Identification of Genes involved in the Intracellular Survival of *Mycobacterium tuberculosis*

Russell James Hobson B.Sc.(Hons) M.Sc.


School of Biological Sciences,
University of Surrey,
Guildford,
Surrey,
GU2 5XH.
ABSTRACT

Little is known about the survival mechanics of *Mycobacterium tuberculosis* within the human host. A number of genes are likely to be activated during infection, especially within macrophages, and these potential virulence determinants will be of great importance in the ultimate development of new vaccines and antimycobacterials.

In an effort to identify some of these virulence determinants, a *M. tuberculosis* H37Rv gene promoter-probe library in a BCG host was constructed using an *E. coli* -mycobacterial shuttle vector with *lacZ* as the reporter gene. 4800 individual clones were arrayed in a 96 well microtitre™ format, enabling the testing of individual clones for promoter activity under a variety of environmental stressing conditions. Two separate *in vitro* screenings of the arrayed clones identified 53 clones that exhibited up-regulation of *lacZ* when stressed. 50 of these clones were bi-directionally sequenced to determine the cloned *M. tuberculosis* H37Rv DNA sequence. Identification of the location of the cloned DNA within the *M. tuberculosis* H37Rv genome identified 32 clones likely to containing potential promoter sequences. By measuring the levels of *lacZ* expression post-infection of macrophages, compared to *in vitro* expression, clones were identified which contained potential promoter sequences up-regulated within the macrophage.

Quantitative Reverse Transcriptase-PCR was employed, using cDNA obtained from *M. tuberculosis* from infected and un-infected macrophages, to measure the levels of expression of 5 *M. tuberculosis* H37Rv genes identified as potentially up-regulated during infection. The results suggest that two of the five genes, genes Rv1265 and Rv2711, are indeed up-regulated during infection. The protein coded for by the gene Rv1265 has an unknown function whilst the gene Rv2711 encodes the protein IdeR, which is an iron-dependent repressor.

Sequence and *lacZ* expression analysis identified four independently selected and screened clones which contained near identical sequences for the same promoter – that of the gene Rv0440 (*groEL2*). Gene Rv0440 codes for a 60KDa heat shock protein previously identified as up-regulated during infection. The identification of four independent clones containing the promoter for this gene validates this technology as a means of identifying *M. tuberculosis* H37Rv genes up-regulated during infection.
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Jeremy Dale, for all his help, guidance and wisdom throughout the course of this study. I would also like to thank Drs. Karen Kempsell and Ken Duncan, both from GlaxoWellcome, and Drs. Alan McBride, Graham Stewart and Grainne Sheridan from the University of Surrey.

Many thanks to all technical staff and members of the Molecular Microbiology Group and the School of Biological Sciences, past and present, who have given me invaluable assistance during the research project. Thanks to Sue Shaw-Hawkins for performing all the automated DNA sequencing and Dr. Mick Frank for the digital photography.

Many thanks must go to AJ and P, Simon, Tim and Tania and Steph for keeping me sane.

Finally, I am eternally grateful to my parents and family for their ever-present support, care and love.
# CONTENTS

**ABSTRACT** ................................................................................................................................. I

**ACKNOWLEDGEMENTS** ............................................................................................................ II

**CONTENTS** ................................................................................................................................... III

**LIST OF FIGURES** ....................................................................................................................... VI

**LIST OF TABLES** ......................................................................................................................... VIII

**ABBREVIATIONS** ....................................................................................................................... IX

**CHAPTER 1 GENERAL INTRODUCTION** .................................................................................. 1

1.1 INTRODUCTION TO MYCOBACTERIA ................................................................................... 2
1.2 MYCOBACTERIAL SPECIES ................................................................................................... 3
1.3 TUBERCULOSIS ....................................................................................................................... 4
1.4 PREVENTION OF TUBERCULOSIS ....................................................................................... 6
1.5 TREATMENT FOR TUBERCULOSIS ....................................................................................... 8
1.6 ENTRY AND SURVIVAL OF *M. TUBERCULOSIS* WITHIN THE MACROPHAGE .......... 10
1.7 MYCOBACTERIAL VIRULENCE FACTORS .......................................................................... 14
1.8 MOLECULAR TOOLS FOR MYCOBACTERIAL INVESTIGATION .................................. 15
   1.8.1 Gene Cloning in Mycobacteria ...................................................................................... 15
   1.8.2 Mycobacteriophages ..................................................................................................... 17
   1.8.3 Mycobacterial Genomics .............................................................................................. 18
   1.8.4 Mutagenesis in Mycobacteria ...................................................................................... 19
   1.8.5 In Vivo Expression Technology (IVET) ......................................................................... 23
   1.8.6 Gene Transfer ................................................................................................................ 24
   1.8.7 Subtractive Hybridization ............................................................................................. 25
   1.8.8 Differential Display Polymerase Chain Reaction (DD-PCR) ...................................... 26
   1.8.9 Gene Fusions ................................................................................................................ 27
1.9 AIMS ........................................................................................................................................ 28

**CHAPTER 2 CONSTRUCTION OF *MYCOBACTERIUM TUBERCULOSIS* H37Rv PLASMID LIBRARY** .......................................................................................................................... 30

2.1 INTRODUCTION ....................................................................................................................... 31
   2.1.1 Strategy .......................................................................................................................... 32
2.2 MATERIALS AND METHODS .............................................................................................. 34
   2.2.1 Oligonucleotide Primers ............................................................................................... 34
   2.2.2 Construction of Promoter Probe Shuttle Vector .......................................................... 34
   2.2.3 Construction of Positive Control for Promoter Probe Vector ...................................... 36
   2.2.4 β-galactosidase Assay with ONPG Substrate ............................................................... 37
   2.2.5 β-galactosidase Assay Optimised for BCG ................................................................. 38
   2.2.6 Preparation of *M. tuberculosis* H37Rv Genomic DNA Fragments .......................... 39
   2.2.7 *M. tuberculosis* H37Rv Library Construction ............................................................. 40
   2.2.8 Generation of *M. tuberculosis* H37Rv Library in *M. bovis* BCG Pasteur ............. 41
2.3 RESULTS .................................................................................................................................. 42
   2.3.1 Construction and Expression Analysis of Plasmids pUS1800 and pUS1808 ............... 42
   2.3.2 Construction of Mycobacterium tuberculosis Promoter Library ......................... 46
   2.3.3 Selection and Arraying of Individual *M. bovis* BCG Pasteur Library Clones ......... 50
2.4 DISCUSSION ............................................................................................................................. 54

**CHAPTER 3 IN VITRO SELECTION OF CLONES EXHIBITING UP-REGULATED PROMOTER ACTIVITY** ......................................................................................................................... 56

3.1 INTRODUCTION ....................................................................................................................... 57
   3.1.1 In Vitro *M. tuberculosis* Stress Stimuli Mimicking Intracellular Conditions .......... 57
   3.1.2 Strategy .......................................................................................................................... 60
CHAPTER 4 MACROPHAGE CELL-LINE INFECTIONS WITH MYCOBACTERIUM BOVIS
BCG PASTEUR LIBRARY

4.1 INTRODUCTION ................................................................. 80
  4.1.1 Strategy ................................................................. 80
4.2 MATERIALS AND METHODS ............................................... 81
  4.2.1 Macrophage Infection ............................................... 81
4.3 RESULTS ........................................................................ 83
  4.3.1 Macrophage Infection with the Pre-Selected Clones ........ 83
  4.3.2 Re-Infection of 5 Clones Showing Up-Regulation during Initial Infection ... 91
  4.3.3 Second Round Screening of M.tuberculosis H37Rv Plasmid Library Clones for Further Investigation ... 86
  4.3.4 Materials and Methods ............................................. 86
  4.3.5 Macrophage Infection .............................................. 90
  4.3.6 Materials and Methods ............................................. 90
  4.3.7 Macrophage Infection .............................................. 94
  4.3.8 Materials and Methods ............................................. 94
  4.3.9 Macrophage Infection .............................................. 98
  4.3.10 Materials and Methods .......................................... 98
  4.3.11 Macrophage Infection ............................................ 102
  4.3.12 Materials and Methods .......................................... 102
  4.3.13 Macrophage Infection ............................................ 106
  4.3.14 Materials and Methods .......................................... 106
  4.3.15 Macrophage Infection ............................................ 110
  4.3.16 Materials and Methods .......................................... 110
  4.3.17 Macrophage Infection ............................................ 114
  4.3.18 Materials and Methods .......................................... 114
  4.3.19 Macrophage Infection ............................................ 118
  4.3.20 Materials and Methods .......................................... 118
  4.3.21 Macrophage Infection ............................................ 122
  4.3.22 Materials and Methods .......................................... 122
  4.3.23 Macrophage Infection ............................................ 126
  4.3.24 Materials and Methods .......................................... 126
  4.3.25 Macrophage Infection ............................................ 130
  4.3.26 Materials and Methods .......................................... 130
  4.3.27 Macrophage Infection ............................................ 134
  4.3.28 Materials and Methods .......................................... 134
  4.3.29 Macrophage Infection ............................................ 138
  4.3.30 Materials and Methods .......................................... 138

CHAPTER 5 DNA SEQUENCE ANALYSIS ...................................... 107

5.1 INTRODUCTION .................................................................. 108
  5.1.1 Strategy .................................................................... 108
5.2 MATERIALS AND METHODS ............................................. 109
  5.2.1 DNA Amplification ................................................... 109
  5.2.2 DNA sequencing ..................................................... 109
  5.2.3 Plasmid recovery from M.bovis BCG Pasteur clones .......................... 109
  5.2.4 Automated DNA sequencing ................................. 220
  5.2.5 Sequence analysis .................................................. 220
  5.3 RESULTS ....................................................................... 113
  5.3.1 Sequence Analysis and the Genome ......................... 113
  5.3.2 Cloned DNA Internal to M.tuberculosis H37Rv Genes ......... 113
  5.3.3 Clones 5, 90 and 302 .................................................. 117
  5.3.4 Clone 188 .................................................................. 117
  5.3.5 Clone 233 .................................................................. 121
  5.3.6 Clone 13 .................................................................... 121
  5.3.7 Complete Promoters .................................................. 126
  5.3.8 Clones containing Probable Promoters ..................... 126
  5.3.9 Sequenced Clones not used during Infection Studies ...... 126
  5.3.10 Analysis of Cloned M.tuberculosis H37Rv DNA in the 3 Library Clones used in Second Round of Macrophage Infection ............................ 128
  5.4 DISCUSSION .................................................................. 132

CHAPTER 6 REAL TIME PCR USING TAQMAN® TECHNOLOGY .......... 138

6.1 INTRODUCTION .................................................................. 139
  6.1.1 Fluorogenic 5′ Nuclease Assay ...................................... 139
  6.1.2 Quantitative Real Time PCR without the use of Molecular Beacons .. 139
  6.1.3 Positive Control Selection ........................................ 143
  6.1.4 Strategy .................................................................... 143
6.2 MATERIALS AND METHODS ............................................ 144
  6.2.1 Primers .................................................................... 144
  6.2.2 Mycobacterium tuberculosis H37Rv RNA Extraction In Vitro .... 145
  6.2.3 Mycobacterium tuberculosis H37Rv RNA Extraction Post-infection of Cultured Human Macrophage. ........................................... 145
  6.2.4 Primer Design ........................................................ 145
  6.2.5 Reverse Transcription of Bacterial RNA ...................... 145
LIST OF FIGURES

Figure 1.1: Graph indicating the number of total tuberculosis notifications in England and Wales over recent years ................................................................. 5
Figure 2.1: Vector map of plasmid pUS1781 ................................................................................................................. 35
Figure 2.2: Vector map of plasmid pUS1800 ................................................................................................................. 36
Figure 2.3: Construction of plasmids pUS1800 and pUS1808 ......................................................................................... 43
Figure 2.4: Three β-galactosidase assays, using method according to Miller, on M. bovis BCG Pasteur harbouring pUS1800 or pUS1808................................................................. 44
Figure 2.5: β-galactosidase assays carried out at elevated temperatures ......................................................................... 45
Figure 2.6: M. tuberculosis plasmid library construction ................................................................................................. 47
Figure 2.7: Agarose gel of PCR products of discrete E. coli DH5α clones harbouring M. tuberculosis plasmid library ............................................................................. 48
Figure 2.8: M. bovis BCG Pasteur harbouring the M. tuberculosis plasmid library grown on 7H11 solid media containing kanamycin and X-Gal ........................................................................... 49
Figure 3.1: Growth curve of M. bovis BCG Pasteur grown in liquid 7H9 medium at different pH values ................................................................. 65
Figure 3.2: Growth curve of M. bovis BCG Pasteur grown in liquid 7H9 medium at pH 5.4 with varying concentrations of sodium nitrite ............................................................................................................. 66
Figure 3.3: Scale of intensity of blue colouration generated by M. bovis BCG Pasteur clones harbouring the M. tuberculosis promoter trap vector when grown on solid 7H11 medium containing X-Gal ................................................................. 68
Figure 3.4: Graph of light units/ 10^6 bacterial cells against the M. bovis Pasteur clone number .......................................... 74
Figure 3.4: (continued) .................................................................................................................................................. 75
Figure 4.1: Bar chart of light units/ 10^6 bacterial cells for 4 clones, numbers 180, 233, 239 and 254 and control pUS1800 ................................................................................................................. 90
Figure 4.2: Bar chart of light units/ 10^6 bacterial cells against time of infection in human THP-1 macrophages for clone 169 ................................................................................................................................................. 95
Figure 4.3: Bar chart of light units/ 10^6 bacterial cells against time of infection in human THP-1 macrophages for clone 175 ................................................................................................................................................. 95
Figure 4.4: Bar chart of light units/ 10^6 bacterial cells against time of infection in human THP-1 macrophages for clone 176 ................................................................................................................................................. 96
Figure 4.5: Bar chart of light units/ 10^6 bacterial cells against time of infection in human THP-1 macrophages for clone 180 ................................................................................................................................................. 96
Figure 4.6: Bar chart of light units/ 10^6 bacterial cells against time of infection in human THP-1 macrophages for clone 239 ................................................................................................................................................. 97
Figure 4.7: Bar chart showing a comparison of the light units/ 10^6 bacterial cells for clone 169 when assayed at different times ....................................................................................................................... 101
Figure 4.8: Bar chart comparing clone 169 samples grown in liquid 7H9 medium and assayed at the same time ................................................................................................................................. 102
Figure 5.1: Map of the M. tuberculosis H37Rv genome ................................................................................................. 114
Figure 5.2: Diagrammatic representation of all sequenced clones ......................................................................................... 117
Figure 5.3: Diagrammatic representation of DNA in the vicinity of cloned DNA contained within clone 7 ................................................................................................................................................. 119
Figure 5.4: Diagrammatic representation of DNA in the vicinity of cloned DNA contained within clone 302 ................................................................................................................................................. 119
Figure 5.5: Diagrammatic representation of M. tuberculosis H37Rv DNA in the plasmid pUS1800 harbouring within clone 188 ................................................................................................................................................. 121
Figure 6.1: Denaturing gel of total RNA isolated from M. tuberculosis H37Rv grown under different culture conditions ................................................................................................................................. 126
Figure 6.2: Graph taken directly from Taqman® analysis software ......................................................................................... 154
Figure 6.3: Graph of Ct value against amount of M. tuberculosis H37Rv genomic DNA added to each real time PCR ................................................................................................................................................. 156
Figure 6.4: Graph of Ct value against amount of M. tuberculosis H37Rv genomic DNA added to each real time PCR ................................................................................................................................................. 157
Figure 6.5: Graph of $C_T$ value against the number of the gene amplified from cDNA generated from mRNA obtained from *M. tuberculosis* H37Rv incubated in 4 different growth conditions.
LIST OF TABLES

Table 1.1: Brief list of mycobacterial species .................................................. 3
Table 3.1: Table of six M.bovis BCG Pasteur clones harbouring the M.tuberculosis promoter trap vector ......................................................... 68
Table 4.1: The clone number is shown against the M.tuberculosis H37Rv DNA inserted into the NarI site of PUS1800 harboured within the clone ........................................ 93
Table 4.2: A copy of the data from Appendix 3 showing the ratio of apparent up-regulation exhibited by clones 169, 175, 176, 180 and 239 ........................................ 105
Table 5.1: Table of all genes whose putative promoters or parts of putative promoters have been cloned into PUS1800 .............................................. 137
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Absorbance</td>
</tr>
<tr>
<td>A</td>
<td>Adenosine</td>
</tr>
<tr>
<td>AIDS</td>
<td>Acquired Immuno-deficiency Syndrome</td>
</tr>
<tr>
<td>BCG</td>
<td>Bacille Calmette-Guerin</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>G</td>
<td>Guanidine</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immuno-deficiency Virus</td>
</tr>
<tr>
<td>IVET</td>
<td><em>In vivo</em> expression technology</td>
</tr>
<tr>
<td>Kan</td>
<td>Kanamycin</td>
</tr>
<tr>
<td>Kb</td>
<td>Kilobase pairs</td>
</tr>
<tr>
<td>KDa</td>
<td>KiloDaltons</td>
</tr>
<tr>
<td><em>lacZ</em></td>
<td><em>Escherichia coli</em> β-galactosidase gene</td>
</tr>
<tr>
<td>LU</td>
<td>Light Units</td>
</tr>
<tr>
<td>nm</td>
<td>nanometers</td>
</tr>
<tr>
<td>ONPG</td>
<td>o-Nitrophenyl-β-D-galactopyranoside</td>
</tr>
<tr>
<td>OriE</td>
<td><em>Escherichia coli</em> origin of replication</td>
</tr>
<tr>
<td>OriM</td>
<td>Mycobacterial origin of replication</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PHLS</td>
<td>Public Health Laboratory Services</td>
</tr>
<tr>
<td>RNI</td>
<td>Reactive nitrogen intermediates</td>
</tr>
<tr>
<td>ROI</td>
<td>Reactive oxygen intermediates</td>
</tr>
<tr>
<td>STM</td>
<td>Signature tagged mutagenesis</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>TB</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>V</td>
<td>volume</td>
</tr>
<tr>
<td>W</td>
<td>weight</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health organisation</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside</td>
</tr>
</tbody>
</table>
CHAPTER 1

GENERAL INTRODUCTION
1.1 Introduction to Mycobacteria

The genus *Mycobacterium* contains bacteria that are non-sporing, non-motile and Gram positive, which exist as straight or curved rods as well as filamentous or branched forms. They have been classified in the actinomycete branch of Gram positive bacteria along with the genera *Nocardia* and *Streptomyces* since they have certain common features such as a primarily aerobic metabolism and DNA with a high GC content (55-70%). Due to an extremely high lipid content (approximately 40% of the total cell dry weight), mycobacteria are impermeable to hydrophilic agents and grow as pellicles in liquid media and hydrophobic colonies on solid media. The high lipid content is also responsible for staining difficulties encountered with conventional stains such as the Gram stain. However, once stained, the cells withstand decolourisation with 95% (v/v) ethanol or with 3% (v/v) hydrochloric acid. This is the basis of the Ziehl-Neelsen stain, which utilises this unique property of acid-fastness in the identification of mycobacteria.

1.2 Mycobacterial species

*Mycobacterium* is the only genus in the Mycobacteriaceae family and contains over forty species, which can be divided into two main groups: the fast-growers and the slow-growers of which a selection of species is shown in Table 1.1. The *Mycobacterium* genus are infamous for their slow growth with doubling times ranging from 3 to 24 hours meaning that the slow-growing species can take up to six weeks for colonies to form on solid media. The slow-growing mycobacteria include most of the pathogenic species with the fast-growers consisting mainly of environmental and saprophytic species. *Mycobacterium leprae* is an exception to
these two groups because, to this day, it has never been cultivated \textit{in vitro} and can only be cultivated in infected animals, specifically the nine-banded armadillo and from the footpads of laboratory mice.

<table>
<thead>
<tr>
<th>Slow growing mycobacteria</th>
<th>Fast growing mycobacteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>The \textit{M.tuberculosis} complex:</td>
<td>\textit{M.fortuitum}</td>
</tr>
<tr>
<td>\textit{M.tuberculosis}</td>
<td>\textit{M.smegmatis}</td>
</tr>
<tr>
<td>\textit{M.bovis}</td>
<td>\textit{M.vaccæ}</td>
</tr>
<tr>
<td>\textit{M.africanum}</td>
<td>\textit{M.aurum}</td>
</tr>
<tr>
<td>\textit{M.microti}</td>
<td></td>
</tr>
</tbody>
</table>

The \textit{M.avium} complex:

| \textit{M.avium} complex: | |
| \textit{M.avium} | |
| \textit{M.paratuberculosis} | |
| \textit{M.lepraemurium} | |
| \textit{M.intracellulare} | |

\textit{M.lepræ}  

Table 1.1: Brief list of mycobacterial species.

On the whole mycobacteria are soil and water dwelling bacteria. However, certain species have adapted to life within a host resulting in infection, the best known being \textit{Mycobacterium tuberculosis} and \textit{Mycobacterium lepræ}, which are the aetiological agents of tuberculosis (TB) and leprosy, respectively. This level of adaptation ranges from rare opportunistic pathogens to obligate parasites. Members of the \textit{M.tuberculosis} complex can also cause TB in animals as well as humans, for example \textit{M.bovis} is responsible for the bovine form of the disease and \textit{M.microti} is known to cause TB specifically in rodents. Many pathogenic species are included in the \textit{M.avium} complex with the most clinically important being \textit{M.avium} itself. \textit{M.avium} is an opportunistic pathogen that has been shown to cause TB in immunocompromised individuals, particularly those with Acquired Immune Deficiency Syndrome (AIDS) (Collins, 1993), but is essentially the aetiological agent for TB in birds. A number of the fast growing mycobacterium species, such as
M. fortuitum and M. smegmatis, have been shown to be rare opportunistic pathogens that are occasionally associated with soft tissue and joint infections in humans (Newton et al., 1993). M. fortuitum and M. smegmatis are more frequently used as model species within the laboratory to study mycobacterial biology and genetics.

1.3 Tuberculosis

Approximately 8 million new cases and 3 million deaths caused by tuberculosis infections are reported annually to the World Health Organisation (WHO) meaning that tuberculosis remains one of the most important infectious bacterial diseases in the world. As much as one third of the world's population may be infected with M. tuberculosis, predominantly in the third world. This means that 30 million people could die from TB in the next 10 years and someone is infected with TB every second.

Due to scientific advancement, chemotherapy, vaccination and improved living standards, the incidence of pulmonary TB in the western world has been declining steadily over the last 150 years. In England and Wales there were 117,139 tuberculosis notifications in 1913, whilst in 1987 the number of notifications had dropped to 5,085 (www.phls.co.uk/facts/TB/). However in recent years, tuberculosis in the western world has increased slightly, particularly among the homeless population, largely as a result of the AIDS epidemic, which renders those immunocompromised individuals highly susceptible to TB (Collins, 1993). Figure 1.1 uses data obtained from the Public Health Service Laboratories (PHLS) website (www.phls.co.uk/facts/TB/) to
indicate the increase in total tuberculosis notifications in England and Wales in recent years.

![Graph indicating the number of total tuberculosis notifications in England and Wales over recent years. Data obtained from the Public Health Service Laboratories (PHLS) website (www.phls.co.uk/facts/TB/)](Figure 1.1)

*Figure 1.1: Graph indicating the number of total tuberculosis notifications in England and Wales over recent years. Data obtained from the Public Health Service Laboratories (PHLS) website (www.phls.co.uk/facts/TB/)*

*M. tuberculosis* is spread by aerosols, normally generated by coughing. Once inhaled by the host, alveolar macrophages phagocytose the bacteria. *M. tuberculosis* is able to replicate intracellularly and then from the macrophages spread throughout the body via the lymphatic system. Most individuals do not develop the disease during this initial stage of infection; only 5% of individuals proceed to develop the symptoms of primary tuberculosis. In the latter cases, the lesions may progress to form cavities within the lung as a result of tissue damage. However, the spread of the
disease within the body can produce lesions in any of the organs resulting in, for example, tuberculous meningitis or genitourinary tuberculosis, which are not particularly common. Multiple lesions throughout the body are caused by disseminated tuberculosis, which arises because of the inadequacy of the host defences in containing the infection. Therefore, in addition to immuno-compromised AIDS sufferers, tuberculosis predominately affects children in the third world presumable due to their immunological immaturity.

Post-primary pulmonary tuberculosis is generally regarded as the most frequent form of the disease. Disease symptoms develop several years following the initial infection usually as a result of temporary immunosuppression due to old age or illness. Bacteria that were previously dormant begin to replicate within the granuloma, which results in the discharge of live, virulent bacteria in the bronchial tree. Coughing expels these bacteria and therefore completes the infection cycle (Hopewell, 1994).

1.4 Prevention of Tuberculosis

Bacille Calmette-Guérin (BCG) is a live attenuated strain of \textit{M.bovis} and is used for vaccination as the current preventative measure against tuberculosis infection in the United Kingdom. North America and the Netherlands, in particular, do not use the BCG vaccine on a national scale (Fine, 1989) but rather rely on skin testing to identify infected individuals, who are then given prophylactic treatment. BCG is administered to populations world-wide with varying levels of efficacy and acquired protective immunity. Although the average efficacy of BCG as a vaccine is stated to
be 50% (Fine, 1998), this figure varies dramatically according to geographic location. For example, a 70-80% efficacy for the vaccine has been reported in the United Kingdom (Hart and Sutherland, 1977) whereas in South India the efficacy is only 0-30% (Tuberculosis Prevention Trial, 1980). The variation in vaccine efficacy has been blamed on a number of factors, including socioeconomic conditions, genetic composition of the population, climate, diet and nutrition and the presence of nontuberculosis mycobacteria (such as the *Mycobacterium avium* complex) in the environment. However the true reason for the variation remains unclear.

A further problem caused by the use of the current vaccine is that it can, indeed generally does, cause the tuberculin skin test to become reactive, which reduces the usefulness of this particular diagnostic test for *M. tuberculosis* infections (Fine, 1989). The tuberculin skin test is routinely used in the identification of those potentially infected with *M. tuberculosis* and a delayed-type hypersensitive (DTH) reaction post-BCG-vaccination reduces the efficacy of this use of the tuberculin skin test. This is presumably why some countries prefer not to use BCG as a vaccine for TB.

The search for a new and improved vaccine against tuberculosis has employed the use of a number of different strategies. Subunit vaccination based on major secretory proteins from *M. tuberculosis* has provided protective immunity in animal models (Horwitz *et al.*, 1995). Strong humoral and cellular immune responses to a variety of *M. tuberculosis* antigens have been demonstrated in murine models (Stover *et al.*, 1991; Bosio and Orme, 1998). The use of DNA as a vaccination method by plasmid vectors to deliver genes coding for antigens is also being investigated
Due to the relatively low efficacy of BCG as a vaccine, the diagnosis and quick and effective treatment of *M. tuberculosis* infection is of paramount importance. The primary chemotherapeutic agents for the treatment of tuberculosis are the antibiotics rifampicin and isoniazid. These antibiotics are favoured for their high efficiency and low toxicity and are used in conjunction with secondary drugs, such as ethambutol, pyrazinamide, *p*-aminosalicylic acid and streptomycin, which are given in combination to prevent the emergence of antibiotic resistant bacilli. This treatment is effective in rendering patients non-infectious, however, treatment must be continued for extended periods to prevent relapse. The standard drug regime requires 12 months of treatment whereas the current preferred regime of treatment is DOTS (Directly Observed Treatment Short course) which lasts for 6-8 months. DOTS is a World Health Organisation (WHO) initiative and has been proven to be successful in countries such as Bangladesh, Vietnam, Peru and the countries of West Africa. There are five elements to the DOTS strategy stated below and obtained from the WHO website (http://www.who.int/gpv-dvacc/diseases/TB.htm):
1) Case detection by sputum smear microscopy among symptomatic patients self-reporting to health services.

2) Standardised treatment regimen of 6-8 months for at least all smear positive cases, with directly observed treatment (DOT) for at least the initial 2 months.

3) A regular, uninterrupted supply of all essential anti-TB drugs.

4) A standardised recording and reporting system that allows assessment of treatment results for each patient and of the TB control programme overall.

5) Government commitment to sustained TB control activities.

The lengthy treatment requirements immediately impose difficult logistic, medical and financial burdens for disease control, particularly for those Third World countries where tuberculosis is endemic and where there is poor compliance to treatment. However, DOTS, is proving successful when it is implemented.

One of the alarming consequences of the lack of compliance is the emergence of multidrug-resistant strains of *M. tuberculosis* (MDR-TB), which are associated with high mortality. Multidrug resistance of *M. tuberculosis* has been defined as resistance to rifampicin and isoniazid (Bradford *et al.*, 1998), but other strains have been shown to be resistant to streptomycin (Edlin *et al.*, 1992). The highest rates of acquired MDR-TB have been reported in Nepal (48.0%), Gujarat, India (33.8%) and, surprisingly, New York City (30.1%), with the figures indicating the percentage of total isolates tested which were found to be multidrug resistant (Cohn *et al.*, 1997). MDR-TB outbreaks are associated with very high death rates and more than 80% of cases occur in patients co-infected with HIV (Dooley *et al.*, 1992), however the high mortality rates may be due to the high predominance of HIV. Delayed recognition of
drug resistance, which results in delayed initiation of effective treatment, is one of the major factors contributing to MDR-TB outbreaks.

With the increased incidence of tuberculosis and of multidrug resistant tuberculosis in particular, the identification of methods employed by \textit{M. tuberculosis} to survive the host defence mechanism is essential in the development of both a new, more effective vaccine and new drug treatments. Therefore the identification of \textit{M. tuberculosis} virulence factors and factors responsible for dormancy in \textit{M. tuberculosis} (thought to be the cause of post-primary pulmonary tuberculosis) need to be identified. In addition, identification of specific factors involved in the host response to \textit{M. tuberculosis} would also be of value for the development of an immuno-potentiating therapy.

1.6 Entry and Survival of \textit{M. tuberculosis} within the Macrophage.

\textit{Mycobacterium tuberculosis} requires entry into the host macrophage in order to initiate infection. However, there is much debate and confusion as to exactly how mycobacteria enter the host cell and the conditions that the bacteria are exposed to once they have gained entry. It is believed that there are a number of different receptors that the bacterium binds to in order to gain entry. \textit{M. tuberculosis} can become sufficiently coated with serum-derived ligands, to allow binding to various types of complement receptors (CR1, CR3 and CR4) and is subsequently phagocytosed and enclosed in membrane-bound endosomes. Mycobacteria also display numerous and diverse ligands on their surface, such as capsular polysaccharides, which are likely to engage a number of different receptor types. \textit{In}
vivo, *M.tuberculosis* is unlikely to be internalised by macrophages via a single receptor-mediated pathway. However, it appears that the extent of survival or the rate of intracellular growth is unaltered by the type of receptor used to gain entry (Ernst, 1998).

Material passing through the endosomal network to the lysosome of higher eukaryotes usually experiences a drop in pH. However, whether pathogenic mycobacteria species remain within phagosomes after uptake is unclear. The majority of evidence suggests that the mycobacteria reside almost exclusively intravacuolar (Clemens and Horwitz, 1995). However, it has also been noted that once internalised, phagosomes, containing mycobacteria, fuse with lysosomes, in fact within 2 hours of phagocytosis 85% of the bacteria reside in fused phagolysosomes (McDonough *et al.*, 1993). Unlike other bacteria such as *Brucella suis* which requires a low pH for survival in murine macrophages, the fused phagolysosome containing *M.tuberculosis* equilibrates to pH 6.3-6.5, due to the exclusion of the vesicular proton-adenosine triphosphatase (ATPase), from the phagosome membrane, which is responsible for phagosome acidification (Porte *et al.*, 1999; Sturgill-Koszycki *et al.*, 1994).

The above suggests that the mycobacterium-containing vacuole behave differently to vacuoles containing other bacteria and that vacuoles containing mycobacteria fail to mature into phagosomes containing a low pH. However Sturgill-Koszycki *et al.* (1996) also noted that vacuoles containing mycobacteria are positive for lysosome associated membrane protein (LAMP-1). LAMP-1 is a sialic acid-rich membrane protein that is usually associated with late endosomes and lysosomes of mammalian cells. It has subsequently been demonstrated that vacuoles
containing mycobacteria do interact with early endosomes by demonstrating the presence of transferrin (used to internalise iron in mammalian cells) on M. avium-infected murine macrophages (Sturgill-Koszycki et al., 1996).

A further discrepancy in vacuole maturation exists. Macrophages possess an innate resistance to bacterial infection that is associated with a single gene encoding Nramp (natural resistance-associated macrophage protein). Arias et al. (1997) used congenic resistant and sensitive murine macrophages (which express or do not express Nramp) and showed that the sensitive macrophages (not expressing Nramp) exhibited a reduced ability to stop the growth of M. tuberculosis. It has been suggested that Nramp plays a role in ion transport and therefore plays a role in decreasing superoxide dismutase activity (Supek et al., 1997). An Nramp rat homologue (DCT-1) has been shown to be an unusually broad-range ion transporter, which is proton-coupled and requires a membrane potential (Gunshin et al., 1997). The Nramp protein has also been shown to be present in endosomal and phagosomal membranes (Gruenheid et al., 1997).

Whether the vacuoles containing mycobacteria are acidified is therefore unclear. Data suggests that the vacuoles do not contain the vesicular proton-adenosine triphosphatase (ATPase), and therefore acidic conditions are believed not to exist within the vacuole, but do contain the Nramp protein which requires a membrane potential in order to function as an ion transporter.

However, the main method employed by the macrophage to kill internalised mycobacteria is believed to be through the use of free radicals such as reactive
oxygen intermediates (ROI) and reactive nitrogen intermediates (RNI). Adams et al. (1997) studied the role of ROI and RNI in the host response to *M. tuberculosis* with the use of two strains of knockout (KO) mice. One strain of KO mouse was unable to produce ROI whilst the second strain lacked an inducible nitric oxide synthase (iNOS) and therefore failed to make RNI. The survival of *M. tuberculosis* in both KO mice was increased compared to control mice. In addition, mice infected with a virulent Erdmann strain of *Mycobacterium tuberculosis* and given inhibitors of nitric oxide synthase showed an increase in morbidity, bacterial burden and pathological tissue damage (Chan et al., 1995). In a separate study, congenic resistant and susceptible murine macrophages were infected with *M. tuberculosis* *in vitro*. The resistant macrophages inhibited the incorporation of radioisotopes by the internalised mycobacteria and were also observed to be producing higher amounts of nitric oxide compared to the susceptible macrophages (Arias et al., 1997).

If ROI and RNI are utilised by the macrophage to kill *M. tuberculosis*, the generation of these compounds would be a rapid response by the macrophage so that the infection can be eliminated as quickly as possible. Nitric oxide can be induced in mice with either live or heat-killed *M. bovis* BCG – with nitrite levels reaching a maximum after 7 days (Saito and Nakano, 1996). With *M. tuberculosis* grown in monocytes cultures *in vitro*, soluble nitrite in the medium can be detected as early as 72 hours post-infection and levels increase progressively up to day 10. The induction of nitrite correlates with an increase in inducible nitric oxide synthase (iNOS) mRNA (Jagannath et al., 1998).
This clearly indicates that there is much confusion and lack of real understanding regarding the conditions internal to mycobacteria-containing vacuoles and shows that there is still a considerable need to research this particular area of mycobacterial infection.

1.7 Mycobacterial Virulence Factors

There is a considerable volume of literature including a number of reviews detailing the different techniques that can be used in the identification of virulence factors in bacteria, and mycobacteria in particular (Clark-Curtiss, 1998; Gordon and Andrew, 1996; Quinn et al., 1997; Shinnick et al., 1995; Marston and Shinnick, 1996; Pelicic et al., 1998; Mahan et al., 1993). However only a few mycobacterial genes associated with virulence have actually been identified. Until recently, only two genes associated with virulence determinants had been identified. Collins et al. (1995) first noted that the sigma factor gene rpoV (subsequently annotated as Rv2703/ sigA (Cole et al., 1998)) of the M.tuberculosis complex was associated with virulence. Since it is believed to be the principal sigma factor and is therefore involved in the expression of both house keeping genes as well as potential important virulence factors, it is perhaps not entirely surprising that a mutation in rpoV causes a loss of virulence. Another gene associated with virulence is katG, since isoniazid resistant strains of M.tuberculosis have a reduced virulence in Guinea pigs (Zhang and Young, 1993). Over the last couple of years and with the aid of new investigative techniques, the flow of reported virulence factors has greatly increased. Biosynthetic enzymes involved in the assembly of virulence-conferring siderophores (Quadri et al., 1998), the macrophage-induced gene (mig) of M.avium (Meyer et al., 1998), the
M. tuberculosis erp gene which encodes an exported protein (Berthet et al., 1998) have all been associated with virulence in mycobacteria, to name but a few.

1.8 Molecular Tools for Mycobacterial Investigation.

The development of a number of effective genetic tools has considerably expanded the knowledge of mycobacteria by allowing the characterisation of genes and proteins. These molecular genetic tools are proving invaluable, particularly for the study of M. tuberculosis, by allowing the identification of potential drug targets, characterisation of virulence determinants and ultimately the construction of attenuated strains.

1.8.1 Gene Cloning in Mycobacteria

The use of mycobacterial plasmids and mycobacterial-E. coli shuttle vectors in mycobacterial research is a relatively new development but their use is now widespread and is required for a number of techniques that are currently being used to identify potential virulence factors.

Mycobacterial plasmids were first reported in the M. avium-intracellulare complex (Crawford and Bates, 1979) and these vectors are now routinely used for gene cloning in mycobacteria. The most frequently used replicon is from the M. fortuitum plasmid pAL5000 (Labidi et al., 1984). A number of selectable antibiotic markers have been inserted into this vector to confer resistance to antibiotics such as...
streptomycin, hygromycin and kanamycin and allow subsequent selection after transformation.

The replicon in pAL5000 (4837bp) is the basis for shuttle vectors in genetic manipulation in both slow and fast growing mycobacteria. The sequence of pAL5000 was first reported by Rauzier et al. (1988). However, the minimum region from pAL5000 required for replication has now been identified as being approximately 1.6Kb, which includes two open reading frames present in tandem and designated repA and repB (Stolt and Stoker, 1996b; Stolt and Stoker, 1996a). These open reading frames are thought to be transcribed as an operon and overlap by 1bp. Located within a region with a high incidence of repeated sequences and directly up-stream of repA is a 435bp segment that contains the origin of replication (ori) (Stolt and Stoker, 1996a). Both RepA and RepB proteins are required for plasmid replication, with RepB binding to a high affinity site within its own promoter region and to a low affinity site within the ori region (Stolt and Stoker, 1996b). Therefore the RepB protein auto-regulates its own expression and also plays a role in the initiation of plasmid replication. Although essential for plasmid replication, the role of the RepA protein is unclear.

