The roles of selected sigma factor genes in

*Mycobacterium tuberculosis*

Bosco How Yeung Chan

A thesis submitted in fulfilment of the requirements of

University College London for the degree of

Master of Philosophy

Department of Infection

University College London

and

Division of Mycobacterial Research

National Institute for Medical Research

Mill Hill

London

August 2006
Declaration

I, Bosco How Yeung CHAN, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in this thesis.

Bosco How Yeung CHAN
Acknowledgements

I would like to express my gratitude to my supervisor Dr. Elaine Davis for her supervision and guidance during the course of this project, and for providing me with the opportunity to conduct the research that contributed towards this thesis.

I would also like to thank the Director of the NIMR, Sir John Skehel, and the Medical Research Council, for granting me special permission to study at the University College London while employed full time at the NIMR.

My sincere thanks to Dr. K G Papavinasasundaram for his continued guidance and support, in particular during the early part of my research, and to Dr. Calvin Boon for his great friendship and advice regarding my study. I would also like to thank Dr. Elena Curti for her constructive suggestions towards this thesis, and everyone at the Division of Mycobacterial Research for their friendship, advice and assistance.

I would like to acknowledge the BμG@S Group at St. George’s Hospital, London, for providing the technical support and materials that contributed towards the microarray study conducted for this thesis. I would also like to thank Dr. Roger Buxton for his advice regarding my microarray work.

Last but not least, I would like to convey my greatest appreciation to my parents and my family for their immeasurable support and encouragement throughout the years of my study. Finally, to my partner Abi, for her unconditional love, support and belief in me that has always been my source of strength throughout these years.
Abstract

*Mycobacterium tuberculosis* is a major cause of mortality worldwide and it has been estimated that a third of the world population is infected with *Mycobacterium tuberculosis*. Each year 8-10 million new active pulmonary cases arise and 2 million patients die of the disease. To improve treatment and prevention of disease it is important to gain further understanding of how this human pathogen adapts to and survives the conditions generated by host defence processes.

Gene regulation is important in enabling the pathogen to adapt to the various environmental conditions it encounters, and the use of sigma factors is one way to achieve this. In this study, phenotypic analyses of an existing *M. tuberculosis* mutant strain lacking *sigC* were undertaken. A *sigC*-complemented strain with a functional *sigC* gene is used to confirm that the phenotype of the *sigC* mutant strain is *sigC* specific. The role of $\sigma^C$ in global regulation of gene expression in *M. tuberculosis* is investigated by a comparative transcriptome study of the wild type and the $\Delta sigC$ mutant strains via microarray analysis. Furthermore, the transcriptional start site for *sigC* is identified via primer extension studies, providing insight into the regulative nature of $\sigma^C$ and the identification of the putative promoter sequence of $\sigma^C$.

In addition, *sigB*, which codes for the principal-sigma factor-like $\sigma^B$, is targeted for mutagenesis in *M. tuberculosis*. This process has been complicated by the discovery of the polycistronic nature of *sigB* with the downstream gene *ideR*, which codes for an essential regulator for iron metabolism in *M. tuberculosis*. This has led to a revision of the gene deletion strategy for *sigB* in *M. tuberculosis*. 
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title Page</td>
<td>1</td>
</tr>
<tr>
<td>Declaration</td>
<td>2</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>3</td>
</tr>
<tr>
<td>Abstract</td>
<td>4</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>5</td>
</tr>
<tr>
<td>List of Figures</td>
<td>9</td>
</tr>
<tr>
<td>List of Tables</td>
<td>12</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>13</td>
</tr>
</tbody>
</table>

## Chapter 1: Introduction

1.1: Tuberculosis – disease and development  
1.1.1: *Mycobacterium tuberculosis*: one of the world’s most successful pathogens  
1.1.2: Latent tuberculosis: the persistent and adaptive nature of *M. tuberculosis*  
1.1.3: Immune-pathological processes  
1.1.4: The genome sequence of *M. tuberculosis*  

1.2: Mycobacterial gene expression  
1.2.1: Regulation of bacterial transcription initiation  
1.2.2: Sigma factors in *M. tuberculosis*  

1.3: Project Aims  

17
Chapter 2: Materials and methods

2.1: Bacterial strains
2.2: Growth media
2.3: PCR amplification of DNA
2.4: Agarose gel electrophoresis
2.5: Restriction enzyme digestion of DNA, gel purification and ligation
2.6: Transformation of E. coli
2.7: Extraction of plasmids from E. coli
2.8: Preparation of electrocompetent mycobacterial cells
2.9: Electroporation of mycobacterial cells
2.10: Extraction of genomic DNA from mycobacterial cells
2.11: Isolation of RNA from mycobacterial cells
2.12: RT-PCR
2.13: Microarray hybridisation
2.14: Microarray analysis
2.15: Primer extension protocol
2.16: Stress studies in M. tuberculosis
2.17: M. tuberculosis mouse infection study
Chapter 3: Functional characterisation of an alternative sigma factor $\sigma^C$

3.1: *in vitro* phenotypic analysis of *M. tuberculosis* $\Delta\sigma^C$

3.1.1: Growth of *M. tuberculosis* $\Delta\sigma^C$ in aerobic conditions

3.1.2: Survival of *M. tuberculosis* $\Delta\sigma^C$ under various stress conditions in exponential growth phase

3.2: *in vivo* phenotypic analysis of *M. tuberculosis* $\Delta\sigma^C$

3.2.1: Verification of the $\sigma^C$-complemented strain $\Delta\sigma^C\text{comp}$

3.2.2: Growth of *M. tuberculosis* $\Delta\sigma^C$ in a mouse infection model

3.3: Transcription regulation of $\sigma^C$ in *M. tuberculosis*

3.3.1: Mapping the transcription start point of $\sigma^C$ in *M. tuberculosis*

3.3.2: Identifying the putative promoter motifs of $\sigma^C$ in *M. tuberculosis*

3.4: Global Gene expression in *M. tuberculosis* $\Delta\sigma^C$ during exponential growth phase

3.4.1: Microarray analysis of *M. tuberculosis* $\Delta\sigma^C$ during exponential growth phase compared to wild-type *M. tuberculosis*

3.4.2: Genes regulated by $\sigma^C$ detected during the microarray analysis

3.5: Discussion: defining the role of $\sigma^C$ in *M. tuberculosis*
Chapter 4: Molecular characterisation of the sigB locus; generation of the sigB gene replacement construct

4.1: Generation of M. tuberculosis ΔsigB::hyg replacement construct 96

4.2: Isolation of single crossover strains for the generation of ΔsigB::hyg 100

4.3: The selection for ΔsigB::hyg from the single recombinants 105

4.4: The influence of the downstream gene, ideR, on the generation of M. tuberculosis ΔsigB::hyg 109

4.5: Revised strategies for the generation of the M. tuberculosis ΔsigB replacement construct 117

4.6: Discussion: the generation of the sigB null mutant in M. tuberculosis 120

Chapter 5: Conclusion and future work 124

References 129

Appendix I: List of plasmids used in this study 148

Appendix II: List of primers used in this study 150

Appendix III: Media, solutions and gel 151
List of Figures

Figure 1.1.2.1: A schematic diagram of the ROI and RNI production in mammalian cells via phox (phagocyte oxidase) and iNOS (Nitric oxide synthase). Adapted from Nathan et al. (Nathan et al. 2000) 23

Figure 1.1.3.1: A schematic circular map of the chromosome of M. tuberculosis H37Rv. Adapted from Cole et al. (Cole et al. 1998) 27

Figure 3.1: (A) Schematic diagram of the sigC gene replacement construct. 410 bp were deleted from the sigC coding sequence. (B) Schematic diagram of the deletion of regions 2 and 4 (Δ region) of the M. tuberculosis σC. Adapted from the NCBI database (NCBI 2006) 58

Figure 3.1.1.1: The growth of wild-type M. tuberculosis H37Rv and ΔsigC strains in aerobic growth conditions. 60

Figure 3.1.2.1: Survival of the wild-type M. tuberculosis H37Rv and ΔsigC mutant strains under various stress conditions. 63

Figure 3.2: A schematic diagram illustrating the construction of the sigC-complemented construct, ΔsigCcomp. 64

Figure 3.2.1.1: RNA purification of the wild-type H37Rv, ΔsigC, and the ΔsigCcomp strains at exponential (log) and stationary (stat) growth phases. 65

Figure 3.2.1.2: A schematic diagram of the primer selection for the RT-PCR studies. 66

Figure 3.2.1.3: RT-PCR analysis of the expression of the M. tuberculosis sigC gene. 68

Figure 3.2.2.1: Growth of the wild-type H37Rv (WT), ΔsigC (sigC mutant) and ΔsigCcomp (sigC complement) strains in the lungs of BALB/c mice. 71
Figure 3.2.2.2: Growth of the wild-type H37Rv (WT), ΔsigC (sigC mutant) and
ΔsigCcomp (sigC complement) strains in the spleens of BALB/c mice.

