis. *Curr Opin Struct Biol* 14(1): 43-9.

Gandhi, N. R., Moll, A., Sturm, A. W., Pawinski, R., Govender, T., Lalloo, U., Zeller, K., Andrews, J. and Friedland, G. (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet* 368(9547): 1575-80.

Gey van Pittius, N. C., Sampson, S. L., Lee, H., Kim, Y., van 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 Evol Biol* 6: 95.

Gilboa, R., Zharkov, D. O., Golan, G., Fernandes, A. S., Gerchman, S. E., Matz, E., Kycia, J. H., Grollman, A. P. and Shoham, G. (2002) Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *J Biol Chem* 277(22): 19811-6.

Glickman, M. S. and Jacobs, W. R., Jr. (2001) Microbial pathogenesis of *Mycobacterium tuberculosis:* dawn of a discipline. *Cell* 104(4): 477-85.

Golan, G., Zharkov, D. O., Feinberg, H., Fernandes, A. S., Zaika, E. I., Kycia, J. H., Grollman, A. P. and Shoham, G. (2005) Structure of the uncomplexed DNA repair enzyme Endonuclease VIII indicates significant interdomain flexibility. *Nucleic Acids Res* 33(15): 5006-16.

Gordhan, B. G., Andersen, S. J., De Meyer, A. R. and Mizrahi, V. (1996) Construction by homologous recombination and phenotypic characterization of a DNA polymerase domain polA mutant of *Mycobacterium smegmatis*. *Gene* 178(1-2): 125-30.

Gordhan, B. G. and Parish, T. (2001) Gene replacement using pretreated DNA. *Methods in molecular medicine Mycobacterium tuberculosis protocols*. Parish, T. and Stoker, N. G. New Jersey, Humana Press**:** 77-90.

Griffiths, J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C. and Gelbart, W. M. (2000). An introduction genetic analysis. New York, W.H. Freeman and Company.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166(4): 557-80.

Hazra, T. K., Izumi, T., Venkataraman, R., Kow, Y. W., Dizdaroglu, M. and Mitra, S. (2000) Characterization of a novel 8-oxoguanine-DNA glycosylase activity in *Escherichia coli* and identification of the enzyme as Endonuclease VIII. *J Biol Chem* 275(36): 27762-7.

Hazra, T. K., Muller, J. G., Manuel, R. C., Burrows, C. J., Lloyd, R. S. and Mitra, S. (2001) Repair of hydantoins, one electron oxidation product of 8-oxoguanine, by DNA glycosylases of *Escherichia coli*. *Nucleic Acids Res* 29(9): 1967-74.

Henle, E. S. and Linn, S. (1997) Formation, prevention, and repair of DNA damage by iron/hydrogen peroxide. *J Biol Chem* 272(31): 19095-8.

Hesseling, A. C., Cotton, M. F., Marais, B. J., Gie, R. P., Schaaf, H. S., Beyers, N., Fine, P. E., Abrams, E. J., Godfrey-Faussett, P. and Kuhn, L. (2007) BCG and HIV reconsidered: moving the research agenda forward. *Vaccine* 25(36): 6565-8.

Hesseling, A. C., Marais, B. J., Gie, R. P., Schaaf, H. S., Fine, P. E., Godfrey-Faussett, P. and Beyers, N. (2007) The risk of disseminated Bacille Calmette-Guerin (BCG) disease in HIV-infected children. *Vaccine* 25(1): 14-8.

Hetherington, S. V., Watson, A. S. and Patrick, C. C. (1995) Sequence and analysis of the *rpoB* gene of *Mycobacterium smegmatis*. *Antimicrob Agents Chemother* 39(9): 2164-6.

Horst, J. P., Wu, T. H. and Marinus, M. G. (1999) *Escherichia coli* mutator genes. *Trends Microbiol* 7(1): 29-36.

Jacobs Jr, W. R. (2000) *Mycobacterium tuberculosis:* a once genetically intractable organism. p. 1-16. *Molecular Genetics of Mycobacteria*. Hatfull, G. F. and Jacobs Jr, W. R. Washington, ASM Press

Jain, R., Kumar, P. and Varshney, U. (2007) A distinct role of formamidopyrimidine DNA glycosylase (MutM) in down-regulation of accumulation of G, C mutations and protection against oxidative stress in mycobacteria. *DNA Repair (Amst)* 6(12): 1774-86.