Plasmid DNA is effectively introduced into electrocompetent M.smegmatis mc²155 by electroporation yielding $10^4 - 10^5$ transformants μg⁻¹ of plasmid DNA (Snapper et al., 1990). Electroporation of slow growing mycobacteria, such as M.bovis BCG, is currently carried out at elevated temperatures, compared to electroporation of M.smegmatis, and can generate as many as $10^6$ transformants μg⁻¹ of plasmid DNA (Wards and Collins, 1996).
1.8.2 Mycobacteriophages

Mycobacteriophages are viruses of mycobacteria and over 250 mycobacteriophages are currently known. The best characterised is the temperate mycobacteriophage L5. The entire L5 genome has been sequenced (Hatfull and Sarkis, 1993) and the genes involved in integration and lysogeny have been identified, including the integrase (int) and attachment (attP) genes and the chromosomal attachment site attB. Mycobacteriophages can be used to deliver DNA to mycobacteria, generate mutant libraries or study the survival characteristics of mycobacteria in the presence of antibiotics. Plasmids can be introduced and stably maintained by site-specific integration into the genomes of *M. smegmatis*, *M. tuberculosis* and BCG by mycobacteriophages and such a system can be used for the construction of recombinant BCG vaccines (DasGupta *et al.*, 1998; Lee *et al.*, 1991).

Phasmids are chimeric molecules capable of replicating as plasmids in *E. coli* and as phages in mycobacteria (Jacobs Jr *et al.*, 1987), which are capable of delivering genetic elements such as reporter genes or transposons to the bacteria. Using luciferase reporter phages, Jacobs *et al.* (1993) showed that *M. tuberculosis* infected with this phage resulted in substantial light production which was sensitive to the presence of antibiotics and could therefore be used in the testing of novel drugs.
1.8.3 Mycobacterial Genomics

Undoubtedly the greatest advancement in mycobacterial research in recent years has been the sequencing of the complete *M. tuberculosis* genome which is 4,411,529bp long and contains around 4000 genes (Cole *et al.*, 1998). The genome sequence can be accessed from the World Wide Web at www.sanger.ac.uk/projects/M_tuberculosis/. These databases can be used to search for homologies of *M. tuberculosis* genes with genes from other organisms, which can enable the identification of the potential function of *M. tuberculosis* proteins. A physical and genetic map of the genome displaying the currently known genes and coding sequences can also be found at the Sanger centre web site. A number of novel techniques are currently being used to exploit the genome sequence data.

High density DNA arrays that contain probes for all potentially expressed *M. tuberculosis* genes can be used for the identification of genes that are differentially expressed under varying environmental conditions which the organism is likely to encounter during infection. Simply, the entire complement of genes, contained within a genome of an organism, can be arrayed in an ordered fashion on slides, membranes or 'chips' and mRNA obtained from the organism can be used for the generation of cDNA probes to be hybridized to the array. Since the cDNA can be labelled with a fluorescence marker, it can be detected via dual-laser scanning confocal microscopy. If the mRNA is obtained from the organism grown in different conditions (such as *in vitro* and intra-macrophage in the case of *M. tuberculosis*), a comparison between the levels of fluorescence of each individual gene can be compared. It would be expected that certain genes would remain constitutively expressed while the expression of
others would be altered by the different growth conditions. This has recently been used to show that HIV alters the expression of human host cells during infection (Geiss et al., 2000).

High density arrays have also been used to identify the genetic composition and gross genome differences between \textit{M.tuberculosis}, \textit{M.bovis} and various \textit{M.bovis} BCG daughter strains (Behr \textit{et al.}, 1999). Microarray technology is now used to investigate the \textit{in vivo} \textit{in vitro} expression differences with organisms for which we have the entire genome sequence. Recently, however, microarray technology has been applied to the parasite \textit{Plasmodium falciparum}, the causative agent of malaria, for which the entire genome sequence is not available (Hayward \textit{et al.}, 2000). A randomly generated nuclease genomic library was arrayed and through differential hybridization and sequencing, large differences in gene expression were identified between the blood stage form and the sexual stage form of the parasite.

Analysis of the total protein content or ‘proteome’ analysis can also be used to study the characterisation of proteins expressed under differing conditions by two-dimensional gel electrophoresis as with the high-density DNA arrays. Two-dimensional gel electrophoresis is used to separate all of the proteins expressed by an organism. Differentially expressed proteins, obtained from the organism grown in different conditions, can be remove from the gel and sequenced using in-gel trypsin digests and time-of-flight mass spectometry. The protein sequences obtained can be used to identify the genes that encode them. Two-dimensional gel electrophoresis has been used to identify proteins differentially expressed by \textit{M.tuberculosis} (Garbe \textit{et al.}, 1996), although at the time the \textit{M.tuberculosis} H37Rv genome sequence was not
available and therefore the reverse identification of the coding genes could not be carried out. Proteomics has also been used to identify genes differentially expressed under conditions of high and low concentrations of iron in *M.tuberculosis* (Wong et al., 1999) and to characterise a sulfate starvation-induced response in *Pseudomonas aeruginosa* PAO1 (Quadroni et al., 1999).

It is clear that the sequence data will remain invaluable in conjunction with new technologies, such as high density microarrays and proteomics, to further the current understanding of *M.tuberculosis*. However these techniques are high technology and are therefore very expensive to set-up and run for the majority of laboratories.

### 1.8.4 Mutagenesis in Mycobacteria

Mutagenesis involves the disruption of genes and the phenotype of the generated mutant is then compared to that of the parental wild-type strain. Mutagenesis has become an important means of characterisation of mycobacterial genes and in particular the identification of virulence determinants.

Allelic exchange exploits the homologous recombination mechanism of the host. Homologous recombination occurs between donor and target DNA molecules that have a certain level of sequence identity. Norman *et al.* (1995) first demonstrated allelic exchange in the *M.tuberculosis* complex by replacing the *accBC* target gene, encoding a subunit of biotin carboxylase, with the *aph* gene on the *M.bovis* BCG chromosome. This was carried out using a replicating shuttle vector harbouring a
counter selectable marker katG (conferred isoniazid resistance) and a copy of the
target gene, which had been inactivated by an aph gene insertion. Homologous
recombination-mediated, double cross-over events were selected for by loss of the
vector generating the replacement of the accBC target gene and therefore kanamycin
resistant clones. A further example of allelic exchange mutagenesis is that of an ideR
mutant of M. smegmatis that exhibited an altered oxidative stress response compared
to the wild-type strain (Dussurget et al., 1996).

An alternative technique used in the generation of mutants for subsequent
classification of genes is that of transposon mutagenesis. This method uses mobile
elements, called transposons, that upon transposition randomly disrupt chromosomal
genes. Mobile genetic elements have been identified in both prokaryotes and
eukaryotes and are discrete DNA sequences that can move from one location on a
dNA molecule to another location on the same or on a different molecule. This is an
illegitimate recombination event, since no sequence homology is required between
the transposon and the site of insertion, and is not dependent on the host cell’s
homologous recombination system. Mycobacteria have a number of different families
of transposons, called insertion (IS) elements, which in addition to transposon
mutagenesis are used for strain-specific probes for RFLP analysis and
epidemiological studies.

The advantage of the transposon mutagenesis technique is that no previous
assumptions are made regarding the identity of the gene responsible for a particular
phenotype, thus a library of random mutants can be generated and analysed.
Transposon mutants have been generated in M. smegmatis using a conditionally
replicating plasmid to deliver the *M.fortuitum* insertion sequence IS6100 (Guilhot et al., 1994), *M.bovis* BCG using a non-replicating plasmid to deliver the *M.smegmatis* insertion sequence IS1096 (Mcadam et al., 1995) and in *M.tuberculosis* using a conditionally replicating mycobacteriophage to deliver the transposon Tn5367 (Bardarov et al., 1997) (an IS1096-derived transposon containing a kanamycin resistance marker (Mcadam et al., 1995)).

Signature-tagged transposon mutagenesis (STM) is a relatively novel technique that can be used to identify genes associated with virulence (Hensel et al., 1995). Each signature-tagged mutation carries a different DNA tag, which allows mutants to be differentiated from each other. Transposons, flanked by highly variant and invariant flanking sequences, are delivered to the bacteria, usually by a temperature-sensitive plasmid. With the elevation of temperature, the plasmid is lost and the transposon inserts itself, randomly, into the genome of the bacteria being studied. Mutants are subsequently pooled and used in infection studies, either cell or whole animal. At an appropriate time after inoculation, bacteria are recovered from the infection and the tags in the recovered pool and the inoculation pool are used to probe filters arrayed with the different tags representing mutants in the inoculum. Mutants with an attenuated virulence are those with tags that are present in the inoculum pool but are absent from the recovered pool of tags.

Using STM, 50 mutants with attenuated virulence were identified in *Staphylococcus aureus* (Mei et al., 1997). Approximately half of the genes identified via this method had no known function in *S.aureus*. STM has also recently been used to identify a cluster of virulence genes located on a 50Kb chromosomal region of
M. tuberculosis (Camacho et al., 1999). The genes identified in M. tuberculosis are thought to be involved in the production of phthiocerol and phenolphthiocerol derivatives, which are a group of molecules restricted to eight mycobacterial species. Seven of these species are either strict or opportunistic pathogens. These molecules are thought to be involved in the biosynthesis of components of the cell envelope.

1.8.5 In Vivo Expression Technology (IVET)

IVET was first described by Mahan et al. (1993). A suicide vector containing promoterless purA gene and promoterless lacZ gene was constructed as a fusion and a partial digest of Salmonella typhimurium DNA was cloned directly up-stream of this gene fusion. When properly positioned, a S. typhimurium promoter drives the expression of both the purA and the lacZ genes. Introduction of this pool of gene fusion plasmids in to a ΔpurA S. typhimurium strain, which has an attenuated growth in vivo, and selection for antibiotic resistance required integration of the plasmid into the chromosome of the bacteria by homologous recombination. The pool of purA-lacZ fusion strains was injected into mice. Bacteria harvested from the mice spleens three days post-inoculation had therefore complemented the purA mutation. Analysis of lacZ expression of each purA-lacZ fusion strain grown in vitro and in vivo indicated that β-galactosidase production was elevated during in vivo growth. With the sequencing of the cloned S. typhimurium DNA, identification of the genes controlled by any cloned promoter was possible. IVET technology has been used to identify genes induced in Pseudomonas aeruginosa (Wang et al., 1996) and Staphylococcus aureus (Lowe et al., 1998) in the human host, Actinobacillus
pleuropneumoniae (Fuller et al., 1999) in swine and in Candida albicans (Staib et al., 1999), a fungal pathogen of humans.

1.8.6 Gene Transfer

It is possible to identify virulence factors by transferring cloned DNA from a pathogenic strain of an organism to a related non-pathogenic species, a mutant with attenuated virulence or a distant genus that can efficiently express the transferred DNA and subsequent screening for the acquisition of pathogenic traits. This methodology has been used to great effect with M. tuberculosis. Collins et al. (1995) constructed a plasmid library of virulent M. bovis genomic DNA, which was transferred into an avirulent strain of M. bovis. A number of clones had regained the ability to cause spleen lesions in guinea pigs and were subsequently found to contain a common M. bovis DNA fragment, which contained the rpoV open reading frame whose product has a high homology to the principal bacterial sigma (σ) factor. It was also noted that the avirulent strain of M. bovis contained two mutations in this gene, either of which could alter the expression of a subset of virulence factors.

The transfer of M. tuberculosis DNA to a non-invasive strain of E. coli has also identified a region of DNA associated with entry and survival of the macrophage (Arruda et al., 1993). The E. coli clone harboured a plasmid containing cloned M. tuberculosis DNA, which conferred the ability to invade HeLa cells, induce macrophage phagocytosis and survival within the macrophage for up to 24 hours. The M. tuberculosis fragment has recently been sequenced (accession number: X70901)
and has been found to contain a gene, \( mcel \), which has no similarities with DNA sequences of other organisms in the databases.

1.8.7 Subtractive Hybridization

The rational behind this technique is to hybridize DNA or cDNA fragments obtained from either a single strain grown in different environmental conditions or from two related strains, one of which is virulent and the other avirulent, and eliminate from the hybridization mixture all double-stranded nucleic acid fragments thus present in both pools representing sequences common to both. After a number of rounds, the resulting subtractive probes can be used to identify genes that may be specific and perhaps important for virulence.

Subtractive hybridization, using DNA, has been used to identify gross DNA differences between \( M.\text{bovis} \) and \( M.\text{bovis} \) BCG (Mahairas \textit{et al.}, 1996). This identified three regions of difference (RD) between the two genomes, representing 30Kb of DNA. When these were sequenced, a number of previously identified TB-specific antigen genes were found to be located within these regions of the genome. In a separate study, cDNA subtractive hybridization was used to identify a macrophage-induced gene (\( \text{mig} \)) in \( M.\text{avium} \) grown in either liquid culture or in human-derived macrophages (Plum and Clark-Curtiss, 1994). It has subsequently been shown that the \( \text{mig} \) gene is not present in the \( M.\text{tuberculosis} \) genome which perhaps underlines different mechanisms of pathogenesis in these two mycobacteria (Cole \textit{et al.}, 1998).
1.8.8 Differential Display Polymerase Chain Reaction (DD-PCR)

Differential display PCR can be used with two different template types. It can be used to either identify gene expression differences between closely related organisms or to identify differences in the levels of expression of an organism growing in different conditions. Unfortunately results obtained using this technique should be viewed with caution as DD-PCR is renowned for generating false positives.

Using reverse transcribed mRNA (cDNA) as a template, genes that are up-regulated during infection can be identified. As stated above, *M.tuberculosis* infects and survives inside macrophages. It would be expected that both the *M.tuberculosis* and the macrophage each react to the infection process by altering the expression of certain genes. Ragno *et al.* (1997) used DD-PCR to study the effect of *M.tuberculosis* infection on the gene expression of the invaded macrophage. Roughly 20 PCR bands were identified as representing differentially expressed products. However, only one PCR product was investigated further - a band of 158bp that was down-regulated during infection - which was found to have a high degree of sequence homology to the murine mitochondrial enzyme cytochrome C oxidase subunit VIIc (COX VIIc). It was suggested that down-regulation of COX VIIc could be either an early apoptotic signalling event or a down-stream consequence of apoptotic signalling.

RNA arbitrarily-primed differential display PCR (RAP-PCR) has been used to identify differentially expressed genes between attenuated and virulent laboratory strains of *M.tuberculosis* (Rivera-Marrero *et al.*, 1998). Using this method, cDNA
fragments with homology to three known mycobacterial genes and six hypothetical genes were identified, four of which were later identified as differentially up-regulated.

1.8.9 Gene Fusions

Potential virulence factors in pathogenic mycobacteria can be identified by fusions to promoterless reporter genes whose expression can be easily monitored. The most useful application of reporter genes in the identification of virulence factors is with reporter genes whose expression can be easily followed intracellularly.

The Green Fluorescent Protein (GFP) from the jellyfish *Aequorea victoria* absorbs ultraviolet light, at a wavelength of 395nm, and emits green light at 510nm. The protein folds without any external influences and does not require any exogeneous substances or co-factors, which means that the excitation of GFP can be seen intracellularly. GFP was first used as a reporter gene in mycobacteria in 1995 (Dhandayuthapani *et al.*, 1995; Kremer *et al.*, 1995) and it has been shown that the presence of GFP does not affect the bacteria’s ability to invade and survive within the host cell (Parker and Bermudez, 1997; Valdivia *et al.*, 1996). Since bacteria expressing GFP fluoresce, they can be separated from non-fluorescing bacteria with a fluorescence-activated cell sorter (FACS). This has been shown in the identification of 14 *Salmonella typhimurium* genes specifically induced in host macrophages and which are therefore potential virulence factors (Valdivia and Falkow, 1997). A similar study identified 12 *Mycobacterium marinum* genes that are differentially expressed in macrophages (Barker *et al.*, 1998).
An alternative reporter system to GFP is that of β-galactosidase, which is coded for by the lacZ gene and requires an external substrate in order to report expression levels. Dellagostin et al. (1995) used a translational fusion of an amino-terminally truncated E.coli lacZ reporter gene with mycobacterial promoters for the M.bovis BCG hsp60 gene, the M.leprae 28KDa gene and the M.leprae 18KDa gene to measure the activity of the promoters during intracellular and extracellular growth. The lacZ gene has also been used to analyse transcription and expression of β-lactamase gene isolated from a natural isolate of M.fortuitum and from a high-level amoxicillin-resistant mutant (Timm et al., 1994). This method identified a single base pair difference in the promoter region was responsible for the high-level production of the β-lactamase in the mutant. The lacZ reporter system has also been applied for methodologies to rapidly identify antimycobacterial drugs (Srivastava et al., 1997a; Srivastava et al., 1997b).

1.9 Aims

The identification of virulence factors is central to the understanding of tuberculosis and its survival within the host. Currently few M.tuberculosis genes have been identified as virulence factors (Gordon and Andrew, 1996). Newly identified virulence factors will be invaluable in furthering our understanding of the process of mycobacterial infection and could be used to generate a new, more effective vaccine (either DNA, subunit or a live, attenuated strain similar BCG).

It is clear from the above that there are a number of different techniques that are currently being used to identify virulence factors and range from low technology,
labour intensive gene fusions through to high technology high density DNA arrays. All of the techniques mentioned have positive virtues regarding their ability to identify potential virulence factors but it is unlikely that one technique will identify all *M. tuberculosis* virulence factors. Even if this were to be the case, other techniques would be required to support any initial findings. The aim of this work is to identify *M. tuberculosis* genes that are up-regulated during infection, using a gene fusion methodology.

To this end, a *M. tuberculosis* H37Rv gene library has been constructed using a promoter probe vector using *lacZ* as a promoterless reporter gene. Once transformed into *M. bovis* BCG Pasteur, discrete clones were randomly selected and arrayed in a 96 well microtitre™ format. Each clone was screened for up-regulated expression of the reporter gene, *lacZ*, by exposure to stressing environmental conditions designed to mimic those inside the macrophage. A smaller number of clones identified via this type of screening have been used in macrophage infection studies and the level of *lacZ* expression measured and compared to the level of expression *in vitro*. The *M. tuberculosis* H37Rv DNA cloned within the plasmid of each clone was sequenced. The genes, whose expression is driven by any of the promoters identified as having increased *lacZ* expression inside the macrophage compared to *in vitro*, were subjected to studies designed to measure the specific mRNA expressed during *in vitro* and intra-macrophage growth.
CHAPTER 2

CONSTRUCTION OF MYCOBACTERIUM TUBERCULOSIS H37RV PLASMID LIBRARY
Despite the advent of the sequenced *Mycobacterium tuberculosis* genome, little is known about the structure of mycobacterial promoters or which genes are critical during the processes of infection and intracellular survival (Cole *et al.*, 1998). A number of methods have been employed to study genes whose expression may be altered under different conditions. Differential display PCR has been used to identify differences in the expression levels of mitochondrial cytochrome *c* oxidase of the host macrophage prior to and post-infection with *M. tuberculosis* and has identified genes that are differentially expressed between virulent and avirulent *M. tuberculosis* strains (Ragno *et al.*, 1998; Rivera-Marrero *et al.*, 1998). Two dimensional gel electrophoresis of *Mycobacterium tuberculosis* cell extracts grown in conditions of either high or low iron concentrations has been used to identify 15 induced proteins and 12 proteins whose expression is decreased (Wong *et al.*, 1999).

The latest method currently being employed to identify differentially regulated genes is microarray technology. The genes of a particular organism, in this case *M. tuberculosis*, are amplified and arrayed on small glass ‘chips’ and position of each arrayed gene recorded. mRNA obtained from the organism grown in different conditions can be hybridised to the ‘chip’-bound DNA. Differentially expressed genes can be detected by measuring the amount of hybridisation to the DNA. The power of microarray technology is best shown in the work carried out by Behr *et al.* (1999) and their identification of the differences between *M. tuberculosis* H37Rv and the *M. bovis* BCG strains at the genome level. This type of high technology is increasingly being used but sadly is not currently available to all laboratories. DNA microarrays have not
as yet been used successfully to identify any potential virulence factors of *M.tuberculosis*.

However, one of the methods most likely to identify differentially expressed *M.tuberculosis* genes is the use of promoter probe vectors. Promoter-less reporter genes have been employed to identify promoter fragments of *Mycobacterium marinum* genes that are differentially expressed *in vitro* and *in vivo* (Barker *et al.*, 1998). The *M.marinum* genes were identified with the use of a promoter-less green fluorescent protein (GFP). To further support the use of promoter-less reporter genes, Dellagostin *et al* (1995) showed that an amino-terminally truncated *E.coli lacZ* reporter gene could be used to follow expression of the *Mycobacterium bovis* BCG *hsp60* gene (Rv0440/ *groEL2*), the *Mycobacterium leprae* 28 kDa gene and the *M.leprae* 18 kDa gene *in vivo*. The ability to be able to follow expression levels *in vivo* is crucial to identify genes involved in the virulence and pathogenicity of *Mycobacterium tuberculosis*.

2.1.1 Strategy

In order to study the activity of mycobacterial promoters, an effective replicating vector system based on pUS1781 (Pers. Comm.: constructed by Stewart, G.R. – See Section 2.2.2) carrying a promoter-less reporter gene was constructed. A promoter-less *E.coli lacZ* gene was amplified from pMC1871 (Promega) with a small multiple cloning site directly up-stream. Using a heat shock protein promoter (*groEL2* – Rv0440), a positive control vector was constructed and analysed for expression
compared to the basal vector. The standard β-galactosidase assay (Miller, 1972) for the reporter gene *lacZ* was optimised for mycobacteria.

Fragments of genomic *Mycobacterium tuberculosis* H37Rv DNA, between 1.0 and 0.3 Kb in size, were inserted into the multiple cloning site directly up-stream of the promoterless *lacZ* and the resulting plasmid promoter library was used to transform *M. bovis* BCG Pasteur. Individual clones were randomly selected and arrayed in a 96 well microtitre™ format.
2.2 Materials and Methods

### 2.2.1 Oligonucleotide Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'-3')</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z-Bgl-Nar</td>
<td>GAAGATCTGCATCGCGCGCATG ATACCGCGAGACCCACG</td>
<td>LacZ gene nt 1-20, 5' BglII SphI and NarI restriction sites</td>
</tr>
<tr>
<td>Z-Bgl-Kpn</td>
<td>GAAGATCTGGTACCTATTATTATGC ACCAGACCAACTG</td>
<td>LacZ gene nt 3290-3267 (complementary strand), 5' BglII and KpnI restriction sites</td>
</tr>
<tr>
<td>RH22</td>
<td>GACTGAGGCACCACCACACGA CGCG</td>
<td><em>M. tuberculosis</em> genome nt 528218-528232, gene Rv0440. 5' NarI restriction site.</td>
</tr>
<tr>
<td>RH23</td>
<td>GACTGAGGCACCACGCTTCTCC GGATCGG</td>
<td><em>M. tuberculosis</em> genome nt 528600-528582 (complementary strand), gene Rv0440. 5' NarI restriction site.</td>
</tr>
<tr>
<td>RH20</td>
<td>CGGTTTACAAGCATAAAG</td>
<td>pUS1800 (and subsequent plasmids) nt 109-126 of plasmid</td>
</tr>
<tr>
<td>RH21</td>
<td>GGTCTCGCGGTATCAGG</td>
<td>pUS1800 (and subsequent plasmids) nt 169-152 (complementary strand) of plasmid</td>
</tr>
</tbody>
</table>

*Under scored sequences represent endonuclease restriction sites in the order stated in details column*

### 2.2.2 Construction of Promoter Probe Shuttle Vector

The plasmid pUS1781 was constructed using pUC19 as the starting plasmid (Pers. Comm.: Constructed by G.R. Stewart) (See Figure 2.1). Into the multiple cloning site of pUC19 a 1.6Kb Mycobacterial origin of replication from pAL5000 and a kanamycin resistance gene (the *aph* gene from Tn903) were inserted using the *KpnI* and the *PstI* restriction sites respectively. The resulting plasmid was digested with *NarI* restriction enzyme and *Bsal* restriction enzymes, in order to remove the ampicillin resistance gene. The resulting ‘sticky ends’ were polished and the DNA
re-circularised. The T4 terminator was PCR amplified with the 5’ primer containing an *SpeI* restriction site and the 3’ primer containing an *XbaI* restriction site. The DNA was cloned into the *XbaI* restriction site of the pUC19 multiple cloning site generating the plasmid pUS1781.

![Vector map of plasmid pUS1781](image)

Figure 2.1: Vector map of plasmid pUS1781. The mycobacterial origin of replication (Myco Ori), the *E.coli* origin of replication (OriE), the selectable kanamycin resistance marker (Kan(R)) and the T4 terminator are all indicated. The *BamHI* restriction site into which the promoterless *lacZ* is cloned is also indicated.

A promoterless *lacZ*, including the translational start site, was amplified using the primers, Z-Bgl-Nar and Z-Bgl-Kpn. Z-Bgl-Nar anneals to the first 20 bp of the *lacZ* gene, and carries a small multiple cloning site up-stream, whilst Z-Bgl-Kpn anneals to the last 23 bp of the *lacZ* gene. These primers were used to generate a 3.3Kb DNA fragment by PCR, using the plasmid pMC1871 (Promega) as the template. PCR amplification reaction conditions consisted of an initial denaturation at 95°C for 2 minutes, followed by 35 cycles with denaturation 95°C, primer annealing at 67°C and extension at 72°C for 1 minute, 1 minute and 4 minutes respectively. The final cycle was extension at 72°C for 5 minutes. The promoter probe Mycobacterial-*E.coli* shuttle vector, pUS1800, was constructed by inserting the 3.3
Kb lacZ gene as a BglII fragment into the BamHI site of pUS1781. The required orientation of the promoterless lacZ with transcription away from the T4 terminator was identified by digestion with EcoRV. Clones containing lacZ in the right orientation generated fragments of 6482 bp and 1348 bp (clones containing lacZ transcribing in the opposite direction generated 5864 bp and 1966bp fragments). Figure 2.2 shows a plasmid map of pUS1800. All cloning was performed in E.coli DH5α. The efficiency of pUS1800 as a promoter probe vector was assessed once a positive control had been constructed.

Figure 2.2: Vector map of plasmid pUS1800. The mycobacterial origin of replication (OriM), the E.coli origin of replication (OriE), the selectable kanamycin resistance marker (Kan(R)), the T4 terminator (T4 Term) and the promoterless lacZ reporter gene are all indicated. The restriction sites indicated are unique.

2.2.3 Construction of Positive Control for Promoter Probe Vector

In order to test the efficacy of the promoter probe Mycobacterial-E.coli shuttle vector in identifying promoters, the promoter for the hsp60 gene (Rv0440/ groEL2) of
M. tuberculosis H37Rv was amplified by PCR using the primers RH22 and RH23, which generated a 0.4 Kb DNA fragment. PCR reaction conditions consisted of an initial denaturation at 95°C for 2 minutes, followed by 35 cycles with denaturation at 95°C, primer annealing at 55°C and extension at 72°C each for 1 minute. The final cycle was extension at 72°C for 5 minutes. The amplified DNA contains the transcription signals and the ribosome binding site of the M. tuberculosis H37Rv hsp60 gene (Rv0440/ groEL2). The positive control promoter probe Mycobacterial-E.coli shuttle vector, pUS1808, was constructed by inserting the 0.4 Kb hsp60 promoter (Rv0440/ groEL2) as a NarI fragment into the same site of the dephosphorylated plasmid pUS1800. The required orientation of the cloned promoter to promote transcription of lacZ was identified via restriction digest with the enzyme AatII. If the hsp60 promoter has been cloned into pUS1800 in the right orientation two fragments are generated - 7032 bp and 1170 bp. With the hsp60 promoter in the wrong orientation fragments of 7332 bp and 870 bp are created. The efficiency of pUS1800 and pUS1808 as a negative and a positive control promoter probe vector were assessed with the use of the β-galactosidase assay (Miller, 1972).

2.2.4 β-galactosidase Assay with ONPG Substrate

Assays were performed as described by Miller (Miller, 1972). Cultures of M. bovis BCG Pasteur were grown to mid-exponential phase (A\textsubscript{600}\text{0.8}) or to stationary phase (A\textsubscript{600}\text{2.0}) from which a given volume (between 0.2 ml and 0.8 ml) was diluted in Z buffer (60 mM Na\textsubscript{2}HPO\textsubscript{4}, 40 mM NaH\textsubscript{2}PO\textsubscript{4}, 10 mM KCl, 1 mM MgSO\textsubscript{4}, 50 mM β-mercaptoethanol) to a final 1 ml volume. Cells were permeabilised with 2.0% (v/v) chloroform and 0.001% (v/v) SDS and the solution equilibrated at 28°C for 5
minutes. The substrate, ONPG (o-nitrophenyl-β-D-galactopyranoside) was then added to a concentration of 0.8 mg ml\(^{-1}\) and the reaction carried out at 28°C until sufficient yellow colour had developed as a result of hydrolysis of ONPG by β-galactosidase. The reaction was stopped by the addition of 500 μl 1M Na\(_2\)CO\(_3\) and the solution centrifuged at 10 000 g for 2 minutes. The \(A_{420}\) of the supernatant was measured and the β-galactosidase activity expressed in Miller Units using the following equation:

\[
A_{420}(1000) / (tv)A_{600} = \beta\text{-galactosidase activity (Miller Units)},
\]

where \(t\) = incubation time (minutes) and \(v\) = volume of culture (ml).

### 2.2.5 β-galactosidase Assay Optimised for BCG

Assays performed as described above identified two problems when used to measure the difference in expression of \(lacZ\) from the two plasmids pUS1800 and pUS1808 harbourered within \(M.\text{bovis}\) BCG Pasteur. Firstly lysis of the mycobacteria was inconsistent. Secondly, due to the relatively low levels of expression of \(lacZ\) by both clones, the activity of the enzyme had to be increased. Therefore β-galactosidase assays were performed as above but with alterations to the assay temperature and method of cell lysis.

Cultures of \(M.\text{bovis}\) BCG Pasteur harbouring the plasmids pUS1800 and pUS1808 were grown to mid exponential phase \((A_{600}0.8)\) and 1ml of culture was added to a Hybaid Ribolyser™ blue tube (a 2 ml skirted screw-capped
microcentrifuge tube with 'O' ring seal, containing acid-washed 0.1 mm silica/ceramic beads) and processed in a Hybaid Ribolyser™ at speed 5.0 for 40 seconds. The Ribolyser tube, containing the lysed bacteria, were cooled on ice for 1 minute and then centrifuged at 14000g for 1 minute to sediment the silica beads.

A known volume of lysed bacteria was diluted in Z buffer (60mM Na$_2$HPO$_4$, 40mMNaH$_2$PO$_4$, 10mM KCl, 1mM MgSO$_4$, 50mM β-mercaptoethanol) to a final 1ml volume and equilibrated to an assay temperature between 28 and 50°C. Temperatures used for incubation were 28°C, 37°C, 42°C, 45°C and 50°C. The substrate, ONPG was then added to a concentration of 0.8 mg ml$^{-1}$ and the reaction carried out at the chosen equilibration temperature until sufficient yellow colour had developed as a result of hydrolysis of ONPG by β-galactosidase. The reaction was stopped by adding 500μl 1M Na$_2$CO$_3$ and β-galactosidase activity calculated and expressed as described previously in section 2.2.4.

2.2.6 Preparation of *M.tuberculosis* H37Rv Genomic DNA Fragments

*M.tuberculosis* H37Rv genomic DNA was digested individually with Accl, HpaII and TaqI restriction enzymes. Genomic DNA (1μg) was digested with 1 unit of enzyme in a final volume of 50μl of water. At time intervals of 0, 5, 10, 20, 40, 60 minutes, 8μl of the reaction was removed and added to loading buffer and stored at -80°C (to stop the reaction) until required at a later date. The digested DNA was analysed as described in Appendix 2 (Agarose gel electrophoresis) and fragments of each restriction digest between 1 and 0.3 Kb were excised and gel purified.
2.2.7 *M.tuberculosis H37Rv* Library Construction

The plasmid pUS1800 was digested with the restriction enzyme *NarI*, dephosphorylated and gel purified. The purified, digested genomic DNA from section 2.2.6 was pooled in equal quantities. The *M.tuberculosis H37Rv* promoter library Mycobacterial-*E.coli* shuttle vector was constructed by insertion of the *AciI*, *HpaII* and *TaqI* digested and pooled genomic *M.tuberculosis H37Rv* DNA into the *NarI* restriction site in pUS1800. The cloning efficacy was estimated by a standard transformation of the *M.tuberculosis H37Rv* library into *E.coli* DH5α. Individual clones were picked and subjected to PCR analysis. Primers RH20 and RH21 were used to amplify across the *NarI* restriction site of the plasmid pUS1800. Reaction conditions consisted of an initial denaturation at 95°C for 2 minutes, followed by 35 cycles with denaturation 95°C, primer annealing at 67°C and extension at 72°C for 1 minute, 1 minute and 2 minutes respectively. The final cycle was extension at 72°C for 5 minutes. The amplified DNA fragments were visualised as described in Appendix 2 (Agarose gel electrophoresis). In order to generate sufficient *M.tuberculosis H37Rv* library plasmid DNA with few or no re-circularised pUS1800, a number of ligation reactions were carried out. The ligation reaction mix showing the least number of re-circularised plasmid was introduced into Epicurian Coli® XL2-Blue ultracompetent cells (Stratagene). The *E.coli* clones harbouring the *M.tuberculosis H37Rv* library plasmid were grown on selective solid media and resuspended in LB broth. The plasmid DNA from the pooled *E.coli* clones was recovered with the Qiagen Midi Kit used as described by the manufacturer.
2.2.8 Generation of *M. tuberculosis* H37Rv Library in *M. bovis* BCG Pasteur.

The *M. tuberculosis* H37Rv library was introduced by electroporation (see appendix 2) into *M. bovis* BCG Pasteur and grown on selective media containing 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) for between 14 and 21 days. Discrete clones were randomly selected and arrayed in microtitre plates and grown in selective 7H9 broth media at 37°C for 7 days. Each clone was subcultured three times, with the use of 96 pin replicator, into selective 7H9 broth media and grown at 37°C to late exponential growth phase. This generated four identical copies of the *M. tuberculosis* H37Rv library harboured in *M. bovis* BCG Pasteur. Three copies were stored at −80°C whilst the fourth copy was maintained for continued work.
2.3 Results

2.3.1 Construction and Expression Analysis of Plasmids pUS1800 and pUS1808

A promoter probe Mycobacterial-E.coli shuttle vector (pUS1800) was constructed containing the amplified E.coli promoterless lacZ gene from pMC1871 (Promega). Into the multiple cloning site directly up-stream of the promoterless lacZ, the hsp60 promoter was cloned so as to generate pUS1808 as a positive control vector for pUS1800. The hsp60 promoter, amplified directly for Mycobacterium tuberculosis H37Rv genomic DNA, was cloned so that the ribosome binding site was separated from the translational start site of lacZ by the same number of bases which separate it from the translational start site of the hsp60 gene. These two plasmids were subsequently used independently to transform M.bovis BCG Pasteur (See Figure 2.3).

M.bovis BCG Pasteur transformants separately harbouring vectors pUS1800 and pUS1808 were cultured in kanamycin-supplemented 7H9 broth at 37°C until the cultures were at mid-log phase. Cultures were harvested and assayed for β-galactosidase activity, using ONPG substrate, as described by Miller (Miller, 1972). The difference in Miller units between the two vectors pUS1800 and pUS1808 was always a factor of a log. However, assays carried out at different times using cultures at similar growth phases generated a wide range of different values making comparison between cultures assayed on different days extremely difficult (See Figure 2.4).
Figure 2.3: Construction of plasmids pUS1800 and pUS1808. The 3.3 kb PCR-amplified *E. coli lacZ* gene from pMC1871 was inserted into pUS1781 on a *BglII/BamHI* site to produce pUS1800. pUS1808 was generated by the insertion of the 0.4 kb PCR-amplified *M. bovis BCG Pasteur* hsp60 promoter (Rv0440/groEL2) on a *NarI* site. *M. bovis* BCG Pasteur transformants harbouring pUS1808 and pUS1800 are shown (left to right).
Bacterial cells in the Miller method are disrupted by the addition of chloroform and 0.001% SDS. The variation, in Miller units, when using *M. bovis* BCG Pasteur cells is most probably due to inconsistent lysis of the mycobacterial cell wall by the chloroform and the SDS. The β-galactosidase assay as described by Miller was therefore altered in two ways – the lysis method of the mycobacterial cells and the temperature at which the assay is carried out. Disruption of the mycobacterial cell was found to be optimal in a Hybaid Ribolyser™. The second alteration to the β-galactosidase assay described by Miller was to elevate the incubation temperature (28°C). The time required for sufficient yellow colour to develop from the hydrolysis of the substrate ONPG by β-galactosidase is included in the calculation of Miller units. The incubation temperature was probably set by Miller to 28°C to slow down the activity of the β-galactosidase enzyme which allows a reasonable length of time to
pass between the addition of ONPG and the stopping of the reaction with sodium carbonate. When the two *M. bovis* BCG Pasteur cultures harbouring pUS1800 and pUS1808 were assayed by incubating at 28°C, the time required for sufficient yellow colour to develop was too long. They generated low values of Miller units signifying low levels of expression of *lacZ*. Elevated incubation temperatures for the β-galactosidase assay were therefore tested. Cultures were ribolysed and constant volumes of the same lysate were incubated at 5 temperatures varying from 28°C to 50°C.

![Figure 2.5: β-galactosidase assays carried out at elevated temperatures. Assays carried out on *M. bovis* BCG Pasteur harbouring pUS1800 and pUS1808 carried at 28, 37, 42, 45, 50°C in triplicate. Values for the plasmid pUS1800 harboured by *M. bovis* BCG Pasteur at 28°C and 42°C are shown but the values are very small. The error bar for the plasmid pUS1800 harboured within *M. bovis* BCG Pasteur incubated at 37°C is also present but is so small that it can not be seen.](image-url)
The results in Figure 2.5 clearly demonstrate that an increase in temperature raised the Miller units observed. At 50°C the activity of the β-galactosidase enzyme is reduced in the lysate from \textit{M. bovis} BCG Pasteur harbouring pUS1808. However at that temperature the activity of the pUS1800 lysate appears to be increased but this is probably due to the substrate ONPG, which needs to be made fresh for each assay, spontaneously hydrolysing. Therefore an assay incubation temperature of 42°C was used subsequently.

\textit{M. bovis} BCG Pasteur harbouring the plasmid pUS1800 displays no significant levels of β-galactosidase expression. The positive control plasmid, whose expression is being driven by the \textit{hsp60} promoter, is exhibiting Miller units in the order of 40 units compared to the 0 units produced by pUS1800. Therefore, pUS1800 has the potential to be used as an effective promoter trap vector.