Figure 3.3.1.1: Primer extension analysis of sigC in M. tuberculosis.

Figure 3.4.1.1: A sample scan image of a microarray slide with the wild-type M.
tuberculosis H37Rv RNA-derived cDNA isolated at the exponential growth
phase versus the wild-type M. tuberculosis H37Rv genomic DNA.

Figure 3.4.1.2: A sample scan image of a microarray slide with the M. tuberculosis
ΔsigC RNA-derived cDNA isolated at the exponential growth phase versus the
wild-type M. tuberculosis H37Rv genomic DNA.

Figure 3.5.1: A hypothetical model of the functional characteristics of σC in M.
tuberculosis.

Figure 4.1.1: A schematic diagram of the strategy previously used to attempt to delete
the M. tuberculosis sigB coding sequence.

Figure 4.1.2: A schematic diagram of the strategy adopted in this study in the attempt
to generate a sigB null mutant in M. tuberculosis.

Figure 4.1.3: Summary of the cloning strategy for the construction of the marked
sigB gene replacement construct.

Figure 4.2.1: A schematic diagram representing the selection process for allelic
exchange mutants in M. tuberculosis.

Figure 4.2.2: Schematic diagrams showing the PCR-based screening method for the
legitimate recombination of sigB.

Figure 4.2.3: Agarose gel showing the PCR product obtained using the genomic
DNA extracted from twelve blue single recombinants from the targeting
construct pBC2 as templates, using primers AS and AT.
Figure 4.2.4: Agarose gel showing the PCR products obtained using the genomic DNA extracted from twelve blue single recombinants from the targeting construct pBC2 as templates, using primers AU and AV.

Figure 4.3.1: A schematic diagram showing the PCR-based screening method for ΔsigB mutant.

Figure 4.3.2: Agarose gel showing the PCR products obtained using genomic DNA extracted from ten white colonies isolated during the first selection for ΔsigB::hyg as templates.

Figure 4.3.3: Agarose gel showing the PCR products obtained using genomic DNA extracted from nine white kanamycin sensitive colonies isolated from the second selection step as templates.

Figure 4.4.1: A schematic diagram of the sigB gene locus within the *M. tuberculosis* genome. Adapted from Tuberculist (Tuberculist 2006)

Figure 4.4.2: A schematic diagram showing the RT-PCR system used to investigate the expression of sigB and ideR within the *M. tuberculosis* genome.

Figure 4.4.3: RT-PCR analysis of the co-transcription of sigB and ideR in *M. tuberculosis*.

Figure 4.4.4: A schematic diagram showing the RT-PCR system used to investigate the co-transcription of Rv2709 and sigB within the *M. tuberculosis* genome.

Figure 4.4.5: RT-PCR analysis of the co-transcription of Rv2709 and sigB in *M. tuberculosis*.

Figure 4.5.1: A schematic diagram of the revised strategy for the deletion of sigB.

Figure 4.5.2: Summary of the cloning strategy for the construction of an in-frame sigB gene replacement construct.
List of Tables

Table 3.2.1.1: Sequences of primers used for the RT-PCR studies  66

Table 3.4.2.1: Table of 23 genes induced in the M. tuberculosis ΔsigC strains at the exponential growth phase, compared to their parental H37Rv strains  80

Table 3.4.2.2: Table of 9 genes repressed in the M. tuberculosis ΔsigC strains at the exponential growth phase, compared to their parental H37Rv strains  81

Table 4.1.1: Sequences of the primers used for the inverse PCR to make a targeting construct for the deletion of sigB (Gopaul 2002).  97

Table 4.2.1: Sequences of primers used for PCR-based screening methods for the legitimate recombination of sigB  103

Table 4.3.1: Sequences of primers used for PCR-based screening method for ΔsigB mutant  106

Table 4.4.1: Sequences of primers used for the RT-PCR system used to investigate the co-transcription of sigB and ideR.  111

Table 4.4.2: Sequences of primers used for RT-PCR system for the co-transcription of Rv2709 and sigB  115

Table 4.5.1: Sequences of primers used for the revised deletion strategy of sigB  117

Table 4.5.2: Sequences of primers used for the sequencing of the sigB coding region of pBC20  119
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMV</td>
<td>avian myeloblastosis virus</td>
</tr>
<tr>
<td>APS</td>
<td>ammonium persulphate</td>
</tr>
<tr>
<td>BCG</td>
<td>Bacille Calmette-Guérin</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td><em>Bacillus subtilis</em></td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CDS</td>
<td>coding sequence</td>
</tr>
<tr>
<td>CFU</td>
<td>colony forming units</td>
</tr>
<tr>
<td>CHMP</td>
<td>conserved hypothetical membrane protein</td>
</tr>
<tr>
<td>CHP</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>CTMP</td>
<td>conserved transmembrane protein</td>
</tr>
<tr>
<td>Cy</td>
<td>cyanine dye</td>
</tr>
<tr>
<td>dATP</td>
<td>2’-deoxyadenosine 5’triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>2’-deoxycytosine 5’triphosphate</td>
</tr>
<tr>
<td>ddH₂O</td>
<td>double-distilled water</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl Pyrocarbonate</td>
</tr>
<tr>
<td>dGTP</td>
<td>2’-deoxyguanosine 5’triphosphate</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle’s Medium</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulphoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>dNTP</td>
<td>2’-deoxynucleoside 5’triphosphate</td>
</tr>
</tbody>
</table>
DOTS  Directly Observed Therapy, Short-course

dsDNA  double stranded DNA

DTT  dithiothreitol

dTTP  2’-deoxythymidine 5’triphosphate

E. coli  Escherichia coli

EDTA  ethylenediaminetetraacetic acid

FCS  Foetal Calf Serum

g  grams/gravity

GmR  gentamycin resistance

H₂O₂  hydrogen peroxide

HCl  hydrochloric acid

HIV  Human Immunodeficiency Virus

HP  hypothetical protein

HygR  hygromycin resistance

KanR  kanamycin resistance

KanS  kanamycin sensitive

kb  kilobase

l  litre

L-agar  Luria-Bertani agar

L-broth  Luria-Bertani broth

log  logarithmic

M  molar

μ-  micro-

μm  micrometre

m-  milli-
MDR = multi-drug resistant
mol = mole
mRNA = messenger RNA
*M. bovis* = *Mycobacterium bovis*
*M. fortuitum* = *Mycobacterium fortuitum*
*M. leprae* = *Mycobacterium leprae*
*M. smegmatis* = *Mycobacterium smegmatis*
*M. tuberculosis* = *Mycobacterium tuberculosis*
n- = nano-
nm = nanometre
NaCl = sodium chloride
NaOH = sodium hydroxide
OADC = oleic acid, albumin, dextrose and catalase
OD$_{600}$ = optical density at 600nm wavelength
p- = pico-
PBS = phosphate buffered saline
PCR = polymerase chain reaction
RNA = ribonucleic acid
RNase = ribonuclease
RNI = reactive nitrogen intermediates
ROI = reactive oxygen intermediates
SDS = sodium dodecyl sulphate
SSC = salt and sodium citrate solution
Suc$^R$ = sucrose resistance
*S. aureus* = *Staphylococcus aureus*
| **S. pneumoniae** | *Streptococcus pneumoniae* |
| **S. coelicolor** | *Streptomyces coelicolor* |
| **S. griseus** | *Streptomyces griseus* |
| **S. hygroscopicus** | *Streptomyces hygroscopicus* |
| **TB** | tuberculosis |
| **TAE** | Tris (hydroxymethyl) aminomethane-acetic acid-ethylenediaminetetraacetic acid buffer |
| **TBE** | Tris (hydroxymethyl) aminomethane-boric acid-ethylenediaminetetraacetic acid buffer |
| **TE** | Tris (hydroxymethyl) aminomethane-ethylenediaminetetraacetic acid buffer |
| **TEMED** | N,N,N',N'-Tetramethylethylenediamine |
| **U** | unit |
| **UV** | ultra-violet light |
| **V** | volts |
| **v/v** | volume to volume ratio |
| **w/v** | weight to volume ratio |
| **WHO** | World Health Organisation |
| **x** | times |
| **X-Gal** | 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside |
| **°C** | degrees Celsius |
| **Δ** | null mutant (deletion) |
| **σ** | sigma |
CHAPTER 1: Introduction

1.1 Tuberculosis – disease and development

1.1.1 Mycobacterium tuberculosis: one of the world’s most successful pathogens

Tuberculosis is an infectious disease transmitted mainly by inhalation of tubercle bacilli. It is manifested by fever, night sweats and cough (if pulmonary) although infection can also spread to other parts of the body. Today, as it has been for centuries, tuberculosis continues to be one of the leading infectious causes of death in the world, accounting for an estimated 8 – 10 million new cases of active infection and 1.75 million deaths annually (WHO 2005). Furthermore, one-third of the world’s population is estimated to harbour a latent infection with Mycobacterium tuberculosis, the causative agent of tuberculosis. Cases of tuberculosis have been on the increase in recent years, largely due to HIV infection, immigration, increased trade and globalisation. If left untreated, each person with the active form of tuberculosis will infect between 10 to 15 people per year (WHO 2005).