Jiang, D., Hatahet, Z., Blaisdell, J. O., Melamede, R. J. and Wallace, S. S. (1997) *Escherichia coli* Endonuclease VIII: cloning, sequencing, and overexpression of the *nei* structural gene and characterization of *nei* and *nei nth* mutants. *J Bacteriol* 179(11): 3773-82.

Karunakaran, P. and Davies, J. (2000) Genetic antagonism and hypermutability in *Mycobacterium smegmatis*. *J Bacteriol* 182(12): 3331-5.

Kaufmann, S. H. (2005) Recent findings in immunology give tuberculosis vaccines a new boost. *Trends Immunol* 26(12): 660-7.

Kaufmann, S. H., Cole, S. T., Mizrahi, V., Rubin, E. and Nathan, C. (2005) *Mycobacterium tuberculosis* and the host response. *J Exp Med* 201(11): 1693-7.

Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. and Collins, J. J. (2007) A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* 130(5): 797-810.

Krokan, H. E., Standal, R. and Slupphaug, G. (1997) DNA glycosylases in the base excision repair of DNA. *Biochem J* 325 ( Pt 1): 1-16.

Larsen, M. H. (2000) Appendix 1. *Molecular genetics of mycobacteria*. Hatfull, G. F. and Jacobs Jr, W. R. Washington, D.C, ASM Press**:** 313-320.

Levin, B. R. and Rozen, D. E. (2006) Non-inherited antibiotic resistance. *Nat Rev Microbiol* 4(7): 556-62.

Maartens, G. and Wilkinson, R. J. (2007) Tuberculosis. *Lancet* 370(9604): 2030-43.

Machowski, E. E., Barichievy, S., Springer, B., Durbach, S. I. and Mizrahi, V. (2007) *In vitro* analysis of rates and spectra of mutations in a polymorphic region of the Rv0746 PE_PGRS gene of *Mycobacterium tuberculosis*. *J Bacteriol* 189(5): 2190-5.

Madden, T. (2003). The BLAST Sequence Analysis Tool. NCBI Handbook. NCBI, http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.ch16.

Manca, C., Paul, S., Barry, C. E., 3rd, Freedman, V. H. and Kaplan, G. (1999) *Mycobacterium tuberculosis* catalase and peroxidase activities and resistance to oxidative killing in human monocytes *in vitro*. *Infect Immun* 67(1): 74-9.

Martinez, J. L. and Baquero, F. (2000) Mutation frequencies and antibiotic resistance. *Antimicrob Agents Chemother* 44(7): 1771-7.

Master, S. S., Springer, B., Sander, P., Boettger, E. C., Deretic, V. and Timmins, G. S. (2002) Oxidative stress response genes in *Mycobacterium tuberculosis*: role of *ahpC* in resistance to peroxynitrite and stage-specific survival in macrophages. *Microbiology* 148(Pt 10): 3139-44.

Matsumoto, Y., Zhang, Q. M., Takao, M., Yasui, A. and Yonei, S. (2001) *Escherichia coli* Nth and human hNTH1 DNA glycosylases are involved in removal of 8-oxoguanine from 8-oxoguanine/guanine mispairs in DNA. *Nucleic Acids Res* 29(9): 1975-81.

Meya, D. B. and McAdam, K. P. (2007) The TB pandemic: an old problem seeking new solutions. *J Intern Med* 261(4): 309-29.

Michaels, M. L., Cruz, C., Grollman, A. P. and Miller, J. H. (1992) Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proc Natl Acad Sci U S A* 89(15): 7022-5.

Michaels, M. L. and Miller, J. H. (1992) The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J Bacteriol* 174(20): 6321-5.

Michaels, M. L., Pham, L., Cruz, C. and Miller, J. H. (1991) MutM, a protein that prevents G.C----T.A transversions, is formamidopyrimidine-DNA glycosylase. *Nucleic Acids Res* 19(13): 3629-32.

Mizrahi, V. and Andersen, S. J. (1998) DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? *Mol Microbiol* 29(6): 1331-9.

Mizrahi, V., Dawes, S. S. and Rubin, H. (2000) DNA replication. *Molecular Genetics of Mycobacteria*. Hatful, G. F. and Jacobs Jr, W. R. Washington, ASM Press.

Morlock, G. P., Plikaytis, B. B. and Crawford, J. T. (2000) Characterization of spontaneous, *In vitro*-selected, rifampin-resistant mutants of *Mycobacterium tuberculosis* strain H37Rv. *Antimicrob Agents Chemother* 44(12): 3298-301.