2.3.2 Construction of \textit{Mycobacterium tuberculosis} Promoter Library

\textit{M. tuberculosis} H37Rv total genomic DNA was digested individually with the restriction enzymes \textit{AciI}, \textit{HpaI} and \textit{TaqI}. DNA fragments of between 0.3 and 1 Kb were gel-purified and cloned into the NarI restriction site of the plasmid pUS1800 (See Figure 2.6). \textit{E. coli} DH5α were transformed with a small sample of the cloned \textit{M. tuberculosis} H37Rv plasmid library. Individual clones were subjected to PCR with the primers RH20 and RH21, amplifying across the NarI cloning site (See Figure 2.7).
M. tuberculosis plasmid library construction. *M. tuberculosis* genomic DNA was digested with *AceI*, *Hpall* and *TaqI* for 0, 5, 10, 20, 40, and 60 minutes. Fragments of between 0.3 and 1 Kb were gel-purified and inserted on the *NarI* site of pUS1800 thus generating a *M. tuberculosis* promoter trap library.

<table>
<thead>
<tr>
<th>Time of digest in minutes</th>
<th><em>AceI</em></th>
<th><em>Hpall</em></th>
<th><em>TaqI</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.6:** *M. tuberculosis* plasmid library construction. *M. tuberculosis* genomic DNA was digested with *AceI*, *Hpall* and *TaqI* for 0, 5, 10, 20, 40, and 60 minutes. Fragments of between 0.3 and 1 Kb were gel-purified and inserted on the *NarI* site of pUS1800 thus generating a *M. tuberculosis* promoter trap library.
Figure 2.7: Agarose gel of PCR products of discrete E.coli DH5α clones harbouring M.tuberculosis plasmid library. Amplification carried out using primers RH20 and RH21. Lanes 1, 15, 29, 30, 44 and 58 are 1 Kb ladder and lanes 2, 16, 31 and 45 are the region amplified from pUS1800.

Figure 2.7 clearly shows that all of the E.coli clones, harbouring the M.tuberculosis library plasmid, subjected to PCR contain cloned DNA since the amplified regions of the clones are larger than the basal plasmid amplification. Indeed 380 discreet clones were subjected to PCR with the primers RH20 and RH21 and none were found be re-circularised pUS1800 plasmid. The maximum amplified fragment seen was 1.5 Kb, with the smallest amplified fragment being just 150 bp. The average insert size was found to be 400bp.
The *M. tuberculosis* H37Rv plasmid library ligation was introduced into Epicurian Coli® XL2-Blue ultracompetent *E.coli* cells with a transformation efficiency greater than $5 \times 10^9 \, \mu g^{-1}$ of plasmid DNA and the plasmid DNA recovered. In excess of $1.5 \times 10^5$ discrete *E.coli* clones were harvested. *M.bovis* BCG Pasteur was transformed with the library plasmid purified from the pooled *E.coli* clones and individual clones selected for on solid media containing kanamycin and X-gal (See Figure 2.8).

Figure 2.8: *M.bovis* BCG Pasteur harbouring the *M.tuberculosis* plasmid library grown on 7H11 solid media containing kanamycin and X-gal. It can be clearly seen that about 15% of the independent clones are expressing *lacZ* and are therefore blue. Red arrows identify two clones, which are expressing *lacZ* but, in the presence of X-gal, are not completely blue in colour. The clone at the top of the plate is blue in the centre of the colony and white at the edges whilst the clone at the bottom of the plate is white but has turned the media below it blue.
2.3.3 Selection and Arraying of Individual *M. bovis* BCG Pasteur Library Clones.

Despite the fact that little is known about the structure and function of *M. tuberculosis* promoters, it is more than likely that a number of genes required during infection will be silent *in vitro*. It is quite clear from Figure 2.8 that roughly 15% of the clones are, to varying degrees, blue, as they are expressing *lacZ* and therefore appear to have some sort of promoter activity. However the *M. tuberculosis* DNA cloned in the other 85% of the clones may be extremely significant.

Before any clones can be arrayed in a 96 well microtitre™ format, the number of clones required to statistically cover the entire *M. tuberculosis* genome had to be estimated by using the equation below.

The number of clones needed = \[
\frac{\ln (1-P)}{\ln (1-1/N)}
\]

where \( P \) = the probability of selecting a specific point on the genome, and

\( N = \text{The number of nucleotide bases in the genome} \)
\( \text{The average size of the cloned fragment (in nucleotide bases)} \)

The size of the genome and the cloned fragment sizes are known to be 4.4x10\(^6\) and 400 bp respectively. If 95% of the *M. tuberculosis* genome is to be statistically covered then the number of clones that need to be arrayed can be calculated thus:

\[
\frac{\ln (1-0.95)}{\ln (1-1/(4.4x10^6/400))}.
\]

=32951
The use of this equation to estimate the number of clones that need to be arrayed in order to statistically cover the entire *M. tuberculosis* H37Rv genome requires an assumption to be made. The equation is designed to be used when a genome is cloned randomly. As described earlier, three restriction enzymes were used to generate the small DNA fragments that were subsequently inserted directly up-stream of the promoterless *lacZ* in pUS1800. Obviously even a partial enzymatic digest of DNA could never be considered to be random and even with three restriction enzymes increasing the randomness, this equation can only be used to suggest a number of clones to be arrayed had the cloning been entirely random.

The use of robotics was investigated to aid in selecting and arraying this number of clones. It was hoped that a robot could be instructed to identify an *M. bovis* BCG Pasteur clone, grown on selective solid 7H11 media containing X-gal, and decide whether it was blue or white. Once the colour had been determined, the clone would be arrayed in a 96 well microtitre™ format with clones of similar colour intensity.

When *E.coli* is grown on solid media, discrete clones are formed which are soft, malleable and the diameter of the clone is relatively constant. *M. bovis* BCG Pasteur clones, however, do not form regular shaped or sized colonies. In addition to their inconsistent size, unless the *M. bovis* BCG clone is expressing large levels of *lacZ*, the bacteria do not always turn blue in the presence of X-gal - the media on which they are growing can be turned blue (See Figure 2.7). The two clones identified with arrows clearly show two of the different colony colourings. The clone in the bottom right hand corner of the plate appears to be white but has a definite blue
colouring to the agar beneath it. In the top half of the plate, the identified clone is blue in the centre and white around the edges of the colony. The reason for some clones expressing \textit{lacZ} and turning blue while others cleave X-gal and turn the media blue is unclear. The reason for the clone in the top half of the plate having a blue centre and a white outside is probably due either to expression late in the growth phase or due to a response to some type of starvation (such as anaerobic growth or nutrient deprivation) of the bacteria in the centre of the colony.

On advice from experts, the use of robotics was never actually tested with \textit{M.bovis} BCG Pasteur due to all of these distinctive properties of \textit{M.bovis} BCG Pasteur clones. A robot would never have been able to discern the different sizes and discriminate between the different intensities of blue of \textit{M.bovis} BCG Pasteur clones grown in the presence of X-gal. However, had the robot had sufficient power to overcome these problems, all 32951 clones required to statistically cover the \textit{M.tuberculosis} genome would have been arrayed.

Thus it was decided that a smaller number of clones generated by the transformation of \textit{M.bovis} BCG Pasteur with the \textit{M.tuberculosis} promoter probe library would be selected and arrayed in a 96 well microtitre™ format manually. The aim of a study of a smaller library sample size was to identify some of the up-regulated promoters of \textit{M.tuberculosis} H37Rv and not to carry out a comprehensive study of up-regulation. Therefore, due to the prohibitive length of time required to array clones, 15\% of the desired maximum 32951 clones were arrayed. Discrete clones were selected randomly and arrayed, where possible, by colour intensity caused by the expression of \textit{lacZ} in the presence of X-gal \textit{in vitro},
under no external stimuli. Clones expressing lacZ (and therefore blue) were arrayed in the same 96 well microtitre™ plate, with white clones being arrayed separately. In total fifty 96 well microtitre™ plates were arrayed - 8 contained blue clones with the rest of the plates contained white clones. These arrayed plates represent 768 blue clones and 4032 white clones, suggesting that 19% of the arrayed clones are expressing lacZ in vitro under no external stimuli.
2.4 Discussion

The use of the β-galactosidase assay as described by Miller (1972) does not generate constant results when used for the detection of \( lacZ \) expression in slow growing Mycobacteria (see Figure 2.3). There are probably two reasons for this. The first is that there is insufficient or inconsistent lysis of the mycobacteria probably due to the thickness of the cell wall. Secondly, \textit{Mycobacterium bovis} BCG is thought to have relatively low levels of gene expression. As such, the amounts of β-galactosidase protein being produced make detection of the enzyme more difficult.

If the β-galactosidase assay is to be used to assay the levels of expression of \( lacZ \) of a number of different, unknown promoters \textit{in vitro} and \textit{in vivo}, the assay must generate consistent results for a known promoter.

The Miller assay was therefore adapted for the slow growing \textit{M.bovis} BCG Pasteur by altering the method of lysis and increasing the activity of the low levels of β-galactosidase enzyme present. Ribolysing the mycobacteria prior to the addition of the ONPG substrate in the Miller assay ensures that the bacteria are completely lysed whilst raising the incubation temperature of the assay increases the activity of the enzyme thus making the enzyme more detectable.

With the ability to detect the β-galactosidase enzyme consistently, the plasmid pUS1808 harboured in \textit{M.bovis} BCG Pasteur generates 40 Miller units compared to the expression of \( lacZ \) from pUS1800 harboured within \textit{M.bovis} BCG Pasteur, which does not produce any detectable Miller Units.
This clearly indicates that the plasmid pUS1800 is a suitable vector for identifying promoter fragments of cloned DNA. Genomic *M. tuberculosis* H37Rv DNA fragments were randomly generated by partial digestion with the 3 enzymes *Acil*, *HpaII* and *TaqI* and cloned into the *NarI* restriction enzyme site directly up-stream of *lacZ* in plasmid pUS1800. Of the 380 discrete *E.coli* clones examined via PCR, no plasmids were found to be re-circularised pUS1800. All of the PCR products generated were different sizes, between 1500 and 150 bp, with an average insert size of 400 bp. With a genome the size of *M. tuberculosis* and an average cloned fragment size of 400 bp, statistically to have cloned the entire *M. tuberculosis* H37Rv genome, about 3.3x10^4 discreet clones would need to be arrayed. However without the ability to use automatic selection, 4800 clones were arrayed, which was a library size sufficient for a pilot study of *Mycobacterium tuberculosis* gene promoter expression *in vivo*.

4800 clones represents 15% of the total number of clones required to cover 95% of the *M. tuberculosis* genome had the generation of the *M. tuberculosis* H37Rv DNA fragments been completely random. As the DNA was digested with three restriction enzymes, the fragments are not random and therefore the number of clones required to statistically cover the entire *M. tuberculosis* genome is fewer than the 32951 previously stated. If promoters involved in the infection process are to be identified then a method to select clones harbouring these promoters must be developed.
CHAPTER 3

*IN VITRO SELECTION OF CLONES EXHIBITING UP-REGULATED PROMOTER ACTIVITY*
With 4800 clones of the Mycobacterium tuberculosis H37Rv promoter trap vector harboured within M.bovis BCG Pasteur (the library clones) selected randomly and arrayed in a 96 well microtitre™ format, the next requirement was to develop a method that identifies genes involved in the intracellular survival of M.tuberculosis. To identify genes up-regulated during infection a screening method would compare the expression of each of the 4800 clones in vitro and in cultured macrophages. This idyllic situation is impractical due to the enormous task required to infect cultured macrophages with all 4800 of these clones and to assay the level of expression of lacZ in vitro and in vivo.

3.1.1 In Vitro M.tuberculosis Stress Stimuli Mimicking Intracellular Conditions.

The conditions within the phagosome, discussed in Section 1.7, presumably control the ability of mycobacteria to survive and grow. Workers have therefore sought to mimic in vitro these intracellular environmental conditions. In 1978, Jackett et al. (1978) investigated the connection between the resistance of 6 strains of M.tuberculosis, with different virulence in guinea pigs, to their sensitivity to low pH, to hydrogen peroxide (H$_2$O$_2$) at different pH values and to superoxide (O$_2$). Low virulence was associated with susceptibility to hydrogen peroxide except for the laboratory-attenuated strain H37Ra. Susceptibility to an acid environment showed no correlation with the virulence of the strain but the tuberculocidal effect of H$_2$O$_2$ was significantly increased at low pH. All strains were uniformly resistant to superoxide. Later experiments by other workers attempted to induce de novo synthesis of proteins
using stress stimuli (Monahan et al., 1994). The stimuli used were heat and oxidative stress. Heat shock induced known heat shock proteins but de novo protein synthesis was not detected in mycobacteria exposed to oxidative stress and in some cases protein synthesis was inhibited. Superoxide had no effect on protein synthesis. Importantly, Monahan et al. acknowledged the fact that heat and oxidative stress were the only stress stimuli studied and the effect of other stimuli associated with intra-macrophage environment, such as low pH, iron deprivation and reactive nitrogen intermediates, were not investigated.

The effects of low pH, iron concentration and reactive nitrogen intermediates on Mycobacteria have been investigated by other groups. The susceptibility of three different strains of the Mycobacterium avium complex (one strain of M. avium and two strains of M. intracellare) to nitric oxide generated chemically by the acidification of nitrite ions, as well as nitric oxide generated by rat alveolar macrophages, showed that the M. avium strain was more susceptible to both forms of generated nitric oxide than the M. intracellare strains (Doi et al., 1993). The in vitro resistance of mycobacteria to reactive nitrogen intermediates, generated by sodium nitrite at an acidic pH, has been found to have a significant positive correlation to virulence of strains in guinea pigs (Obrien et al., 1994).

In vitro infection of human THP-1 cells with M. tuberculosis has identified 16 induced proteins and 28 repressed proteins (Lee and Horwitz, 1998). In the same study the effects of stress conditions including heat-shock, low pH and hydrogen peroxide were also investigated with the pattern of induced and repressed proteins
being unique to each stress condition. However, 10 of the *M.tuberculosis* proteins induced by the macrophages were also induced with the stress conditions.

One of the molecular methods of showing the effect of *in vitro* stress stimuli on *de novo* *M.tuberculosis* protein synthesis is two-dimensional (2D) gel electrophoresis. Garbe *et al.* (1996) used both reactive oxygen and nitrogen intermediates generated by chemical donors *in vitro* to investigate the effect on *M.tuberculosis* protein synthesis. Menadione, a redox cycling compound, caused the enhanced synthesis of seven polypeptides – six of these proteins were heat shock proteins. However the chemical release of nitric oxide induced 8 polypeptides of which only 1 could be identified as a heat shock protein. Heat shock proteins routinely appear in 2D gel analysis as proteins induced by *in vitro* stress stimuli (Garbe *et al*., 1999). Iron concentration has been shown to be another stress stimuli (Wong *et al*., 1999). Again with the use of two-dimensional (2D) gel electrophoresis, in *M.tuberculosis* Erdmann at least 15 proteins were induced and 12 repressed under low iron conditions.

Infected macrophages with all the library clones would prove an enormous task. The intracellular environment that *M.tuberculosis* is exposed to is evidently a complex environment, which is still not entirely understood. However key elements of the bactericidal abilities of the macrophage have been identified.

Reactive nitrogen and oxygen intermediates have been identified as chemicals likely to be utilised by the macrophage to kill intracellular *M.tuberculosis*. Iron deprivation is also a mechanism by which macrophages inhibit the growth of
ingested bacilli. It has been suggested that acidic conditions do not play a role in the microbicidal ability of the macrophage when infected with *M. tuberculosis* (Sturgill-Koszycki *et al.*, 1994). Phagosomes containing mycobacteria failed to acidify below pH6.3 – pH6.5. However the study only measured the pH for only a short length of time (100 minutes) post-infection and this may not reflect the early or steady state events. This may be important considering the previously mentioned work by Jackett (1978) that showed that the bactericidal effect of hydrogen peroxide was enhanced by the presence of low pH. Obviously, due to our poor understanding of the phagosome environment, there are likely to be other conditions that *M. tuberculosis* could be exposed to. All of these conditions undoubtedly act in unison against an invading microorganism but each environmental condition probably causes a specific response from *M. tuberculosis*. The response by *M. tuberculosis* is likely to involve an alteration in the level of expression of certain genes and it is reasonable to expect that proteins generated by the expression of these genes would be involved in the survival of *M. tuberculosis* intracellularly.

### 3.1.2 Strategy

The ability of *Mycobacterium bovis* BCG Pasteur to tolerate low pH and high reactive nitrogen intermediates, generated by acidified sodium nitrite, was investigated by exposing *in vitro* grown *M. bovis* BCG Pasteur to these stresses. Initially, the survival of *M. bovis* BCG Pasteur grown in media at low pH and containing a high concentration of reactive nitrogen intermediates was measured. This identified the lowest pH and highest concentration of RNI which would allow *M. bovis* BCG Pasteur to grow and these conditions were subsequently used with the
Mycobacterium tuberculosis H37Rv promoter trap vector library harboured within M.bovis BCG Pasteur. If the 4800 M.bovis BCG Pasteur promoter trap clones were to be exposed to one of these suspected environmental conditions in vitro, it may be possible to identify clones harbouring promoter-like M.tuberculosis DNA which regulate gene expression related to that specific condition. If this were the case, the expression of lacZ in these clones would be up-regulated in the presence of this environmental chemical stimuli compared to the control.

Therefore all of the 4800 clones of the library were exposed to two conditions which may be part of the defence mechanism employed by the macrophage. Each clone was grown in control media, media at a low pH and media at a low pH containing sodium nitrite. At acid pH, nitrite ions are converted to nitrous acid which subsequently decomposes to generate other reactive nitrogen intermediates including nitric oxide (Taylor et al., 1927). Bacteria grown in acidified media did not only identify promoter fragments up-regulated by the low pH but also acted as a control for the RNI medium which is at the same low pH. The amount of β-galactosidase enzyme produced by each clone would directly report the level of expression of lacZ being driven by the M.tuberculosis DNA cloned within the plasmid. As the amount of β-galactosidase produced by each clone can be measured, the expression of lacZ in the low pH and low pH with sodium nitrite media could be compared to the control medium and give a direct indication of the promoter activity. This preliminary in vitro screening mimicking in vivo conditions reduced the M.tuberculosis library clones to a smaller number containing the most reactive promoter fragments. These few clones were subsequently be used to infect cultured macrophages.
Chapter 3: In Vitro Selection of Clones Exhibiting Up-regulated Promoter Activity

3.2 Materials and Methods

3.2.1 Initial Selection of *M. tuberculosis* H37Rv Library Clones of Interest.

The 4800 library clones arrayed in fifty 96 well microtre™ plates were subcultured with the use of a 96 pin replicator into selective 7H9 broth media at pH 6.6 (control), pH 5.4 and pH 5.4 containing fresh sodium nitrite (NaNO₂)(0.5mM) and also onto solid selective 7H11 media at pH 6.6 (control), pH 5.4 and pH 5.4 containing fresh NaNO₂ (0.5mM). The pH of both solid and liquid media was reduced to the required level by the addition of concentrated hydrochloric acid before autoclaving. All media contained 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) and clones were incubated at 37°C. Each of the clones in the liquid media were grown for 7 days and scored, by eye, on a scale of 0-3 on the level of blueness seen (with 0=no blue colouration and 3=very blue). The clones on solid media were grown for 20 days and scored, by eye, on a scale of 0-10 on the level of blueness seen (with 0=no blue colouration and 10=very blue).

3.2.2 Selection of *M. tuberculosis* H37Rv Library Clones for Further Investigation

Clones identified as being of interest from section 3.2.1 were subcultured into a larger volume of selective 7H9 broth media and grown at 37°C, without shaking, for 4 days. Each clone was then subcultured into selective 7H9 broth media at pH 6.6 (control), pH 5.4 and pH 5.4 containing fresh NaNO₂ (0.5mM) as before.
Four days post-inoculation, a 0.5ml aliquot was removed from each culture, (diluted in 7H9 liquid media if necessary) and the absorbance at 600nm of each sample was measured. A further 0.5ml was removed and the activity of the β-galactosidase assayed. The culture was placed in a Hybaid Ribolyser™ blue tube containing 0.5ml DTT (0.5mM) in PBS, processed in a Hybaid Ribolyser™ at speed 5.0 for 40 seconds, cooled on ice for 1 minute and frozen at −80°C to be assayed for β-galactosidase activity with Galacto-light plus. The β-galactosidase assay with Galacto-Light Plus substrate was carried out as described in Appendix 2.
3.3 Results

3.3.1 Growth of *M. bovis* BCG Pasteur at Low pH and in the Presence of Reactive Nitric Intermediates (RNI).

*M. bovis* BCG Pasteur was grown in 40ml liquid 7H9 media at pH 6.6, pH 5.4, pH 5.0, pH 4.8, pH 4.6, pH 4.4, pH 4.2, pH 4.0 and incubated at 37°C, without shaking. It was already known that *M. bovis* BCG Pasteur was capable of growing in medium at pH 5.4 (pers. comm.: Stewart, G.R.). However, for this study it was necessary to stress the bacteria as much as possible whilst still allowing them to grow at, or near, the rate of mycobacteria grown in liquid medium at pH 6.6. Every 6 to 12 hours, a 0.5ml aliquot was removed and the absorbance of the culture was measured at 600nm (diluting confluent samples in complete liquid 7H9 medium if necessary). The absorbance of each aliquot was plotted on a log scale against time. *M. bovis* BCG Pasteur grows in 7H9 medium at pH 5.4 almost as well as it grows in the same medium at pH 6.6 (Figure 3.1). However, when grown in more acidic media, *M. bovis* BCG Pasteur appears not to replicate as indicated by the lack of increase in OD$_{600}$.

At pH 5.4, the acidity is low enough to generate reactive nitrogen intermediates from sodium nitrite. Therefore, *M. bovis* BCG Pasteur was grown in 40ml 7H9 liquid media at pH 5.4 with final concentrations of 0.25mM, 0.5mM, 1.0mM, 2mM, 4mM, 8mM and 12mM sodium nitrite. The sample aliquots were treated as described previously and the absorbance at 600nm measured. At concentrations of sodium nitrite above 2mM there is little if any bacterial growth (See Figure 3.2). *M. bovis* BCG Pasteur grown in medium containing 1mM sodium nitrite do grow but, within the time scale of this experiment, only reach about half the cell
density of bacteria grown in the control medium at pH 6.6. Figure 3.2 also shows that there is little difference in the growth of *M. bovis* BCG Pasteur when grown in liquid media at pH 6.6, pH 5.4 and liquid media at pH 5.4 containing 0.25 and 0.5mM sodium nitrite.

The object of this short study was to identify the lowest pH value and highest sodium nitrite concentration that *M. bovis* BCG Pasteur can grow ‘normally’ at. With the knowledge that *M. bovis* BCG Pasteur is capable of growing in 7H9 liquid media at pH 5.4 and 7H9 liquid media at pH 5.4 containing 0.5mM sodium nitrite as well as it grows in the control media at pH 6.6, the entire library will be subjected to these conditions.
Chapter 3: In Vitro Selection of Clones Exhibiting Up-regulated Promoter Activity

3.3.2 First Round In Vitro Screening of Library Clones.

All 4800 library clones were subcultured with a 96 pin replicator into both solid and liquid selective media containing X-gal, which will turn blue on expression of β-galactosidase by any clone. Solid and liquid media were both used because some clones grew better in one or other of the media. Additionally (as mentioned in Chapter 2) some clones grown on solid 7H11 media do not turn blue but rather turn the media blue. Clones that do not become blue but turn the medium blue or clones that are weakly expressing lacZ are easier to detect in liquid rather than solid media.

Each clone was subcultured in both solid 7H11 and liquid 7H9 media at pH 6.6 (control), pH 5.4 and pH 5.4 containing 0.5mM sodium nitrite. Differential
expression of promoter activity can be directly related to the intensity of the blue
colouration of the clone. This is the simplest method of identifying potentially
up-regulated promoters for such a large number of clones. It is possible to measure
levels of blue colour produced digitally by using an image analyser. However
observational determination was as good at assigning an intensity of blue colouration
in addition to being able to identify differences in clone morphology. In order to
maintain consistency, the scale below was used to score clones grown on the three
different types of selective solid media (see Figure 3). Each clone was subcultured
and incubated at 37°C on each of the three solid 7H11 media types. After 21 days
incubation, each clone was independently compared with the colouration scale and
allocated a number representing the level of blueness.

Determination of blue colouration of clones in liquid medium was not without
technical difficulty. Liquid 7H9 medium, despite being fairly clear, has a faint yellow
colour. Therefore a colour scheme analysis was extremely difficult to devise. To
minimise any inconsistencies a reduced scale was used with a range of only 0-3 with
0 representing no blue colour at all and 3 being very blue. Each clone was subcultured
and incubated at 37°C in each of the three liquid media types. After 7 days incubation,
each clone was independently compared with the colouration scale and allocated a
number representing the level of blueness.
Figure 3.3: Scale of intensity of blue colouration generated by *M. bovis* BCG Pasteur clones harbouring the *M. tuberculosis* promoter trap vector when grown on solid 7H11 medium containing X-gal.

The scores, directly related to the expression of *lacZ*, for each clone grown on the control solid 7H11 medium were compared with the score given to the clone grown on solid media at pH 5.4 and pH 5.4 containing 0.5mM sodium nitrite. The scores for each clone grown in the three liquid 7H9 media were compared in a similar manner. For example, Table 4.1 shows the scores, for the three different liquid media, for 6 clones from the same 96 well microtitre™ plate.

<table>
<thead>
<tr>
<th>Clone Number</th>
<th>Liquid 7H9</th>
<th>Liquid 7H9 pH 5.4</th>
<th>Liquid 7H9 pH5.4 0.5mM NaNO₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>A10</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>B10</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C10</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>D10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E10</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>F10</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3.1: Table of six *M. bovis* BCG Pasteur clones harbouring the *M. tuberculosis* promoter trap vector. All clones from the same 96 well microtitre™ plate and grown in liquid 7H9 medium at pH 6.6, pH 5.4 and pH 5.4 containing 0.5mM sodium nitrite. For example, clone A10 clearly expresses more *lacZ* when grown in medium at pH 5.4 than it does in the control medium. Clone A10 does not express any *lacZ* when grown in liquid 7H9 medium at pH 5.4 containing 0.5mM sodium nitrite.

A total of 333 clones were identified as showing a difference in the level of *lacZ* expression, either higher or lower compared to the solid or liquid control media, in either or both of the stressing media. These were selected for further investigation. Clones exhibiting up-regulation of *lacZ* expression when grown on stressing media, compared to the control medium, numbered 240, whilst 80 clones showed...
down-regulation. Thirteen clones grew and showed levels of expression when stressed but failed to grow on the control media. Since each of the clones were exposed to the stressing conditions in both solid and liquid media, there should be some correlation between the results of the two media types. Of the clones exhibiting up-regulation of lacZ expression only 4 correlated between solid and liquid media, whilst 14 down-regulated clones correlated. Twenty clones exhibited different regulation between solid and liquid media. The correlation between solid and liquid media is therefore quite low, however this may be due in part to technical difficulties when transferring clones to solid media using a 96 pin replicator. Indeed nearly 70% of the clones selected for further investigation were identified from growth in the liquid media. Roughly 4.5% of the 4800 clones failed to re-grow on any media, accordingly these clones were not re-tested.

When the 4800 clones were initially selected and arrayed from the transformation plates containing X-gal, each of the fifty 96 well microtitre™ plates contained either blue or white clones. The selection process described above identified clones that required further investigation from 32 different 96 well microtitre™ plates. Of the 333 clones identified, 207 had been initially arrayed as being blue.

### 3.3.3 Second Round Screening of \textit{M.tuberculosis} H37Rv Plasmid Library Clones for Further Investigation

The 333 clones selected as having an altered level of lacZ expression when stressed, compared to unstressed, were identified previously using a qualitative assay.
In order to reduce the number of clones to be progressed towards an in vitro macrophage infection experiment and to provide a quantitative verification of the results above, the 333 clones were individually subcultured into selective liquid 7H9 broth at pH 6.6 (control), pH 5.4 and pH 5.4 containing 0.5mM sodium nitrite. Media contained no X-gal, as the level of expression of lacZ would be measured by using a β-galactosidase assay. Four days post-inoculation, the number of mycobacterial cells present in each culture was estimated by measuring the absorbance of the culture at 600nm. 0.5ml of the culture was also ribolysed in 0.5ml 0.5mM DTT in PBS and frozen. (DTT solution in PBS stabilises any β-galactosidase enzyme present.)

To measure the β-galactosidase expressed in this subset of clones, a different assay procedure was adopted, taking advantage of the availability of a more sensitive product to measure lacZ expression. The use of Galacto-Light Plus (Tropix Inc.) is described in the methods appendix. This chemiluminescent assay exhibits over three orders of magnitude greater sensitivity compared to colorimetric β-galactosidase assays, such as the Miller test. There is also a large signal-to-noise ratio making it suitable for monitoring low levels of reporter activity and thus it was used in preference over of the conventional Miller β-galactosidase assay.

Ideally for reliable quantitative results, each clone would have been grown in the three media types and assayed in triplicate in the three liquid media generating 8991 samples. Handling this number of samples is impractical in the circumstances of this project. The primary objective at this stage was to reduce the number of clones to a manageable number for studying the lacZ expression in cultured macrophages.
Dispensing with replicate assays is a necessary compromise that still allows a rational selection of an appropriate number of clones. Therefore each clone and the two clones of \textit{M. bovis} BCG Pasteur harbouring pUS1800 and pUS1808 respectively, as negative and positive controls, were subcultured into selective liquid 7H9 broth at pH 6.6 (control), pH 5.4 and pH 5.4 containing 0.5mM sodium nitrite once and assayed only once.

Each 96 well microtitre\textsuperscript{TM} plate was set up with a blank well of 20\textmu l of 0.5mM DTT in PBS, with the rest of the wells containing samples. With the luminescence measured, the sample readings were compensated to reduce any background effect of the DTT and the PBS. As the absorbance at 600nm of each culture was measured prior to ribolysation, the number of cells added to each assay could be calculated by assuming that an \textit{M. bovis} BCG Pasteur culture with an OD\textsubscript{600nm} of 1.0 contains 1x10\textsuperscript{8} cells/ml. Using the value, light units (LU)/10\textsuperscript{6} \textit{M. bovis} BCG Pasteur cells were calculated and relative light units expressed in this fashion.

Appendix 3 contains a table that shows the LU/10\textsuperscript{6} \textit{M. bovis} BCG Pasteur cells of each clone grown in the control medium and ratios of LU/10\textsuperscript{6} \textit{M. bovis} BCG Pasteur cells of the clones grown in low pH and high nitric oxide media compared to the control medium.

A number of clones, highlighted in Appendix 3 with green ink, showed no Light Units (LU)/10\textsuperscript{6} \textit{M. bovis} BCG Pasteur cells when grown in control medium but exhibited luminescence when grown either or both of the stressing media. It these cases a ratio of regulation can not be generated. In many case this may be
insignificant, especially when the Light Units (LU)/$10^6$ *M. bovis* BCG Pasteur cells generated from growth in stressing media is low. However this is not the case with a number of clones, such as clones 295 and 322, which exhibit considerable Light Units (LU)/$10^6$ *M. bovis* BCG Pasteur cells when grown in the stressing media. These clones may contain promoter like cloned DNA, which may be of interest.

Once all 333 clones grown in all three media had been assayed, the next step was to decide which clones were to be used in macrophage infection experiments. By dividing each of the LU/$10^6$ bacterial cells for both the stressing media by the LU/$10^6$ bacterial cells produced by the clone grown in the control medium, a ratio of 'up-regulation' was generated. All clones that generated an 'up-regulation' ratio from growth in the nitric oxide stressing media greater than 3 or more were selected for further investigation. This process was repeated using the ratio for growth in the pH stressing and control media for each clone. This process identified 53 clones, from the previous 333 clones, that were considered eligible by these selection criteria for future investigation.

It is important to note the wide variation in expression being exhibited by the clones identified for further investigation. The majority of the clones have a basal expression of between 100 and 1000 Light Units (LU)/$10^6$ *M. bovis* BCG Pasteur cells. This is in contrast to the clone harbouring pUS1808, containing the *hsp60* promoter, which exhibits 8945 Light Units (LU)/$10^6$ *M. bovis* BCG Pasteur cells. However it was observed with clone 233 that more than 2 fold more LU/$10^6$ *M. bovis* BCG Pasteur cells than the pUS1808 clone were expressed and this clone exhibits up-regulation of 4.7 when stressed by low pH medium. This implies that this is a very
strongly expressing clone as the \( hsp60 \) promoter is generally considered a strong promoter. There are a number of clones which exhibit ratios of up-regulation greater that 10, but the majority of clones are up-regulated by between 3 and 5.

It can be clearly seen from Figure 3.4 that the clones selected by the process of a ratio of ‘up-regulation’ exhibit great variation in their relative levels of expression up to a 5 log variation in LU/10^6 bacterial cells. Some clones are ‘up-regulated’ by the presence of either low pH or a high concentration of reactive nitrogen intermediates, some clones by both.
Chapter 3: In Vitro Selection of Clones Exhibiting Up-regulated Promoter Activity

Figure 3.4: Graphs of Light Units/10^6 bacterial cells against the *M. bovis* BCG Pasteur clone number. The clones shown are the 53 clones which showed ‘up-regulation’ of the cloned *M. tuberculosis* DNA driving expression of *lacZ*, in either of the *in vivo* mimicking conditions, by a factor of at least 3 compared to the control 7H9 medium.
Chapter 3: In Vitro Selection of Clones Exhibiting Up-regulated Promoter Activity

Figure 3.4 (continued)
3.4 Discussion

The object of this study was to identify clones exhibiting greater promoter activity when grown in stressing conditions compared to growth in control medium.

It is believed that *M. tuberculosis* is exposed to high concentrations of reactive nitrogen intermediates in macrophages. However unlike other microbes, *M. tuberculosis*-containing phagosomes fail to acidify early on in the infection process. There are no studies of this in late infection. If the macrophage is incapable of acidifying the phagosome, this effect may, in part, be due to the actions of *M. tuberculosis* itself, which perhaps has a compensatory early protective response to ensure survival.

The 4800 clones were exposed to two conditions *in vitro* that are thought to play a pivotal role in arresting the infection of *M. tuberculosis*; low pH and high concentrations of reactive nitrogen intermediates. Clones were exposed to these conditions and the level of expression of lacZ of each clone was estimated by eye and compared to the control. This identified 333 clones which showed altered expression in one or both of the stressing conditions compared to the control medium. In the second screen these 333 clones were exposed to the same conditions and a more accurate measurement of lacZ expression was carried out using the Galacto Light Plus β-galactosidase assay.

It was originally thought that each screening regime might discard approximately 90% of the clones. The visual comparison reduced the number of
clones carried forward to the β-galactosidase assay from 4800 to 333, whilst the β-galactosidase assay reduced the number of clones to be screened in macrophage infection to 53 clones, representing a 93% and a 84% discard rate respectively.

Each of 4800 *M. bovis* clones were selected at random but had been arrayed in a 96 well microtitre™ format on the basis of whether they were blue or white when grown on solid 7H11 medium containing X-gal. The 53 clones selected for macrophage infection experiments come from 13 different 96 well microtitre™ plates. Of the clones selected for further investigation, 46 came from 96 well microtitre™ plates that were arrayed as containing blue clones and the remainder coming from the white arrayed plates. This is not really surprising as any genes up-regulated during infection will more than likely have some activity in vitro and therefore cloned promoters controlling these genes will be capable of driving *lacZ* expression in vitro without being stressed, hence being present in blue 96 well microtitre™ plates. However, perhaps more interesting are the clones originally arrayed as having no *lacZ* expression when arrayed but which have subsequently been selected as expressing *lacZ* under challenge conditions. The reason that these clones may be of greater interest is that, if they contain cloned promoters, the genes controlled by these promoters may only be activated by stressing conditions designed to mimic those in vivo.

During both of the stressing selection processes, not only were clones that potentially contain up-regulated promoters identified but also clones whose *lacZ* expression was reduced. These clones contain potential promoters which appear to be...
down-regulated and these may be just as important as any of those which appear up-regulated. However, within the remit of this study, these clones were not investigated any further.

One limitation of this study is that the growth phase of the assayed clones is not considered. The promoters of different genes are activated at different stages in the growth phase according to their function in the cell cycle. This is shown by the positive control plasmid, pUS1808. It has been previously identified that the Hsp60 protein (GroEL2) is expressed at low levels during the \textit{in vitro} exponential growth phase but expression is increased during the late-exponential and stationary phases (Batoni \textit{et al.}, 1998). Batoni \textit{et al.} also found that \textit{hsp60} expression increased when \textit{M.bovis} BCG was exposed to elevated temperatures but no up-regulation was observed when exposed to hydrogen peroxide (expression levels were not measured when \textit{M.bovis} BCG was exposed to low pH or reactive nitrogen intermediates). However in this study, the pUS1808 clone containing the \textit{hsp60} promoter (expression being measured in 3 different growth conditions) has a lower level of expression of \textit{lacZ} when the clone is stressed. This most likely due to the growth phase of the \textit{M.bovis} BCG Pasteur cultures harbouring pUS1808 in the three different conditions (see Figure 3.4).

Clone 295, which comes from a 96 well microtitre\textsuperscript{TM} plate initially selected as being blue, is another example of where care should be taken in interpreting the results from this \textit{in vivo} mimicking study. This clone exhibited no \textit{lacZ} expression when grown in the 7H9 liquid control medium. However when the expression of \textit{lacZ} was measured, for the clone grown in stressing condition, significant LU/ 10\textsuperscript{6}
bacterial cells were measured. This may or may not be due to the absence of \( \text{lacZ} \) expression when grown in the control medium or may be due to the fact that the each clone was only measured once and thus the measurement of \( \text{lacZ} \) expression is not actually correct. Since there was no detected expression when the clone was not stressed, a ratio of induction can not be calculated but clone 295 exhibited significant levels of \( \text{lacZ} \) expression when it was stressed. Clone 295 will therefore be included in further macrophage infection studies.

Importantly, the rationalising process of exposing the 4800 \( M.\text{bovis} \) library clones to conditions designed to mimic two of the environmental conditions believed to be within the \( M.\text{tuberculosis} \)-containing phagosome has reduced the number of clones to be infected by 99%. These clones were selected for progression into the macrophage challenge assay to determine whether they would also differentially express \( \text{lacZ} \) \textit{in vivo}. 
CHAPTER 4
MACROPHAGE CELL LINE INFECTIONS WITH \textit{M. bovis} BCG PASTEUR LIBRARY
4.1 Introduction

In Chapter 3, 53 of the 4800 *M. bovis* BCG Pasteur clones were identified as harbouring cloned *M. tuberculosis* DNA with promoter-like activity, which was up-regulated when stressed *in vitro*. The object of stressing the clones *in vitro* was to generate a smaller number of clones which could subsequently be used in macrophage infection studies.