The first breakthrough in combating tuberculosis was the development of the Bacille Calmette- Guérin (BCG) vaccine in the 1920s, which is a live attenuated M. bovis strain related to M. tuberculosis. M. bovis was passaged on culture medium for 13 years until it had reached a stage which had a significant decrease in virulence in animals (Calmette et al. 1929). The BCG vaccine had recorded a resounding success of reducing mortality rates by approximately 90% when administered to infants in France from 1921, leading to a recommendation by the League of Nations in 1928 for its use in the prevention of tuberculosis (Bloom et al. 1994). However, this level of efficacy has not been consistently reproduced worldwide (Bloom et al. 1994).
There are many factors that influence the efficacy of the BCG vaccine; for instance, as the BCG vaccine is given as a live vaccine, variations in the proportion of viable bacilli within the vaccine due to poor vaccine preparation, transportation and administration may contribute to the decreased potency of the vaccine. The nutritional status of the BCG-recipient population may also influence the efficacy, as well as pre-existing immunity acquired from exposure to environmental mycobacteria (Roche et al. 1995). A small number of cases of BCGosis have been reported among children who received BCG vaccine and were subsequently found to be HIV positive. Despite extensive use of BCG in Africa administered to HIV positive children there are few subsequent cases from this route, as there appears to be a window in time allowing safe vaccination of newborns before congenitally HIV-infected infants become so immunodeficient that BCGosis becomes a substantial risk (Bloom et al. 1994).

The tuberculosis chemotherapy undertaken today was evolved from numerous experimental and clinical studies conducted between the 1940s and 1970s (Fox et al. 1999). During that time limited success against active tuberculosis was achieved when a novel active tuberculosis treatment through a “triple therapy” was introduced; this consisted of the anti-tuberculosis agents isoniazid, streptomycin and para-aminosalicylic acid. Streptomycin treatment had to be discontinued after several months due to its significant toxicity to the recipient, though the intake of the remaining two antibiotics continued for a total course of 18 to 24 months (Crofton 1959). This combination chemotherapy was initially successful in combating tuberculosis, and over the next few decades improved drug combinations were tested, leading to shorter durations of therapy with the same efficacy (Zhang 2005).
However, poor compliance with drug intake by patients has led to the emergence of multi-drug resistant tuberculosis (Espinal 2003). The multi-drug resistant *M. tuberculosis* strains encountered today are resistant to isoniazid and rifampicin, as well as additional antimycobacterial agents (Espinal 2003). Subsequent action by the World Health Organisation (WHO) implemented an improved and effective tuberculosis treatment control through DOTS (Directly Observed Therapy, Short-course), a six-month chemotherapy regime consisting of an initial two-month phase of treatment with four antituberculosis drugs, isoniazid, rifampicin, pyrazinamide and ethambutol, followed by a continuation phase of treatment with isoniazid and rifampin for another four months. The responsibility for monitoring drug intake is undertaken by a health worker or trained person who is not a family member, thus shifting the responsibility for cure from the patients to the healthcare system. However, DOTS alone may not work in areas where there is a high incidence of MDR-TB (De Cock et al. 1999). Thus in such areas the World Health Organisation recommends the use of DOTS-Plus, which is DOTS plus second-line tuberculosis drugs, for example kanamycin, amikacin, capreomycin and cycloserine, for the treatment of MDR-TB. However, treatment of MDR-TB with DOTS-Plus may take up to two years, which is highly costly and would introduce significant toxicity to the recipient.

There have not been any new tuberculosis drugs developed in the last 40 years, and with the emergence of multi-drug resistant *M. tuberculosis* strains there is an urgent need for new drug development (Anonymous 2001). Further understanding of the molecular interactions within *M. tuberculosis* may provide potential answers to the functional characteristics and pathogenicity of the pathogen, thus contributing to the
identification of potential antimicrobial targets leading to the eradication of *M. tuberculosis*, one of the most successful pathogens in the world.

1.1.2 Latent tuberculosis: the persistent and adaptive nature of *M. tuberculosis*

*Mycobacterium tuberculosis* is capable of causing both an acute disease and an asymptomatic latent infection. Latent *M. tuberculosis* resides in the host cell and can persist for many years. The infected individuals carry a 2-23% chance of reactivation of the disease within a life-time (Parrish *et al.* 1998), with an average life-time risk of 10%. However, the risk can increase to 5-10% per year if the host is co-infected with HIV (Parrish *et al.* 1998). The antimycobacterial drugs available today are primarily targeted towards killing mycobacteria that are actively replicating, and hence additional emphasis on developing new drugs that are specifically targeted for the latent, persistent form of *M. tuberculosis* is essential for the eradication of the disease.

During the initial host infection *M. tuberculosis* and other microbial pathogens would encounter the resistance of the host immune system. A key cell involved in this defence system is the macrophage, which recognizes the foreign microbes and engulfs them into vacuoles called phagosomes. Subsequent to the phagocytosis, the cytokine interferon-γ (IFN-γ) activated macrophages makes life within the phagosome even more hostile by lowering the pH, fusing the phagosome with lysosomes, and stimulating the production of nitric oxide (McKinney *et al.* 2003). A key to the success of *M. tuberculosis* and related pathogenic mycobacteria lies in their ability to persist and replicate within the hostile environment of the host macrophage. *M. tuberculosis* interferes with phagosome maturation by blocking the fusion of nascent phagosomes with lysosomal compartments and by causing alterations in membrane
proteins that normally promote the formation of an acidic phagolysosome (Monack et al. 2004).

1.1.3 Immune – pathological processes

The mycobacterial-infected area can often be recognized as a granulomatous lesion or tubercle, which comprises a central area of infected cells surrounded by other, non-infected phagocytic cells and foamy giant cells (giant multinucleate macrophages loaded with liquid), with lymphocytes found at the periphery (Russell 2001). The lesion is sealed off and contained from the surrounding tissue by a fibrotic capsule to prevent further infection within the host, as well as to concentrate the immune response directly to the site of infection. However, the formation of the granuloma could also be advantageous to the mycobacteria, as the sealed lesion is screened from further bactericidal attacks from the host. If the infection is successfully contained, the granuloma shrinks and may eventually calcify. If however, the immune response does not successfully control the bacterial replication which may result from malnutrition and/or HIV co-infection, the granulomas will increase in size and cellularity. Eventually, cell death in the hypoxic centre of the granuloma leads to necrosis. If the granuloma is close to the surface of the lung, the tissue destruction caused by necrosis can breach the mucosal surface, a process referred to as cavitation, giving rise to the prototypic symptom of tuberculosis, a persistent cough with blood in the sputum. At this point the patient becomes highly infectious, spreading the bacteria by aerosol (Russell 2001).
Phagolysosome formation and the subsequent degradation by hydrolytic enzymes are not the only mechanisms that contribute to the destruction of the internalized microbes. Reduction of pH and the increase of local concentrations of reactive oxygen and nitrogen intermediates (ROI and RNI) that accompany lysosomal delivery also aid microbe destruction. The advantage of using ROI and RNI for defence is that these chemically reactive micromolecules do not discriminate the genomic source of their chemical targets, and their targeted key molecules usually confer essential chemical functions for the invading pathogens. Macrophages have the ability to produce superoxide (O$_2^-$) and nitric oxide (NO) in nearly equimolar amounts and thus can be prolific generators of the toxic peroxynitrite (OONO') (Zhu et al. 1992)(see Figure 1.1.2.1).