Murphy, T. M. and George, A. (2005) A comparison of two DNA base excision repair glycosylases from *Arabidopsis thaliana*. *Biochem Biophys Res Commun* 329(3): 869-72.

Musser, J. M. (1995) Antimicrobial agent resistance in mycobacteria: molecular genetic insights. *Clin Microbiol Rev* 8(4): 496-514.

Nathan, C. and Shiloh, M. U. (2000) Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. *Proc Natl Acad Sci USA* 97(16): 8841-8.

Newton, G. L., Ta, P. and Fahey, R. C. (2005) A mycothiol synthase mutant of *Mycobacterium smegmatis* produces novel thiols and has an altered thiol redox status. *J Bacteriol* 187(21): 7309-16.

Nguyen, L. and Thompson, C. J. (2006) Foundations of antibiotic resistance in bacterial physiology: the mycobacterial paradigm. *Trends Microbiol* 14(7): 304-12.

Ó Gaora, P. S., Barnini, C., Hayward, E., Filley, G., Rook, D., Young, J. and Thole, J. (1997) Mycobacteria as immunogens: development of expression vectors for use in multiple mycobacterial species. *Med. Princ. Pract.* 6: 91-96.

Parish, T. and Stoker, N. G. (2000) Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA plcABC* mutant by gene replacement. *Microbiology* 146( Pt 8): 1969-75.

Quan, S., Venter, H. and Dabbs, E. R. (1997) Ribosylative inactivation of rifampin by *Mycobacterium smegmatis* is a principal contributor to its low susceptibility to this antibiotic. *Antimicrob Agents Chemother* 41(11): 2456-60.

Ramaswamy, S. and Musser, J. M. (1998) Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber Lung Dis* 79(1): 3-29.

Rawat, M., Uppal, M., Newton, G., Steffek, M., Fahey, R. C. and Av-Gay, Y. (2004) Targeted mutagenesis of the *Mycobacterium smegmatis mca* gene, encoding a mycothiol-dependent detoxification protein. *J Bacteriol* 186(18): 6050-8.

Roback, P., Beard, J., Baumann, D., Gille, C., Henry, K., Krohn, S., Wiste, H., Voskuil, M. I., Rainville, C. and Rutherford, R. (2007) A predicted operon map for *Mycobacterium tuberculosis*. *Nucleic Acids Res* 35(15): 5085-95.

Rosche, W. A. and Foster, P. L. (2000) Determining mutation rates in bacterial populations. *Methods* 20(1): 4-17.

Rosner, J. L. and Storz, G. (1994) Effects of peroxides on susceptibilities of *Escherichia coli* and *Mycobacterium smegmatis* to isoniazid. *Antimicrob Agents Chemother* 38(8): 1829-33.

Russell, D. G. (2007) Who puts the tubercle in tuberculosis? *Nat Rev Microbiol* 5(1): 39-47.

Saito, Y., Uraki, F., Nakajima, S., Asaeda, A., Ono, K., Kubo, K. and Yamamoto, K. (1997) Characterization of endonuclease III (*nth*) and endonuclease VIII (*nei*) mutants of *Escherichia coli* K-12. *J Bacteriol* 179(11): 3783-5.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular cloning A Laboratory Manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press.

Scortecci, K. C., Lima, A. F., Carvalho, F. M., Silva, U. B., Agnez-Lima, L. F. and Batistuzzo de Medeiros, S. R. (2007) A characterization of a MutM/Fpg ortholog in sugarcane--A monocot plant. *Biochem Biophys Res Commun* 361(4): 1054-60.

Sherman, D. R., Sabo, P. J., Hickey, M. J., Arain, T. M., Mahairas, G. G., Yuan, Y., Barry, C. E., 3rd and Stover, C. K. (1995) Disparate responses to oxidative stress in saprophytic and pathogenic mycobacteria. *Proc Natl Acad Sci USA* 92(14): 6625-9.

Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. and Jacobs, W. R., Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* 4(11): 1911-9.

Springer, B., Sander, P., Sedlacek, L., Hardt, W. D., Mizrahi, V., Schar, P. and Bottger, E. C. (2004) Lack of mismatch correction facilitates genome evolution in mycobacteria. *Mol Microbiol* 53(6): 1601-9.

Storz, G. and Imlay, J. A. (1999) Oxidative stress. *Curr Opin Microbiol* 2(2): 188-94.