Ideally both a positive and negative control should be used to compare the level of lacZ expression with the 53 selected clones during a macrophage infection experiment. In Chapter 2 two plasmids were constructed – the basal plasmid, pUS1800, and the positive control plasmid, pUS1808. However the positive control plasmid pUS1808 was constructed with a heat shock promoter (*hsp60*) driving the expression of lacZ and, when constructed, it was thought the expression of the Hsp60 protein was constitutive and unaltered during infection. However, during the time frame of this study it was shown that the *hsp60* gene (Rv0440/ *groEL2*) is up-regulated during infection and is therefore unsuitable for use as a positive control in this instance (Batoni *et al.*, 1998). pUS1800 was used as a negative control whilst pUS1808 was not used in the macrophage infection studies.

4.1.1 Strategy

All 53 clones, with the addition of Clone 295, (identified previously in Chapter 3) were used in cultured macrophage infection experiments. Two types of cultured macrophages cell lines, mouse J774 and human THP-1 macrophages, were
infected with each of the clones (Unkeless, 1979; Auwerx, 1991). Both cell lines were
used as it is not clear whether there would be any difference in the environmental
conditions inside the vacuoles of the two different macrophages. It is thought that the
two cell lines may well differ in their ability to generate reactive nitrogen and oxygen
intermediates.

At pre-selected time intervals, the internalised *M. bovis* BCG Pasteur clones
were recovered from the macrophages and β-galactosidase assays carried out using
the Galacto Light Plus assay, along with any relevant controls. The Light Units of
each clone post-infection was compared with any relevant controls and a small
number of clones that still show up-regulated *lacZ* expression were selected for
further investigation.
4.2 Materials and Methods

4.2.1 Macrophage Infection

4.2.1.1 J774 mouse macrophage cell line (Unkeless, 1979)

The mouse macrophage cell line J774 was maintained as described in Appendix 2. The pre-selected promoter probe and control recombinant clones were grown to mid-exponential growth phase and the bacterial numbers were estimated by measurement of the absorbance at 600nm ($A_{600}$). The mycobacterial cultures were harvested by centrifugation at 3000g for 10 minutes. The medium was discarded and replaced with Dulbecco's modified Eagles medium containing 10% foetal calf serum. Mycobacteria were added to adherent macrophages with a multiplicity of infection (MOI) of 10. Infection was allowed to proceed for the required length of time, either 2, 12 or 24 hours, and stopped by the removal of the medium. Any non-phagocytosed mycobacteria were removed by washing the J774 mouse macrophage cell line twice with 0.5 ml PBS. Mycobacteria were extracted from the macrophages by the addition of 0.6 ml SDS (0.0025% w/v). The number of internalised bacterial cells was estimated by serially diluted 100μl aliquots of the macrophage lysate, containing the internalised $M. bovis$ BCG Pasteur clones. A range of dilutions (between $10^{-2}$ and $10^{-5}$) were plated on kanamycin selective solid 7H11 media and incubated at 37°C for 21 days. After this time interval, colonies were counted.

The remaining mycobacteria (in 0.5 ml SDS (0.0025% w/v), extracted from each mouse macrophage infection, were transferred to Hybaid Ribolyser™ blue tube containing 0.5ml DTT (0.5mM) in PBS and were processed in a Hybaid Ribolyser™
at speed 5.0 for 40 seconds, cooled on ice for 1 minute then frozen at -80°C. ß-galactosidase activity was then measured with Galacto-light plus.

4.2.1.2 THP-1 human macrophage cell line (Auwerx, 1991)

The human macrophage cell line THP-1 was maintained and differentiated as described in Appendix 2. Mycobacterial numbers were estimated by absorbance at 600nm ($A_{600}$), after growth to mid-exponential growth phase in 7H9 broth. These were harvested by centrifugation at 3000g for 10 minutes, the medium discarded and replaced with Dulbecco's modified Eagles medium containing 10% foetal calf serum. Mycobacteria were added to adherent macrophages with a multiplicity of infection (MOI) of 10. Infection was allowed to proceed for the required length of time, either 2, 12 or 24 hours, and stopped by the removal of the medium. Any non-phagocytosed mycobacteria were removed by washing the THP-1 human macrophage cells twice with 0.5 ml PBS. Mycobacteria were extracted from the macrophages by the addition of 0.6 ml SDS (0.0025% w/v). The number of internalised bacterial cells was estimated by serial diluting 100µl of the macrophage lysate, containing the internalised *M. bovis* BCG Pasteur clones. A range of dilutions (between $10^{-2}$ and $10^{-5}$) were plated on kanamycin selective solid 7H11 media and incubated at 37°C for 21 days. After this time interval, colonies were counted.

The remaining mycobacteria, extracted from each mouse macrophage infection, were transferred to Hybaid Ribolyser™ blue tube containing 0.5ml DTT (0.5mM) in PBS, processed in a Hybaid Ribolyser™ at speed 5.0 for 40 seconds,
cooled on ice for 1 minute and frozen at -80°C. β-galactosidase activity was then measured with Galacto-light plus.
4.3 Results

4.3.1 Macrophage Infection with the Pre-Selected Clones.

All 53 selected *M. bovis* BCG Pasteur, with the addition of clone 295, were subcultured into selective 7H9 liquid medium. Each clone was to be used in macrophage infection experiments of both human and mouse cell lines. However, 12 clones were found not to be viable upon subculture making further study not possible. These clones were numbers 5, 62, 137, 173, 174, 194, 220, 229, 240, 282, 297 and 313. Despite repeated attempts at reculture, it was not possible to re-start growth with any of these clones or to recover the clones from the frozen cultures. These clones were therefore not tested further. Therefore the number of clones used in infection studies in human and murine macrophages totalled 42 – 41 clones up-regulated by *in vivo* mimicking stressing conditions and clone 295.

Each well of a six well microitre™ plate containing macrophage cells was inoculated with an individual clone (MOI=10) and was also subcultured in complete Dulbecco's modified Eagles medium and into selective 7H9 liquid medium as a control. All cultures were incubated at 37°C. At 2, 12 and 24 hours post-infection macrophages were washed and lysed. Control samples of BCG clones in complete Dulbecco's modified Eagles medium were taken at 2 and 24 hours whilst one liquid 7H9 control sample was taken at 24 hours. For each time point, a 100μl aliquot of the macrophage lysates and the *M. bovis* BCG Pasteur clone incubated in both complete Dulbecco's modified Eagles medium and liquid 7H9 media were serially diluted and selective plates (kanamycin) inoculated with a range of dilutions. These plates were incubated at 37°C for 21 days and colony numbers counted.
0.5ml of the remaining lysate was added to a Hybaid Ribolyser™ blue. Each ribolyser tube was ribolysed immediately, cooled on ice for 1 minute and stored at -80°C. The lysis of the macrophages, addition of the lysate to the ribolyser tubes and ribolysation was carried out as quickly as possible to minimise any protein degradation.

Each of the 42 \textit{M. bovis} BCG Pasteur clones was assayed in triplicate after inoculation into macrophage cell lines. As two cell lines were infected with each clone for three different lengths of time and that each liquid 7H9 media control was conducted twice (once for each macrophage cell line), this infection experiment was conducted once. As with the \textit{in vivo} mimicking condition experiments described in Chapter 3, these initial infections provided a means of screening clones for use in further detailed macrophage infection studies.

When using mammalian cell lines, the Galacto-Light Plus assay buffer has been designed to discriminate between any endogenous \(\beta\)-galactosidase enzyme and the bacterial form of the enzyme. The buffer is supplied at pH 8.0 which enables sensitive detection even in cell lines with relatively high levels of endogenous \(\beta\)-galactosidase activity (described in protocol supplied with Galacto-Light Plus assay). In order to ensure that the macrophage \(\beta\)-galactosidase is not affecting the detectable \(\beta\)-galactosidase being generated by each clone, each cell line was cultured for 24 hours, not infected, lysed and assayed using Galacto-Light Plus to measure the macrophage basal \(\beta\)-galactosidase activity. \(\beta\)-galactosidase enzyme activity could not be detected in either of the macrophage types.
Figure 4.1 is a graphical depiction of Light Units (LU)/10^6 bacterial cells for 4 different clones, and the negative control harbouring pUS1800, post-infection of both macrophage types. The clones in question were selected as they represent a number of key points. The results obtained from the infections of both macrophage types with all 42 clones can be seen in Appendix 4. The graph for each clone is almost the same regardless of the macrophage type infected – the graph of expression during the infection of a clone into murine macrophages is in general similar to that in THP-1 cells. Importantly each clone has a fairly consistent level of expression in each of the controls at the different time points.

It is clear from Figure 4.1 that, as before, the clones show a wide range of levels of expression of \( \text{lacZ} \). The negative control (\textit{M. bovis} BCG Pasteur harbouring the basal plasmid pUS1800) has a very low level of \( \text{lacZ} \) expression. Clone 180 appears to have greater expression of \( \text{lacZ} \) post-infection in both murine and human macrophages. In both macrophage types, \( \text{lacZ} \) expression from clone 180 has generated more LU/10^6 bacterial cells than both the controls. The greater the length of time that the clone is within the macrophage the higher the level of \( \text{lacZ} \) expression. This suggests that clone 180 potentially contains a promoter whose activation is increased during infection.

Clones 233 and 254 have been included in Figure 4.1 as examples of clones which do not appear to be up-regulated during infection. Clone 233 shows a little up-regulation of expression of \( \text{lacZ} \) after 2 hours when infected in murine macrophages but shows no induction after 12 and 24 hours infection, compared to the complete Dulbecco’s modified Eagles medium controls. Conversely when infected
into human macrophages, clone 233 is induced after 12 and 24 hours infection compared to the complete Dulbecco’s modified Eagles medium control at 24 hours but there is no up-regulation after 2 hours compared to the control. The graphs representing clone 254 show a similar trend.

Expression from clone 239 is perhaps the most straightforward to interpret. In both macrophage types, there is a considerable induction (about 2 logs) of lacZ expression 2 hours post-infection compared to the complete Dulbecco’s modified Eagles medium control at 2 hours. The expression of lacZ at the subsequent time points decreases in both murine and human macrophages, whilst the lacZ expression of the clone grown in the complete Dulbecco’s modified Eagles medium control is induced after 24 hours growth.
Figure 4.1: Bar Chart of Light Units/ $10^6$ bacterial cells for 4 clones, numbers 180, 233, 239 and 254 and control pUS1800. Graph is Light units/ $10^6$ bacterial cells against clone number and macrophage type infected. These clones were chosen as they identify a number of key interesting points. The negative control, \textit{M. bovis} BCG Pasteur harbouring pUS1800, is shown for comparison (not all bars can be seen due to the small values). Each bar represents the \(\beta\)-galactosidase detected in one infection or subculturing, assayed in triplicate using the Galacto-Light Plus assay.
It is evident that the interpretation of Figure 4.1, and Appendix 4, requires great care to identify any clones that should be investigated further. The increase in LU/10^6 bacterial cells of clones 180 and 239 post-infection of both macrophages, compared to the clone incubated in complete Dulbecco's modified Eagles medium for the same length of time, suggest that they may be up-regulated during infection. Figure 4.1 also suggests that the expression of lacZ in clones 233 and 254 are not up-regulated.

Clone 295 was included within the clones for further investigation as an example of a clone whose expression of lacZ could not be detected when not stressed but could be detected when the clone was grown in stressing media. The clone exhibits relatively low levels of β-galactosidase activity when incubated in the control media and in the complete Dulbecco's modified Eagles medium after 2 hours growth (see Appendix 4). However during the infection of both murine and human macrophages the lacZ expression of clone 295 is increased during the initial stages of the infection, up to 12 hours. At 24 hours post-infection, the expression of the clone within the macrophages is similar to the expression of the clone incubated in complete Dulbecco's modified Eagles medium after 24 hours. This indicates that clones showing elevated expression of lacZ when grown in low pH and/or raised nitric oxide intermediates but have undetectable levels of expression when not stressed merit investigation via macrophage infection.

There are a number of clones which exhibit high levels of lacZ expression when incubated in complete Dulbecco's modified Eagles medium, compared to the liquid 7H9 control medium. Taking clone 279 as an example, the expression of lacZ
after 2 hours incubation in complete Dulbecco's modified Eagles medium and at all
time points during the infection is not dissimilar to the \( \text{lacZ} \) expression of the clone
when grown in the liquid 7H9 control medium. However, after 24 hours incubation in
complete Dulbecco's modified Eagles medium, the \( \beta \)-galactosidase activity of clone
279 is considerably elevated compared to the growth in the liquid 7H9 control
medium. This may be a significant stress response to serum factors as mycobacteria
are not thought to grow in complete Dulbecco's modified Eagles medium.

The results of the infection experiments, shown in Appendix 4, were not
considered in themselves sufficient to decide which clones should be selected for
further investigation, therefore a second criterion of selection was introduced. The
ultimate aim of this work was to identify promoters, and therefore their genes, which
are up-regulated during infection. The best way of identifying clones carrying
genuine putative promoters responsible for up-regulation is to sequence the
\( \text{M.tuberculosis} \) H37Rv DNA inserted at the \( \text{NarI} \) site of pUS1800. Therefore all 42
clones were sequenced in both directions and homology to the \( \text{Mycobacterium}
tuberculosis \) H37Rv genome, held at the Sanger centre, determined
(http://www.sanger.ac.uk/Projects/M_tuberculosis/). This is discussed in more detail
in Chapter 5

Once the homologies of the 42 clones had been identified, a number of clones
were identified in which \( \beta \)-galactosidase expression was more likely to be due to an
artefact of the cloned piece of DNA rather than a genuine promoter. In one clone the
insert contains a potential promoter that is in the wrong orientation for driving the
expression of \( \text{lacZ} \), compared to its genomic orientation. Another clone contains an
intergenic region between two converging genes and is there for unlikely to contain a potential promoter sequence. When sequenced, 15 clones were identified as containing *M.tuberculosis* H37Rv DNA internal to genes, which would suggest that they are unlikely to be promoters. A further 25 clones were identified as containing potential promoter sequences. These clones contain *M.tuberculosis* H37Rv DNA in the correct orientation directly up-stream of a gene, contain the end of one gene and the start of the next gene or contain regions of *M.tuberculosis* H37Rv DNA covering more than two genes (and therefore potentially contained two or more promoters).

I decided that 5 clones was a suitable number for further investigation and those selected were clones 169, 175, 176, 180 and 239. These clones were selected on the basis of containing only one promoter-like sequence whilst still exhibiting up-regulation of lacZ post macrophage infection compared to the complete Dulbecco’s modified Eagles medium controls. Table 4.1 gives brief descriptions of the *M.tuberculosis* H37Rv DNA inserted within these clones but a more detailed description can be seen in Chapter 5.

<table>
<thead>
<tr>
<th>Clone Number</th>
<th><em>M.tuberculosis</em> H37Rv DNA inserted at the NarI site of pUS1800.</th>
</tr>
</thead>
<tbody>
<tr>
<td>169</td>
<td>Intergenic between genes Rv1043 and Rv1044 through to within gene Rv1044</td>
</tr>
<tr>
<td>175</td>
<td>Internal to gene Rv0301 through to within gene Rv0302</td>
</tr>
<tr>
<td>176</td>
<td>Internal to gene Rv2710 through to within gene Rv2711</td>
</tr>
<tr>
<td>180</td>
<td>Internal to gene Rv1164 through to within gene Rv1165</td>
</tr>
<tr>
<td>239</td>
<td>Intergenic between genes Rv1264 and Rv1265 through to within gene Rv1265</td>
</tr>
</tbody>
</table>

Table 4.1: The clone number is shown against the *M.tuberculosis* H37Rv DNA inserted into the NarI site of pUS1800 harboured within the clone. The harboured plasmid was sequenced across the inserted *M.tuberculosis* DNA bi-directionally and homologies identified against the *M.tuberculosis* H37Rv genome held at the Sanger centre. The Rv numbers correspond to the annotations of Cole *et al.* (1998).
4.3.2 Re-Infection of 5 Clones Showing Up-Regulation during Initial Infection.

The five clones above, selected on the basis of apparent up-regulation during infection of murine and human macrophages and due to the fact that the sequence data suggested the presence of a genuine promoter, were re-infected into human macrophages and studied in greater detail. As the data in Appendix 4 suggests that the type of macrophage infected with a clone makes little difference to the level of expression of $\text{lacZ}$ by that clone, only one macrophage type was used subsequently. *M. tuberculosis* is a human pathogen and it was therefore decided that human THP-1 macrophages would be used in the more detailed study.

Each clone was either infected or subcultured (in complete Dulbecco’s modified Eagles medium or liquid 7H9 medium) in triplicate for 2, 4, 6, 12 and 24 hours. If a gene is responsible, in part, for survival within a macrophage, it would be expected to be up-regulated soon after phagocytosis and therefore the time scale is concentrated towards the initial stages of infection. Infections and sample collection were carried out as described previously and each sample was assayed in triplicate. Following are Figures 4.2-4.6 which are bar charts that show the Light Units/ $10^6$ bacterial cells against time of growth or time of infection for the Clones 169, 175, 176, 180 and 239.
Figure 4.2: Bar Chart of Light Units/10^6 bacterial cells against time of infection in human THP-1 macrophages for clone 169. Each bar represents the average of three independent infections or subcultures. Error bars represent triplicate samples assayed in triplicate using the Galacto-Light Plus assay as standard deviations. 2 hour infection P<0.05, 24 hour infection P<0.01.

Figure 4.3: Bar Chart of Light Units/10^6 bacterial cells against time of infection in human THP-1 macrophages for clone 175. Each bar represents the average of three independent infections or subcultures. Error bars represent triplicate samples assayed in triplicate using the Galacto-Light Plus assay as standard deviations. 2 hour and 24 hour infection P<0.001, 12 hour infection P<0.01.
Figure 4.4: Bar Chart of Light Units/ $10^6$ bacterial cells against time of infection in human THP-1 macrophages for clone 176. Each bar represents the average of three independent infections or subcultures. Error bars represent triplicate samples assayed in triplicate using the Galacto-Light Plus assay as standard deviations. 4 hour and 6 hour infection P<0.005, 12 hour infection P<0.01.

Figure 4.5: Bar Chart of Light Units/ $10^6$ bacterial cells against time of infection in human THP-1 macrophages for clone 180. Each bar represents the average of three independent infections or subcultures. Error bars represent triplicate samples assayed in triplicate using the Galacto-Light Plus assay as standard deviations. 4 hour infection P=0.02.
Figures 4.2-4.4 and figure 4.6 suggest that the clones 169, 175, 176 and 239 show evidence of up-regulation of the reporter gene $\text{lacZ}$ at some of the time points during the course of infection. The interpretation of the data in figure 4.5 is complicated although there is some suggestion of up-regulation of $\text{lacZ}$ expression at 12 hours. Figure 4.3 represents the expression of $\text{lacZ}$ for clone 175, which appears to be up-regulated at most time points during the infection in the order of at least a log of LU/ $10^6$ bacterial cells compared to the complete Dulbecco’s modified Eagles medium control. Clone 175 exhibits only a very small level of up-regulation during the first 12 hours of the infection but by 24 hours post-infection the expression of $\text{lacZ}$ of the clone from within the macrophage is at least 2 logs greater than that of the clone grown in the complete Dulbecco’s modified Eagles medium control.

Conversely, Figure 4.6 shows that clone 239 exhibits considerably more...
up-regulation of $\text{lacZ}$ expression during the first 4 hours of the infection with complete Dulbecco’s modified Eagles medium control expression rising towards the end of the infection time scale. There was also an apparent sudden increase of $\text{lacZ}$ expression at 12 hours post-infection with clone 180. Clone 176, whose expression is depicted in Figure 4.4, indicates up-regulation of the promoter throughout the time scale of the macrophage infection.

The expression of $\text{lacZ}$ in the liquid 7H9 control medium for all clones, except clone 169, is reasonably consistent. The reason for the considerable variation in the liquid 7H9 control medium for clone 169 is unclear but does need further investigation since it may be due to growth phase. However, the $\text{lacZ}$ expression of clone 169 post-infection also fluctuates but when studied as a whole Figure 4.2 suggests that the clone may well indeed be up-regulated during infection. $M.\text{bovis}$ BCG Pasteur harbouring pUS1800 and $M.\text{bovis}$ BCG Pasteur were used to infect the human macrophage cell line but did not generate any LU/ $10^6$ bacterial cells and therefore no data is shown. Un-infected macrophage controls were also carried out as controls and $\beta$-galactosidase could not be detected as found previously.

4.3.3 Replicate Assays for Clones carried out at Different Times.

The Galacto-Light Plus assays carried out for each clone grown in control 7H9 control medium during the in vivo mimicking (Chapter 3) and macrophage infection studies (this chapter) exhibited varying levels of $\text{lacZ}$ expression. This was an unexpected observation. It would be expected that a clone subcultured in the same
growth medium over a long period and assayed for lacZ expression at different times would exhibit similar levels of lacZ expression.

The luminometer, (Packard Lumicount Model AL10001) and PlateReader Software (v2.1) measures the quantity of β-galactosidase enzyme present, by scanning the 96 well microtitre™ plate to find the well producing the greatest amount of light. Once this well has been identified, the plate reader and software set this well as the maximum light emission. The main benefit of this internal adjustment is to broaden the linear range of the luminometer and thus makes the equipment more versatile. If multiple 96 well microtitre™ plates are being read, and it is expected that all readings will fall within the set linear range, then the maximum reading of the first 96 well microtitre™ plate can be used for all subsequent 96 well microtitre™ plate readings. This makes the Galacto-light Plus assay a quick assay when analysing a large number of samples.

Within this study, three Galacto-light Plus assays were carried out at different times, described in sections 3.3.3, 4.3.1 and 4.3.2. Therefore each assay was carried out with a different internal setting on the luminometer and its software. This means that without the inclusion of a positive control, the three different experiments can not be compared. The obvious positive control to be included would be pure β-galactosidase enzyme. However, the inclusion of a positive control would be difficult because, for each concentration of the standard, a separate 96 well microtitre™ plate would have to be set up.
During each assay, in order to account for the effect of the PBS and DTT that the *M. bovis* BCG Pasteur lysates where stored in, each 96 well microtitre™ plate was assayed with a negative control of 0.5mM DTT in PBS. This allowed the Light Units/10⁶ bacterial cells for each sample to be calculated to take account of the presence of PBS and DTT. Therefore all Light Units/10⁶ bacterial cells shown in Chapters 3 and 4 have been corrected so as to indicate only the level of β-galactosidase enzyme present.

Therefore despite the fact that the assays were carried out in the same manner, each of the results have to be viewed separately (see Figures 3.4, the graphs in Appendix 4 and Figures 4.1 – 4.6). The graphs in Chapters 3, Appendix 4 and Chapter 4 can not be compared. If they are compared it soon becomes apparent that there is one key problem. If an individual *M. bovis* BCG Pasteur clone were to be cultured over a long period of time it would be expected that the Light Units/10⁶ bacterial cells would remain relatively constant, ignoring the effect of growth phase on expression. Figure 4.7 is a graph that shows the Light Units/10⁶ bacterial cells for liquid 7H9 controls for clone 169 taken during the three different studies. Figure 4.7 suggests that clone 169 is not expressing the same amount of β-galactosidase protein over the period of this entire study.

This apparent difference in lacZ expression is due to the Galacto-Light Plus assays for each of the studies being carried out at different times and therefore with different internal settings on the luminometer. The reason for different internal settings in the luminometer is that different numbers of cells are being assayed.
This being the case, if all samples were to be re-assayed at the same time it would be expected that the liquid 7H9 controls for each clone assayed at different times should exhibit similar levels of Light Units/10^6 bacterial cells. In order to show that this is indeed true, the liquid 7H9 controls were re-assayed on a small scale. Since only 5 clones had been studied at every stage of the second screening process one of these clones was selected for this comparison – clone 169. Samples containing the lysate of clone 169 grown in liquid 7H9 control medium were diluted, where necessary, so that roughly the same number of cells were assayed for β-galactosidase activity with the Galacto-Light Plus assay.

<table>
<thead>
<tr>
<th>Indication of which experimental liquid 7H9 control came from</th>
<th>Light Units/10^6 * Bacterial cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>in vitro stress</td>
<td>1E+02</td>
</tr>
<tr>
<td>J774</td>
<td>1E+03</td>
</tr>
<tr>
<td>THP-1 2 Hour</td>
<td>1E+04</td>
</tr>
<tr>
<td>THP-1 4 Hour</td>
<td>1E+05</td>
</tr>
<tr>
<td>THP-1 6 Hour</td>
<td>1E+06</td>
</tr>
<tr>
<td>THP-1 12 Hour</td>
<td>1E+05</td>
</tr>
<tr>
<td>THP-1 24 Hour</td>
<td>1E+04</td>
</tr>
</tbody>
</table>

Figure 4.7: Bar chart showing a comparison of the Light units/10^6 bacterial cells for clone 169 when assayed at different times. Clone grown in liquid 7H9 medium during the three separate studies described in Sections 3.3.3, 4.3.1 and 4.3.2. β-galactosidase assays carried out at the same time are shown in the same colour, with pink, blue and purple representing sections 3.3.3, 4.3.1 and 4.3.2 respectively.
Figure 4.8 represents all of the liquid 7H9 medium samples for clone 169 collected during the studies described in sections 3.3.3, 4.3.1 and 4.3.2. The important overall picture is that, when assayed at the same time, clone 169 grown in liquid 7H9 medium on different occasions does in fact exhibit a similar level of $\text{lacZ}$ expression.

The three different studies described in Sections 3.3.3, 4.3.1 and 4.3.2 can not be directly compared with one another. However it is possible to compare the results of each study relative to the level of $\text{lacZ}$ expression of the clone grown in the 7H9 control medium.

Figure 4.8: Bar chart comparing clone 169 samples grown in liquid 7H9 medium and assayed at the same time. Assayed samples taken from sections 3.3.3, 4.3.1 and 4.3.2. The colour scheme for lanes 1-8 is as in Figure 4.7 - pink, blue and purple representing sections 3.3.3, 4.3.1 and 4.3.2 respectively. Samples 1-8 were assayed with Galacto-Light Plus at the same time with the same luminometer settings. Lane 1 is a $10^2$ dilution of the sample described in section 3.3.3. Lane 2 and 3 are the undiluted in vitro controls for the initial infections of murine and human macrophages, respectively, described in Section 4.3.1. Lanes 4-8 are 0.2, 0.2, 10$^{-1}$, 0.25 and 10$^{-2}$ dilutions of samples taken at 2, 4, 6, 12 and 24 hours post-infection of human macrophages, respectively, described in section 4.3.2. Lane 9 is an average LU/10$^6$ bacterial cells for lane 1-8.
4.4 Discussion

Both human and murine macrophages were infected with the 42 library clones (identified as appearing to have greater promoter activity when stressed by *in vivo* mimicking conditions). The level of *lacZ* expression, driven by the cloned *M. tuberculosis* H37Rv DNA, of each phagocytosed clone was compared with the *lacZ* expression of the clone incubated in complete Dulbecco’s modified Eagles medium and 7H9 control medium.

It is clear from Appendix 4 that care must be taken in the interpretation of the results as intra-experiment comparisons are extremely complex. There are some clones where post-macrophage infection expression of *lacZ* is greater than its expression of *lacZ* in the liquid controls (for example see clone 180). However, the majority of clones appear to be up-regulated at certain times during the length of the infection but show no difference in expression, compared to the complete Dulbecco’s modified Eagles medium control, at other times – such as clone 239. A further complexity is added by using two different macrophage types. For example, infecting murine macrophages with clone 292 appears to causes the expression of *lacZ* to be up-regulated toward to end of the infection time scale, compared to the complete Dulbecco’s modified Eagles medium control. However, when used to infect the human macrophage cell line THP-1, clone 292 shows no significant up-regulation compared to the complete Dulbecco’s modified Eagles medium control.

To summarise the data of this Chapter, of the 42 clones used to infect macrophages 20 clones appear to be up-regulated when used to infect both
macrophage cell lines, 8 clones are apparently down-regulated and 4 clones indicate that there is no alteration in the level of expression of lacZ once internalised within macrophages, compared to the complete Dulbecco’s modified Eagles medium control. Whether the other 11 clones are up-regulated or down-regulated is unclear because the clones appear to be up-regulated during the infection of one macrophage cell line and down-regulated during the infection of the other. These clones require further more detailed investigation as to their specific pattern of expression during infection of cultured macrophages.

With the promoter activity of the 42 clones characterised, each clone was bi-directionally sequenced and homology against the *M.tuberculosis* H37Rv genome determined (see Chapter 5). Two criteria, the initial infection data and the sequence data, were used in conjunction to select a small number of clones for further investigation – the clones numbered 169, 175, 176, 180 and 239. These clones appear to show up-regulation of lacZ expression during the initial murine and human macrophage infection study and the sequence data indicates that a genuine promoter sequence had been cloned (see Chapter 5).

The re-infection data supported the initial findings that these clones contained promoter sequences that are up-regulated during infection. Table 4.2 links these 5 clones with the *in vivo* mimicking study in Chapter 3 and shows that these clones appear to exhibit relatively high ratios of induction in both the low pH and the low pH containing sodium nitrite media. Interestingly, all of these 5 clones have been identified as being potentially up-regulated in both media types.
The ability to be able to link the second *in vitro* screen and the two macrophage infection studies is not critical and, due to the way that the light units are measured by the luminometer, difficult. Each study has to be viewed as a separate entity and data from either of the other two studies can not be used to validate the other experimental observations. However, the second *in vitro* screen and both macrophage screening processes contained controls of each clone grown in liquid 7H9 liquid medium. This being the case and ignoring any growth phase differences, it would be expected that the expression levels of *lacZ* would be relatively constant for each clone. Figure 4.8 clearly shows that this is indeed the case. There are minor differences in the level of expression of *lacZ* but this would be expected. It is clear from figure 4.8 that, despite the considerable differences in LU/10^6 bacterial cells assayed during each study, the controls in the three different studies do actually express similar levels of β-galactosidase expression.

<table>
<thead>
<tr>
<th>Clone Number</th>
<th>LU/Million bacterial cells in control medium</th>
<th>Ratio of LU for Low pH to control</th>
<th>Ratio of LU for High sodium nitrite to control</th>
</tr>
</thead>
<tbody>
<tr>
<td>169</td>
<td>264.9</td>
<td>7.6</td>
<td>4.8</td>
</tr>
<tr>
<td>175</td>
<td>121.6</td>
<td>4.4</td>
<td>2.9</td>
</tr>
<tr>
<td>176</td>
<td>240.7</td>
<td>3.9</td>
<td>2.4</td>
</tr>
<tr>
<td>180</td>
<td>29.5</td>
<td>3.2</td>
<td>3.9</td>
</tr>
<tr>
<td>239</td>
<td>267.0</td>
<td>3.8</td>
<td>7.4</td>
</tr>
</tbody>
</table>

Table 4.2: A copy of the data from Appendix 3 showing the ratio of apparent up-regulation exhibited by clones 169, 175, 176, 180 and 239. All clones exhibit up-regulation of *lacZ* in both stressing media types compared to the control medium. The ratios have been generated by dividing the LU/10^6 bacterial cells for the clone grown in low pH medium and raised nitric oxide medium by the LU/10^6 bacterial cells for the clone grown in the control medium.

The data of *lacZ* expression in the clones 169, 175, 176, 180 and 239 thus far must be viewed with caution. The sequence data (in Chapter 5) suggests that these clones contain *M. tuberculosis* H37Rv promoters cloned in the correct orientation to
drive *lacZ* expression, but flanking regions of DNA were also cloned. Promoters, which are cloned out of their natural environment, can exhibit drastically different expression levels compared to the expression from the genome. Therefore the genes associated with the promoters cloned in the five clones need to have their expression studied when being driven from the genome. This will be done using the real time reverse transcriptase PCR (see Chapter 6).
CHAPTER 5

DNA SEQUENCE ANALYSIS
5.1 Introduction

The sequencing of the entire *Mycobacterium tuberculosis* H37Rv genome has led to significant advancements in the study of this organism and tuberculosis as a disease. It is an enormously powerful tool. As all previously identified genes have been annotated and a considerable number of previously unknown genes have been identified (Cole *et al.*, 1998). Where possible, homologies to genes from other organisms have been used in order to predict the function of *M. tuberculosis* H37Rv genes. However there are still a number of hypothetical genes that do not show homology to any previously identified genes from other organisms. Sequence homologies of a gene to those of other bacteria can give some suggestion to potential function of unknown proteins.

5.1.1 Strategy

All 42 clones that were identified by previous screenings and infected into murine and human macrophages were bi-directionally sequenced, across the NarI cloning site of pUS1800, and were identified by sequence searching against the *Mycobacterium tuberculosis* H37Rv genome held at the Sanger centre (http://www.sanger.ac.uk/Projects/M_tuberculosis/). A number of other clones that were identified during the *in vivo* mimicking study as having elevated levels of lacZ expression when stressed were also sequenced for comparison.
5.2 Materials and Methods

5.2.1 Oligonucleotide Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'→3')</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH20</td>
<td>CGGTTTACAAGCATAAAG</td>
<td>pUS1800 (and subsequent plasmids) nt 109-126</td>
</tr>
<tr>
<td>RJH1800REV</td>
<td>GGACCACTTCTCGCCTCGG</td>
<td>pUS1800 (and subsequent plasmids) nt 239-221</td>
</tr>
</tbody>
</table>

5.2.2 DNA amplification for sequencing

Primers RH20 and 1800REV were used to amplify across the *NarI* restriction site of the pUS1800-based plasmids harboured within the *M.bovis* BCG Pasteur clones. Reaction conditions consisted of an initial denaturation at 95°C for 5 minutes, followed by 35 cycles with denaturation 95°C, primer annealing at 55°C and extension at 72°C for 1 minute, 1 minute and 2 minutes respectively. The final cycle was extension at 72°C for 5 minutes. The amplified DNA fragments were visualised as described in Appendix 2 (Agarose gel electrophoresis).

5.2.3 Plasmid recovery from *M.bovis* BCG Pasteur clones

When sequencing of PCR products was unsuccessful, cultures of individual clones (10ml) were centrifuged at 3000g for 10 minutes. The supernatant was discarded and the pellet resuspended in 1ml water. Each of the *M.bovis* BCG Pasteur clones were transferred to a Hybaid Ribolyser™ blue tube and processed in a Hybaid Ribolyser™ at speed 5.0 for 40 seconds. Each of the lysates were subjected to phenol/ chloroform extraction of the total DNA, including the plasmid, which was subsequently subjected to DNA precipitation (see Appendix 2). The precipitated
DNA was resuspended in 30μl water. Each DNA preparation was used to transform *E.coli* DH5α, then individual clones were subjected to PCR analysis with primers RH20 and 1800REV, as described in section 5.2.2, to ensure that the selected clone did in fact contain the pUS1800 based plasmid. Plasmid preparations were then carried out from the transformed *E.coli* DH5α cells as described in Appendix 2 and were subsequently sequenced using the plasmid as the template.

### 5.2.4 Automated DNA sequencing

DNA was sequenced by the Core Sequencing Facility at Glaxo Wellcome Research and Development PLC, Stevenage, England. All sequencing was carried out on Perkin-Elmer ABI Prism 377 Sequencers with BigDye Terminator Sequencing Chemistry. All chemicals were supplied by Perkin-Elmer. BigDye 3 cycling conditions consist of DNA denaturing at 96°C for 5 minutes and 25 cycles of primer annealing at 56°C for 15 seconds, primer extension at 64°C for 4 minutes and DNA melting at 95°C for 30 seconds. The sequencing reaction mixture consisted of 8μl neat BigDye premix, 5pmol of primer, 2μl of template, 1μl DMSO to final volume of 20μl. Samples were purified after sequence cycling using 96 well filtration blocks (manufactured by Glaxo Wellcome Research and Development PLC, Stevenage, England) containing hydrated sephadex G50, dried, then 2μl of formamide/blue dextran added to dissolve the DNA. The DNA was denatured at 96°C for 2 minutes before 1.5μl loaded onto the sequencing gel.
5.2.5 Sequence analysis

Sequence alignments of *M.tuberculosis* library clones were performed on the world wide web using the *Mycobacterium tuberculosis* genome sequence held at the Sanger centre (http://www.sanger.ac.uk/Projects/M_tuberculosis/) and the BCM Search Launcher located at the Human Genome Center, Baylor College of Medicine, Houston Texas (http://kiwi.imgen bcm.tmc.edu:8088/search-launcher/launcher.html). Sequence comparisons were also performed using a BLAST search located at the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/). Additional sequence analysis was carried out using bioinformatics computer programs specifically written for use with the *M.tuberculosis* H37Rv genome at Glaxo Wellcome Research and Development, Stevenage, England.
Chapter 5: DNA Sequence Analysis

5.3 Results

5.3.1 Sequence Analysis and the Genome.

All 42 clones used to infect murine and human macrophages in Chapter 4 were PCR amplified across the NarI cloning site of pUS1800, with the primers RH20 and 1800REV. An additional 8 clones from the 333 clones shown in Appendix 3 were also subjected to PCR in the same manner. Each amplicon was bi-directionally sequenced with the same primers (RH20 and 1800REV). The majority of sequences obtained from the pUS1800-based plasmids covered the entire insert sequence. However the sequences obtained from 12 clones did not noticeably overlap and therefore the entire M.tuberculosis H37Rv DNA sequence cloned into these plasmids had to be deduced from database sequences. A number of clones generated multiple PCR products, which were gel purified (see Appendix 2) and when sequenced independently each of the PCR products from each clone generated the same sequence. The M.tuberculosis H37Rv DNA insert in 6 clones proved very hard to amplify and/or sequence from the PCR product. The plasmids from these clones were recovered into E.coli and subsequent plasmid preparations were bi-sequenced with the primer RH20 and 1800REV, as outlined previously.

The sequences obtained in either direction from each clone, with the primers RH20 and 1800REV, were paired then independently identified against the complete genome sequence of M.tuberculosis H37Rv held at the Sanger institute (http://www.sanger.ac.uk/Projects/M_tuberculosis/). In all cases, except clone 188, the pairs of sequences identified the same region of cloned DNA confirming a single fragment insert. The M.tuberculosis H37Rv genome sequences that were identified in
each clone were copied from the Sanger institute. These *M. tuberculosis* H37Rv DNA sequences are shown in Appendix 5. In addition, information regarding the sequences and the clones harbouring the sequences is also given, such as the size of the *M. tuberculosis* H37Rv DNA insert for each clone. The genome co-ordinates of the insert have also been included. All sequences are shown in the orientation that they appear in pUS1800 and therefore the translational start site of *lacZ* is at the 3’ end of all sequences shown. The restriction enzyme sites (*AciI*, *HpaII* or *TaqI*) used to generate each of the clones are identified, as are the translational start and stop codons of any genes or portions of genes. Appendix 5 shows that 27 of the *M. tuberculosis* H37Rv DNA inserts have been cloned using the restriction enzyme *AciI*. Digestion of genomic *M. tuberculosis* H37Rv DNA with restriction enzymes *HpaII* and *TaqI* have generated 12 and 10 clones, respectively. The average size of the cloned *M. tuberculosis* H37Rv DNA inserts is 370bp.