The general oxidative stress response in *M. tuberculosis* is the induction of a series of protective enzymes to confront the ROI, which include catalases, peroxidases, and superoxide dismutases (Dussurget et al. 1998). Hence instead of interfering with host cell production of ROI and RNI, the mycobacterial pathogens catabolize the intermediates or repair their damage from the host attack (Nathan et al. 2000). Surprisingly, specific adaptive responses in *M. tuberculosis* and other members of the *M. tuberculosis* complex against ROI have diminished during their course of evolution. In particular *M. tuberculosis* lacks a functional oxidative-stress regulatory gene oxyR (Deretic et al. 1997), which is involved in the response to hydrogen peroxide stress by activating the catalase-peroxidase KatG (Li et al. 1998), and is a central regulator for ROI responses in enteric bacteria. The presence of more efficient defence mechanisms may have replaced the conventional systems in pathogenic mycobacteria in response to the phagocyte oxidase (phox) and its subsequent reactive
Figure 1.1.2.1  A schematic diagram of the ROI and RNI production in mammalian cells via phox (phagocyte oxidase) and iNOS (Nitric oxide synthase). Adapted from Nathan et al. (Nathan et al. 2000)
oxygen intermediates that are induced within the host cells (Deretic et al. 1997; Garbe et al. 1996).

In macrophages, nitric oxide is generated by inducible nitric oxide synthase (iNOS), which is a cytosolic enzyme that catalyzes the conversion of L-arginine to L-citrulline and nitric oxide (James 1995). Nitric oxide, in the mildly acidic activated macrophage environment (pH~5.5), is extremely bactericidal (MacMicking et al. 1997). The pathogenic mycobacteria have evolved ways of countering the destructive effects of nitric oxide, including the DNA repair mechanism involving the uvrB gene (Darwin et al. 2005). Recently it has been suggested that the mycobacterial proteasome also serves as a defence against oxidative or nitrosative stress (Darwin et al. 2003; Darwin et al. 2005), although the precise nitric-oxide mediated bactericidal mechanisms and the mycobacterial adaptive response involving the proteasomes remain to be determined.

Indeed, the ability of *M. tuberculosis* to lie dormant and undetected by the host immune system requires considerable adaptation by *M. tuberculosis*, and hence latency can be viewed as an adaptive virulence mechanism against its host. There are many factors that lead to the latency adaptive response in *M. tuberculosis* – the oxidative, nitrosative and pH stresses mentioned above, anaerobiosis or hypoxia, and the entry into stationary growth phase due to nutrient limitation and/or cell-cell signalling. Knowledge of the transcriptional regulatory mechanisms in *M. tuberculosis* would provide a platform for the better understanding of how the pathogenic bacilli adapt and persist within the host at different phases of infection.
1.1.4 The genome sequence of *M. tuberculosis*

Since its original isolation in 1905, the H37Rv bacillus has been the model for the biomedical research in *M. tuberculosis* as it has retained virulence in animal models, unlike some clinical isolates; it is also susceptible to drugs and amenable to genetic manipulation. The complete genome sequence of the *M. tuberculosis* H37Rv strain was published in 1998 (Cole *et al.* 1998)(Figure 1.1.2.1), and its annotation was revised in 2002 (Camus *et al.* 2002). Additionally, the recently published genome sequence of *M. bovis*, the causative agent of tuberculosis in a range of animal species, has been shown to be more than 99.95% homologous to that of *M. tuberculosis*, but deletion of genetic information has led to a reduced genome size (Garnier *et al.* 2003). The *M. tuberculosis* H37Rv strain has around 4000 genes encoded in 4411 kb of DNA with a G+C content of 65.6%, with 52% of the proteins predicted from the sequence being assigned specific functions (TubercuList 2006). From the sequence analysis there are 376 proteins that share no homology with known proteins, and therefore may be unique to *M. tuberculosis* (Camus *et al.* 2002). Another unique feature identified from the H37Rv genome was that *M. tuberculosis* contains numerous examples of lipid and polyketide biosynthetic systems, with over 250 genes involved in fatty acid metabolism compared with around 50 in *E. coli*. This may explain the nature of the lipid-rich cell envelope that is observed in pathogenic mycobacteria. In summary, the additional genetic information acquired from the whole *M. tuberculosis* genome sequence have contributed to a refined focus for strategies designed to understand the functional characteristics of specific genes that may contribute to the pathogenicity of *M. tuberculosis*.
Additionally, knowledge of the complete gene set of *M. tuberculosis* H37Rv has also enabled global gene expression of *M. tuberculosis* to be studied using whole genome DNA microarray technology. Microarrays represent one of the new functional genomics technologies exploiting genome sequence information to address the physiology, host-pathogen interactions, mechanisms of drug action, *in vitro* and *in vivo* gene expression, host responses, comparative genomics, and functional analysis of particular genes in *M. tuberculosis* (Butcher 2004). An understanding of gene function can be obtained through observing the consequences of its deletion, thus providing further insight into possible interactions between genes with altered levels of expression and the gene of interest. Thus, the microarray technology is a novel and important tool for investigating the physiology and pathogenesis of mycobacteria.
Figure 1.1.3.1

A schematic circular map of the chromosome of *M. tuberculosis* H37Rv. The outer circle shows the scale in Mb, with 0 representing the origin of replication. The first ring from the exterior denotes the positions of stable RNA and the direct repeat region (pink cube); the second ring inwards shows the coding sequence by strand (clockwise, dark green; anticlockwise, light green); the third ring depicts repetitive DNA (insertion sequences, orange; 13E12 REP family, dark pink; prophage, blue); the fourth ring shows the positions of the PPE family members (green); the fifth ring shows the PE family members (purple, excluding PGRS); and the sixth ring shows the positions of the PGRS sequences (dark red). The histogram (centre) represents G + C content, with <65% G + C in yellow, and >65% G + C in red. Adapted from Cole et al. (Cole et al. 1998).
1.2 Mycobacterial gene expression

The *M. tuberculosis* genome encodes about 190 transcriptional regulators, including 13 RNA polymerase sigma factors, 11 two-component regulators, 14 protein kinases or phosphatases and over 140 other putative transcriptional regulators (Bishai 1998; Cole *et al.* 1998). Although the overall picture is still incomplete, these findings suggest very complex regulatory systems exist with overlapping functions and redundancies in *M. tuberculosis*. Several of these regulators have been characterized, with some of them responding to environmental stresses like heat shock, cold shock, hypoxia, iron starvation, surface stress and oxidative stress, while others respond to unknown environmental conditions (Manganelli *et al.* 2004). Such regulatory complexity is consistent with the intricate host infection process of *M. tuberculosis*.

1.2.1 Regulation of bacterial transcription initiation

During infection *M. tuberculosis* is exposed to many different environmental conditions depending upon the stage and severity of the disease. The ability of *M. tuberculosis* to adapt to different environments in the infected host suggests a major role for the regulation of gene expression. The regulation of genome expression involves a complex network of control systems operating at different stages of the expression pathways for different genes. Of all these regulatory systems, transcriptional control is the primary mechanism regulating gene expression in prokaryotes (Helmann 2002). Transcription initiation is the primary step of gene expression, as it controls the level of regulation that determines which genes are expressed and the rates of expression of those genes (Ebright 2000). Later steps in the
pathway may respond to secondary regulation in the form of translational regulation, which involves modulation of the amount of protein that is synthesized, or post-translational modifications whereby the nature of the protein is altered leading to a change in functionality, for example, by phosphorylation or protein-protein interactions, which can be exploited to control functions such as the reversible antagonistic mechanism of an anti-σ factor on its cognate σ factor.

Transcription is mediated by the DNA-dependent RNA polymerase (RNAP), which catalyses RNA synthesis. The bacterial RNAP core enzyme consists of five subunits (α2, β, β', and ω) (Minakhin et al. 2001), and is fully capable of elongation and termination of transcription, but is incapable of promoter recognition and specific transcription initiation. Promoter recognition and the promoter melting to form an open complex for RNA synthesis are dependent upon the binding of a σ factor to the core RNAP to form the RNAP holoenzyme.

RNAP is conserved in all living organisms throughout evolution (Ebright 2000). The bacterial RNAP, archaeal RNAP, and eukaryotic RNAP I, II, and III belong to a conserved multi-subunit RNAP family. The bacterial RNAP is the smallest member of this protein family, and has been extensively studied in *E. coli* (Ebright 2000). The bacterial core enzyme has a dual cleft structure and is defined by the assembly of the α'β and α''β' subunits (Zhang et al. 1999).

The two α subunits, α' and α'', are encoded by the gene *rpoA*, and are involved in RNAP assembly. Each α subunit consists of two domains; an N-terminal domain (αNTD), which is responsible for the interactions with the β and β' subunits, and the
C-terminal domain (αCTD), which is responsible for sequence specific protein-DNA interactions with the upstream promoter DNA and protein-protein interactions with upstream bound transcriptional regulators when the RNAP is in the holoenzyme conformation. The αCTD is connected to αNTD through a long flexible linker, which allows the αCTD to bind to the promoter region (Busby et al. 1994).