Strachan, T. and Read, A. (2003) Human Molecular Genetics. London, Garland Science.

Suvarnapunya, A. E., Lagasse, H. A. and Stein, M. A. (2003) The role of DNA base excision repair in the pathogenesis of *Salmonella enterica serovar Typhimurium*. *Mol Microbiol* 48(2): 549-59.

Suvarnapunya, A. E. and Stein, M. A. (2005) DNA base excision repair potentiates the protective effect of *Salmonella* Pathogenicity Island 2 within macrophages. *Microbiology* 151(Pt 2): 557-67.

Telenti, A., Honore, N., Bernasconi, C., March, J., Ortega, A., Heym, B., Takiff, H. E. and Cole, S. T. (1997) Genotypic assessment of isoniazid and rifampin resistance in *Mycobacterium tuberculosis:* a blind study at reference laboratory level. *J Clin Microbiol* 35(3): 719-23.

Termini, J. (2000) Hydroperoxide-induced DNA damage and mutations. *Mutat Res* 450(1-2): 107-24.

Toossi, Z. and Ellner, J. J. (1998) Host response to *Mycobacterium tuberculosis*. *Front Biosci* 3: e133-40.

Voet, D. and Voet, J. (1995) Chapter 31.5. *Biochemistry*. New York, John Wiley & Sons, Inc.

Walkup, L. K. and Kogoma, T. (1989) *Escherichia coli* proteins inducible by oxidative stress mediated by the superoxide radical. *J Bacteriol* 171(3): 1476-84.

Wallace, S. S. (2002) Biological consequences of free radical-damaged DNA bases. *Free Radic Biol Med* 33(1): 1-14.

Wallace, S. S., Bandaru, V., Kathe, S. D. and Bond, J. P. (2003) The enigma of endonuclease VIII. *DNA Repair (Amst)* 2(5): 441-53.

Wang, D., Kreutzer, D. A. and Essigmann, J. M. (1998) Mutagenicity and repair of oxidative DNA damage: insights from studies using defined lesions. *Mutat Res* 400(1-2): 99-115.

Warner, D. F. (2005). DNA metabolism in mycobacteria. Molecular Mycobacterial Research Unit. Johannesburg, University of the Witwatersrand. *PhD Thesis*.

WHO (2007). Global tuberculosis control - surveillance, planning, financing. WHO Report 2007. Geneva, Switzerland World Health Organization.

Wiederholt, C. J., Patro, J. N., Jiang, Y. L., Haraguchi, K. and Greenberg, M. M. (2005) Excision of formamidopyrimidine lesions by endonucleases III and VIII is not a major DNA repair pathway in *Escherichia coli*. *Nucleic Acids Res* 33(10): 3331-8.

Wilson III, D. M., Engleward, B. and Samson, L. (1998) Prokaryotic Base Excision Repair. *DNA Damage and Repair*. Nickoloff, J. A. and Hoekstra, M. F. Totowa, New Jersey, Humana Press.

Young, D. and Dye, C. (2006) The development and impact of tuberculosis vaccines. *Cell* 124(4): 683-7.

Young, D. and Robertson, B. (2001) Genomics: leprosy - a degenerative disease of the genome. *Curr Biol* 11(10): R381-3.

Zahrt, T. C. and Deretic, V. (2002) Reactive nitrogen and oxygen intermediates and bacterial defenses: unusual adaptations in *Mycobacterium tuberculosis*. *Antioxid Redox Signal* 4(1): 141-59.

Zhang, Q. M., Miyabe, I., Matsumoto, Y., Kino, K., Sugiyama, H. and Yonei, S. (2000) Identification of repair enzymes for 5-formyluracil in DNA. Nth, Nei, and MutM proteins of *Escherichia coli*. *J Biol Chem* 275(45): 35471-7.

Zharkov, D. O., Golan, G., Gilboa, R., Fernandes, A. S., Gerchman, S. E., Kycia, J. H., Rieger, R. A., Grollman, A. P. and Shoham, G. (2002) Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *Embo J* 21(4): 789-800.

Zharkov, D. O., Shoham, G. and Grollman, A. P. (2003) Structural characterization of the Fpg family of DNA glycosylases. *DNA Repair (Amst)* 2(8): 839-62.

Zhou, C., Yang, Y. and Jong, A. Y. (1990) Mini-prep in ten minutes. *Biotechniques* 8(2): 172-3.