Figure 5.1 is a depiction of the *M. tuberculosis* H37Rv genome (Cole et al., 1998). The circle represents the *M. tuberculosis* H37Rv genome with the position of all of the sequences in Appendix 5 represented by the clone number and red arrows to identify their position on the genome. It can be clearly seen that the positioning of the sequences is well spread through out the genome. However there are three portions of the genome which have not been represented by the sequenced clones. The genome co-ordinates for these 3 portions are 2101154-2614036, 2623969-2925356 and 3698200-3967697. In addition there are some small clusters of hits on the genome, the reason for this are not known. The 4800 clones picked, and subsequently assayed, to date represent 15% of the total number of clones required to statistically cover 95% of the *M. tuberculosis* H37Rv genome. Thus this study represents a small scale pilot
study to investigate the feasibility of this technology in identifying mycobacterial promoters and should not be viewed as a definitive study to cover the differential expression of all known reading frames.

Figure 5.1: Map of the *M. tuberculosis* H37Rv genome. The red arrows indicate the position of DNA inserted at the *NarI* site of the plasmid pUS1800 and harboured within each of the numbered clones.

It has been argued that bacterial genes involved in virulence are grouped together in 'pathogenicity islands' (Hacker *et al.*, 1997). Pathogenicity islands have
been identified in a number of bacteria including uropathogenic *Escherichia coli* (Ritter *et al.*, 1997; Dobrindt *et al.*, 1998; Ritter *et al.*, 1995), *Shigella flexneri* (Vokes *et al.*, 1999; Moss *et al.*, 1999), *Staphylococcus aureus* (Lindsay *et al.*, 1998), *Yersinia pseudotuberculosis* (Buchriesser *et al.*, 1998) and in *Salmonella* spp. (Wood *et al.*, 1998). If pathogenicity islands were to be present in *M. tuberculosis*, then the small clusters and 3 portions of the genome not covered in figure 5.1 may not be entirely surprising, as this study should only have identified genes associated with intracellular survival, a key requirement of the pathogenicity of *M. tuberculosis*.

As previously stated, each clone contains a different random segment of cloned *M. tuberculosis* H37Rv DNA including different genes, parts of genes or intergenic regions. During the identification of the DNA contained within each clone, the sequence position on the *M. tuberculosis* H37Rv genome and the annotation of each sequence was noted. This included the Rv numbers, which have been given to each gene by Cole *et al.* (1998) as well as indicating any genes or intergenic regions. All cloned genes or parts of cloned genes are identified in the sequences of Appendix 5 by their translational start and stop positions. This has enabled the identification of potential promoters that may be contained within each library clone. The criterion for deciding what was and what was not a potential promoter was dependent on whether any intergenic DNA directly up-stream of a gene was cloned in the correct orientation to drive the expression of *lacZ*.

Figure 5.2 represents all sequenced clones diagrammatically. Rectangles represent *M. tuberculosis* H37Rv genes containing the Rv numbers, with the direction of translation being represented by the red arrowheads. The green arrows represent
the DNA cloned within the plasmid pUS1800 harboured within each clone with the arrowhead pointing towards the start of the promoterless reporter gene lacZ contained within pUS1800. The size of all rectangles and arrows is not representative of relative size.

Of all of the clones sequenced, 16 clones contain \textit{M.tuberculosis} H37Rv DNA that is internal to genes and are therefore unlikely to contain promoters. Clone 90 contains intergenic DNA situated between 2 \textit{M.tuberculosis} H37Rv genes where the direction of transcription of the two genes converges and is therefore unlikely to contain a potential promoter. Two clones (5 and 302) contain potential promoter fragments where the orientation with regard to \textit{lacZ} expression is opposite to that driving gene expression on the genome. Six clones contain potential promoters cloned into pUS1800 in the correct orientation to drive \textit{lacZ} expression but are missing part of the 3' putative promoter DNA. Clone 188 contains a double insert whilst clone 13 contains 2 potential promoter fragments and one complete gene. The remaining 23 clones all contain probable \textit{M.tuberculosis} H37Rv promoters in the same orientation that they appear in the \textit{M.tuberculosis} H37Rv genome and the promoters are in correct orientation to drive the expression of \textit{lacZ}.
Figure 5.2: Diagrammatic representation of all sequenced clones. All *M. tuberculosis* H37Rv genes are represented by hollow rectangles containing the respective gene number (Rv). The direction of the red arrow-head within each rectangle indicates the direction of translation of each gene. Green arrows represent the region of *M. tuberculosis* H37Rv DNA contained within pUS1800 of each clone. The direction of the arrow indicates the direction of cloning, with the arrow-head representing the transcriptional fusion of the cloned DNA with the start of *lacZ*. The size of rectangles or arrows is not representative of relative size.
5.3.2 Cloned DNA Internal to *M.tuberculosis* H37Rv Genes.

DNA cloned into pUS1800, which appears in the *M.tuberculosis* H37Rv genome positioned within a gene, was considered unlikely to contain potential promoters. In total 16 clones contained *M.tuberculosis* H37Rv DNA internal to genes.

To ensure that cloning of a promoter was unlikely in these fragments, the *M.tuberculosis* H37Rv genome in the close vicinity of the cloned DNA was checked for the presence of any other open reading frames. If a second gene had been identified in the near vicinity, and its orientation on the *M.tuberculosis* H37Rv genome had been the same as the cloned DNA in pUS1800, it would have been feasible that the promoter for the down-stream gene was contained within the coding region of the first gene. In all cases where DNA internal to *M.tuberculosis* H37Rv genes was cloned, other genes in the near vicinity were either in the wrong orientation or were too distant to be likely to have their expression driven by the fragment of cloned DNA.

As an example, clone 7 contains a portion of the coding region for the *M.tuberculosis* H37Rv gene Rv1508c. The genes directly up and down-stream of Rv1508c are both situated on the direct strand of the genome. Figure 5.3 is a diagrammatic representation of the portion of the *M.tuberculosis* H37Rv genome containing genes Rv1507, Rv1508c and Rv1509. Despite the ability of the cloned DNA in clone 7 to drive the expression of *lacZ*, due to its location and orientation it is unlikely that the cloned DNA contains a true *M.tuberculosis* H37Rv promoter for any of the genes depicted in Figure 5.3.
Chapter 5: DNA Sequence Analysis

Figure 5.3: Diagrammatic representation of DNA in the vicinity of cloned DNA contained within clone 7. Genes are represented by hollow rectangles containing the respective gene numbers. Red arrowheads indicate the direction of translation of individual genes. The green arrow represents the region of *M.tuberculosis* H37Rv DNA contained within pUS1800. The direction of the arrow indicates the direction of cloning with the arrowhead representing the start of *lacZ*.

5.3.3 Clones 5, 90 and 302

Clone 5, 90 and 302 all contain intergenic *M.tuberculosis* H37Rv DNA. Clone 90 contains part of the intergenic region between the *M.tuberculosis* H37Rv genes, Rv0609 and Rv0610c (see Figure 5.2). Due to the fact that these two genes converge, it is unlikely that the cloned DNA contains a putative promoter. Clones 5 and 302 are different to clone 90 in that they do contain the potential promoters for the genes Rv0822c and Rv1853, respectively (see Figure 5.2). However, in both cases, the cloned putative promoters in pUS1800 are orientated to drive expression away from the start of *lacZ*. Figure 5.4 shows this diagrammatically for clone 302.

![Diagram 5.4: DNA in clone 302](image)

Figure 5.4: Diagrammatic representation of DNA in the vicinity of cloned DNA contained within clone 302. Genes are represented by hollow rectangles containing the respective gene numbers. Red arrowheads indicate the direction of transcription of individual genes. The green arrow represents the region of *M.tuberculosis* H37Rv DNA contained within pUS1800 and the direction of the arrow indicates the direction of cloning with the arrowhead representing the start of *lacZ*. The blue arrow represents the potential promoter fragment responsible for the expression of Rv1853.
Clone 188 requires special note as it is unusual and a cloning artefact. The amplicon spanning the NarI insert site of this clone has been generated by DNA amplification a number of times and each amplification has been bi-directionally sequenced. The plasmid was also recovered from the M.bovis BCG clone and transformed into E.coli and the purified plasmid also being bi-directionally sequenced several times. On every occasion, sequence homology with two separate regions of the M.tuberculosis H37Rv genome sequence were identified, suggesting that clone 188 contains a double insert. The primer sequencing up-stream of lacZ (1800REV) identifies a 344bp insert with the M.tuberculosis H37Rv genome co-ordinates being 4328658 – 4329001bp. The sequence obtained using the RH20, which reads towards lacZ, identifies a 482bp insert with the M.tuberculosis H37Rv genome co-ordinates being 1623341 – 1623822bp. Both of the M.tuberculosis H37Rv DNA fragments have been generated by AcI digest and both of the sequences, obtained from pUS1800 with RH20 and 1800REV, carried the junction sequence between the two fragments.

Figure 5.5 is a graphical representation of the M.tuberculosis H37Rv DNA fragment in clone 188. It shows that the putative promoter for Rv1444c is cloned in the correct orientation to drive the expression of lacZ. However, between the end of the potential promoter and the translational start site of lacZ, there are portions of both Rv1444c and Rv3856c. This has the effect of putting 699bp between the potential promoter of Rv1444c and the translational start site of lacZ. However, the translation start codons for Rv1444c and lacZ are in frame and therefore a
translational fusion has been generated. The fusion protein would consist of the initial portion of Rv1444c, a middle section of Rv3856c reading into the entire β-galactosidase protein. The function of the protein coded for by Rv1444c is unknown.

![Diagram](image)

Figure 5.5: Diagrammatic representation of *M.tuberculosis* H37Rv DNA in the plasmid pUS1800 harboured within clone 188. Genes are represented by hollow rectangles containing the respective gene numbers. Red arrowheads indicate the direction of translation of individual genes. The green arrow represents the region of *M.tuberculosis* H37Rv DNA contained within pUS1800 and the direction of the arrow indicates the direction of cloning with the arrowhead pointing towards the start of *lacZ*. The start of gene Rv1444c and the middle of gene Rv3856c are between the potential promoter of Rv1444c and the translational start of *lacZ*.

**5.3.5 Clone 233**

Clone 233 is also worthy of special note. The cloned *M.tuberculosis* H37Rv DNA is located at co-ordinates 1446292 – 1446456 and has been cloned such that the complementary strand of the *M.tuberculosis* H37Rv genome is driving the expression of *lacZ*. The cloned DNA includes the first 79bp of gene Rv1292, *argS*, which codes for an arginyl-tRNA synthase, in the opposite orientation to that of *lacZ* in the plasmid. Up-stream of Rv1292 is *argV*, which is an arginyl-tRNA, which is located on the complementary strand of the *M.tuberculosis* H37Rv genome. The end of the cloned DNA directly up-stream of *lacZ* is probably part of the promoter for *argV*. 121
However there are 24bp missing between the end of the cloned DNA and the start of \( \text{argV} \). This is diagrammatically represented in figure 5.2.

5.3.6 Clone 13

Clone 13 contains two potential promoter sequences and one complete gene, Rv0460, and is diagrammatically represented in Figure 5.2. The \( M.\text{tuberculosis} \) H37Rv DNA cloned into pUS1800 straddles 3 genes – Rv0459, Rv0460 and Rv0461 and has the genome co-ordinates 551479-552343. The 5’ portion of the gene Rv0459 and the 3’ end of gene Rv0461 have not been cloned. Gene Rv0459 has an unknown function but is similar to unidentified open reading frames in \( Xanthobacter autotrophicus \) (gene: AF029733_2 - 139aa) and \( Rhodococcus erythropolis \) (gene: REREUTPBC_1 - 186aa) – these open reading frames have been linked to aldehyde dehydrogenase genes. Both Rv0460 and Rv0461 are unknown hydrophobic proteins with the start site of Rv0461 being uncertain.

The two intergenic regions that have been cloned probably contain the promoters responsible for driving the expression of the two genes Rv0460 and Rv0461 from the \( M.\text{tuberculosis} \) H37Rv genome. The second potential promoter cloned, that of gene Rv0461, is 320bp away from the transcriptional start site of \( \text{lacZ} \). However the transcriptional start sites of gene Rv0461 and \( \text{lacZ} \) are in frame and therefore a translational fusion has been formed. Whether the transcriptional start site of gene Rv0460 is in frame with the start site of \( \text{lacZ} \) is unclear since the start site of gene Rv0460 is uncertain.
5.3.7 Incomplete Promoters

The plasmid pUS1800 is a transcriptional promoter trap vector and therefore does not require *M.tuberculosis* H37Rv promoters and down-stream genes to be in frame with the translational start site of *lacZ*. However a cloned promoter would be required to supply both the transcription initiation signals as well as the ribosome binding site (RBS) in order to initiate translation.

The potential promoters in seven of the clones (clones 166, 220, 233, 278, 289, 291 and 330) have been cloned so that not all of the intergenic region between two genes has been cloned. Clone 233 has been discussed previously. Clone 330 harbours part of the 5' coding region of the *M.tuberculosis* H37Rv gene Rv2338c as well as part of the intergenic region between Rv2338c and Rv2339. However the cloned DNA stops 59bp short of the start site of gene Rv2339. Both the transcription initiation signals and the RBS may be contained within the portion of the intergenic region that has been cloned. However this would position the RBS a considerable distance from the start point of translation of the gene Rv2339. The gene Rv2339 codes for a probable membrane protein with strong similarities to other proteins in *M.leprae* and *M.tuberculosis*.

The DNA within clone 289 contains a portion of the intergenic DNA between the *M.tuberculosis* H37Rv genes Rv2401 and Rv2402. The intergenic region between these two genes is orientated on the genome to drive the expression of gene Rv2402, which codes for a hypothetical 74.5KDa protein and contains a helix-turn-helix motif (helix-turn-helix motifs are routinely found in DNA binding proteins such as
repressors and activators). The portion of the intergenic region of DNA that has been cloned into pUS1800 is orientated, with respect to lacZ, in the same orientation that it is on the genome to drive the expression of Rv2402. However, not all of the intergenic region between Rv2401 and Rv2402 has been cloned, a 132bp region between the end of the cloned DNA and the start of the gene Rv2402 has not been cloned. This *M.tuberculosis* H37Rv DNA fragment may contain transcription initiation signals. This is diagrammatically represented in Figure 5.2.

The clones 166, 220, 278 and 291 all contain inserts of DNA up-stream of the gene Rv0440 (*groEL2*) and are diagrammatically represented in Figure 5.2. Clone 166 harbours an insert containing the 3’ end of the gene Rv0439c and the DNA up-stream of the gene Rv0440, but the cloned DNA does not contain the first 10bp of DNA directly up-stream of the gene. The *M.tuberculosis* H37Rv genome co-ordinates of this cloned DNA fragment are 528230-528690. Clones 220, 278 and 291 harbour identical inserts (the *M.tuberculosis* H37Rv genome co-ordinates of the cloned DNA are 528230-528532) containing the 3’ end of the gene Rv0439c and the up-stream region of the *M.tuberculosis* H37Rv gene Rv0440. However, in these latter clones, the DNA stops 68bp short of the gene Rv2339. Both the transcription initiation signals and the RBS may well be contained within the portion of the intergenic region that has been cloned.

It is possible that during the transformation with the plasmid library and subsequent recovery of *M.bovis* BCG Pasteur that sibling transformants could have been generated, since clones 220, 278 and 291 all contain identical regions of cloned *M.tuberculosis* H37Rv DNA. All clones were selected and arrayed at random but
clones 278 and 291 were arrayed in the same 96 well microtitre™ plate, at positions E8 and D11. Clone 220 was arrayed at a much earlier time and is arrayed in a 96 well microtitre™ plate that appears much earlier on in the 96 well microtitre™ plate numbering system. As more than one transformation of *M. bovis* BCG Pasteur, using the same plasmid library preparation, was carried out to generate the 4800 clones that have been arrayed, clones 278 and 291 were probably generated by the same transformation and are more than likely sibling clones with clone 220 being generated by a separate transformation and is therefore an independent clone. However in the context of this study the presence of sibling clones may be expected but is not particularly significant.

Clone 220 was not used during the infection of murine and human macrophages but there is expression data for clones 166, 278 and 291. Clones 278 and 291 contain an identical section of *M. tuberculosis* H37Rv DNA and do exhibit similar levels of *lacZ* expression (see Appendix 4). Clone 166, harbouring more of the probable promoter region for Rv0440, exhibits a greater level of expression than clones 278 and 291.

It is interesting to note that the gene Rv0440 codes for the chaperonin GroEL2. In Chapter 2, not only was the promoter probe vector pUS1800 constructed but pUS1808 was also constructed. pUS1808 was constructed as a positive control plasmid for pUS1800 and has the promoter of a heat shock protein (*hsp60*) cloned at the *NarI* site directly up-stream of the promoterless *lacZ* reporter gene. The gene *hsp60* and Rv0440 (*groEL2*) are different nomenclature for the same gene.
5.3.8 Clones containing Probable Promoters

In Figure 5.2, there are 22 clones that contain intergenic DNA, between two *M.tuberculosis* H37Rv genes, which is in the correct orientation within pUS1800 to drive the expression of *lacZ*. These promoters can be divided into two groups. In all cases, not only was an intergenic region cloned but also a portion of the gene directly down-stream of the intergenic region and in most cases a portion of the up-stream region.

Clone 63 is the first clone in Figure 5.2 that fits in to the group of cloned potential promoters that are flanked, within the plasmid pUS1800, by portions of both the up- and down-stream *M.tuberculosis* H37Rv genes. It contains the intergenic region between genes Rv0940c and Rv0941c as well as the 3’ coding region of Rv0941c and the 5’ coding region of Rv0940c. Both genes are situated on the complementary strand of the *M.tuberculosis* H37Rv genome and the intergenic region of DNA is orientated in pUS1800 to drive the expression of *lacZ*. Gene Rv0940c has been annotated as being similar to several *M.tuberculosis* H37Rv hypothetical proteins whilst Rv0941c has an unknown function. Therefore clone 63 probably contains the promoter for the gene Rv0940c which is driving the expression of *lacZ* in the promoter trap vector. In addition to the intergenic region of DNA, the initial bases of gene Rv0940c are also present in clone 63. Indeed 121bp of Rv0940c have been cloned between the intergenic region and the translational start site of *lacZ*. The start sites of Rv0940c and *lacZ* are in frame and thus a translational fusion between the start of Rv0940c and *lacZ* has been created.
The second group of clones is typified by clone 174 (see Figure 5.2). Clone 174 contains a section of the intergenic regions between the *M. tuberculosis* H37Rv genes Rv0129c and Rv0130 and a portion of the 5’ coding region of gene Rv0130. The intergenic region between Rv0129c and Rv0130 is 249bp however, only 161bp, before the start of Rv0130, of this region has been cloned into pUS1800 along with the first 139bp of gene Rv0130. The cloned intergenic region in pUS1800 is in the correct orientation to drive the expression *lacZ*. As in clone 63, a translational fusion has been generated between the 5’ coding region of Rv0130 and the start position of *lacZ*. The function of the protein coded for by Rv0130 is uncertain but it does have a 44% homology to nodulation proteins from *Rhizobium* ssp. and is also similar to a monoamine oxidase regulatory protein from *Archaeoglobus fulgidus*.

Of the 22 clones that contain cloned intergenic regions of *M. tuberculosis* H37Rv DNA and down-stream genes, 13 of the down-stream genes are in frame with the translational start site of *lacZ* and have therefore formed translational fusions. Since pUS1800 is a transcription promoter trap vector the presence or absence of translational fusions is not critical. However the absence of a translational fusion with *lacZ* means that the distance between the potential promoter sequence and the start of *lacZ* becomes more important since the promoter may have become too distal to be responsible for the transcription of *lacZ*.

5.3.9 Sequenced Clones not used during Infection Studies.

Clones 5, 137, 173, 174, 194, 220, 229 and 282 were identified in Chapter 3 as being potentially up-regulated during exposure to *in vivo* mimicking conditions but
were not used during any infection studies. All clones are diagrammatically represented in Figure 5.2. Of these clones, only clones 5 and 194 are unlikely to contain promoters. Clone 5 has been described previously. The \textit{M.tuberculosis} H37Rv DNA that is cloned into pUS1800 of clone 194 is internal to gene Rv3530c, which is similar to \textit{cis-1,2-dihydroxy-3,4-cyclohexadiene-1-carboxylate dehydrogenase} coding genes.

However, interestingly, the rest of these clones contain probable promoters. Clones 137, 173, 174 and 282 contain the probable promoters for genes with no known function. As previously stated, clone 220 contains the promoter for Rv0440, \textit{(groEL2 or hsp60)}. The probable promoter in clone 229 is directly up-stream of gene Rv3077 \textit{(atsF)}, which is similar to aryl- and steryl-sulfatases. Clones 173 and 229 are worthy of special note since both inserts are internal to two over-lapping genes.

5.3.10 Analysis of Cloned \textit{M.tuberculosis} H37Rv DNA in the 5 Library Clones used in Second Round of Macrophage Infection.

The 5 clones (169, 175, 176, 180 and 239) used to re-infect human macrophages in section 4.3.2 are of special interest in this study. They were studied in detail and the expression of the \textit{M.tuberculosis} H37Rv genes, whose expression may be driven by the cloned potential promoter fragments, was additionally investigated with the use of quantitative reverse transcribed PCR (see Chapter 6).

Clone 169 contains the putative promoter region (163bp) for the \textit{M.tuberculosis} H37Rv gene Rv1044 and the first 360bp of this gene. The portion of
the gene that has been cloned is out of frame with the translational start site of \textit{lacZ}. There is no portion of the DNA fragment containing the up-stream gene, Rv1043c. The gene Rv1044 has no suggested function but is similar to another hypothetical \textit{M.tuberculosis} H37Rv protein.

Clone 175 contains cloned \textit{M.tuberculosis} H37Rv DNA that includes the end of Rv0301, the start of Rv0302 and the region of DNA separating the two genes. The cloned end of the gene Rv0301 is 242bp long and the cloned region of Rv0302 is just 14bp. However all of the 139bp separating the two genes has been cloned. The translational start site of the cloned Rv0302 is in frame with the translational start site of \textit{lacZ}. Gene Rv0302 has an unknown function but is similar to several repressors and regulatory proteins. Rv0302 has a 22.7% identity with the \textit{E.coli} repressor \textit{acrR} which modulates the \textit{acrAB} operon when stressed by ethanol, high sodium chloride and in stationary phase growth (Ma \textit{et al.}, 1996). Rv0302 also has a similarity to the N-terminus of another \textit{M.tuberculosis} H37Rv hypothetical regulator and contains a probable helix-turn-helix.

Clone 176 contains a fragment of \textit{M.tuberculosis} H37Rv DNA that includes the end of Rv2710, the start of Rv2711 and the region of DNA separating the two genes. The cloned end of gene Rv2710 is 67bp long and the cloned region of Rv2711 is 55bp long. However all of the 136bp separating the two genes has been cloned. The translational start site of the cloned Rv2711 is in frame with the translational start site of \textit{lacZ}. The gene Rv2711 codes for the iron dependent repressor \textit{ideR}, which has strong homology to \textit{dtxR} coding for the diptheria toxin repressor (Schmitt \textit{et al.}, 1995).
Clone 180 contains cloned *M. tuberculosis* H37Rv DNA that includes the end of Rv1164, the start of Rv1165 and the region of DNA separating the two genes. The cloned end of the gene Rv1164 is 234bp long and the cloned region of Rv1165 is just 11bp. However all of the 25bp separating the two genes has been cloned. The translational start site of gene Rv1165 out of frame with the translational start site of lacZ. Rv1165 is similar to hypothetical proteins from a number of other organisms, such as the hypothetical protein GTP-binding protein from Syneochocystis sp. and to the elongation factor G from *Micrococcus luteus*. However no function has been assigned to Rv1165.

Clone 239 is particularly interesting since the cloned *M. tuberculosis* H37Rv DNA spans only 17bp of DNA up-stream of gene Rv1265 and the first 235bp of the gene Rv1265. The translational start site of the cloned Rv1265 is in frame with the translational start site of lacZ. The number of base pairs cloned containing the potential promoter is very small and therefore the fact that the clone shows up-regulation during the infection of human macrophages may be fortuitous. The function of the protein translated from gene Rv1265 is unknown.

Of the five clones used to re-infect human macrophages, three clones (175, 176 and 239) have formed translational fusions between a cloned 5' portion of the cloned gene and lacZ. The other two clones, 169 and 180, contain a portion of the start of a gene, Rv1044 and Rv1165 respectively, that has not formed a translational fusion with lacZ. This may not be relevant in clone 180 since the end of the putative promoter for Rv1165 ends 13 bp before the start of lacZ. The presence of 13bp is
unlikely to interfere with the initiation of transcription of \textit{lacZ}. This may not be the case with clone 169 since there are 362bp between the start of Rv1044 and \textit{lacZ}. The presence of 362bp between a putative promoter and \textit{lacZ} may generate a transcriptional fusion, with RNA polymerase generating a transcript containing both the 362 bases of the cloned Rv1044 gene and the \textit{lacZ} gene. Ribsomes could attach to the transcript at both the TTG of Rv1044 and the ATG of \textit{lacZ}.
Sequence comparison of cloned DNA with the *M. tuberculosis* H37Rv genome database shows that this promoter probe library has a good coverage of the *M. tuberculosis* H37Rv genome despite the fact that the library was not generated randomly. Indeed there are only 3 gaps in the genome coverage of the *M. tuberculosis* H37Rv of any real significance with the genome co-ordinates are 2101154-2614036, 2623969-2925356 and 3698200-3967697. The reason for these 3 gaps is unclear but may be entirely due to the statistically small number of cloned sequences analysed to date. Sequencing of other clones may identify cloned regions within these portions of the genome.

Interestingly there are only 2 sequences that have been cloned more than once. Clones 303 and 311 contain exactly the same section of the *M. tuberculosis* H37Rv genome, with the genome co-ordinates 4306285 – 4306610. These co-ordinates place the DNA fragments within gene Rv3822c and these two clones may be siblings. The second sequence that has been repeated is the probable promoter sequence for gene Rv0440, *groEL2*, which is harboured in clones 166, 220, 278 and 291. Clones 220, 278 and 291 contain identical fragments cloned into pUS1800 whilst clone 166 contains a slightly larger fragment. It is important to point out that these 4 clones were originally selected as being potentially up-regulated in Chapter 3 from 3 different 96 well Microtitre™ plates. Clones 278 and 291 were both originally arrayed in the same 96 well Microtitre™ plate but were arrayed in different wells. The protein coded for by *groEL2* is a chaperonin that has previously been annotated as Hsp60 – a heat shock protein with a molecular weight of 60KDa.
The hsp60 promoter has been cloned into pUS1800 to generate pUS1808 (see Chapter 2). A M. bovis BCG Pasteur clone harbouring the plasmid pUS1808 was subjected to in vivo mimicking conditions (see Chapter 3), but was not used during macrophage infection studies. The reason that this clone was not used during the macrophage infection studies was that gene Rv0440 had been shown to be up-regulated during infection and therefore would not be a good constitutive promoter positive control (Batoni et al., 1998). In a number of other investigations designed to study differential expression, heat shock proteins are routinely found to have an elevated expression during infection (Monahan et al., 1994; Lee and Horwitz, 1998; Garbe et al., 1996).

Sequence alignments, using ClustalW multiple sequence alignments (http://dot.imgen.bcm.tmc.edu:9331/multi-align/clustalw.html), have been carried out in order to identify any consensus sequences. The alignments were carried out using M. tuberculosis H37Rv sequences contained within all clones, only the sequences of clones containing potential promoters or with cloned intergenic regions. No consensus sequences could be identified. It has been suggested that M. tuberculosis does not have a conserved -35 transcription signal but the sequence for the conserved -10 signal is T(80%), A(90%), Y(60%), G(40%), A(60%) and T(100%), where Y indicates a pyrimidine residue (Bashyam et al., 1996). Mulder et al. (1997) have reviewed mycobacterial promoters and have calculated the mycobacterial consensus sequences to be: -35: T(91.7%), T(52.9%), G(62.3%), A(44.4%), C(56.9%) and G(30.6%)/A(26.4%); and the -10: T(68.4%), A(75.8%), T(41.1%), A(35.8%)/G(26.3%), A(33.7%)/C(28.5%) and T(75.8%).
The clones identified in Chapter 3 and 4 and sequenced within this chapter contain *M. tuberculosis* H37Rv DNA that has been identified as being potential promoter sequences that appear to cause up-regulated gene expression during infection. However it is necessary to determine that the identification of these probable promoter fragments has not occurred via random events but rather by the ordered rational selection described therein.

In order to show that the *M. tuberculosis* H37Rv DNA was not isolated due to random events it was necessary to generate 50 random, virtual clones. 50 numbers between 1 and 4411529 were randomly generated (using Excel). (4411529 is the number of nucleotide bases contained with the *M. tuberculosis* H37Rv genome.) The average insert size of the library has been identified as being about 400bp. To continue the randomness of these virtual clones, half of the initial virtual co-ordinates had 400bp added to generate a second related virtual co-ordinate whilst the other half had 400bp deducted to generate the second co-ordinate. This generated 50 random virtual clones each containing 400bp. Half of the 50 clones were designated as having the direct strand of the cloned *M. tuberculosis* H37Rv DNA driving the expression of *lacZ* on the pUS1800-based plasmid whilst the other half were designated as having the *lacZ* expression driven from the complementary strand. The genes and intergenic regions contained within these virtual clones were then determined from the *M. tuberculosis* H37Rv genome held at the Sanger centre (www.sanger.ac.uk/projects/M_tuberculosis/).

Of these randomly generated virtual clones, 27 contained *M. tuberculosis* H37Rv DNA situated internal to genes and 3 contained *M. tuberculosis* H37Rv
situated between converging genes. These clones are therefore unlikely to contain potential promoter fragments. 20 of the virtual clones contain intergenic *M.tuberculosis* H37Rv DNA between genes that may contain potential promoter sequences. However of these 20 virtual clones, 10 would have been cloned into pUS1800 in an orientation to drive the expression away from *lacZ*. Therefore only 10 of these virtual clones contain potential promoter sequences that are in the correct orientation to drive the expression of *lacZ*. This is in contrast to the 33 clones that contain probable promoter sequences, identified during this work.

Over half of the 50 clones sequenced to date contain probable promoters in the correct orientation to drive the expression of *lacZ*, i.e. regions of *M.tuberculosis* H73Rv DNA directly up-stream of a known gene. Table 5.1 gives data describing the function of proteins whose putative promoters or parts of putative promoters have been cloned into pUS1800 and are diagrammatically depicted in Figure 5.2. Of the genes contained in Table 5.1, 9 code for proteins with no known function, 5 code for proteins with some homologies to proteins with known functions, 8 genes code for proteins with an identified function, 4 genes code for proteins with described possible functions and 3 genes code for proteins that are described as ‘similar to hypothetical proteins’.

Of the clones used to re-infect human macrophages in Chapter 4, only one contains the probable promoter for a gene with a known function. Clone 176 contains the promoter for the gene Rv2711, which is the iron dependent repressor *ideR* (Schmitt *et al.*, 1995). Clone 169 contains the probable promoter for the hypothetical protein Rv1044. Clones 175, 180 and 239 contain the probable promoters for genes
Rv0302, Rv1165 and Rv1265, all of which code for proteins with no known function. Recent searches for homologies for any of the hypothetical proteins or proteins with unknown functions against proteins with known functions stored in various databases have failed to identify any new homologies. Further analysis and discussion of the promoters of these 5 clones are contained in Chapter 6.
<table>
<thead>
<tr>
<th>Gene Number</th>
<th>Clone Number</th>
<th>Protein Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0130</td>
<td>Clone 174</td>
<td>Unknown function, similar to nudulation protein in from Rhizobium ssp. and is also similar to a monoamine oxidase regulatory protein from Archaeoglobus fulgidus.</td>
</tr>
<tr>
<td>Rv0183</td>
<td>Clone 254</td>
<td>Possible oxidoreductase.</td>
</tr>
<tr>
<td>Rv0302</td>
<td>Clone 175</td>
<td>Unknown function – similar to repressors and regulatory proteins.</td>
</tr>
<tr>
<td>Rv0347</td>
<td>Clone 292</td>
<td>Unknown function</td>
</tr>
<tr>
<td>Rv0440</td>
<td>Clone 166, 220, 278, 291</td>
<td>GroEL2, chaperonin.</td>
</tr>
<tr>
<td>Rv0457c</td>
<td>Clone 320</td>
<td>Probable peptidase</td>
</tr>
<tr>
<td>Rv0460</td>
<td>Clone 13</td>
<td>Unknown hydrophobic protein.</td>
</tr>
<tr>
<td>Rv0461</td>
<td>Clone 13</td>
<td>Unknown hydrophobic protein.</td>
</tr>
<tr>
<td>Rv0596c</td>
<td>Clone 137</td>
<td>Unknown function, has chaperonin signature.</td>
</tr>
<tr>
<td>Rv0613c</td>
<td>Clone 173</td>
<td>Unknown function.</td>
</tr>
<tr>
<td>Rv0767c</td>
<td>Clone 205</td>
<td>Unknown function.</td>
</tr>
<tr>
<td>Rv0940c</td>
<td>Clone 63</td>
<td>Similar to hypothetical protein</td>
</tr>
<tr>
<td>Rv1044</td>
<td>Clone 169</td>
<td>Similar to hypothetical protein.</td>
</tr>
<tr>
<td>Rv1109c</td>
<td>Clone 325</td>
<td>Unknown function</td>
</tr>
<tr>
<td>Rv1165</td>
<td>Clone 180</td>
<td>Similar to hypothetical proteins from different organisms, ie gtp-binding protein SLR1105 from Synechocystis sp.</td>
</tr>
<tr>
<td>Rv1265</td>
<td>Clone 239</td>
<td>Unknown function.</td>
</tr>
<tr>
<td>Rv1444c</td>
<td>Clone 188</td>
<td>Unknown function.</td>
</tr>
<tr>
<td>Rv1630</td>
<td>Clone 279</td>
<td>RpsA, ribosomal protein S1.</td>
</tr>
<tr>
<td>Rv1658</td>
<td>Clone 281</td>
<td>ArgG, argininosuccinate synthase.</td>
</tr>
<tr>
<td>Rv1852</td>
<td>Clone 302</td>
<td>UreG, urease accessory protein.</td>
</tr>
<tr>
<td>Rv2339</td>
<td>Clone 330</td>
<td>Probable membrane protein.</td>
</tr>
<tr>
<td>Rv2345</td>
<td>Clone 277</td>
<td>Probable membrane protein.</td>
</tr>
<tr>
<td>Rv2402</td>
<td>Clone 289</td>
<td>Unknown function, contains helix-turn-helix motif.</td>
</tr>
<tr>
<td>Rv2594c</td>
<td>Clone 226</td>
<td>Holliday junction resolvase.</td>
</tr>
<tr>
<td>Rv2711</td>
<td>Clone 176</td>
<td>IdeR, iron dependent repressor.</td>
</tr>
<tr>
<td>tRNAArgV</td>
<td>Clone 233</td>
<td>tRNA</td>
</tr>
<tr>
<td>Rv2921c</td>
<td>Clone 234</td>
<td>FtsY, probable cell division protein.</td>
</tr>
<tr>
<td>Rv3077</td>
<td>Clone 229</td>
<td>AtsF, similar to aryl- and steryl- sulfatases.</td>
</tr>
<tr>
<td>Rv3174</td>
<td>Clone 222</td>
<td>Unknown oxidoreductase.</td>
</tr>
<tr>
<td>Rv3603c</td>
<td>Clone 282</td>
<td>Unknown function</td>
</tr>
</tbody>
</table>

Table 5.1: Table of all genes whose putative promoters or parts of putative promoters have been cloned into pUS1800. The function of proteins is stated where possible. All gene and protein details obtained from the Sanger institute *M. tuberculosis* H7Rv genome site ([http://www.sanger.ac.uk/Projects/M_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)).
Chapter 6: Real time PCR using Taqman® Technology

6.1 Introduction

The data presented in Chapter 4 suggests that the promoters harboured within the clones 169, 175, 176, 180 and 239 (representing the putative promoters for the \textit{M.tuberculosis} H37Rv genes Rv1044, Rv0302, Rv2711, Rv1165 and Rv1265, respectively) are up-regulated during infection. However this needs to be viewed with a certain amount of caution. These promoters have been cloned out of their natural environment and therefore may not behave as they would do \textit{in vivo}. It is therefore feasible that the 5 promoters, contained within Clones 169, 175, 176, 180 and 239, will behave differently driving the expression of \textit{lacZ} on a plasmid rather than the expression of their own genes on the \textit{M.tuberculosis} H37Rv genome. This being the case, the activity of these promoters needs to be studied when driving expression of their own open reading frames. Since \textit{M.tuberculosis} H37Rv genomic DNA was used to create the library harboured in \textit{M.bovis} BCG Pasteur, this strain should therefore be used to study endogenous expression from these promoters.

There are several methods available to study the relative amount of mRNA being expressed from potentially differentially regulated genes. Northern blots or reverse transcriptase PCR (RT-PCR) could be used but both methods are unsuitable. Northern blotting, it would appear, would be ideal for measuring and comparing the amount of mRNA being expressed under different growth conditions. This is true provided that sufficient mRNA can be recovered from each growth condition. In order to measure the level of mRNA being expressed from any of the 5 promoters identified in Chapter 4, a comparison of mRNA levels \textit{in vitro} and in cultured human macrophages needs to be carried out. Gathering sufficient mRNA from \textit{in vitro} grown
M. tuberculosis H37Rv is relatively simple. However, recovering mRNA from M. tuberculosis H37Rv post-infection is difficult and yields are low. This would mean that enormous numbers of human macrophages (in excess of $10^{10}$) would need to infected in order to recover sufficient M. tuberculosis H37Rv mRNA for Northern blotting.

RT-PCR could also be used to measure the level of mRNA being expressed by the 5 genes identified as being potentially up-regulated in Chapter 4. Despite the fact that RT-PCR is sequence specific and thus unambiguous, it is only semi-quantitative and therefore can not be used to irrefutably show that these 5 genes are indeed up-regulated during infection.