The β and β' subunits, encoded by the rpoB and rpoC genes respectively, represent the two largest subunits in the RNAP, and are involved in catalysis (Minakhin et al. 2001). In vitro deletion studies of the β' subunit have resulted in the inability of the RNAP core enzyme to bind to sigma factor (Luo et al. 1996). Further analysis indicated that the sigma-binding site is located at region 260-309 within the β' subunit (Burgess et al. 2001; Young et al. 2001). Rifampin, a powerful antimicrobial drug, has the ability to inhibit transcription in bacterial RNA polymerase by binding to the β subunit of the enzyme (McClure et al. 1978), and has been used against *M. tuberculosis* infection since the 1970s (Musser 1995). The crystal structure of *Thermus aquaticus* core RNAP complexed with Rifampin has been identified, demonstrating that Rifampin binds to a pocket of the RNAP β subunit deep within the DNA/RNA channel, but more than 12 Å away from the active site (Campbell et al. 2001). The structure, combined with biochemical results, explains the effects of Rifampin on RNAP function and indicates that the inhibitor acts by directly blocking the path of the elongating RNA when the transcript becomes 2 to 3 nucleotides in length. Rifampin-resistant *M. tuberculosis* isolates were found to have missense mutations in a defined 81-bp region in the rpoB gene (Telenti et al. 1993).
The ω subunit, encoded by the gene *rpoZ*, is present in all sequenced bacterial genomes (Minakhin *et al.* 2001). It promotes RNAP assembly and/or increases RNAP stability by latching the N- and C-terminal regions of the β' subunit together thereby facilitating its interaction with the β and α subunits (Ghosh *et al.* 2001; Ghosh *et al.* 2003; Mathew *et al.* 2005).

The dissociable sixth subunit σ (Burgess *et al.* 1969) interacts transiently with the core enzyme to form the RNAP holoenzyme and directs the binding of RNAP to specific promoter sequences, thus allowing specific initiation of transcription. If the σ subunit is absent the core enzyme is still able to bind to DNA, but in a more random fashion at a variety of sites and not specifically to promoter sequences. In the absence of the RNAP core enzyme, σ factors are unable to recognize promoter DNA in either double- or single-stranded form (Dombroski *et al.* 1992). During this inactive state the DNA-binding function of the σ subunit may possibly be autoinhibited by its own N-terminal region (region 1.1) with the conserved region 4.2, which binds to the -35 promoter element (Camarero *et al.* 2002). Upon association with core RNAP, many bacterial σ factors are believed to convert from an autoinhibited state, which cannot interact with promoter DNA, to an active form that can bind to the promoter DNA (Dombroski *et al.* 1992).

In *E. coli*, of the seven sigma factors identified within its genome, σ70 is the major housekeeping sigma factor. Promoter binding is determined by direct contact of the sigma factor and the DNA sequences that, in the case of σ70, are hexamers with consensus sequences TTGACA and TATAAT centred at the -35 and -10 regions upstream of the transcriptional start site respectively (Rosenberg *et al.* 1979). These
hexamers contact key amino acids in a helical structure and a helix-turn-helix motif, present in two regions of the σ factor that have been termed 2.4 and 4.2 respectively (Lonetto et al. 1992).

Most bacteria produce a principal sigma factor and a group of alternative sigma factors which recognise different promoter motifs. The ability to induce or activate alternative sigma factors that can replace the principal sigma factor and change RNAP promoter specificity is a mechanism that allows the expression of different sets of genes, named regulons, in response to various physiological and environmental stimuli.

Bacteria produce two major families of sigma factors: the σ70 and σ54 families (Wosten 1998). Although sigma factors from both families bind to the same core polymerase to initiate transcription, there is no sequence homology between the two families (Merrick 1993). The members of the σ54 family are non-essential, and are not present in mycobacteria. The σ70 family can be divided into three major groups, depending upon their structure and function: principal σ factors, principal-like σ factors and alternative σ factors (Wosten 1998).

The principal sigma factors are usually essential and allow the transcription of housekeeping genes. *E. coli* σ70, *B. subtilis* σ^A^ and *S. coelicolor* σ^{HdB} belong to this category. In *M. smegmatis* and *M. tuberculosis*, this group is represented by σ^A^ (Doukhan et al. 1995; Gomez et al. 1998).
The principal-like σ factors are closely related to the principal σ factors, but are dispensable for bacterial growth. The possible role for principal-like σ factors may include the regulation of expression of a precise set of genes in response to environmental conditions, while maintaining expression of most housekeeping functions normally dependent on the principal σ factor (Helmann 2002). The σ factors in this group include the E. coli σS protein, and three σHrd proteins (σHrdA, σHrdC and σHrdD) in S. coelicolor (Buttner et al. 1992). It is represented by σB in M. smegmatis (Mukherjee et al. 2005) and in M. tuberculosis (Hu et al. 1999).

The third group of the alternative σ factors is the most heterogeneous and can be divided into at least four subgroups, responsible for regulating the heat shock response, sporulation, motility and extracytoplasmic functions (ECFs) (Lonetto et al. 1994); indeed there are incidents of bacterial virulence determinants whose expression is regulated by an alternative sigma factor (Finlay et al. 1997). ECF σ factors are environmental stimuli regulators, controlling a variety of functions in response to specific extracellular signals, such as the presence of misfolded proteins in the periplasmic space, cell wall adaptations and the presence of toxic molecules in the external environment (Missiakas et al. 1998). Examples of some of the σ factors represented in the ECF σ factor subgroup are the E. coli σE, which controls the response to periplasmic stress; P. aeruginosa AlgU, which controls alginate biosynthesis; E. coli FeCl, which controls iron uptake; and B. subtilis σX, which is involved in the stress response (Missiakas et al. 1998). In M. tuberculosis, 10 ECF σ factors have been identified – σC-E, σG-M (Cole et al. 1998; Manganelli et al. 1999). The specific roles of the mycobacterial sigma factors will be discussed further in the following section.
In addition to the $\sigma$ factors, regulation of transcription can be further controlled by transcriptional activators and repressors (Ishihama 2000; Vicente et al. 1999), sigma-binding antisigmas (Helmann 1999; Hughes et al. 1998), and even small RNAs (Wassarman et al. 2000). Studies investigating the transcriptional circuitry of \textit{M. tuberculosis} would offer the potential of deciphering the pathogen’s strategies for its adaptive capabilities within the host.

1.2.2 \textbf{Sigma factors in \textit{M. tuberculosis}}

Within the \textit{M. tuberculosis} genome, 13 sigma factors have been identified, of which two are principal and principal-like sigma factors ($\sigma^A$ and $\sigma^B$), one is the stress and sporulation type sigma factor ($\sigma^F$) and the remaining 10 are ECF sigma factors ($\sigma^{C-E}$ and $\sigma^{G-M}$) (Cole et al. 1998).

$\sigma^A$ is considered to be the principal sigma factor of \textit{M. tuberculosis} because of its close resemblance to other principal sigma factors of the $\sigma^{70}$ family (Predich et al. 1995), and is found to be essential in \textit{M. smegmatis} (Gomez et al. 1998). The \textit{sigA} promoter sequences are highly conserved in \textit{M. smegmatis}, \textit{M. tuberculosis} and \textit{M. leprae} (Gomez et al. 1998), and the promoters resemble a consensus for mycobacterial promoters that are expressed under normal growth conditions (Bashyam et al. 1996). Although $\sigma^A$ is responsible for the expression of essential genes, it is also involved in virulence; a missense mutation in the \textit{M. bovis} $\sigma^A$ resulted in significant attenuation of the strain, although it did not affect growth \textit{in vitro} (Collins et al. 1995). However, it was noted that this mutation occurred within the
region of $\sigma^A$ that that was involved in the interaction between $\sigma$ factors and transcription activators (Baldus et al. 1995), and hence the attenuation of the strain may have resulted from the defective expression of a set of virulence genes dependent upon a particular transcriptional activator. Complementation of the mutant strain with a wild type copy of the gene restored its virulence, fulfilling Koch’s postulate and indicating that $\sigma^A$ is a virulence factor (Collins et al. 1995).