6.1.1 Fluorogenic 5’ Nuclease Assay

Fluorogenic 5’ Nuclease assay, commonly called Taqman®, is a relatively novel method of following DNA amplification during the PCR reaction. Holland et al. (1991) first utilised the 5’→3’ exonuclease activity of the DNA polymerase from Thermus aquaticus (Taq) in the detection of specific PCR products. The 5’→3’ exonuclease activity of Taq cleaves 5’ terminal nucleotides of double-stranded DNA and releases mono- and oligonucleotides but preferentially cleaves displaced single-stranded DNA. If an additional complementary probe with a radioactive phosphate ($^{32}$P) at the end of a non-complementary 5’ tail is added to a PCR mix then the 5’ tail, and therefore the $^{32}$P, is removed from the probe by the Taq. The amplified DNA and the 5’ tail of the radioactive probe can be separated by thin layer chromatography. The amount of probe degradation products can be measured and
thus the amount of product DNA measured. Holland et al. envisaged “the use of this technique to create a truly homogeneous assay, in which amplification and actual detection (of DNA) occur in one tube”. This has indeed become the case however the theory is now more complex but generates more specific results.

Currently, Taqman® routinely utilises a probe, internal to the amplification primers, called a molecular beacon which carry labels at both 5’ and 3’ ends of the DNA. The 3’ end of the probe is labelled with a quencher dye whilst the 5’ end of the DNA carries a fluorescent reporter molecule. When the probe is single-stranded but coiled, the quencher and fluorescent molecule are sufficiently close for the quencher dye to suppress the fluorescence of the reporter molecule. If the target sequence is present, the probe anneals down-stream of one of the primer site and is cleaved by the 5’ nuclease activity of Taq DNA polymerase as the primer is extended. This cleavage of the probe separates the reporter dye from the quencher dye, increasing the reporter dye signal. Cleavage removes the fluorescent reporter molecule and the quencher dye from the probe associated with the target strand, allowing the primer extension to continue to the end of the template strand. The inclusion of the probe during the extension of the primer does not inhibit the overall PCR process. An increase in fluorescence intensity is proportional to the amount of amplicon produced.

Taqman® PCR has a number of different applications. It was initially used for the detection of bacterial DNA sequences (Whitham et al., 1996; Livak et al., 1995; Bassler et al., 1995). Taqman® has subsequently been used in assessing the development of hereditary diseases, such as familial breast and ovarian cancer (Chen
Reverse transcriptase PCR (RT-PCR) can also be used to measure the level of mRNA in a sample using Taqman® (Manganelli et al., 1999).

### 6.1.2 Quantitative Real Time PCR without the use of Molecular Beacons.

An adaptation of the quantitative real time PCR process allows the DNA to be measured, in a similar manner but without the use of a probe. In 1992, Higuchi et al. (1992) noted that if ethidium bromide was included in a PCR, the amplification of target DNA could be continuously monitored because ethidium bromide only fluoresces in the presence of double-stranded DNA. Progress has allowed the removal of the carcinogenic compound ethidium bromide from the PCR tube, to be replaced by another fluorescent chemical, such as SYBR Green (Perkin Elmer).

The use of an internal probe during real time PCR is more sequence specific than the use of SYBR Green (Perkin Elmer) by reducing any mis-priming of the amplification primers. However, provided that the amplification primers do not mis-prime, there is little difference in sensitivity between the two methods. Due to the increased sequence specificity, Taqman® utilising the internal probe is generally preferred.

### 6.1.3 Positive Control Selection

When using real time PCR, a positive control must be used and the selection of the positive control is critical especially if using cDNA as a template. The samples being subjected to real time PCR will obviously contain differing amounts of starting
template. When using cDNA as a template, a constitutively expressed gene whose expression is unaltered by different growth conditions is an ideal candidate as a control. In *M. tuberculosis* H37Rv the ideal candidate gene for use as a control is *sigA* (Rv2703 or *rpoV*), believed to be the essential housekeeping sigma factor for *M. tuberculosis*. Indeed, using Taqman®, the expression of Rv2703 (*sigA*/*rpoV*) was shown to be unaltered in *M. tuberculosis* when grown in a number of different stressing environments (Manganelli *et al.*, 1999). This makes Rv2703 (*sigA*/*rpoV*) an excellent control when comparing the expression of Rv302, Rv1044, Rv1165, Rv1265, and Rv2711 grown *in vitro* and post-infection of cultured human macrophages.

### 6.1.4 Strategy

mRNA was isolated from *M. tuberculosis* H37Rv incubated in control 7H9 liquid medium and complete Dulbecco's modified Eagles Medium. mRNA was also isolated from *M. tuberculosis* H37Rv post-infection of human macrophages. cDNA was generated via reverse transcription of the isolated mRNA. Primers for each of the 5 genes, identified in Chapter 4 as being potentially up-regulated, and the control gene Rv2703 (*sigA*/*rpoV*) were designed. Internal probes, for Fluorogenic 5' Nuclease assay, were not designed due to potential problems over ownership rights with Perkin Elmer (who own the Taqman® licence). Quantitative reverse transcriptase PCR amplification of the 6 genes using cDNA from each of the growth conditions was subsequently followed using SYBR Green (Perkin Elmer).
### 6.2.1 Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>302FOR</td>
<td>GGGAGAGGTCACGAGAATCG</td>
<td><em>M.tuberculosis</em> genome nt 364637-364657. Rv0302</td>
</tr>
<tr>
<td>302REV</td>
<td>AGCCCTTGGTCGCCATC</td>
<td><em>M.tuberculosis</em> genome nt 364700-364680 (complementary strand). Rv0302</td>
</tr>
<tr>
<td>1044FOR</td>
<td>GTCGACGGAATACCACGCAC</td>
<td><em>M.tuberculosis</em> genome nt 1167471-1165491. Rv1044</td>
</tr>
<tr>
<td>1044REV</td>
<td>TAAGGATCCGTGCCCCTCT</td>
<td><em>M.tuberculosis</em> genome nt 1165542-1165523 (complementary strand). Rv1044</td>
</tr>
<tr>
<td>1165FOR</td>
<td>TTGTTCTGCCTAAAGCGCTG</td>
<td><em>M.tuberculosis</em> genome nt 1294506-1294526. Rv1165</td>
</tr>
<tr>
<td>1165REV</td>
<td>GCCGGTCTGTCTTGGTTGACC</td>
<td><em>M.tuberculosis</em> genome nt 1294569-1294549 (complementary strand). Rv1165</td>
</tr>
<tr>
<td>1265FOR</td>
<td>GACTGGTTCGACGGCGAC</td>
<td><em>M.tuberculosis</em> genome nt 1413483-1413501. Rv1265</td>
</tr>
<tr>
<td>1265REV</td>
<td>GCTTGCCAATCACCAGTAGGTTTTC</td>
<td><em>M.tuberculosis</em> genome nt 1413547-1413525 (complementary strand). Rv1265</td>
</tr>
<tr>
<td>2711FOR</td>
<td>AGCACGTGATGAGCGAGC</td>
<td><em>M.tuberculosis</em> genome nt 3023875-3023895. Rv2711</td>
</tr>
<tr>
<td>2711REV</td>
<td>TGGTCGGTGGTGGTGGAGCAC</td>
<td><em>M.tuberculosis</em> genome nt 3023935-3023916 (complementary strand). Rv2711</td>
</tr>
<tr>
<td>SIGAFOR</td>
<td>TTCGCACCTACCTCAACAG</td>
<td><em>M.tuberculosis</em> genome nt 3018514-3018534. Rv2703 <em>(sigA/ rpoV)</em></td>
</tr>
<tr>
<td>SIGAREV</td>
<td>GCTAGCTCGACCTTCTTCTTCG</td>
<td><em>M.tuberculosis</em> genome nt 3018581-3018560 (complementary strand). Rv2703 <em>(sigA/ rpoV)</em></td>
</tr>
</tbody>
</table>
6.2.2 *Mycobacterium tuberculosis* H37Rv RNA Extraction *In Vitro*

10 ml mid-exponential phase *Mycobacterium tuberculosis* H37Rv incubated in liquid media (either 7H9 or complete Dulbecco's modified Eagles medium) were harvested by centrifugation at 3000g for 10 minutes. The media was discarded and the cells were resuspended in 500μl 4M guanidinium isothiocyanate. The bacteria were added to a Hybaid Ribolyser™ blue tube containing 500μl acid phenol, 100 μl chloroform/isoamyl alcohol (24:1 (v/v)) and 200 μl acidified Divolab No.1 (Diversey Ltd., UK). The cells were processed immediately in a Hybaid Ribolyser™ at speed 4.0 for 40 seconds and cooled on ice for 1 minute. Cell debris was removed by centrifugation at full speed in a microcentrifuge for 5 minutes. The aqueous layer was removed to a fresh tube and re-extracted with an equal volume of chloroform/isoamyl alcohol (24:1 (v/v)). The aqueous layer was again removed to a fresh tube and the RNA precipitated with 500 μl isopropanol/salt solution (0.3ml 3M sodium acetate pH4.0, 49.7 ml isopropanol) and 50ng linear acrylamide at -70°C for at least 1 hour. The RNA sample was pelleted by microcentrifugation for 20 minutes, air-dried and resuspended in an appropriate volume of DEPC-treated water (Hybaid Ribolyser™ Kit – blue protocol).

6.2.3 *Mycobacterium tuberculosis* H37Rv RNA Extraction Post-infection of Cultured Human Macrophage.

THP-1 human macrophage cell line was infected with *Mycobacterium tuberculosis* H37Rv as described in section 4.2.1.2 and lysed with the addition of 4M guanidinium isothiocyanate. *Mycobacterium tuberculosis* H37Rv were separated
from the monocyte cell debris by microcentrifuge for 10 minutes. The macrophage
cell debris was discarded, and the pelleted *Mycobacterium tuberculosis* H37Rv
pooled and resuspended in 500 µl guanidinium isothiocyanate. RNA was isolated
from mycobacteria as described in section 6.2.2

6.2.4 Primer Design

All primers for Taqman PCR were designed on Primer Express software v. 1.0
(Perkin Elmer), which is specifically designed for Taqman® primer design.

6.2.5 Reverse Transcription of Bacterial RNA

8ng of RNA were incubated at 37°C with RNase-free DNase for 30 minutes.
The enzyme was subsequently inactivated at 65°C for 10 minutes. Reverse
transcription was carried out described by the manufacturer (Gibco BRL). 5ng of
RNA was added to 3 mM each dNTP (dATP, dGTP, dCTP, dTTP), 1x DTT, 1x
buffer and 100 fmol random hexamers (Roche). Incubation of the mixture at 72°C and
cooling on ice caused the random hexamers to anneal to the RNA. Superscript II
reverse transcriptase (Gibco BRL) was then added and the mixture incubated at 42°C
for 1 hour. The efficacy of the reverse transcription was checked by PCR.

6.2.6 Real Time PCR

All reactions were carried out using the SYBR Green (Perkin Elmer) reporter
system as described by the manufacturer. All PCRs were carried on an ABI PRISM
7700 Sequence Detection System (Perkin Elmer) using MicroAmp® Optical 96-Well Reaction Plate and Optical Caps (PE Biosystems). The SYBR Green PCR Master Mix is 2X in concentration and contains sufficient reagents to perform 200 50-mL reactions. The mix is optimized for SYBR Green reactions and contains SYBR Green I Dye, AmpliTaqGold® DNA Polymerase, dNTPs with dUTP, Passive Reference (ROX), and optimized buffer components. The Passive Reference (ROX) is a dye included in the SYBR Green PCR Master Mix that does not participate in the PCR amplification and provides an internal reference to which the SYBR Green/dsDNA complex signal can be normalized during data analysis. Initial calibration experiments were carried using *Mycobacterium tuberculosis* H37Rv genomic DNA so as not to utilise valuable reverse transcribed cDNA. PCR reaction conditions consist of an initial AmpliTaqGold® DNA Polymerase activation at 95°C for 5 minutes followed by 40 cycles with primer annealing and extension at 60°C for 30 seconds and denaturation at 95°C for 30 seconds.

6.2.6.1 Calculation of Standard Curves for DNA Template

Using the optimal concentration of each set of primers, the amount of template *Mycobacterium tuberculosis* H37Rv genomic DNA added to each primer set was altered as follows: 10 ng, 1 ng, 0.5 ng, 0.2 ng, 0.1 ng, 0.05 ng, 0.02 ng and 0.01 ng. A standard curve of comparative threshold (C_T) values against amount of DNA added was calculated. See figures 6.3 and 6.4.
6.2.6.2 Measurement of Levels of mRNA Present in Different Growth Conditions of *Mycobacterium tuberculosis* H37Rv

Constant amounts of reverse transcribed cDNA (0.4 ng) from each of the growth conditions were amplified with each set of primers. Comparative threshold (C\textsubscript{T}) values were plotted against growth conditions. See figure 6.5.
6.3 Results

6.3.1 mRNA Isolation and Reverse Transcription

Total RNA was extracted from *M. tuberculosis* H37Rv incubated in liquid 7H9 control media, complete Dulbecco's modified Eagles Medium and in cultured human macrophages. The quality of the total RNA extracted was determined visually via denaturing gel electrophoresis and is shown in Figure 6.1.

Isolating large quantities of mRNA from *M. tuberculosis* H37Rv incubated in liquid 7H9 media and complete Dulbecco's modified Eagles Medium is relatively simple because large numbers of mycobacterial cells can be cultured. This is not the case with mRNA isolated from *M. tuberculosis* H37Rv harvested from within cultured human macrophages. Not only does the isolation of *M. tuberculosis* H37Rv from the macrophage lysate complicate matters but also gathering sufficient mycobacterial cells post-infection is dependant on the number of macrophages used during the infection process and the number of bacilli ingested.

Figure 6.1 clearly identifies the problems associated with isolation of total RNA from *M. tuberculosis* H37Rv post-infection of cultured human macrophages. Lanes 2 and 3 are 4 hour post-infection samples collected at the same time and lane 4 is a 4 hour complete Dulbecco's modified Eagles Medium control. Lanes 5, 6 and 7 are post-infection, complete Dulbecco's modified Eagles Medium control and 7H9 control samples collected after 24 hours growth, respectively.
Lanes 2, 3 and 5 are the total RNA extracted from \textit{M.\text{tuberculosis}} H37Rv post-infection of $1\times10^8$ cultured human macrophages infected with a multiplicity of infection of 10. Despite the fact that $1\times10^8$ cultured human macrophages were used for each infection, the samples in lanes 2, 3 and 5 contain virtually undetectable levels of total RNA.

![Denaturing gel of total RNA isolated from \textit{M.\text{tuberculosis}} H37Rv grown under different culture conditions. RNA visualised with SYBR gold. Lane 1 is RNA marker (0.28-6.58kb) (Promega). Lanes 2 and 3 are independent \textit{M.\text{tuberculosis}} H37Rv total RNA samples obtained 4 hours post-infection of cultured human macrophages. Lane 4 is \textit{M.\text{tuberculosis}} H37Rv total RNA sample isolated after 4 hours incubation in complete Dulbecco's modified Eagles medium. Lane 5, 6 and 7 are total RNA samples isolated after 24 hours incubation post-infection of cultured human macrophages, in complete Dulbecco's modified Eagles medium and in liquid 7H9 medium, respectively.](image)

Knowing that 3μg of marker was loaded into lane 1, it was estimated that lanes 6 and 7 contained 1μg of total \textit{M.\text{tuberculosis}} H37Rv RNA, lane 4 contained 100ng of total RNA and lane 5 contained 10ng of total RNA. This means that the total RNA concentrations of the 4 hour complete Dulbecco's modified Eagles Medium control, the 24 hour infection, the 24 hour complete Dulbecco's modified Eagles Medium control and the 24 hour 7H9 liquid medium control are approximately 20
ng/µl, 2 ng/µl, 1µg/µl and 1µg /µl, respectively. Since the total RNA in lanes 2 and 3 can not be detected, the total RNA concentration of the two 4 hour infections can not be estimated and it was therefore assumed that the concentration of total RNA was no more than 1ng /µl. As there was no difference between the conditions used to infect the human macrophages, the two *M.tuberculosis* H37Rv total RNA samples collected at 4 hours were pooled together. Each of the 5 samples was subjected to reverse transcription as described in Section 6.2.5.

Due to the fact that the *M.tuberculosis* H37Rv total RNA obtained from the 7H9 liquid medium control was most plentiful, a PCR, using the pairs of primers to amplify Rv0302, Rv1044, Rv1165, Rv1265, Rv2711 and Rv2703 (sigA/ rpoV), was carried out to check the efficacy of the reverse transcription. Amplification products were seen with the primer sets for Rv0302, Rv1265, Rv2711 and Rv2703 (sigA/ rpoV), but was absent with the primer sets for Rv1044 and Rv1165. Despite the lack of amplification of genes Rv1044 and Rv1165, reverse transcription was considered sufficiently successful for use with real time PCR. The reason for the failure to amplify DNA with the primers for Rv1044 and Rv1165 is not known. This could be due to a problem with the reverse transcription of the total RNA or with the amplification of the target sequences. Alternatively the level of expression of these genes could be expressed at very low levels when *M.tuberculosis* is grown in liquid 7H9 medium.
6.3.2 Calculation of Optimal Primer Concentration

When using real time PCR and SYBR Green to analyse PCR amplification, the first requirement is to minimise any primer dimers. Since DNA preferentially forms double stranded molecules, single stranded primers that are heated and cooled during the amplification cycles can form dimers, even if the DNA is not homologous. SYBR Green is similar to ethidium bromide in that it only chelates to double stranded DNA. Any primer dimers present would therefore interfere with detection of the required DNA sequences following amplification.

Therefore the pairs of primers (e.g. 302FOR and 302REV, 1044FOR and 1044REV, etc) were mixed at a range of different concentrations. The concentrations were 50nM:50nM, 50nM:300nM, 50nM:900nM, 300nM:50nM, 300nM:300nM, 300nM:900nM, 900nM:300nM and 900nM:900nM, with the initial concentration representing the forward primer. Full 40 cycle amplifications were carried out with and without 1ng of *M.tuberculosis* H37Rv genomic DNA template, in order to identify the mixture of primers concentrations that generate the least number of primers dimers, using SYBR green to measure both the amplification of the target sequence and the number of primer dimers. With all pairs of primers, the optimal concentration of primers pairs, generating the least number of primer dimers and the greatest amplification of target sequence, was 900nM:900nM.

The primers designed to amplify genes Rv1044 and Rv1165 formed a greater number of primer dimers than the primers designed to amplify genes Rv0302,
Rv1265, Rv2711 and Rv2703 \((\text{sigA/ rpoV})\). In addition to the increased number of primer dimers, amplification of the target DNA was not seen. The reason for the lack of amplification with the primers for Rv1044 and Rv1165 is not known. Quantitative RT-PCR analysis of the genes Rv1044 and Rv1165 was therefore not conducted.

### 6.3.2.2 Calculation of Standard Curves for DNA Template

With the ideal concentration of mixed primers identified, the efficiency of amplification with each set of primers had to be determined. As previously stated, in order to relate RNA samples containing different amounts of starting template, a constitutively expressed gene which is not up- or down-regulated is required as a control. This control is called a comparator and in this case Rv2703 \((\text{sigA/ rpoV})\) was selected.

*\textit{M.tuberculosis*} H37Rv genomic DNA was used to calculate the efficiency of amplification with each primer set. The genomic DNA was diluted to 10 ng, 1 ng, 0.5 ng, 0.2 ng, 0.1 ng, 0.05 ng, 0.02 ng and 0.01 ng of template and PCR amplification was carried out with each set of primers at their optimal concentration. Figure 6.2 is a copy of the immediate output from the Taqman® equipment and analysis software. It shows a series of graphs that are generated during the amplification process with the range of diluted *\textit{M.tuberculosis*} H37Rv genomic DNA with the primers set designed to amplify a short region of the gene Rv2711. All of the graphs are horizontal during the initial stages of the amplification, until sufficient template DNA has been generated. Once sufficient template DNA has been generated, the signal accelerates exponentially and each of the graphs curves away from the baseline, shown with the
dark black line close to 0.000. The point where each graph leaves the baseline is
called the Ct value. The Ct value represents the point in time when there is sufficient
template for the detection of the exponentially increasing signal. The lower the Ct
value the greater the number of template molecules present at the start of the
amplification. Taking the dark blue and the red graphs as examples, which are at
opposite ends of the scale, the dark blue graph has a much lower Ct value than the red
line. This means that PCR represented by the dark blue graph contains more starting
template that the PCR represented by the red line. The dark blue line depicts the
amplification of M.tuberculosis H37Rv genomic DNA with 10ng of DNA and the red
line represents 0.01ng of starting genomic DNA.

Figure 6.2: Graph taken directly from Taqman® analysis software. The dark blue, dark green, blue,
 pink, light blue, yellow, light green and red lines represent PCR amplifications with 10 ng, 1 ng, 0.5
 ng, 0.2 ng, 0.1 ng, 0.05 ng, 0.02 ng and 0.01ng of M.tuberculosis H37Rv genomic DNA, respectively.
The dark black horizontal line close to 0.000 represents the threshold at which DNA amplification
accelerates exponentially.
Figure 6.3 shows $C_T$ value for each genomic DNA concentration for 4 sets of primers designed to amplify regions of the genes Rv0302, Rv1265, Rv2711 and Rv2703 (sigA/ rpoV). These graphs show the efficiency each of the amplifications.

If the amplification of cDNA from *M. tuberculosis* H37Rv grown in different growth conditions is to be related, the efficacy of PCR, with each set of primers, needs to be similar. The efficacy of a PCR is related to the slope of the standard curve for the primer sets used in the amplification. Thus, if the linear best-fit trendlines are correct, in Figure 6.3, the yellow, dark blue and pink curves which show the amplification from the *M. tuberculosis* H37Rv genes Rv2711, Rv0302 and Rv1265, respectively, are quite close to being parallel meaning that the amplification efficiency is roughly the same. However, the amplification of the gene Rv2703 (sigA/ rpoV), which is represented by the pale blue line, is not parallel to the other 3 lines and the angle of the slope is shallower. This means that the PCR represented by the green line (the amplification of Rv2703) is less efficient than the PCRs of the genes Rv0302, Rv1265 and Rv2711.
Since the amplification efficacy of the comparator (Rv2703 (sigA/ rpoV)) is lower than the efficacy of the other 3 PCRs (Rv0302, Rv1265 and Rv2711 – the samples), the amplification from cDNA generated from \textit{M.tuberculosis} H37Rv grown in different growth conditions for these genes can not be accurately compared. It is known that the expression of Rv2703 (sigA/ rpoV) is unaltered by the growth of \textit{M.tuberculosis} H37Rv in different conditions (Manganelli et al., 1999). Therefore if the efficiencies of the 4 amplifications had been comparable, it would have been possible to definitively calculate a ratio of the level of expression of each sample gene in each of the growth conditions, with Rv2703 (sigA/ rpoV) being used as a standard control.
It is clear from figure 6.3 that the linear best-fit trendlines do not accurately represent the observed data, which is non-linear over this range of DNA concentrations. However, the range of template concentrations and $C_T$ values shown in figure 6.3 is wider than the actual range used subsequently to generate in figure 6.5; it is therefore possible to examine the effect of re-plotting figure 6.3 without the 10ng point, which otherwise has an excessive influence on the slope of the line plotted. Figure 6.4 indicates that over this range the PCR efficiencies of the $sigA$ (Rv2703/ $rpoV$) and Rv1265 amplifications are similar. With this in mind, it may be possible to use figure 6.4 to extrapolate greater significance from quantitative PCR using $M.tuberculosis$ H37Rv cDNA as a template.

![Figure 6.4: Graph of $C_T$ value against amount of $M.tuberculosis$ H37Rv genomic DNA added to each real time PCR. $M.tuberculosis$ H37Rv DNA added was 1 ng, 0.5 ng, 0.2 ng, 0.1 ng, 0.05 ng, 0.02 ng and 0.01 ng. Each line represents a linear best-fit trendline. Individual amplifications are represented by dots.](image)
6.3.2.3 Measurement of Levels of mRNA Present in Different Growth Conditions of *Mycobacterium tuberculosis* H37Rv

Although the absence of a standard control was a limitation, it was still possible to carry out a semi-quantitative comparison of real time RT-PCR with the four sets of primers. In order to generate some consistency, a constant amount of cDNA was added to each amplification. This was difficult since the amount of cDNA produced can not be accurately measured. It was assumed that the reverse transcription of all *M.tuberculosis* H37Rv RNA samples were equally successful and therefore the amount of cDNA present was related to the starting amount of mRNA. For each cDNA preparation, the amount of starting mRNA was adjusted to be constant.

Therefore cDNA (0.4 ng) generated from each of the RNA preparations from *M.tuberculosis* H37Rv after 4 and 24 hour macrophage infections and 24 hours incubation in complete Dulbecco's modified Eagles Medium and liquid 7H9 medium was amplified independently with each primer set. Each set of primers was used to amplify each of the cDNA samples in triplicate. In Figure 6.5, the average $C_T$ value for each primer set with each cDNA sample is depicted graphically.

It is apparent from figure 6.5 that the aforementioned problem with the comparator is significant. It would be expected that different preparations of mRNA would be at different concentrations. Despite the fact that Rv2703 (*sigA*/*rpoV*) is not up- or down-regulated during infection, the proportion of the *sigA* molecules, compared to the total mRNA molecules, would be expected to change. Therefore the
number of $\text{sigA}$ molecules in each of the RNA samples would be different. This is indeed the case since the $C_T$ values for the 4 cDNA preparations are different.

Normally, if the real time PCR comparator exhibits a similar PCR amplification efficiency to the sample amplifications (in this case Rv0302, Rv1265 and Rv2711), the $C_T$ values can be adjusted so that the comparator $C_T$ value is constant. This would mean that the number of initial template molecules for the comparator are constant in all samples. This would also have the effect of altering the $C_T$ values for all other sample amplifications.
Figure 6.5: Graph of $C_T$ value against the number of the gene amplified from cDNA generated from mRNA obtained from *M. tuberculosis* H37Rv incubated in 4 different growth conditions. Estimated constant amounts of cDNA (0.4ng) were added to each amplification.
As previously discussed this type of comparison can not be carried out because the PCR efficiencies are not similar. Therefore the only method available to suggest any relationship between the generation of mRNA, from the genes in question, in the different growth conditions is to amplify DNA from constant amounts of cDNA.

Figure 6.5 suggests that, with constant amounts of cDNA at the start of the amplification, the *M.tuberculosis* H37Rv expresses both Rv1265 and Rv2711 at a higher rate post-infection than it does when grown *in vitro* since the C_T values during infection are lower than the C_T values generated by *in vitro* cDNA. However, as in Chapter 4, the *M.tuberculosis* H37Rv incubated in complete Dulbecco's modified Eagles medium is the direct control for *M.tuberculosis* H37Rv grown in macrophages. In the case of Rv1265, the expression of the gene appears to be greater during both the 4 and the 24 hour infections compared to the level of expression of the gene in the complete Dulbecco's modified Eagles medium control. This is also the case in the expression of Rv2711, but to a lower extent. It appears that the expression of Rv2711 is also up-regulated during the incubation of *M.tuberculosis* H37Rv in complete Dulbecco's modified Eagles medium compared to growth in liquid 7H9 medium. Figure 6.5 suggests that the expression of Rv0302 is not up-regulated during macrophage infection or incubation in complete Dulbecco's modified Eagles medium.

However figure 6.4, which shows the PCR efficiencies within the range of *M.tuberculosis* H37Rv cDNA concentrations being amplified, suggests that the amplification efficiencies of Rv2703 (*sigA* or *rpoV*) and Rv1265 are in fact similar. This enables a greater interpretation of the results in figure 6.5 with respect to the
potential induction of Rv1265. Since the expression of \textit{sigA} is unaltered during infection (Manganelli \textit{et al.}, 1999), there is a strong case suggesting that Rv1265 is induced during infection. Quantitative RT-PCR calculations for fold induction of Rv1265 require the accommodation of the \textit{sigA} expression and the expression of Rv1265 \textit{in vitro} (grown in liquid 7H9 medium). This suggests that Rv1265 is induced after 24 hours post-macrophage infection by 10-20 fold.
Total RNA was extracted from *M. tuberculosis* H37Rv incubation in liquid 7H9 medium, complete Dulbecco's modified Eagles Medium and in cultured human macrophages. Reverse transcription of the isolated mRNA was carried out and the efficiency of the reverse transcription tested by PCR with each of the 6 sets of primers.

The primers designed to amplify short regions of Rv1044 and Rv1165 failed to amplify any DNA during the test of the efficiency of cDNA generation. There was also no amplification from genomic *M. tuberculosis* H37Rv DNA during the calculation of optimal primer concentration and the primers designed to amplify the genes Rv1044 and Rv1165 also generated a greater number of primer dimers than expected. There are 2 possible reasons for the lack of amplification of the genes Rv1044 and Rv1165. The first is that each of the primer pairs may not be very good primers and therefore may need to be re-designed. Secondly, the dimerization of each primer set may be strong enough to interfere with the annealing of the primer to the target sequence.

It is disappointing that the amplification efficiencies of PCRs for Rv0302, Rv1265, Rv2711 and Rv2703 (*sigA/ rpoV*) are not similar in figure 6.3. Since the efficiency of the Rv2703 (*sigA*) PCR is low, in order to cause the 4 PCRs to have similar efficiencies, the efficiencies of PCRs for Rv0302, Rv1265 and Rv2711 would have to be reduced. Due to the short length of time available it was not possible to alter the efficiencies of any of the PCRs.
Since the efficiency of the Rv2703 (sigA) PCR (the comparator) in figure 6.3 is lower than the efficiencies of the PCRs for Rv0302, Rv1265 and Rv2711 (the samples), a definitive quantitative measurement of the up-regulation of the sample genes is not possible. In this instance, quantitative RT-PCR can only be used to suggest any possible up-regulation. However, using figure 6.4 to suggest comparable PCR efficiencies between the Rv2703 (sigA) and Rv1265 primers and an extrapolation of the cDNA amplification results in figure 6.5, there is strong evidence that Rv1265 is indeed up-regulated 10-20 fold during infection. However, the truncation of the graph in figure 6.3, to generate figure 6.4, can not be used to calculate definitive quantitative values of the degree of up-regulation of Rv1265.

It has been shown that Rv2711 is the gene responsible for the production of an iron dependent repressor (which has been annotated as ideR) and is a homologue for the diphtheria toxin repressor (dtxR) (Schmitt et al., 1995). An interesting result of this work is that the detection of the Rv2711 mRNA is greater in both samples obtained post-infection and the sample obtained from incubation in complete Dulbecco's modified Eagles Medium than it is the sample grown in liquid 7H9 medium. This is interesting since the protein product of gene Rv2711 is an iron dependent repressor. Not only is the iron availability low within a macrophage but it is also limited in complete Dulbecco's modified Eagles Medium.

Iron is required by almost every living organism as a protein cofactor. In response to low iron availability in the environment, microorganisms have developed a number of methods for iron uptake. Microorganisms are known to secrete siderophores, which are low molecular weight, high-affinity iron chelators that
solubilize iron complexed in the environment or bound to compounds from their host, such as transferrin and lactoferrin. It has been shown that the IdeR protein is responsible for repressing siderophore biosynthesis in the presence of iron and that *M. smegmatis* ideR mutants are more sensitive to oxidative stress (Dussurget *et al.*, 1996). Therefore Dussurget *et al.* suggested that IdeR couples iron metabolism to the oxidative stress response. It has subsequently been suggested that IdeR positively controls the expression of *fxbA* (Dussurget *et al.*, 1999), which encodes a putative formyltransferase necessary for exochelin biosynthesis. Exochelins are siderophores found in fast growing mycobacteria. Dussurget *et al.* suggested that IdeR could act directly as a positive regulator for the *fxbA* gene or as an activator of a second activator for this gene. Alternatively, IdeR could act as a repressor of a repressor. Any of these potential positive functions of IdeR could provide a reason for the up-regulation of Rv2711 during infection. IdeR is a homolog of the *Corynebacterium diphtheriae* DtxR protein and a similar cascade mechanism has been hypothesised for the induction of a heme oxygenase (Schmitt, 1997).

The genes Rv0302 and Rv1265 have been annotated by Cole *et al.* (1998) as coding for proteins with unknown functions. However the gene Rv0302 has sequence similarity to a number of repressors and regulatory proteins from other organisms including the *E.coli acrAB* operon repressor (Ma *et al.*, 1996). The gene Rv1265 has no homologies to any currently known proteins. The results for the amplification of the gene Rv0302 are unclear since the error bars of the triplicate amplifications overlap and suggest that the expression of this gene is not up-regulated during infection compared to the expression during incubation in complete Dulbecco’s
modified Eagles medium and liquid 7H9 medium. The expression of gene Rv0302 in different growth condition does need further investigation.

The evidence of the up-regulation of Rv1265 and Rv2711 contained within this chapter needs to be carefully considered. There is undoubtedly evidence of up-regulation but certain assumptions had to be made in order to generate data. The amount of RNA present in preparations obtained from *M. tuberculosis* was estimated as opposed to measured. To compound this issue, the *M. tuberculosis* RNA isolated 4 hours post-macrophage infection could not be detected using denaturing gel electrophoresis and therefore a maximum quantity of RNA present was estimated. It was also assumed that different mRNA preparations, containing different quantities of mRNA, were reverse transcribed with equal efficacy. These assumptions may well be affecting the detected up-regulation of Rv1265 and Rv2711. However, the apparent up-regulation of Rv1265, in particular, is considerable and therefore, despite the assumptions mentioned above, the evidence strongly suggests that this gene is indeed up-regulated during infection.

The potential power of real time PCR has not been fully realised with analysis on these particular genes. However, these results do suggest that the genes Rv1265 and Rv2711 may be up-regulated during the infection of cultured human macrophages, which supports the β-galactosidase assay results in Chapter 4. Further investigation of the genes Rv1265 and Rv2711 is required to obtain quantitative estimates of the degree of up-regulation of these genes during infection.
7.1 General Discussion

The aim of this study was to identify genes that are potentially involved in the virulence of *M. tuberculosis* and are therefore up-regulated during the infection of the host. This study has shown that it is possible to mimic *in vitro* the conditions that *M. tuberculosis* may be exposed to *in vivo* and use them to identify potentially up-regulated genes. In this study, the *in vivo* mimicking conditions used were a reduced pH (pH 5.4) and a reduced pH (pH 5.4) with sodium nitrite. Sodium nitrite at a reduced pH decomposes to generate reactive nitrogen intermediates (Taylor *et al.*, 1927). Therefore the intention of this study was to identify gene expression responses to acidification and/or reactive nitrogen intermediates. However the true environmental conditions that *M. tuberculosis* is exposed to during infection are unclear and gene expression will undoubtedly respond to other signals, such as reactive oxygen intermediates, not covered in this study.

The clones that were identified as up-regulated by the *in vivo* mimicking conditions were subsequently used to infect human and murine macrophages and these experiments identified a number of clones whose expression of *lacZ* appeared to be up-regulated inside the macrophage compared to *in vitro*. Sequence data showed that over half of the clones used to infect macrophages contained cloned *M. tuberculosis* H37Rv DNA that contained potential promoters. mRNA analysis, using Taqman® technology, of a small number of these potentially up-regulated genes supported the findings of up-regulation in two of the five genes tested.
The *M. tuberculosis* genes Rv1265 and Rv2711 (*ideR*) do appear to be up-regulated during infection and the quantitative reverse transcriptase PCR findings are consistent with this interpretation if not fully conclusive. Superficially, it may be regarded as not surprising if the *ideR* gene were to be up-regulated during infection as its role as an expression regulator would probably be required during the process of infection, especially considering its potential activator role in siderophore biosynthesis (Dussurget *et al.*, 1996). However, the real significance of an up-regulation of the repressor itself, as opposed to the genes that it controls, remains to be investigated. The apparent up-regulation of Rv1265, which codes for a protein with an unknown function is perhaps more intriguing, especially considering the strong evidence, obtained using real time PCR, of its macrophage-infection induction. The possible intra-macrophage up-regulation of genes Rv0302, Rv1044 and Rv1165 needs to be re-examined.

The strongest evidence supporting the success of the gene fusion methodology employed here to identify genes up-regulated during mycobacterial infection was the repeated finding of the promoter sequence for Rv0440 (*groEL1*) in 4 independently tested clones. Other studies designed to investigate differential gene expression in mycobacteria have identified the gene Rv0440 (*groEL2*) as being up-regulated during infection or exposure to stressing conditions (Monahan *et al.*, 1994; Lee and Horwitz, 1998; Garbe *et al.*, 1996). The *groEL2* gene codes for a chaperonin heat shock protein with a molecular weight of 60KDa. The up-regulation of the gene Rv0440 (*groEL2*) may be a general stress response rather than a specific response to acidification or reactive nitrogen intermediates.
7.2 Direct Comparison of Promoter Probe Methodology with Alternative Techniques.

During this study, a number of clones were identified as being apparently up-regulated during infection but upon sequencing were found to contain portions of *M. tuberculosis* H37Rv DNA that are unlikely to contain promoter sequences, such as regions internal to genes. These false positive results are to be expected, as they are also found in other studies using technologies such as differential display PCR (DD-PCR) (Garbe *et al*., 1996; Lee and Horwitz, 1998; Monahan *et al*., 1994; Ragno *et al*., 1997; Rivera-Marrero *et al*., 1998). However the fact that approximately half of the sequenced, potentially up-regulated genes contain probable promoter sequences suggests that the presence of these false positives is acceptable as they are easily identifiable.

During the use of subtractive hybridization to identify the mig gene of *M. avium* as induced during phagocytosis, Plum and Clark-Curtiss (1994) commented that their protocol would be especially useful in identify induced genes “for organisms that do not presently have a well-defined genetic system, such as the mycobacteria”. This may well be the case but perhaps the biggest problem with using RNA based procedures, such as subtractive hybridization, microarrays and the real-time reverse-transcriptase PCR used during this study, to identify potential virulence factors is whether they have the ability to identify genes with transient expression or low level expression. The expression of a particular virulence factor may be extremely short lived and, unless mRNA is extracted during this spike of expression, it may not be reverse transcribed and therefore can not subsequently be amplified. If a gene were to be expressed in small quantities, it may be difficult to generate sufficient
amplicon to detect it during the hybridization step of subtractive hybridization or microarrays. An additional consideration is the half-life of mRNA. Some mRNAs have very short half-lives and therefore, despite considerable levels of expression, detection of these mRNAs can be difficult.

The completion of the *M.tuberculosis* H37Rv genome sequence (Cole *et al.*, 1998) is undoubtedly an enormous advancement in mycobacterial research in general and the study of tuberculosis specifically. The use of techniques such as high density DNA arrays and proteome analysis, which exploit the availability of the genome sequence to great effect, are ideally suited to identifying virulence factors. However the huge cost involved in this state of the art technology precludes a significant portion of the research community from making use of these techniques. The work described here is very labour intensive but does not require expensive laboratory equipment, apart from the Taqman® equipment, which can be replaced with lower technology, lower cost alternatives. Although cost should not an argument for or against a specific technique, the availability of that technique is. However, perhaps more importantly, the methodology described here is not only capable of identifying novel virulence factors but could also be used to support data obtained from genomic techniques, such as high density DNA arrays and proteome analysis. Genomics and proteomics are undoubtedly one of the main methods that will be employed to identify virulence factors in the future, not only of *M.tuberculosis* but of other organisms as well, such as the causative agent of malaria, *Plasmodium falciparum* (Hayward *et al.*, 2000).
It is important to note that genomics and proteomics can detect all of the changes in gene control but do have certain limitations. Transcriptional methods, such as the promoter probe methodology described here and genomics, are not capable of identifying any post-transcriptional regulation, such as mRNA stability, mRNA secondary structure or initiation and termination of translation. Problems can be caused by post-translational modification when using proteomics. Proteins that are post-translationally modified (i.e. glycosylated) during certain growth conditions will occupy different positions in a two dimensional polyacrylamide gel to the un-modified protein and therefore appear to be de novo synthesised proteins. The detection of low abundance proteins, such as repressors, is particularly difficult using proteomics. One advantage that the promoter probe methodology has over genomics and proteomics is that it can be used in advance of the availability of genome sequence data. This methodology could therefore be used to identify virulence determinants of other microbial pathogens.

Both in vivo expression technology (IVET) and signature-tagged mutagenesis (STM) have an enormous advantage over the methodology used here to identify potential virulence factors. During these two techniques all of the screening for virulence factors is carried out in vivo. The use of animal models, as opposed to cultured macrophages, would expose M. tuberculosis to an environment that could never be reproduced in vitro – such as serum proteins, the presence or absence of chemicals (such as iron availability) and the effect of these on the macrophage. An additional advantage of STM and IVET are that gene expression is studied in situ on the chromosome. This negates any alteration in gene expression experienced when studying expression with a plasmid construct. The methodology described here could
be adapted to use animal models by altering the reporter gene used (*lacZ*) to a selectable reporter gene and mutant strain of *M.tuberculosis* (as in the IVET methodology). However, the problem of expression of the reporter gene still being initiated from the plasmid, as opposed to the chromosome, would still exist.

There is little doubt that each of these methods are capable of identifying potential virulence factors or generating additional data regarding potential virulence factors. Identification of virulence factors via one technique will require the support of data generated by other techniques, such as allelic exchange mutagenesis. Allelic exchange mutagenesis was first carried out in the *M.tuberculosis* complex by Norman *et al.* (1995), with viable mutants being reported later that year (Reyrat *et al.*, 1995). However, allelic exchange mutagenesis is not, in itself, capable of identifying virulence factors but is better suited to identifying the effect on mycobacteria of removal of virulence factor proteins identified via other methods. This being the case, allelic exchange mutagenesis is well suited to validating virulence factor expression data previously identified via other means. Allelic exchange mutagenesis will probably be most useful in the mutation of specific genes in *M.tuberculosis*, if an alternative live vaccine to BCG were ever to be created.
The clones that were sequenced and identified as containing the probable promoters of genes potentially up-regulated during infection need further investigation. Perhaps the simplest method of definitively identifying which genes are up-regulated during infection is using Taqman® technology, as described in Chapter 6, provided that an appropriate positive control can be utilised. Other lower-technology techniques could be used identified which genes are up-regulated during infection, such as Northern blotting or RT-PCR. The problem with these two techniques is that Northern blots require large amounts of mRNA, which is difficult to obtain and RT-PCR is only semi-quantitative. There are currently 23 sequenced clones from this work, in addition to those already used during quantitative RT-PCR experiments, containing potential promoters whose genes should be investigated further.

As stated above the overall process of promoter library generation, in vitro selection, in vivo selection and sequencing has been successfully used to identify potentially up-regulated promoters. Therefore the size of the library, currently at 4800 clones, should be increased in order to identify other M. tuberculosis H37Rv genes up-regulated during infection. Ideally it would be increased to the 32951 clones required to statistically cover the entire M. tuberculosis H37Rv genome. In addition to increasing the size of the library, it may also be profitable to use different in vitro stimuli designed to mimic the intra-macrophage conditions, such as reactive oxygen intermediates, or anaerobic conditions. The library could also be used to identify
genes expressed during different growth conditions in vitro, or expressed at different stages of growth, such as stationary phase.

The logical progression of the separate studies carried out during this work was to mimic the effect of in vivo conditions on each clone, a study of the expression levels of each clone post-infection, sequence analysis and ultimately quantitative RT-PCR. With a complete view of the work carried out here and in retrospect, this may not have been the best utilisation of resources. In future, having carried out the in vivo mimicking study, I think that the next step should be the sequence analysis. This would further reduce the number of clones to be used in macrophage infection studies since only clones containing potential promoters would need to be infected. Since a low number of clones would then to used to infect macrophages, not only could each clone be studied in greater detail (i.e. with more time points and in triplicate infections), but the macrophages could also be pre-activated with interferon γ. Pre-activated macrophages are more capable of suppressing *M.tuberculosis* infections (Sato et al., 1998).

This study was designed to identify genes that are up-regulated during infection and are therefore likely to be associated with the pathogenicity of *M.tuberculosis*. Clearly the immediate requirement of the bacterium post-infection internalisation within the macrophage is survival. The effect of long-term infection has not been investigated during this study. However the methodology described here could easily be adapted to measure the level of lacZ expression over a greater number of days. This may have identified different clones associated with the longevity of the *M.tuberculosis* infection.
The identification of further virulence factors will better our understanding of *M.tuberculosis* infection and potentially lead to an alternative and better vaccine than the current BCG vaccine. The methodology described here has identified genes that are potentially up-regulated during infection and a continuation of this work will consolidate the findings described here.
Appendix 1 – Materials

All media were obtained from Difco and chemicals from Sigma unless otherwise stated. Appropriate media and chemicals were manufactured using double distilled water and autoclaved at 121°C under 15 psi pressure for 15 minutes. Alternatively, solutions were sterilised by filtration through a 0.2 μm Ministart filter unit.

Growth Media

LB broth: Bacto tryptone 1.0% (w/v), yeast extract 0.5% (w/v), NaCl (85 mM) pH7.5.
LB agar: As LB broth with bacteriological agar 1.5% (w/v).
7H9 broth: 5.2g/l in water, with tween and glycerol.
7H11 agar: 21g/l in water, with glycerol.
SOC: Bacto tryptone 2.0% (w/v), yeast extract 0.5% (w/v), NaCl (8.0 mM) KCl (2.5 mM) MgSO₄ (20 mM) glucose (20 mM) pH7.5. Sterile MgCl₂ (10 mM) prior to use.
Dulbecco's modified Eagles medium (Life-Tech)(DMEM GMAX): 4 mM L-glutamine adjusted to contain 1.5 g/l sodium bicarbonate, 4.5 g/L glucose and 1.0 mM sodium pyruvate.

Media Supplements

Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside): 40 μg ml⁻¹ dissolved in dimethyl formamide. Filter sterilised, stored at -20°C in the dark.

Tween 80: 0.05% (v/v).
Glycerol: 0.02% (v/v).
Kanamycin: 50 $\mu$g ml$^{-1}$ for *Escherichia coli*, 15$\mu$g ml$^{-1}$ for *Mycobacterium bovis* BCG. Filter sterilised, stored at -20°C.

Heat inactivated Foetal Calf Serum (Life-Tech): 10% (v/v)

Phorbol Myristate Acetate (PMA): 10mg/ml in DMSO. Stored at -20°C.

**General Buffers**

6X Loading buffer: 40.0% (w/v) sucrose, 0.25% (w/v) Orange G, 5mM EDTA.

Buffered phenol/ chloroform: 50:50 ratio (v/v).

Phosphate-buffered saline: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na$_2$HPO$_4$, 1.4 mM KH$_2$PO$_4$ (pH7.3).

50X TAE: 270 mM tris, 7.0 mM EDTA, 1.15% (v/v) glacial acetic acid.

TE: 10 mM tris/HCl pH8.0, 1 mM EDTA pH8.0.

10X MOPS: 0.2 M 3-(N-morphino) propanesulfonic acid, 50 mM sodium acetate, 5 mM EDTA

**Buffers Required for RNA Isolation**

guanidinium Isothiocyanate (GITC): 4M guanidinium isothiocyanate, 20mM Tris (pH7.6), 10mM EDTA, 2% Sarkosyl and 1% Beta mercaptoethanol.

Acidified Divolab Detergent: 9.6ml Divolab No1 (Diversy Lever), 24ml 500mM sodium acetate (pH 4.0), 66.4ml RNAse free water.

Acid phenol: Pro analysis phenol (Merek) saturated with water and equilibrated with 50mM sodium acetate (pH 4.0).

RNA denaturant: Glyoxal/DMSO (1.5mls/2.5mls). Stored at -80°C.

SYBR Gold (Cambridge Bioscience): diluted 1:10000 in MOPS
Plasmid DNA Mini-preparation Solutions

Solution 1: 50 mM glucose, 25 mM tris/HCl pH8.0, 10 mM EDTA.
Solution 2: 0.2M NaOH, 1% (w/v) SDS.
Solution 3: 60% (v/v) 5M potassium acetate, 11.5% (v/v) glacial acetic acid.

Bacterial Strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em> DH5α</td>
<td><em>supE44, ΔlacU169 (φ80lacZΔ15), hsdR17, recA1 endA1, gyrA96, thi-1, relA1</em> (Hanahan 1983).</td>
</tr>
<tr>
<td>Epicurian Coli® XL2-Blue ultracompetent cells</td>
<td>recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac [F’ proAB lacIqZAM15 Tn10 (Tet’) Amy Cam’]</td>
</tr>
<tr>
<td><em>Mycobacterium</em> bovis BCG</td>
<td>ATCC Number: 35748</td>
</tr>
<tr>
<td>Pasteur</td>
<td></td>
</tr>
<tr>
<td><em>Mycobacterium</em> tuberculosis H37Rv</td>
<td>ATCC Number: 27294</td>
</tr>
</tbody>
</table>

Macrophage Cell Lines

<table>
<thead>
<tr>
<th>Macrophage Cell Line</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>J774 (Unkeless, 1979)</td>
<td>ATCC Number: HB-197</td>
</tr>
<tr>
<td>THP-1 (Auwerx, 1991)</td>
<td>ATCC Number: TIB 202</td>
</tr>
</tbody>
</table>
## Plasmids

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Details</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAL5000</td>
<td>Mycobacterial origin of replication</td>
<td>Labidi et al., 1985</td>
</tr>
<tr>
<td>pUC18</td>
<td><code>bla</code>, <code>colE1</code>, <code>lacZα</code></td>
<td>Yanisch-Perron et al., 1985</td>
</tr>
<tr>
<td>pUC19</td>
<td><code>bla</code>, <code>colE1</code>, <code>lacZα</code></td>
<td>Yanisch-Perron et al., 1985</td>
</tr>
<tr>
<td>pMC1871</td>
<td></td>
<td>Promega</td>
</tr>
</tbody>
</table>
Appendix 2 - Methods

General molecular biology techniques were performed according to Current Protocols in Molecular Biology unless stated otherwise (1998).

Small Scale Preparation of Plasmid DNA

Plasmid DNA was routinely extracted from \textit{E.coli} by alkaline lysis. A 3 ml 16-18 hour culture was centrifuged at 1400g for 10 minutes and the supernatant discarded. The remaining pellet was resuspended in 150 μl of solution I followed by addition of 300 μl of freshly prepared solution II. The contents were mixed by inversion and placed on ice for 5 minutes. Subsequent addition of 225 μl of solution III was followed by inversion mixing and incubation on ice for a further 5 minutes. The mixture was centrifuged at 15000g for 5 minutes to remove precipitate. The supernatant was extracted and incubated with RNAseA (final concentration 20 μg ml$^{-1}$) at 37°C for 20 minutes. Plasmid DNA in solution was purified by phenol/chloroform extraction precipitated and resuspended in 30 μl water (stored at -20°C).

Plasmid DNA to be used for sequencing or for transformation of \textit{M.bovis} BCG Pasteur was extracted using commercially available kits (Hybaid Recovery$^\text{TM}$ Plasmid Midi Prep kit or Qiagen Mini and Midi kits) according to manufacturers' instructions.
Appendix 2 - Methods

Chromosomal DNA Extraction from Mycobacteria

A 10 ml culture of *M. tuberculosis* H37Rv was grown at 37°C to late exponential phase after which cycloserine was added to a final concentration of 1mg ml⁻¹ and the culture grown for a further 12 hours. The bacteria were pelleted by centrifugation at 3000g for 15 minutes then the supernatant discarded. The remaining pellet was resuspended in 400 µl 0.3M sucrose/TE and incubated at 80°C for 1 hour to heat kill the bacteria. Freshly prepared lysozyme was added to a concentration of 2 µg ml⁻¹ and incubated at 37°C for 1 hour. EDTA (50mM), SDS (1%v/v) and pronase (0.2mg ml⁻¹) were and the mixture incubated at 55°C for 1 hour. This was followed by 2 phenol/chloroform extractions and treatment with RNAseA (final concentration 20 µg ml⁻¹). After incubation at 37°C for 1 hour, chromosomal DNA in solution was again purified by phenol/chloroform extraction, ethanol precipitated and resuspended in an appropriate volume of water.

Agarose Gel Electrophoresis

Routine DNA analysis was performed using agarose gel electrophoresis. Gels of appropriate concentration (w/v) were prepared in 1X TAE and subjected to electrophoresis at 1-5 V cm⁻¹ in a horizontal submarine gel tank (BioRad). Gels were then stained in an ethidium bromide bath (0.5 µg ml⁻¹ in 1X TAE). DNA was observed using an ultra-violet transilluminator and photographed on gel documentation systems Imagestore 5000 (UVP) or ImageMaster® (Pharmacia Biotech)
Formaldehyde Gel Electrophoresis

RNA was mixed 1:10 with RNA denaturant loading buffer and incubated at 50°C for 30 minutes. Gels of appropriate concentration (w/v) were prepared in 1X MOPS (10X MOPS added post agarose melting in water) and subjected to electrophoresis at 1-5 V cm⁻¹ in a horizontal submarine gel tank (BioRad). Gels were then stained in a SYBR Goldbath (1:20000 MOPS). RNA was observed an ultra-violet transilluminator and photographed on a gel documentation system (UVP Imagestore 5000) or ImageMaster® (Pharmacia Biotech).

Purification of DNA from Agarose Gels

Specific DNA fragments were excised from agarose under UV transillumination and purified using the QIAEX II Gel extraction kit (Qiagen) or Advantage™ PCR-Pure Kit (Clonetech) according to manufacturer’s instructions.

Phenol/Chloroform Extraction of DNA

Phenol/chloroform and DNA solution were mixed in equal volumes. Samples were centrifuged at 15000g for 5 minutes. The upper aqueous layer containing the DNA was extracted and retained.
DNA Precipitation

To a given volume of DNA in solution, was added 0.1 volume 2M NaCl and 2 volumes 100% (v/v) ethanol. Samples were placed at -20°C for 18-24 hours and centrifuged at 15000g for 10 minutes at 4°C. The supernatant was discarded and the DNA pellet was washed with 70% (v/v) ethanol. The resulting pellet was air dried and resuspended in an appropriate volume of water.

Restriction Endonuclease Digestion of DNA

DNA was digested with the required restriction endonuclease in the appropriate 1X restriction buffer and incubated at the enzyme’s optimum temperature. The total reaction volume was altered according to the amount of DNA to be digested. Restriction endonucleases and buffers were supplied by Roche and New England Biolabs.

Deposphorylation of DNA

DNA was dephosphorylated using calf intestinal alkaline phosphatase (CIP) in the supplied 1X buffer (Roche). The reaction was first incubated at 37°C for 15 minutes and then at 55°C for a further 15 minutes. Another 1 unit of CIP was added to the reaction and the first two incubation steps repeated. The reaction was terminated by inactivation of the enzyme at 75°C for 10 minutes.
Ligation of DNA Fragments

DNA fragments were ligated in a 3:1 molar ratio of insert DNA to vector using T4 DNA ligase (Bethesda Research Laboratories) in the supplied 1X buffer. The reaction was incubated at 14°C for 18-24 hours.

Synthesis of Oligonucleotides

Synthetic oligonucleotides were purchased from Sigma Genosys Limited.

Polymerase Chain Reaction (PCR)

PCR amplification was performed using the Perkin Elmer GeneAmp 9700 thermal cycler. A standard reaction contained 200 μM each dNTP (dATP, dGTP, dCTP, dTTP), 0.5 μM each oligonucleotide primer, 0.5 ng DNA template and 1U Taq DNA polymerase in 1X PCR buffer (Stratagene). Typical PCR reaction conditions consisted of an initial denaturation step at 95°C for 5 minutes, followed by 35 cycles with denaturation at 95°C, primer annealing at a temperature appropriate for the primers and extension at 72°C for a time dependant on the length of the expected product. The final cycle was extension at 72°C for 5 minutes.
Preparation of Competent Cells

*E.coli* for Routine Transformations.

*E.coli* DH5α cells were cultured in 100 ml LB broth at 37°C with shaking at 240 rpm to an OD$_{600}$ of 1.0 and chilled to 4°C for 20 minutes. Cells were harvested at 1400g for 15 minutes at 4°C and washed in 50 ml sterile water. This procedure was repeated twice and the cells resuspended in 2 ml 10% (v/v) glycerol. Aliquots were stored at -70°C.

*M.bovis* BCG Pasteur

*M.bovis* BCG Pasteur cells were prepared for transformation as described by Wards and Collins (Wards and Collins, 1996). Aliquots were either used immediately or stored at -70°C.

Transformation of Competent Cells

*E.coli* – Routine Transformations

Up to 10 ng of DNA was added to 20 µl of competent *E.coli* DH5α cells and placed in ice for 45 minutes. Cells were then heat-shocked at 42°C for 45 seconds and immediately placed in ice for 2 minutes. Recovery of cells was achieved by addition of 80 µl SOC and incubation at 37°C with shaking for 60 minutes prior to selection of transformants on LB agar supplemented with the appropriate selective antibiotics.
E. coli – Generation of Mycobacterium tuberculosis Plasmid Library.

Epicurian Coli® Ultracompetent Cells (Stratagene) were used according to manufacturer’s instructions with selection of transformants on LB agar supplemented with the appropriate antibiotics.

*M. bovis* BCG Pasteur

An aliquot (100 µl) of competent *M. bovis* BCG Pasteur cells were electroporated with 0.1-1 µg of plasmid DNA using the BioRad GenePulser (voltage 2.5 kV, capacitance 25 µFD, resistance 600Ω). Cells were then recovered in 900 µl of 7H9 broth at 37°C for 6-12 hours and transformants selected at 37°C on 7H11 agar supplemented with the appropriate antibiotics.

β-galactosidase Assay with Galacto-Light Plus Substrate

Galacto-Light Plus (Tropix Inc.) is a Chemiluminescent Reporter Gene Assay System for the detection of β-galactosidase which exhibits three orders of magnitude greater sensitivity compared to colourimetric β-galactosidase assays. The Galacto-Light Plus assay was used according to manufacturer’s instructions but with one alteration. Due to the nature of the mycobacterial cell wall, the lysis solution supplied with the kit does not disrupt the mycobacteria sufficiently. In order to cause sufficient cell disruption, 0.5ml of mycobacterial culture (unless otherwise stated) was added to a Hybaid Ribolyser™ blue tube containing 0.5ml PBS, with fresh Dithiothreitol (DTT) to 0.5mM, and ribolysed in a Hybaid Ribolyser™ at speed 5.0
for 45 seconds. The Ribolyser tubes, containing the lysed bacteria, were cooled on ice for 1 minute and then stored at -80°C to be used at a later date. The lysed bacteria were thawed slowly at 4°C and then centrifuged at 15000g for 5 minutes to sediment the cell debris. In all cases the Galacto-Light Plus assay was carried out using 20μl of lysed bacteria and the reaction was carried out in a microtitre plate incubated at 25°C. After 1 hour the luminescence of each well was measured using Packard Lumicount (Model AL10001) and analysed with PlateReader Software (v2.1). In all cases a cell count was carried out (by either measuring the absorbance at 600nm (A600) or by plating out small samples of cultures on selective 7H11 media) and the light units expressed as Light Units (LU) per 1x10⁶ bacterial cells.

Cell Culture

J774 Mouse Macrophage Cell Line (Unkeless, 1979)

Start cultures at 2x10⁵ cells/ml were incubated at 37°C with 5% CO₂ and maintained between 1x10⁵ and 1x10⁶ cells/ml. Every 3 days the complete Dulbecco’s modified Eagles medium was discarded and the attached cells washed twice with 10ml Versine (Gibco BRL), which was discarded. Adherence of the macrophages was reduced with the addition of 5mg/ml Trypsine in Versine. Complete Dulbecco’s modified Eagles medium was added to inactivate the trypsin and scraping brought the cells into suspension. The cells were centrifuged at 1000g for 5 minutes, the supernatant discarded and the cells resuspended in a small volume, counted and the cultures restarted.
THP-1 Human Macrophage Cell Line (Auwerx, 1991)

Start cultures at 2x10^5 cells/ml were incubated at 37°C with 5% CO₂ and maintained, by addition or replacement of fresh medium. Cell numbers were not allowed to exceed 1x10^6 cells/ml. Every 3 days the cells were passaged by centrifugation at 1000g for 5 minutes, the media discarded and the cells resuspended in a small volume, counted and the cultures restarted.

To cause differentiation, THP-1 cells were passaged, restarted at 1x10^6 cells/ml and incubated at 37°C with 5% CO₂ for 3 days in Dulbecco's modified Eagles Medium containing 5 ng/ml Phorbol Myristate Acetate (PMA) to induce monocyte differentiation. After this incubation time had elapsed the medium was discarded, the adherent monocytes were washed twice with PBS to remove any residual PMA, then fresh Dulbecco's modified Eagles medium containing 10% foetal calf serum was added and the cells incubated at 37°C with 5% CO₂.

Statistical Analysis

All statistical analysis was carried out using SPSS for Windows (SPSS Inc.) and are analysis of variance (AnoVa) tests.
Appendix 3 – Tables of β-galactosidase Assays of All Clones Exposed to Stressing Conditions In Vitro in the Second Round of Screening.

Shown overleaf is a table depicting the Light Units (LU)/10⁶ bacterial cells, indicating the level of lacZ expression of the 333 M. bovis BCG Pasteur library clones and the negative and positive control plasmids of pUS1800 and pUS1808, harboured within M. bovis BCG Pasteur, grown in selective 7H9 broth media at pH 6.6 (control). Columns 3 and 4 are ratios of the LU/10⁶ bacterial cells generated by the clone grown in liquid media at pH 5.4 and liquid media at pH 5.4 containing sodium nitrite divided by the LU/10⁶ bacterial cells generated by the clone grown in the control media. Therefore column 3 and 4 represent the factors of activation/repression of each of the clones when grown in the stressing media compared to the control media.

The clones shown in red have a factor of ‘up-regulation’ in either the low pH and/or the high nitric oxide media greater than 3 compared to the expression of the clone grown in the control medium. These clones were used in subsequent macrophage infection studies. Figures in green are the LU/10⁶ bacterial cells generated by clones grown in the low pH and/or the high nitric oxide media - a factor of ‘up-regulation’ can not be calculated for these clones since the clone grown in the control media did not generate measurable LU/10⁶ bacterial cells.
<table>
<thead>
<tr>
<th>Clone Number</th>
<th>RLU/Million bacterial cells in control medium</th>
<th>Low pH:control</th>
<th>High sodium nitrite:control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>684.8</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>697.6</td>
<td>0.6</td>
<td>2.3</td>
</tr>
<tr>
<td>3</td>
<td>2773.8</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>620.3</td>
<td>2.1</td>
<td>1.3</td>
</tr>
<tr>
<td>5</td>
<td>1066.8</td>
<td>3.4</td>
<td>3.1</td>
</tr>
<tr>
<td>6</td>
<td>313.0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>7</td>
<td>5945.7</td>
<td>4.0</td>
<td>1.6</td>
</tr>
<tr>
<td>8</td>
<td>173.5</td>
<td>1.4</td>
<td>1.3</td>
</tr>
<tr>
<td>9</td>
<td>1188.3</td>
<td>1.9</td>
<td>0.7</td>
</tr>
<tr>
<td>10</td>
<td>885.9</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td>11</td>
<td>2784.9</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>12</td>
<td>33.4</td>
<td>0.3</td>
<td>1.1</td>
</tr>
<tr>
<td>13</td>
<td>2195.1</td>
<td>3.3</td>
<td>2.5</td>
</tr>
<tr>
<td>14</td>
<td>773.7</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>15</td>
<td>5223.9</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>16</td>
<td>739.9</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>17</td>
<td>1494.4</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>18</td>
<td>1573.3</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>19</td>
<td>1329.3</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>20</td>
<td>2377.7</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>21</td>
<td>975.8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>22</td>
<td>1857.7</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>23</td>
<td>263.8</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>24</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>25</td>
<td>7265.2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>26</td>
<td>273.2</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>27</td>
<td>918.8</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>28</td>
<td>130.1</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>29</td>
<td>2285.9</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>30</td>
<td>797.8</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>31</td>
<td>434.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>32</td>
<td>1124.7</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>33</td>
<td>6630.4</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>34</td>
<td>753.5</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>35</td>
<td>1044.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>36</td>
<td>3173.3</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>37</td>
<td>4597.1</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>38</td>
<td>4403.1</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>39</td>
<td>976.2</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>40</td>
<td>4780.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>41</td>
<td>20.8</td>
<td>0.3</td>
<td>2.5</td>
</tr>
<tr>
<td>42</td>
<td>422.4</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>43</td>
<td>86.3</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>44</td>
<td>70.6</td>
<td>2.4</td>
<td>1.0</td>
</tr>
<tr>
<td>45</td>
<td>889.5</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>46</td>
<td>823.7</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>47</td>
<td>1165.9</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>48</td>
<td>744.7</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>49</td>
<td>281.5</td>
<td>1.7</td>
<td>0.6</td>
</tr>
<tr>
<td>50</td>
<td>81.3</td>
<td>0.7</td>
<td>1.8</td>
</tr>
</tbody>
</table>
## Appendix 3 - Tables of β-galactosidase Assays of All Clones Exposed to Stressing Conditions In Vitro in the Second Round of Screening

<table>
<thead>
<tr>
<th>Clone Number</th>
<th>RLU/Million bacterial cells in control medium</th>
<th>Low pH:control</th>
<th>High sodium nitrite:control</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>279.8</td>
<td>1.2</td>
<td>0.4</td>
</tr>
<tr>
<td>52</td>
<td>70.6</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>53</td>
<td>1395.1</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>54</td>
<td>30.4</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>55</td>
<td>17.0</td>
<td>1.1</td>
<td>0.0</td>
</tr>
<tr>
<td>56</td>
<td>35.1</td>
<td>2.0</td>
<td>0.6</td>
</tr>
<tr>
<td>57</td>
<td>202.8</td>
<td>1.4</td>
<td>0.5</td>
</tr>
<tr>
<td>58</td>
<td>223.2</td>
<td>0.3</td>
<td>1.5</td>
</tr>
<tr>
<td>59</td>
<td>31.3</td>
<td>1.7</td>
<td>1.9</td>
</tr>
<tr>
<td>60</td>
<td>104.9</td>
<td>0.6</td>
<td>1.5</td>
</tr>
<tr>
<td>61</td>
<td>483.6</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>62</td>
<td>74.8</td>
<td>3.4</td>
<td>0.7</td>
</tr>
<tr>
<td>63</td>
<td>652.5</td>
<td>0.8</td>
<td>3.2</td>
</tr>
<tr>
<td>64</td>
<td>219.8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>65</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>66</td>
<td>47.8</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>67</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>68</td>
<td>171.4</td>
<td>0.3</td>
<td>2.0</td>
</tr>
<tr>
<td>69</td>
<td>11.7</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>70</td>
<td>0.0</td>
<td>0.0</td>
<td>10.3</td>
</tr>
<tr>
<td>71</td>
<td>0.0</td>
<td>5.53</td>
<td>16.86</td>
</tr>
<tr>
<td>72</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>73</td>
<td>0.0</td>
<td>13.58</td>
<td>0.0</td>
</tr>
<tr>
<td>74</td>
<td>456.3</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>75</td>
<td>221.1</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>76</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>77</td>
<td>49.6</td>
<td>0.9</td>
<td>0.0</td>
</tr>
<tr>
<td>78</td>
<td>36.6</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>79</td>
<td>21.2</td>
<td>2.1</td>
<td>0.9</td>
</tr>
<tr>
<td>80</td>
<td>427.9</td>
<td>1.2</td>
<td>0.2</td>
</tr>
<tr>
<td>81</td>
<td>120.8</td>
<td>1.0</td>
<td>0.1</td>
</tr>
<tr>
<td>82</td>
<td>117.6</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>83</td>
<td>0.0</td>
<td>43.82</td>
<td>0.0</td>
</tr>
<tr>
<td>84</td>
<td>632.7</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>85</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>86</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>87</td>
<td>809.9</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>88</td>
<td>47.5</td>
<td>1.9</td>
<td>2.6</td>
</tr>
<tr>
<td>89</td>
<td>42.1</td>
<td>4.6</td>
<td>0.7</td>
</tr>
<tr>
<td>90</td>
<td>8.9</td>
<td>7.4</td>
<td>3.8</td>
</tr>
<tr>
<td>91</td>
<td>259.0</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>92</td>
<td>770.7</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>93</td>
<td>956.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>94</td>
<td>1444.3</td>
<td>0.9</td>
<td>0.5</td>
</tr>
<tr>
<td>95</td>
<td>1673.0</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>96</td>
<td>587.9</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>97</td>
<td>5821.8</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>98</td>
<td>2885.1</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>99</td>
<td>39.7</td>
<td>0.2</td>
<td>0.9</td>
</tr>
<tr>
<td>100</td>
<td>166.3</td>
<td>0.4</td>
<td>0.1</td>
</tr>
</tbody>
</table>
## Appendix 3 Tables of β-galactosidase Assays of All Clones Exposed to Stressing Conditions In Vitro in the Second Round of Screening

<table>
<thead>
<tr>
<th>Clone Number</th>
<th>RLU/Million bacterial cells in control medium</th>
<th>Low pH:control</th>
<th>High sodium nitrite:control</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>385.7</td>
<td>2.0</td>
<td>0.6</td>
</tr>
<tr>
<td>102</td>
<td>370.5</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>103</td>
<td>243.0</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>104</td>
<td>325.1</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>105</td>
<td>0.0</td>
<td>24.19</td>
<td>0.0</td>
</tr>
<tr>
<td>106</td>
<td>675.8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>107</td>
<td>1342.0</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>108</td>
<td>0.0</td>
<td>5.74</td>
<td>0.0</td>
</tr>
<tr>
<td>109</td>
<td>0.0</td>
<td>18.99</td>
<td>0.71</td>
</tr>
<tr>
<td>110</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>111</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>112</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>113</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>114</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>115</td>
<td>1604.2</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>116</td>
<td>2009.7</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>117</td>
<td>113.4</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>118</td>
<td>96.3</td>
<td>1.2</td>
<td>0.5</td>
</tr>
<tr>
<td>119</td>
<td>107.0</td>
<td>2.6</td>
<td>1.2</td>
</tr>
<tr>
<td>120</td>
<td>2144.4</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>121</td>
<td>406.2</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>122</td>
<td>48.6</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>123</td>
<td>308.6</td>
<td>1.8</td>
<td>0.9</td>
</tr>
<tr>
<td>124</td>
<td>413.5</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>125</td>
<td>460.7</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>126</td>
<td>1103.8</td>
<td>1.5</td>
<td>1.0</td>
</tr>
<tr>
<td>127</td>
<td>9219.5</td>
<td>1.3</td>
<td>0.8</td>
</tr>
<tr>
<td>128</td>
<td>78.4</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>129</td>
<td>0.0</td>
<td>12.82</td>
<td>9.66</td>
</tr>
<tr>
<td>130</td>
<td>1109.7</td>
<td>1.5</td>
<td>0.8</td>
</tr>
<tr>
<td>131</td>
<td>194.6</td>
<td>0.8</td>
<td>1.0</td>
</tr>
<tr>
<td>132</td>
<td>2071.3</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>133</td>
<td>547.1</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>134</td>
<td>6866.6</td>
<td>0.6</td>
<td>1.1</td>
</tr>
<tr>
<td>135</td>
<td>2120.0</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>136</td>
<td>493.8</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>137</td>
<td>1559.9</td>
<td>2.3</td>
<td>3.0</td>
</tr>
<tr>
<td>138</td>
<td>2877.0</td>
<td>2.1</td>
<td>1.9</td>
</tr>
<tr>
<td>139</td>
<td>504.3</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>140</td>
<td>2319.1</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>141</td>
<td>215.1</td>
<td>1.4</td>
<td>1.6</td>
</tr>
<tr>
<td>142</td>
<td>2568.1</td>
<td>2.1</td>
<td>2.3</td>
</tr>
<tr>
<td>143</td>
<td>283.5</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>144</td>
<td>2337.0</td>
<td>2.2</td>
<td>2.0</td>
</tr>
<tr>
<td>145</td>
<td>378.8</td>
<td>1.0</td>
<td>2.4</td>
</tr>
<tr>
<td>146</td>
<td>1068.9</td>
<td>1.1</td>
<td>0.8</td>
</tr>
<tr>
<td>147</td>
<td>2368.9</td>
<td>1.5</td>
<td>1.8</td>
</tr>
<tr>
<td>148</td>
<td>1726.0</td>
<td>2.3</td>
<td>1.9</td>
</tr>
<tr>
<td>149</td>
<td>822.7</td>
<td>1.4</td>
<td>2.0</td>
</tr>
<tr>
<td>150</td>
<td>28.8</td>
<td>0.4</td>
<td>1.0</td>
</tr>
</tbody>
</table>
### Appendix 3 Tables of β-galactosidase Assays of All Clones Exposed to Stressing Conditions In Vitro in the Second Round of Screening