The gene encoding $\sigma^B$, $\text{sig}B$, is located about 3 kb downstream of $\text{sig}A$ (TuberculList 2006). This genomic organisation is conserved in all mycobacterial species (Doukhan et al. 1995; Predich et al. 1995). However, unlike $\sigma^A$, it is non-essential for normal growth in $M. \text{smegmatis}$ (Mukherjee et al. 2005). The sequence homology of $\text{sig}B$ to the principal sigma factor $\text{sig}A$ and the dispensable nature of $\text{sig}B$ have categorized it into the principal sigma factor-like group of sigma factors. The $\text{sig}B$ mutant in $M. \text{smegmatis}$ was found to be more sensitive to $\text{H}_2\text{O}_2$ and superoxide generating compounds (Mukherjee et al. 2005). Moreover, the amount of $\text{sig}B$ transcript in $M. \text{tuberculosis}$ were found to be increased upon exposure to high temperature, SDS, low aeration and during stationary phase (Hu et al. 1999; Manganelli et al. 1999). These findings suggest that $\text{sig}B$ may be involved in the regulation of specific environmental conditions, such as stationary growth. The transcriptional start site and the putative promoter regions of $\text{sig}B$ have been mapped upstream of the $M. \text{tuberculosis}$ $\text{sig}B$ coding sequence (Gomez et al. 2000; Hu et al. 1999), although the results were somewhat varied, and have to be interpreted with care. Additionally, overexpression of $\sigma^B$ in $M. \text{smegmatis}$ and $M. \text{bovis}$ BCG were found to cause an increase in $\text{katG}$ expression, which is believed to be involved in the mycobacterial defence against ROI and RNI within host macrophages, although the precise regulation of $\text{katG}$ by $\sigma^B$ is
not known (Mulder et al. 1999). In summary, the regulation of \textit{sigB} in \textit{M. tuberculosis} is complex; its transcription under physiological conditions and its induction after surface stress are dependent upon $\sigma^E$ (Manganelli et al. 2002), while during heat shock and oxidative stress, its induction is dependent upon $\sigma^H$ (Raman et al. 2001). Therefore $\sigma^B$ could function as a secondary regulatory device to maintain the transcription of essential housekeeping genes during stress conditions, when $\sigma^A$ could be inactive or its levels could be lowered. The role of $\sigma^B$ will be further understood once more of the genes that are regulated by $\sigma^B$ have been identified.

The gene $\textit{sigF}$, which encodes $\sigma^F$, appears to occur only in slow-growing pathogenic mycobacteria (DeMaio et al. 1997). It is the only alternative $\sigma$ factor in \textit{M. tuberculosis} that does not belong to the ECF sigma factor subgroup. \textit{M. tuberculosis} $\sigma^F$ is homologous to many sigma factors from other bacteria. Among these are sporulation specific sigma factors in \textit{S. coelicolor} ($\sigma^F$) and in \textit{B. subtilis} ($\sigma^F$), stress response sigma factors in \textit{B. subtilis} ($\sigma^B$) and \textit{S. aureus} ($\sigma^B$), and the stationary phase sigma factor $\sigma^S$ in \textit{E. coli} (DeMaio et al. 1997; Gomez et al. 1997). The \textit{M. tuberculosis} $\textit{sigF}$ is part of a genetic cluster with an organisation similar to that of the \textit{B. subtilis} $\textit{sigB}$ and $\textit{sigF}$ operons (DeMaio et al. 1997), with the putative anti-sigma factor encoding gene \textit{rsbW} located directly upstream of $\textit{sigF}$, and these two genes are believed to be polycistrally transcribed (DeMaio et al. 1997). The discovery of two anti-$\sigma$ factor antagonists (RsfA and RsfB) that can regulate the activity of the $\sigma^F$-specific anti-$\sigma$ factor RsbW suggests a certain level of functional redundancy of the regulatory system in \textit{M. tuberculosis} (Beauchet et al. 2002). A mutant lacking the $\textit{sigF}$ gene was generated in \textit{M. tuberculosis} (Chen et al. 2000), which was not
defective for *in vitro* growth in culture or in human macrophages, but was more sensitive to exposure to rifampin and did not exhibit the same level of virulence as the wild-type by time of death analysis (Chen et al. 2000). Further studies have revealed that the ΔsigF mutant strain can persist at a higher bacillary load in mouse lungs when compared with the wild-type and the complemented strains, but was delayed in causing lethality in mice (Geiman et al. 2004). Additionally, Michele et al. observed a significant increase in sigF expression in late stationary phase (Michele et al. 1999). All of these findings suggest that *M. tuberculosis* σ^F^ is involved in the regulation of late stages of infection.

Of the 10 ECF sigma factors in *M. tuberculosis*, the roles of σ^E^ and σ^H^ are best understood. Transcription of σ^E^ occurs during exponential growth but increases after exposure to stresses such as heat, detergent, low aeration, and entry into stationary phase, with a similar expression pattern to sigB (Manganelli et al. 1999). A mutant of *M. tuberculosis* lacking sigE was shown to more sensitive to high temperature, detergent and oxidative stress (Manganelli et al. 2001). In addition, microarray results showed significant levels of reduction in sigB transcription in the ΔsigE mutant under both standard physiological conditions and with exposure to SDS (Manganelli et al. 2001), thus suggesting a possible hierarchy in σ regulation. Moreover, the sigE mutant also showed impaired ability to grow within macrophages and was more sensitive to killing by activated macrophages than the wild type (Manganelli et al. 2001). Other studies have revealed that the ΔsigE mutant had demonstrated a similar histopathological pattern to the ΔsigF mutant in aerosol infected mice, with a higher bacillary load in the mouse lungs when compared with the wild-type and the complemented strains, and was partially attenuated in time of death studies (Ando et
al. 2003). The ΔsigE mutant was also found to be attenuated in intravenously infected mice (Manganelli et al. 2004). The results from these studies showed that σ^E plays an important role in the pathogenesis of *M. tuberculosis*. Furthermore, downstream of the genetic locus of *sigE* is *Rv1222 (rseA)*, which is a putative anti-σ factor for σ^E, as it is significantly homologous to the RsrA anti-σ factor protein found in *S. coelicolor* (Kang et al. 1999), suggesting possible mechanisms for the posttranslational regulation of σ^E.

*M. tuberculosis* σ^H has significant sequence homology to the ECF σ factor σ^R in *S. coelicolor*, which responds to intracellular formation of disulfide bonds due to oxidative stress (Paget et al. 1998). σ^R activity is regulated at the posttranslational level by a cysteine rich zinc-binding anti-σ factor RsrA (Paget et al. 2001), which contains a HXXXXCXXC motif and belongs to the ZAS (zinc binding anti-sigma factor) subfamily. The deletion of *rsrA* resulted in high constitutive levels of σ^R dependent transcription, confirming the negative regulatory nature of RsrA on σ^R. *In vitro*, purified *S. coelicolor* thioredoxin can reduce oxidized RsrA, thus creating a homeostatic loop in which the σ^R regulon is regulated in response to the changes in the cellular thiol-disulphide redox status (Kang et al. 1999). The *M. tuberculosis* σ^H gene is induced after heat shock, and also after exposure to thiol-specific oxidative stress (Raman et al. 2001). Deletion of *sigH* in *M. tuberculosis* has been achieved (Kaushal et al. 2002; Manganelli et al. 2002; Raman et al. 2001), producing mutants that are more sensitive to heat shock and oxidative stress. The *sigH* mutant strain also showed a similar histopathological pattern in the lung of the mouse in infection studies to that which was observed with *sigE* and *sigF* mutants (Kaushal et al. 2002).
Expression of $\sigma^H$ seems to be autoregulated, since transcription from the $\text{sig}H$ promoter fails to be induced in the $M. \text{smegmatis} \text{sig}H$ mutant (Fernandes et al. 1999). Additionally, RshA, a RsrA-like anti-$\sigma$ factor is located immediately downstream of $\text{sig}H$ in $M. \text{tuberculosis}$, reminiscing the antagonistic nature of RsrA on $\sigma^R$ in $S. \text{coelicolor}$ (Song et al. 2003). Overall, these data suggest that $\sigma^H$ plays a central role in a network that regulates both heat and oxidative stress responses in $M. \text{tuberculosis}$.

Upon commencing the research work conducted for this thesis, little information was available for the remaining 8 ECF $\sigma$ factors encoded by the $M. \text{tuberculosis}$ genome. However, recent publications have provided limited information into the roles of $\sigma^C$, $\sigma^D$, $\sigma^I$ and $\sigma^L$. As for $\sigma^G$, $\sigma^l$ and $\sigma^M$, RT-PCR studies conducted by Manganelli et al. showed that the expression of $\text{sig}G$ is absent in stationary phase, but elevated levels of $\text{sig}I$ and $\text{sig}M$ expressions were observed upon exposure to moderate cold shock and heat shock respectively (Manganelli et al. 1999). Currently there is no information available about the role of $\sigma^K$ in $M. \text{tuberculosis}$.

For $M. \text{tuberculosis} \sigma^C$, Sun et al. also successfully generated a mutant lacking $\text{sig}C$ (Sun et al. 2004). Their time of death studies in mice provided valuable additional information about the intracellular behaviour of $\sigma^C$. In their aerosol infection time of death study, their $\Delta \text{sig}C$ mutant caused no death during the 235 days test period, while all the mice infected with the wild-type and their $\text{sig}C$-complemented strain died within the test period, with a median time-to-death of 170 days. In their intravenous infection time of death study, no mice infected with the $\Delta \text{sig}C$ mutant died over the first 300 days, while all the mice infected with the wild-type and all but one of the mice infected with the $\text{sig}C$-complemented strains had died by day 107. Therefore
their data indicate that their \(\Delta\text{sigC}\) mutant strain was able to persist in mice but was attenuated in its ability to elicit lethal immunopathology (Sun et al. 2004).

*M. tuberculosis* \(\sigma^D\) mutants were successfully generated by allelic exchange (Calamita et al. 2005; Raman et al. 2004). The \(\sigma^D\) mutants showed similar levels of reduced virulence in mouse infection models along with the histopathology results as observed in \(\sigma^E\), \(\sigma^F\) and \(\sigma^H\). It has been postulated that \(\sigma^D\) may play a role in the regulation of genes encoding ribosomal proteins during stationary growth phase (Calamita et al. 2005).

The *M. tuberculosis* mutant strain lacking sigJ has been shown to be more sensitive to hydrogen peroxide stress (Hu et al. 2004), and there are indications of a possible role of sigJ in the prolonged stationary phase (Hu et al. 2001), although these results have yet to be confirmed.

Recently a \(\sigma^L\) mutant was generated in *M. tuberculosis* and *M. smegmatis* (Hahn et al. 2005). \(\sigma^L\) is highly homologous to \(\sigma^E\) in *S. coelicolor*, which controls the structure of the cell wall. *Rv0736 (rslA)*, which encodes a membrane protein, is located downstream of sigL in the *M. tuberculosis* genome and is co-transcribed with sigL; its gene product is found to be an anti \(\sigma\)-factor that controls the regulation of \(\sigma^L\). The sigL mutant was found to be attenuated in the time of death study in mice, suggesting its role in virulence (Hahn et al. 2005). Through *in vitro* transcription and primer extension studies Hahn *et al.* demonstrated a \(\sigma^L\)-dependent transcript was initiated from the sigB promoter, although \(\sigma^L\) does not seem to have a role in sigB expression in response to oxidative and nitrosative stress. These results suggest a possible
alternate regulatory cascade mechanism between $\sigma^B$ and $\sigma^L$ amongst *M. tuberculosis* sigma factors.

1.3 Project Aims

The overall aim of this study is to investigate the roles of two sigma factors, $\sigma^B$ and $\sigma^C$, in *M. tuberculosis*.

The role of $\sigma^C$ in the survival and pathogenesis of *M. tuberculosis* is examined via phenotype analyses of an existing *M. tuberculosis* mutant strain lacking *sigC* made within the department. A *sigC*-complemented strain with a functional *sigC* gene is used to confirm that the phenotype of the *sigC* mutant strain is *sigC* specific. The role of $\sigma^C$ in global regulation of gene expression in *M. tuberculosis* is investigated by a comparative transcriptome study of the wild type and the $\Delta$*sigC* mutant strains via microarray analysis. Finally, the transcriptional start site for *sigC* is identified via primer extension studies, providing insight into the regulative nature of $\sigma^C$ and the identification of the putative promoter sequence of $\sigma^C$.

To address the role of $\sigma^B$, the *sigB* gene is targeted for mutagenesis in *M. tuberculosis* via homologous recombination. Co-transcription of *sigB* with the essential gene *ideR* is established and a revised strategy is employed.
CHAPTER 2: Materials and methods

2.1 Bacterial strains

In this study the following bacteria strains were used. For the purposes of general cloning and production of plasmids the *Escherichia coli* DH5α (Invitrogen, Strathclyde, UK) strain was used (Sambrook *et al.* 1989). In the case of the mycobacterial work carried out in this study, an isolate of the sequenced *Mycobacterium tuberculosis* strain H37Rv was obtained from the Institut Pasteur (Cole *et al.* 1998).

2.2 Growth media

*E. coli* DH5α (Invitrogen, Strathclyde, UK) was grown in Luria-Bertani broth (L-broth) (Sambrook *et al.* 1989) for liquid culture at 37°C in a shaking incubator (Bellco, New Jersey, USA) with gentle agitation. The cultures were grown in conical flasks of a volume 5 times that of the amount of broth used. In the case of solid media, Luria-Bertani agar (L-agar) was used. For antibiotic selection in *E. coli* one of the following compounds, as appropriate, was added at the prescribed concentration to the growth media: kanamycin (50μg/ml), hygromycin (250μg/ml), gentamycin (20μg/ml), or ampicillin (100μg/ml). For blue/white selection 100μg/ml of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-GAL) was added to the media.

*M. tuberculosis* H37Rv liquid culture was grown in Modified Dubos Broth (Difco, Plymouth, UK), supplemented with 0.2% glycerol (v/v) and 4% Dubos Medium
Albumin (v/v, Difco, Plymouth, UK). For standing cultures, bacteria were inoculated into 10 ml of medium in a 50 ml square based bottle (Nalgene, Fisher, Loughborough, UK) and leave at 37°C for 2-3 weeks. For rolling cultures, bacteria were sub-cultured into 100 ml of medium in a 500 ml screw top roller bottle (Nalgene, Fisher, Loughborough, UK), and incubated in a roller incubator (Bellco, New Jersey, USA) at 37°C with gentle rolling. In the case of solid media for *M. tuberculosis*, 7H11 agar (Difco, Plymouth, UK) supplemented with 4% OADC (oleic acid, albumin, dextrose and catalase)(v/v)(Becton Dickinson, Plymouth, UK) was used. Antibiotic selection was done by the addition of one of the following compounds, as appropriate, at the prescribed concentration: kanamycin (25 μg/ml), hygromycin (50 μg/ml) or gentamycin (15 μg/ml). For blue/white selection, 100 μg/ml of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-GAL) was added, and for sucrose selection, 2% sucrose (v/v) was added.

2.3 PCR amplification of DNA

10-100 ng of genomic DNA of *M. tuberculosis* was used in 50μl of reactions. A standard PCR reaction consist of: 12.5 μM dNTPs (Amersham Pharmacia Biotech, Amersham, UK), 0.5 μM each of forward and reverse primer (Eurogentec, Seraing, Belgium), 1 x concentration of *Pfu* Ultra buffer (containing 2mM MgCl2), 2.5 U *Pfu* Ultra (Stratagene, La Jolla, USA), DMSO (5%, Sigma, Shaftesbury, UK). The standard PCR cycling conditions used was: 5 min at 95°C, then 35 cycles of (15 seconds at 95°C, 30 seconds at 60°C (varied optimal temperatures for primer pairs), 1 min/kb at 72°C for PCR products <3kb or 1 min/kb at 68°C for PCR products >3kb),
and finally 5 minutes at 72°C (for PCR products <3kb), or 68°C (for PCR products >3kb). For storage, the PCR products were kept at 4°C.

2.4 Agarose gel electrophoresis

Plasmid DNA or DNA fragments generated by PCR or restriction digestion were routinely separated by horizontal agarose gel electrophoresis. Agarose gels of either 0.8% (DNA fragments >1kb) or 1% (DNA fragments <1kb) were prepared by melting molecular biology grade agarose (Bio-Rad, Hemel Hempstead, UK) in 1 x TAE buffer (for composition see appendix II). Ethidium bromide (Bio-Rad, Hemel Hempstead, UK) was then added to a final concentration of 1 μg/ml. After allowing the gel to set it was transferred to a gel tank, and the DNA was electrophoresed either in a mini gel tank at 80 V constant or midi tank at 120 V constant. DNA fragments were visualised under an ultraviolet transilluminator at 302 nm (UVP, Esslab, Benfleet, UK), and the gel images were taken using a video graphic printer (Sony, Thatcham, UK).

2.5 Restriction enzyme digestion of DNA, gel purification and ligation

Restriction nuclease digestion was used to cleave DNA at specific sites as an analytical tool or for cloning purposes. DNA digestions were performed using restriction enzymes (NEB, Hitchin, UK/Roche, Lewes, UK) at a concentration of 1U/μl in the appropriate restriction buffer according to the manufacturer recommendation. The reactions were supplemented with 1 x BSA (NEB, Hitchin,
UK) where needed, and incubated at 37°C for a minimum of 2 hours. Following digestion, the fragments were separated by electrophoresis though an agarose gel (see 2.4). When required for subsequent cloning steps, the desired band was cut out using a clean scalpel under UV illumination. The DNA was then extracted and purified from the gel piece using the Gel extraction kit (Qiagen, Crawley, UK), and eluted either in elution buffer (Qiagen, Crawley, UK) or in 1 x dilution buffer (Roche, Lewes, UK) for subsequent ligation.