<table>
<thead>
<tr>
<th>Clone Number</th>
<th>RLU/Million bacterial cells in control medium</th>
<th>Low pH:control</th>
<th>High sodium nitrite:control</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>0.0</td>
<td>0.0</td>
<td>1.80</td>
</tr>
<tr>
<td>152</td>
<td>19.1</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>153</td>
<td>40.7</td>
<td>2.4</td>
<td>2.5</td>
</tr>
<tr>
<td>154</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>155</td>
<td>0.0</td>
<td>0.94</td>
<td>15.62</td>
</tr>
<tr>
<td>156</td>
<td>804.7</td>
<td>1.3</td>
<td>1.1</td>
</tr>
<tr>
<td>157</td>
<td>81.3</td>
<td>1.3</td>
<td>1.0</td>
</tr>
<tr>
<td>158</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>159</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>160</td>
<td>12.8</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>161</td>
<td>0.0</td>
<td>0.0</td>
<td>4.97</td>
</tr>
<tr>
<td>162</td>
<td>19.9</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>163</td>
<td>11.6</td>
<td>1.5</td>
<td>2.4</td>
</tr>
<tr>
<td>164</td>
<td>380.2</td>
<td>2.1</td>
<td>1.8</td>
</tr>
<tr>
<td>165</td>
<td>492.9</td>
<td>0.8</td>
<td>1.0</td>
</tr>
<tr>
<td>166</td>
<td>662.4</td>
<td>2.3</td>
<td>3.6</td>
</tr>
<tr>
<td>167</td>
<td>103.9</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>168</td>
<td>31.0</td>
<td>1.4</td>
<td>2.2</td>
</tr>
<tr>
<td>169</td>
<td>264.9</td>
<td>7.6</td>
<td>4.8</td>
</tr>
<tr>
<td>170</td>
<td>102.4</td>
<td>1.6</td>
<td>1.7</td>
</tr>
<tr>
<td>171</td>
<td>0.0</td>
<td>6.97</td>
<td>21.94</td>
</tr>
<tr>
<td>172</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>173</td>
<td>83.1</td>
<td>3.2</td>
<td>2.6</td>
</tr>
<tr>
<td>174</td>
<td>42.0</td>
<td>2.1</td>
<td>3.4</td>
</tr>
<tr>
<td>175</td>
<td>121.6</td>
<td>4.4</td>
<td>2.9</td>
</tr>
<tr>
<td>176</td>
<td>240.7</td>
<td>3.9</td>
<td>2.4</td>
</tr>
<tr>
<td>177</td>
<td>610.4</td>
<td>2.2</td>
<td>0.9</td>
</tr>
<tr>
<td>178</td>
<td>1110.3</td>
<td>2.0</td>
<td>0.9</td>
</tr>
<tr>
<td>179</td>
<td>569.6</td>
<td>1.2</td>
<td>0.6</td>
</tr>
<tr>
<td>180</td>
<td>29.5</td>
<td>3.2</td>
<td>3.9</td>
</tr>
<tr>
<td>181</td>
<td>3677.5</td>
<td>2.8</td>
<td>1.8</td>
</tr>
<tr>
<td>182</td>
<td>191.3</td>
<td>1.5</td>
<td>1.1</td>
</tr>
<tr>
<td>183</td>
<td>1328.5</td>
<td>2.3</td>
<td>1.6</td>
</tr>
<tr>
<td>184</td>
<td>252.5</td>
<td>1.5</td>
<td>2.3</td>
</tr>
<tr>
<td>185</td>
<td>0.0</td>
<td>34.01</td>
<td>52.63</td>
</tr>
<tr>
<td>186</td>
<td>172.8</td>
<td>1.2</td>
<td>1.7</td>
</tr>
<tr>
<td>187</td>
<td>0.0</td>
<td>0.0</td>
<td>2.06</td>
</tr>
<tr>
<td>188</td>
<td>873.4</td>
<td>4.5</td>
<td>2.0</td>
</tr>
<tr>
<td>189</td>
<td>118.5</td>
<td>2.6</td>
<td>1.1</td>
</tr>
<tr>
<td>190</td>
<td>111.1</td>
<td>2.7</td>
<td>1.6</td>
</tr>
<tr>
<td>191</td>
<td>207.6</td>
<td>2.3</td>
<td>1.6</td>
</tr>
<tr>
<td>192</td>
<td>35.1</td>
<td>1.4</td>
<td>1.1</td>
</tr>
<tr>
<td>193</td>
<td>507.7</td>
<td>1.4</td>
<td>0.8</td>
</tr>
<tr>
<td>194</td>
<td>2754.8</td>
<td>3.4</td>
<td>1.9</td>
</tr>
<tr>
<td>195</td>
<td>310.7</td>
<td>1.4</td>
<td>0.4</td>
</tr>
<tr>
<td>196</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>197</td>
<td>2578.8</td>
<td>2.0</td>
<td>2.3</td>
</tr>
<tr>
<td>198</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>199</td>
<td>26.4</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>200</td>
<td>19.4</td>
<td>2.0</td>
<td>6.5</td>
</tr>
<tr>
<td>Clone Number</td>
<td>RLU/Million bacterial cells in control medium</td>
<td>Low pH:control</td>
<td>High sodium nitrite:control</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------</td>
<td>----------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>201</td>
<td>2.6</td>
<td>13.5</td>
<td>5.8</td>
</tr>
<tr>
<td>202</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>203</td>
<td>2436.8</td>
<td>2.9</td>
<td>2.5</td>
</tr>
<tr>
<td>204</td>
<td>51.9</td>
<td>2.7</td>
<td>2.1</td>
</tr>
<tr>
<td>205</td>
<td>581.1</td>
<td>2.9</td>
<td>3.0</td>
</tr>
<tr>
<td>206</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>207</td>
<td>526.3</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>208</td>
<td>610.9</td>
<td>1.5</td>
<td>0.6</td>
</tr>
<tr>
<td>209</td>
<td>193.3</td>
<td>2.0</td>
<td>0.8</td>
</tr>
<tr>
<td>210</td>
<td>65.5</td>
<td>1.1</td>
<td>0.6</td>
</tr>
<tr>
<td>211</td>
<td>1183.0</td>
<td>2.1</td>
<td>2.5</td>
</tr>
<tr>
<td>212</td>
<td>53.3</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>213</td>
<td>1265.1</td>
<td>2.8</td>
<td>0.8</td>
</tr>
<tr>
<td>214</td>
<td>1351.5</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>215</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>216</td>
<td>16.1</td>
<td>1.6</td>
<td>1.1</td>
</tr>
<tr>
<td>217</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>218</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>219</td>
<td>358.1</td>
<td>1.8</td>
<td>1.9</td>
</tr>
<tr>
<td>220</td>
<td>2118.1</td>
<td>3.4</td>
<td>0.8</td>
</tr>
<tr>
<td>221</td>
<td>823.9</td>
<td>1.6</td>
<td>1.4</td>
</tr>
<tr>
<td>222</td>
<td>1094.3</td>
<td>3.9</td>
<td>10.6</td>
</tr>
<tr>
<td>223</td>
<td>1689.3</td>
<td>2.1</td>
<td>1.5</td>
</tr>
<tr>
<td>224</td>
<td>1240.2</td>
<td>1.5</td>
<td>0.9</td>
</tr>
<tr>
<td>225</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>226</td>
<td>489.0</td>
<td>2.9</td>
<td>3.1</td>
</tr>
<tr>
<td>227</td>
<td>68.6</td>
<td>1.3</td>
<td>0.9</td>
</tr>
<tr>
<td>228</td>
<td>12.1</td>
<td>0.7</td>
<td>0.5</td>
</tr>
<tr>
<td>229</td>
<td>19.9</td>
<td>3.1</td>
<td>0.7</td>
</tr>
<tr>
<td>230</td>
<td>95.3</td>
<td>3.4</td>
<td>2.8</td>
</tr>
<tr>
<td>231</td>
<td>208.1</td>
<td>1.7</td>
<td>1.6</td>
</tr>
<tr>
<td>232</td>
<td>92.2</td>
<td>2.0</td>
<td>0.8</td>
</tr>
<tr>
<td>233</td>
<td>2044.1</td>
<td>4.7</td>
<td>2.1</td>
</tr>
<tr>
<td>234</td>
<td>8428.6</td>
<td>9.2</td>
<td>11.7</td>
</tr>
<tr>
<td>235</td>
<td>213.4</td>
<td>1.2</td>
<td>2.1</td>
</tr>
<tr>
<td>236</td>
<td>144.2</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>237</td>
<td>93.3</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td>238</td>
<td>1297.4</td>
<td>2.0</td>
<td>0.6</td>
</tr>
<tr>
<td>239</td>
<td>267.0</td>
<td>3.8</td>
<td>7.4</td>
</tr>
<tr>
<td>240</td>
<td>917.0</td>
<td>2.9</td>
<td>3.4</td>
</tr>
<tr>
<td>241</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>242</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>243</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>244</td>
<td>7563.9</td>
<td>0.2</td>
<td>1.8</td>
</tr>
<tr>
<td>245</td>
<td>90.0</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>246</td>
<td>7.8</td>
<td>0.0</td>
<td>2.6</td>
</tr>
<tr>
<td>247</td>
<td>24.3</td>
<td>10.1</td>
<td>2.7</td>
</tr>
<tr>
<td>248</td>
<td>0.0</td>
<td>5.50</td>
<td>0.0</td>
</tr>
<tr>
<td>249</td>
<td>0.0</td>
<td>4.16</td>
<td>8.16</td>
</tr>
<tr>
<td>250</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
### Appendix 3 Tables of β-galactosidase Assays of All Clones Exposed to Stressing Conditions In Vitro in the Second Round of Screening

<table>
<thead>
<tr>
<th>Clone Number</th>
<th>RLU/Million bacterial cells in control medium</th>
<th>Low pH:control</th>
<th>High sodium nitrite:control</th>
</tr>
</thead>
<tbody>
<tr>
<td>251</td>
<td>0.0</td>
<td>2.52</td>
<td>16.28</td>
</tr>
<tr>
<td>252</td>
<td>17.9</td>
<td>13.2</td>
<td>3.3</td>
</tr>
<tr>
<td>253</td>
<td>461.0</td>
<td>2.1</td>
<td>1.1</td>
</tr>
<tr>
<td>254</td>
<td>1674.2</td>
<td>5.2</td>
<td>2.2</td>
</tr>
<tr>
<td>255</td>
<td>135.1</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>256</td>
<td>0.0</td>
<td>39.02</td>
<td>30.88</td>
</tr>
<tr>
<td>257</td>
<td>65.0</td>
<td>1.2</td>
<td>1.4</td>
</tr>
<tr>
<td>258</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>259</td>
<td>90.0</td>
<td>1.7</td>
<td>1.3</td>
</tr>
<tr>
<td>260</td>
<td>90.8</td>
<td>1.4</td>
<td>1.2</td>
</tr>
<tr>
<td>261</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>262</td>
<td>0.0</td>
<td>2.69</td>
<td>17.02</td>
</tr>
<tr>
<td>263</td>
<td>0.0</td>
<td>0.0</td>
<td>17.25</td>
</tr>
<tr>
<td>264</td>
<td>521.5</td>
<td>1.6</td>
<td>2.8</td>
</tr>
<tr>
<td>265</td>
<td>0.0</td>
<td>10.22</td>
<td>0.0</td>
</tr>
<tr>
<td>266</td>
<td>3828.6</td>
<td>1.6</td>
<td>0.6</td>
</tr>
<tr>
<td>267</td>
<td>84.5</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>268</td>
<td>1884.1</td>
<td>1.9</td>
<td>2.3</td>
</tr>
<tr>
<td>269</td>
<td>50.5</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>270</td>
<td>1634.6</td>
<td>2.5</td>
<td>1.7</td>
</tr>
<tr>
<td>271</td>
<td>1984.2</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>272</td>
<td>115.4</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>273</td>
<td>190.0</td>
<td>2.7</td>
<td>2.3</td>
</tr>
<tr>
<td>274</td>
<td>147.0</td>
<td>1.6</td>
<td>2.2</td>
</tr>
<tr>
<td>275</td>
<td>2015.6</td>
<td>2.8</td>
<td>1.2</td>
</tr>
<tr>
<td>276</td>
<td>18.9</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>277</td>
<td>273.6</td>
<td>4.0</td>
<td>2.2</td>
</tr>
<tr>
<td>278</td>
<td>2359.4</td>
<td>4.3</td>
<td>5.5</td>
</tr>
<tr>
<td>279</td>
<td>1592.9</td>
<td>4.3</td>
<td>1.7</td>
</tr>
<tr>
<td>280</td>
<td>146.8</td>
<td>2.1</td>
<td>1.4</td>
</tr>
<tr>
<td>281</td>
<td>44.9</td>
<td>9.0</td>
<td>15.7</td>
</tr>
<tr>
<td>282</td>
<td>239.7</td>
<td>1.4</td>
<td>3.0</td>
</tr>
<tr>
<td>283</td>
<td>150.4</td>
<td>1.3</td>
<td>1.5</td>
</tr>
<tr>
<td>284</td>
<td>4276.5</td>
<td>1.4</td>
<td>2.4</td>
</tr>
<tr>
<td>285</td>
<td>4305.4</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>286</td>
<td>4230.1</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>287</td>
<td>4580.8</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>288</td>
<td>6651.4</td>
<td>0.9</td>
<td>1.7</td>
</tr>
<tr>
<td>289</td>
<td>477.1</td>
<td>5.2</td>
<td>3.5</td>
</tr>
<tr>
<td>290</td>
<td>29.8</td>
<td>7.6</td>
<td>2.5</td>
</tr>
<tr>
<td>291</td>
<td>2837.7</td>
<td>5.4</td>
<td>6.1</td>
</tr>
<tr>
<td>292</td>
<td>548.4</td>
<td>2.7</td>
<td>3.4</td>
</tr>
<tr>
<td>293</td>
<td>151.0</td>
<td>5.4</td>
<td>3.8</td>
</tr>
<tr>
<td>294</td>
<td>79.9</td>
<td>0.9</td>
<td>2.6</td>
</tr>
<tr>
<td>295</td>
<td>0.0</td>
<td>810.64</td>
<td>54.25</td>
</tr>
<tr>
<td>296</td>
<td>12.4</td>
<td>7.0</td>
<td>7.4</td>
</tr>
<tr>
<td>297</td>
<td>194.5</td>
<td>4.8</td>
<td>3.2</td>
</tr>
<tr>
<td>298</td>
<td>5153.4</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>299</td>
<td>905.8</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>300</td>
<td>223.4</td>
<td>1.8</td>
<td>0.2</td>
</tr>
<tr>
<td>Clone Number</td>
<td>RLU/Million bacterial cells in control medium</td>
<td>Low pH:control</td>
<td>High sodium nitrite:control</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------</td>
<td>----------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>301</td>
<td>0.0</td>
<td>41.83</td>
<td>62.22</td>
</tr>
<tr>
<td>302</td>
<td>22.5</td>
<td>11.1</td>
<td>6.3</td>
</tr>
<tr>
<td>303</td>
<td>8.9</td>
<td>5.4</td>
<td>0.0</td>
</tr>
<tr>
<td>304</td>
<td>26.1</td>
<td>2.0</td>
<td>1.6</td>
</tr>
<tr>
<td>305</td>
<td>95.4</td>
<td>5.3</td>
<td>2.8</td>
</tr>
<tr>
<td>306</td>
<td>170.3</td>
<td>2.2</td>
<td>1.2</td>
</tr>
<tr>
<td>307</td>
<td>1976.3</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>308</td>
<td>766.0</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>309</td>
<td>143.4</td>
<td>0.3</td>
<td>1.5</td>
</tr>
<tr>
<td>310</td>
<td>566.5</td>
<td>0.5</td>
<td>0.9</td>
</tr>
<tr>
<td>311</td>
<td>845.8</td>
<td>3.0</td>
<td>2.2</td>
</tr>
<tr>
<td>312</td>
<td>3860.5</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>313</td>
<td>142.9</td>
<td>7.1</td>
<td>6.1</td>
</tr>
<tr>
<td>314</td>
<td>28.4</td>
<td>1.6</td>
<td>2.1</td>
</tr>
<tr>
<td>315</td>
<td>1433.8</td>
<td>1.1</td>
<td>0.5</td>
</tr>
<tr>
<td>316</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>317</td>
<td>940.4</td>
<td>1.0</td>
<td>2.1</td>
</tr>
<tr>
<td>318</td>
<td>658.2</td>
<td>1.3</td>
<td>2.8</td>
</tr>
<tr>
<td>319</td>
<td>0.0</td>
<td>6.84</td>
<td>8.52</td>
</tr>
<tr>
<td>320</td>
<td>0.4</td>
<td>30.2</td>
<td>116.5</td>
</tr>
<tr>
<td>321</td>
<td>0.0</td>
<td>36.46</td>
<td>36.28</td>
</tr>
<tr>
<td>322</td>
<td>0.0</td>
<td>134.61</td>
<td>245.05</td>
</tr>
<tr>
<td>323</td>
<td>395.8</td>
<td>7.6</td>
<td>9.8</td>
</tr>
<tr>
<td>324</td>
<td>140.2</td>
<td>1.6</td>
<td>2.5</td>
</tr>
<tr>
<td>325</td>
<td>106.0</td>
<td>4.3</td>
<td>2.6</td>
</tr>
<tr>
<td>326</td>
<td>822.3</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>327</td>
<td>242.2</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>328</td>
<td>81.0</td>
<td>1.6</td>
<td>2.1</td>
</tr>
<tr>
<td>329</td>
<td>98.5</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>330</td>
<td>12.2</td>
<td>3.9</td>
<td>3.9</td>
</tr>
<tr>
<td>331</td>
<td>2555.1</td>
<td>2.2</td>
<td>1.5</td>
</tr>
<tr>
<td>332</td>
<td>14.0</td>
<td>2.6</td>
<td>2.3</td>
</tr>
<tr>
<td>333</td>
<td>0.0</td>
<td>27.86</td>
<td>0.0</td>
</tr>
<tr>
<td>pUS1800</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>pUS1808</td>
<td>8945.8</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Appendix 3  Tables of β-galactosidase Assays of All Clones Exposed to Stressing Conditions In Vitro in the Second Round of Screening.
Appendix 4 - Graphs of β-galactosidase Assays of All Clones Post-Infection in Murine and Human Macrophages.

Below are bar charts of β-galactosidase activity levels from each of the 42 *M.bovis* BCG Pasteur clones, harbouring the *M.tuberculosis* promoter probe vector, and the negative control plasmid of pUS1800, harboured within *M.bovis* BCG Pasteur, obtained post-infection of murine J774 and human THP-1 macrophages. Also shown are the β-galactosidase activity levels from each of the 42 *M.bovis* BCG Pasteur clones, harbouring the *M.tuberculosis* promoter probe vector, and the negative control plasmid of pUS1800, harboured within *M.bovis* BCG Pasteur, incubated in either complete Dulbecco’s modified Eagles medium for 2 and 24 hours or liquid 7H9 medium for 24 hours. The graphs are Light Units (LU)/ $10^6$ bacterial cells against the clone number and are on different logarithmic scales.
Appendix 4: Graphs of ß-galactosidase Assays of All Clones Post-Infection in Murine and Human Macrophages.

Clone Number and Macrophage Type Infected

Number of Bacterial Cells
Appendix 4: Graphs of β-galactosidase Assays of All Clones Post-Infection in Murine and Human Macrophages.
Appendix 4: Graphs of α-galactosidase Assays of All Clones Post-Infection in Murine and Human Macrophages.

<table>
<thead>
<tr>
<th>Clone Number and Macrophage Type Infected</th>
<th>Serum T2</th>
<th>Infection T2</th>
<th>Infection T12</th>
<th>Infection T24</th>
<th>Serum T24</th>
<th>7H9 T24</th>
</tr>
</thead>
<tbody>
<tr>
<td>188 J774</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>188 THP-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 J774</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 THP-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201 J774</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201 THP-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>205 J774</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>205 THP-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>222 J774</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>222 THP-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 4: Graphs of β-galactosidase Assays of All Clones Post-Infection in Murine and Human Macrophages.
Appendix 4: Graphs of β-galactosidase Assays of All Clones Post-Infection in Murine and Human Macrophages.
Appendix 4: Graphs of β-galactosidase Assays of All Clones Post-Infection in Murine and Human Macrophages.
Appendix 4: Graphs of P-galactosidase Assays of AH Clones Post-Infection in Murine and Human Macrophages.
Appendix 4: Graphs of β-galactosidase Assays of All Clones Post-Infection in Marine and Human Macrophages.
Appendix 5 - Sequence Data

All sequences are shown in the orientation that they are cloned in pUS1800 with the translational start site of the lacZ reporter gene at the 3’ end of the DNA sequence shown. The 37 clones printed in blue are those infected into murine and human macrophages. The 5 clones printed in red were screened twice by re-infecting human macrophages. The 8 clones printed in black were assayed as described in Chapter 3.

The size of the cloned DNA, the \( M. tuberculosis \) H37Rv genome bases cloned and the restriction site used to digest genomic \( M. tuberculosis \) H37Rv DNA for cloning in to pUS1800 are detailed prior to the sequence. All genes or parts of genes (start or end) are depicted with arrows with the direction of the arrow indicating the direction of translation. Shaded regions of the sequence represent intergenic regions.

Clone 5
Size of cloned \( M. tuberculosis \) H37Rv DNA: 499bp
Nucleotides of \( M. tuberculosis \) H37Rv genome cloned: 915974bp–916472bp
\( M. tuberculosis \) genome base pair before lacZ in pUS1800: 916472bp
Restriction site used to digest genomic \( M. tuberculosis \) H37Rv DNA for cloning in to pUS1800: TaqI

```
CTGACTCGGAGACCCGCGTGCCCAAGCGGCCGGCGCTCGTGCGTCTGCTCGTGCTGCTGACCGCTGCCTCGGGGACCTGGGGGCCGCGGCCGGATGGCTCGGGGTTCGGGGCGACATGGCGGTGCGTGGGGAGGTCAGGAAAAGCGGGGCCGGAGCCTGGCGATCAGATCGGCCACACTAACGGCGCCGGTGGCATGACAGCCGACATTCTGGGTGTCCCGCGGACCCTGGGCTGCCACCCATGTGGCGGGCGGTACCGTGATCCATCGGTCAACACCATCGGGGAAT
```

```
Rv0822c<---------1
GCTGACTCGGAGACCCGCGTGCCCAAGCGGCCGGCGCTCGTGCGTCTGCTCGTGCTGCTGACCGCTGCCTCGGGGACCTGGGGGCCGCGGCCGGATGGCTCGGGGTTCGGGGCGACATGGCGGTGCGTGGGGAGGTCAGGAAAAGCGGGGCCGGAGCCTGGCGATCAGATCGGCCACACTAACGGCGCCGGTGGCATGACAGCCGACATTCTGGGTGTCCCGCGGACCCTGGGCTGCCACCCATGTGGCGGGCGGTACCGTGATCCATCGGTCAACACCATCGGGGAAT
```

6bp to stop codon of Rv0823c

208
Appendix 5: Sequence Data

Clone 7
Size of cloned *M. tuberculosis* H37Rv DNA: 233bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1699624bp-1699856bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 1699624bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*
Sequence internal to Rv1508.

Clone 13
Size of cloned *M. tuberculosis* H37Rv DNA: 865bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 551479bp-552343bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 552343bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*

Clone 63
Size of cloned *M. tuberculosis* H37Rv DNA: 406bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1051336bp-1051741bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 1051336bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *TaqI*
Appendix 5: Sequence Data

Clone 89
Size of cloned *M. tuberculosis* H37Rv DNA: 243bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 905242bp-905484bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 905484bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *HpaII*
Internal to Rv0811c
CCGGCCGGCACAGCGCTGACGAGCCGCCGCAGACCCCGAGTCCGACG

Clone 90
Size of cloned *M. tuberculosis* H37Rv DNA: 251bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 704074bp-704324bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 704324bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *HpaII*
Intergenic to Rv0609 and Rv0610c - converging genes
CCGGCTGAGCGCGTCGAGCTCAACTATCCAGTCGACGCTCCGCGCCGGTT
ACCTCAATGCCGGGTACGCTGTCGACGCTCACGCGACCGGACTGTCGACCG

Clone 137
Size of cloned *M. tuberculosis* H37Rv DNA: 126bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 695401bp-695526bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 695401bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *TaqI*
CCGGTGACCACAACGACGCGCCCTTTGATCGGGGACGTCTGCGGCCGACCATTTACGGGTCTTGTTG
Rv0597c

Clone 166
Size of cloned *M. tuberculosis* H37Rv DNA: 361bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 528230bp-528590bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 528590bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *HpaII*
TCGGTGACCAACACCGGCCGCGACCGTCTCCGACGTCGACGCTCCGAGG
TCCTGCAACCACACCGGCCGCGACCGTCTCCGAGG

210
Appendix 5: Sequence Data

Clone 169
Size of cloned *M. tuberculosis* H37Rv DNA: 524bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1166888bp-116741bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 116741bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: AciI

Clone 173
Size of cloned *M. tuberculosis* H37Rv DNA: 314bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 709301bp-709614bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 709301bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: AciI

Clone 174
Size of cloned *M. tuberculosis* H37Rv DNA: 322bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 157673bp-157984bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 157984bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: TagI

Clone 175
Size of cloned *M. tuberculosis* H37Rv DNA: 395bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 364222bp-364616bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 364616bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: AciI

211
Appendix 5: Sequence Data

Clone 176
Size of cloned *M. tuberculosis* H37Rv DNA: 259bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 3023359bp-3023617bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 3023617bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *TaqI*

Clone 180
Size of cloned *M. tuberculosis* H37Rv DNA: 271bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1293907bp-1294177bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 1294177bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*

Clone 188
When this clone is bi-directionally sequenced with the two primers and homologies identified against the *M. tuberculosis* H37Rv genome held at the Sanger Centre, different *M. tuberculosis* H37Rv DNA sequences are identified depending on the direction of the sequencing.

Primer RH20 routinely identifies homology as:
Number of base with homology to *M. tuberculosis* H37Rv genome: 482 bases
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1623341bp - 1623822bp
*M. tuberculosis* genome base pair closest to lacZ in pUS1800: 1623341bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI"

Primer 1800Rev routinely identifies homology as:
Number of base with homology to *M. tuberculosis* H37Rv genome: 344 bases
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 4328658bp - 4329001bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 4329001bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI"
Appendix 3: Sequence Data

Clone 194
Size of cloned *M.tuberculosis* H37Rv DNA: 189bp
Nucleotides of *M.tuberculosis* H37Rv genome cloned: 3967509bp-3967697bp
*M.tuberculosis* genome base pair before lacZ in pUS1800: 3967697bp
Restriction site used to digest genomic *M.tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*
Internal to Rv3530

Clone 200
Size of cloned *M.tuberculosis* H37Rv DNA: 709bp
Nucleotides of *M.tuberculosis* H37Rv genome cloned: 987661 bp-988370bp
*M.tuberculosis* genome base pair before lacZ in pUS1800: 987661bp
Restriction site used to digest genomic *M.tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*
Internal to Rv0888

Clone 201
Size of cloned *M.tuberculosis* H37Rv DNA: 183bp
Nucleotides of *M.tuberculosis* H37Rv genome cloned: 4191841bp-4192023bp
*M.tuberculosis* genome base pair before lacZ in pUS1800: 4191841bp
Restriction site used to digest genomic *M.tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*
Internal to Rv3740

Clone 205
Size of cloned *M.tuberculosis* H37Rv DNA: 543bp
Nucleotides of *M.tuberculosis* H37Rv genome cloned: 860532bp-861074bp
*M.tuberculosis* genome base pair before lacZ in pUS1800: 860532bp
Restriction site used to digest genomic *M.tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*

---

**189bp fragment:**

<table>
<thead>
<tr>
<th>Nucleotide Location</th>
<th>Sequence</th>
</tr>
</thead>
</table>
| 3967509-3967697     | CCGCTGCTGGCCGCCAATCGTGCCTCGCGGAGGCAGGCAGACCGCGGCGGCGGAATTCTTCG...

**709bp fragment:**

<table>
<thead>
<tr>
<th>Nucleotide Location</th>
<th>Sequence</th>
</tr>
</thead>
</table>
| 987661-988370       | TCCCGGGCCGCCCTCGGATGATGTTGTCGTTGCGGCCCTCCCGACAGTTGGAGTGTAACCGCGAAGGCAGTCCCGGC...

**183bp fragment:**

<table>
<thead>
<tr>
<th>Nucleotide Location</th>
<th>Sequence</th>
</tr>
</thead>
</table>
| 4191841-4192023     | CCGGGCCGCCAATCGTGCCTCGCGGAGGCAGGCAGACCGCGGCGGCGGAATTCTTCG...

**543bp fragment:**

<table>
<thead>
<tr>
<th>Nucleotide Location</th>
<th>Sequence</th>
</tr>
</thead>
</table>
| 860532-861074       | CCGCGGGCCGCCCTCGGATGATGTTGTCGTTGCGGCCCTCCCGACAGTTGGAGTGTAACCGCGAAGGCAGTCCCGGC...

---

213
Appendix 5: Sequence Data

Clone 220
Size of cloned *M. tuberculosis* H37Rv DNA: 303bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 528230bp-528532bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 528532bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: HpaII

Clone 222
Size of cloned *M. tuberculosis* H37Rv DNA: 298v
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 3541887bp-3542184bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 3542184bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: Acil

Clone 226
Size of cloned *M. tuberculosis* H37Rv DNA: 261bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 2925356bp-2925616bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 2925356bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: HpaII

Clone 229
Size of cloned *M. tuberculosis* H37Rv DNA: 341bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 3439365bp-3439705bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 3439705bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: HpaII

214
Appendix 5: Sequence Data

Clone 230
Size of cloned *M. tuberculosis* H37Rv DNA: 238bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 4398121bp-4398358bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 4398121bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*
Internal to Rv3910

```
GCGGGGTCACAGAACCTCATCCTATGCGGCTAGGGTGTTGGGCTTGCAATTCCGAGGCTAGCTCAAGCCAGTCA
```

Clone 233
Size of cloned *M. tuberculosis* H37Rv DNA: 165bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1446292bp-1446456bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 1446292bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*

```
GCGGAGGCATCGAGGCCGCGCTCGGCCAGCACCGCGGCCGCGTGGATGCCGTGCGAGCCCGAATCAGGAGCG
```

Clone 234
Size of cloned *M. tuberculosis* H37Rv DNA: 781bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 3233754bp-3234534bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 3233754bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*

```
CCGCCCGCCGGGCAAGAAGATTACCCGACTGTCTTTGCTCTCCGGTGGCGAGAAGGCGCTGACCGCGGTGGCG
```

Clone 239
Size of cloned *M. tuberculosis* H37Rv DNA: 253bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1413241bp-1413493bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 1413493bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *TagI*

```
TCGAATGGCTGATCACAATGGTTCTCGCCAGGCCGGACGCTGTTTTCGCGCCAGGAACCGGTTGCT
```

215
Appendix 5: Sequence Data

Clone 247
Size of cloned *M. tuberculosis* H37Rv DNA: 643bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 3209562bp-3210204bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 3209562bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: HpaII
Internal to Rv2900c

Clone 252
Size of cloned *M. tuberculosis* H37Rv DNA: 390bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 638179bp-638568bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 638179bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: AcI
Internal to Rv0547c

Clone 254
Size of cloned *M. tuberculosis* H37Rv DNA: 662bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 213879bp-214540bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 214540bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: AcI
Internal to Rv0183

216
Appendix 5: Sequence Data

Clone 277
Size of cloned *M. tuberculosis* H37Rv DNA: 261bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 2623709bp-2623969bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 2623969bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *Acil*

\[
\text{Rv2334c} \xrightarrow{\text{Acil}} \text{CGGCCTCAGACGCTGTCCGCTCATCGCGGCTTTCGTCCAGCGTCGTCG}
\]

Clone 278
Size of cloned *M. tuberculosis* H37Rv DNA: 303bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 528230bp-528532bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 528532bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *HpalI*

\[
\text{CGGCTTGGCGTGACCACAACGACGCGCCCGCTTTGATCGGGGACGTCTGCGGCCGACCATTTACGGGTCTTGTTG} \\
\text{ATGCCCGTGCTCGGCTTTCTTGACTCGCCGCTACGCGAGTAAGCTAAGATTTTGGACACTCGGCAGCATCGCGGAAGGACACGGTCGACGGCGCGGCGGCTTCTTGCACTCGGCATAGGCGAGTGCTAAGAATAACGTTGGCACTCGCGACC}
\]

Clone 279
Size of cloned *M. tuberculosis* H37Rv DNA: 316bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1833418bp-1833733bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 1833733bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *Acil*

\[
\text{CCGCCCTGAGTTCACGCTCGGCGCAATCGGGACCGAGTTTGTCCAGCGTGTACCCGTCGAGTAGCCTCG} \\
\text{TCTCATCTCGCGGCGACGGTGATGACACCATCCTGGTGGTGGCCCGCA} \\
\text{TCTGCAGATCAGGCTAACAAGACCGCGGTCTGGATCCCCGC}
\]

Clone 281
Size of cloned *M. tuberculosis* H37Rv DNA: 295bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1871235bp-1871529bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 1871529bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *Acil*

\[
\text{CCCGCGCGGCCCTGCCGCCCCAGGTCGTCGGCACCATCGCCGGTGATGACACCATCCTGGTGGTGGCCCGCA} \\
\text{TCACTCGCCGCGGAGAGGAGAGAGTCTACGCGGACCGATAGCTGCGCGCAATAAGAGAAAA} \\
\text{CGATCGATCTGCTACAGCACGTCGCAAGGGCAGCACATCGCTCTAGGTGACGGCGAGGCGTGGTGGCGCTTGCACCGGGCGGCCCGG}
\]

217
Appendix 5: Sequence Data

Clone 282
Size of cloned \textit{M. tuberculosis} H37Rv DNA: 233bp
Nucleotides of \textit{M. tuberculosis} H37Rv genome cloned: 4046042bp-4046274bp
\textit{M. tuberculosis} genome base pair before \textit{lacZ} in pUS1800: 4046042bp
Restriction site used to digest genomic \textit{M. tuberculosis} H37Rv DNA for cloning in to pUS1800: \textit{Acil}
\begin{verbatim}
TGGACGTGCAAAATCGTACAGTCTTTTGGTAGCGGCGGTACCGGCACCGGCATGGGACGAGCAACCACAA
\end{verbatim}

Clone 289
Size of cloned \textit{M. tuberculosis} H37Rv DNA: 196bp
Nucleotides of \textit{M. tuberculosis} H37Rv genome cloned: 2698200bp-2698395bp
\textit{M. tuberculosis} genome base pair before \textit{lacZ} in pUS1800: 2698395bp
Restriction site used to digest genomic \textit{M. tuberculosis} H37Rv DNA for cloning in to pUS1800: \textit{Taql}
Intergenic to \textit{Rv3603c}
\begin{verbatim}
AAACCTGTGAGGCCGTCTGCTATGGAGCGGTTCGACGGTTTGCGTCCGGCCAGGCTCAAGGTGGGGTAC
\end{verbatim}

Clone 290
Size of cloned \textit{M. tuberculosis} H37Rv DNA: 252bp
Nucleotides of \textit{M. tuberculosis} H37Rv genome cloned: 1288947bp-1289198bp
\textit{M. tuberculosis} genome base pair before \textit{lacZ} in pUS1800: 1288947bp
Restriction site used to digest genomic \textit{M. tuberculosis} H37Rv DNA for cloning in to pUS1800: \textit{Acil}
Internal to \textit{Rv1161}
\begin{verbatim}
CCGCCAAGTACCCCGCGCCGCGGCTTTGATCGGGGACGTCTGCGGCCGACCATTTACGGGTCTTGTTG
\end{verbatim}

Clone 291
Size of cloned \textit{M. tuberculosis} H37Rv DNA: 303bp
Nucleotides of \textit{M. tuberculosis} H37Rv genome cloned: 528230bp-528532bp
\textit{M. tuberculosis} genome base pair before \textit{lacZ} in pUS1800: 528532bp
Restriction site used to digest genomic \textit{M. tuberculosis} H37Rv DNA for cloning in to pUS1800: \textit{Hpal}
\begin{verbatim}
TCGTTGCGGCTCATGGGCCAAATACTCCCGCGGATCGAGGGCGCACAGCGGACCGGCGTCAAGGAGGCAC
\end{verbatim}
Appendix 5: Sequence Data

Clone 292
Size of cloned *M. tuberculosis* H37Rv DNA: 494bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 417058bp-417551bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 417551bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*

Clone 293
Size of cloned *M. tuberculosis* H37Rv DNA: 214bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1901144bp-1901357bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 1901357bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*

Clone 295
Size of cloned *M. tuberculosis* H37Rv DNA: 485bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 3279691bp-3280175bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 3279691bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*

Clone 296
Size of cloned *M. tuberculosis* H37Rv DNA: 284bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1541494bp-1541777bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 1541777bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *HpaII*

219
Appendix 5: Sequence Data

Clone 302
Size of cloned *M. tuberculosis* H37Rv DNA: 346bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 2100809bp-2101154bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 2100809bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *HpaII*

\[\text{CGGAGATTCTCATGCTGTCCACGCCCAGGGCGTCCCGCCGCCACAGCAGTGTCGCGGCCG}\]

Clone 303
Size of cloned *M. tuberculosis* H37Rv DNA: 326bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 4306285bp-4306610bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 4306285bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *TaqI*
Internal to Rv2832c

\[\text{TCGATGTTGTCGCGCGTACCGCTAGGCGGCGCCGCAATGCAGCCAGGTCAGTCTTGTTGACTACCAACAAATCCGAATAGGTCACCCCCGG}\]

Clone 305
Size of cloned *M. tuberculosis* H37Rv DNA: 252bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 3423885bp-3424136bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 3423885bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *HpaII*
Internal to Rv3061c

\[\text{CCGGGTGGGTGATCTTGACGCTGATTCAGCACGGAACGCCTTGGCAGATTGAAAGATTCGTCGAGAAAGCGCTGCGCCAGCAGGAGATATGGTGCCAACTGTTCTCCGAACCTGACGCGGGGTCTGACGCGGCCTCGG}\]

Clone 311
Size of cloned *M. tuberculosis* H37Rv DNA: 326bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 4306285bp-4306610bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 4306285bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *TaqI*
Internal to Rv3832c

\[\text{TCGATGTTGTCGCGCGTACCGCTAGGCGGCGCCGCAATGCAGCCAGGTCAGTCTTGTTGACTACCAACAAATCCGAATAGGTCACCCCCGG}\]
Appendix 5: Sequence Data

Clone 320
Size of cloned *M. tuberculosis* H37Rv DNA: 230bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 549556bp–549785bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 549556bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *Taql*

TCGAAGTAGCGCCCATGGACCAGCGGACCGGACCCCCTGAGATTTTGGTACCAGGGATTGATAG

Rv0458

GACATCACGCGGGCAGGGAACCCGGGACGGGAAAGAGCAGCTCATGCTCTGGCTCTGCTGATAATTCA

Rv0457c

Clone 323
Size of cloned *M. tuberculosis* H37Rv DNA: 449bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 601327bp–601775bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 601327bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*
Internal to Rv0509

CCGCTTGGCACCGGGTAATTCCACTGAGTGGCGACCGCATCGACGCGTCTGCTGACGCTACCGAGAG

AGTTCCGCGACCGCCTCCTGCAGTACGGTCACGCACGGCCGGCCGACGCTGTTGATCCAGCGGCTAA

GGCCGCTGACACCTGGTGCAAGCTGCACCGAGCTGCTGTCGGCTGACCCACGCACGGAACTACCTGGCC

AGGCCGGCCAGCGGTTTGTCAAGCCGACCAATCTCCGCTTGGACACCTGGCCTGGCGGCGAACGGG

GCCGTACAGGTTGGGTACCTCGGGCTGTCATCGCCGGGCTGCCCACCAAGATGCTGCAACCTCGGAGCG

ATGTGGCGGCGGCAATCAGGTCAGCGCGCAGCGGGCGGATGTCCAGGTTCAATGTGGCAGCTACCGAC

Clone 325
Size of cloned *M. tuberculosis* H37Rv DNA: 503bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 1235973bp–1236475bp
*M. tuberculosis* genome base pair before lacZ in pUS1800: 1235973bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AciI*

CCGCCTGCTGGCCAGGGTAATTCCACTGAGTGGCGACCGCATCGACGCGTCTGCTGACGCTACCGAGAG

AGTTCCGCGACCGCCTCCTGCAGTACGGTCACGCACGGCCGGCCGACGCTGTTGATCCAGCGGCTAA

GGCCGCTGACACCTGGTGCAAGCTGCACCGAGCTGCTGTCGGCTGACCCACGCACGGAACTACCTGGCC

AGGCCGGCCAGCGGTTTGTCAAGCCGACCAATCTCCGCTTGGACACCTGGCCTGGCGGCGAACGGG

GCCGTACAGGTTGGGTACCTCGGGCTGTCATCGCCGGGCTGCCCACCAAGATGCTGCAACCTCGGAGCG

ATGTGGCGGCGGCAATCAGGTCAGCGCGCAGCGGGCGGATGTCCAGGTTCAATGTGGCAGCTACCGAC

221
Appendix 5: Sequence Data

Clone 330
Size of cloned *M. tuberculosis* H37Rv DNA: 597bp
Nucleotides of *M. tuberculosis* H37Rv genome cloned: 2614036bp-2614632bp
*M. tuberculosis* genome base pair before *lacZ* in pUS1800: 2614632bp
Restriction site used to digest genomic *M. tuberculosis* H37Rv DNA for cloning in to pUS1800: *AclI*

Rv2338c

```
CCGCACCCGGCTCGCATCCATCATCTCCACAGGACGGGCGCTCGATCGCTTGCGCCATTGCAATGTA
TAGATACCCGGGCGCGGCAGGGCGCACTGCGATTCAAACACGGTGTAACACGGTGATAGTGTAC
AGATGGGCTCTGATCAACCGTGAAACCCGGTTACGAGTACGCTGACGTGACGATTTACACAGGC
CGAGCATTCCCGCCCGGGCTACCGACATCTGGCCGCGCTCCCGCAACCGTGTCGCCAAGATTA
CACCCCTGGGTTTACCCAACTTCCGGGCATGCGGCAAGTGCAGCAGCAACAATGGCGTGCTGTA
AGGCCATTACGGGAGCGAGCGCTTACGGAGAAGCGGTACGATTCAATCTCCATATGAGCGGTGCGG
GCCGTTGTCAGGTATCGTTGAAACCCGCTGCGGATCGGGTGCCGGCAGGTTGGTCTTCTCTGTGAT
AGC

59bp to start of Rv2339
```
References


Buchriesser, C., Brosch, R., Bach, S., Guillyyoule, A. and Carniel, E. (1998) The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn tRNA* genes. Molecular Microbiology 30, 5965-978.


224


226


227


Quadroni, James, DaineseHatt, Kertesz and Ma (1999) Proteome mapping, mass spectrometric sequencing and reverse transcription-PCR for characterization of the sulfate starvation- induced response in Pseudomonas aeruginosa PA01. European Journal Of Biochemistry 266, 3986-996.


Wang, Jy, Ramphal, Jin and Sg (1996) Isolation and characterization of Pseudomonas aeruginosa genes inducible by respiratory mucus derived from cystic fibrosis patients. Molecular Microbiology 22, 51005-1012.