For the ligation step, DNA fragments were ligated into restriction digested and hence linearized vector plasmid DNA using a Rapid DNA Ligation kit (Roche, Lewes, UK) according to the manufacturer instructions. To avoid self-religation of the digested and linearized vector plasmid, it was treated with 1μl (1U) of calf alkaline phosphatase (Roche, Lewes, UK) was added for the last 30 minutes of the incubation period.

2.6 Transformation of *E. coli*

Ligated reactions were transformed into subcloning efficiency or library efficiency *E. coli* DH5α competent cells (Invitrogen, Strathclyde, UK) as previously described (Sambrook *et al.* 1989). The ratio of DNA to competent cells was 1:20, and the mixture was incubated on ice for 25 minutes. The mixture was then heat shocked at 37°C for 90 seconds for subcloning efficiency cells, or at 42°C for 30 seconds for the library efficiency cells. The mixture was then cooled on ice for 5 minutes, and transferred into 1 ml of S.O.C medium (Invitrogen, Strathclyde, UK). After incubating with agitation (Belco, New Jersey, USA) for 1 hour at 37°C, the cells
were harvested at 15000xg (Philip Harris, Shaftesbury, UK) for 1 minute, then 900 μl of supernatant was removed, and the cell pellet was resuspended in the remaining solution and plated on the appropriate antibiotic selection L-agar plate, and incubated at 37°C (Bellco, New Jersey, USA) overnight.

2.7 Extraction of plasmids from E. coli

Plasmid DNA was extracted from 10 ml cultures of E. coli grown overnight using a miniprep spin column kit (Qiagen, Crawley, UK) according to the manufacturer instructions. When larger quantity of DNA was required a 100ml of culture E. coli was grown overnight and plasmid DNA was extracted using a Hi-speed plasmid midi kit (Qiagen, Crawley, UK) according to the manufacturer instructions. Plasmid DNA was quantified using the Nanodrop ND-1000 Spectrophotometer (Labtech, Ringmer, UK). In order to determine the concentration and purity of DNA isolated, its optical density at 260 nm and 280 nm was measured. Its purity was determined by the ratio of its 260nm:280nm readings with a value of 1:1.8 indicating a pure DNA sample. Its concentration was calculated using the following formula:

\[
\text{DNA concentration (µg/ml)} = \text{A}_{260} \times \text{dilution factor} \times 50 \text{ (constant)}
\]

2.8 Preparation of electrocompetent mycobacterial cells

*M. tuberculosis* was grown to an OD\textsubscript{600} of 0.8 - 1.0 at 37°C. 1/10\textsuperscript{th} volume of 2 M glycine was then added into the culture for 24 hours to weaken the cell wall. The cells were then harvested at 10,000xg (Sorvall, DJB Labcare, Newport Pagnell, UK) for 30
minutes at 4°C. The supernatant was removed and the cell pellets were resuspended in ice-cold 10% glycerol, harvested and washed again for a total of 4 washes to remove salt. Finally, the cell pellets were resuspended in ice-cold 10% glycerol at 1/10th of the culture volume and stored at 4°C.

2.9 Electroporation of mycobacterial cells

For the electroporation of M. tuberculosis competent cells, 400 µl of the electrocompetent mycobacterial cells were mixed with the plasmid DNA (1 µg for gene deletion studies, 10-100 ng for plasmid insertion studies) in a microtube, and transferred into a 0.2cm path-length electroporation cuvette (Bio-Rad, Hemel Hempstead, UK). A gene-pulser unit (Bio-Rad, Hemel Hempstead, UK) was used for the electroporation at 25µF, 2.5kV and 1000 Ω. The electroporated bacteria was then transferred into 3.6 ml Dubos broth with albumin and glycerol supplements using a sterile plastic Pasteur pipette, and incubated overnight at 37°C. For the gene deletion studies, 1 ml of the incubated culture was harvested at 15,000xg (Jouan, DJB Labcare, Newport Pagnell, UK) at room temperature for 1 minute and the cell pellet resuspended in 100 µl supernatant. For plasmid insertion studies, direct plating from the incubated culture is sufficient to ensure bacterial growth. The cells were then plated on 7H11 agar with the appropriate antibiotic selections and supplements.

2.10 Extraction of genomic DNA from mycobacterial cells

Mycobacterial cells were both harvested from ¼ of the 7H11 agar plate or 1 ml liquid culture, and resuspended in 400 µl TE buffer. The cell suspension was then incubated
at 80°C for 45 minutes. A solution of lysozyme and lipase was added to give a final concentration of 2 mg/ml each, as well as adding Tween 80 (0.1% v/v final concentration) and 5 μg DNase-free RNase (Roche, Lewes, UK). The mixture was then incubated at 37°C for 2-3 hours, and snap-frozen on dry ice for 30 minutes, then thawed at 50°C for 10 minutes. Following the addition of proteinase K (80 μg/ml final concentration) and SDS (0.8% final concentration), the sample was incubated at 50°C for 1 hour before being extracted twice with phenol:chloroform:isoamyl alcohol (Sigma, Shaftesbury, UK). The aqueous phase was collected after 15,000xg centrifugation in a microfuge (Philip Harris, Shaftesbury, UK) for 15 minutes at room temperature. The DNA was precipitated by the addition of 1/50th volume of 5M NaCl and 2.5 volume of 100% ethanol to the aqueous phase and mixed. The sample pellet was harvested at 15,000xg (Philip Harris, Shaftesbury, UK) for 15 minutes at room temperature and washed twice with 70% ethanol, and left to air dry. The pellet was then resuspended in 200 μl TE buffer, and stored at 4°C. The genomic DNA was quantified using the Nanodrop ND-1000 Spectrophotometer (Labtech, Ringmer, UK) as described in section 2.7.

2.11 Isolation of RNA from mycobacterial cells

*M. tuberculosis* were either grown to an OD$_{600}$ of 0.35 - 0.4 (exponential growth phase) or an OD$_{600}$ of >2 (stationary growth phase). Mycobacterial RNA was isolated from the bacterial culture using the RiboPure yeast kit (Ambion, Huntingdon, UK). 120 ml mycobacterial cell culture was harvested at 10,000xg Sorvall, DJB Labcare, Newport Pagnell, UK) for 30 minutes at 4°C, the supernatant was removed and the cell pellet was resuspended in 2 ml lysis buffer, then add 200 μl 10% SDS and 2 ml of
Phenol:Chloroform. The mixture was then divided into 4 x 2 ml screw-cap tubes with 750 µl ice-cold Zirconia Beads (Ambion, Huntingdon, UK), and then ribolyzed once at the speed setting 6.5 for 45 seconds in a ribolyser (Hybaid, Qbiogene, Carlsbad, USA). After spinning for 5 minutes at 15,000xg (Jouan, DJB Labcare, Newport Pagnell, UK) at room temperature, the aqueous phase was transferred to a new microtube. 1.9 ml binding buffer (Ambion, Huntingdon, UK) and 1.25 ml 100% ethanol was added to each aqueous phase sample, and the mixture was then loaded onto a binding column supplied from the RiboPure kit. The column was washed once with 500 µl Wash solution I (Ambion, Huntingdon, UK) and twice with 500 µl Wash solution II (Ambion, Huntingdon, UK) with 1 minute spin at 15,000xg (Philip Harris, Shaftesbury, UK) at room temperature as recommended, then the RNA was eluted by the addition of 25 µl pre-heated (95°C) Elution solution (Ambion, Huntingdon, UK) and centrifuged at 15,000xg (Philip Harris, Shaftesbury, UK) for 1 minute at room temperature. The eluted sample was then treated with DNase I as follows: 1 x DNase I buffer (Ambion, Huntingdon, UK), and 4 µl of DNase I (Ambion, Huntingdon, UK) were added and the sample was incubated at 37°C for 30 minutes. Following this, 0.1 volume of DNase I inactivation buffer (Ambion, Huntingdon, UK) was mixed into the reaction, and the mixture was then left at room temperature for 5 minutes, then centrifuged for 3 minutes at 15,000xg (Philip Harris, Shaftesbury, UK) at room temperature. The supernatant was transferred to a new microtube, and stored at -80°C. The RNA was quantified using the Nanodrop ND-1000 Spectrophotometer (Labtech, Ringmer, UK). In order to determine the concentration and purity of RNA isolated, its optical density at 260nm and 280nm was measured. Its purity was determined by the ratio of its 260nm:280nm readings with a value of 1:2.2 indicating a pure RNA sample. Its concentration was calculated using the following formula:
RNA concentration (µg/ml) = A_{260} \times \text{dilution factor} \times 40 \ (\text{constant})

2.12 RT-PCR

Superscript One-Step RT-PCR kit (Invitrogen, Strathclyde, UK) was used for the RT-PCR study, following the manufacturer instructions, and using *M. tuberculosis* H37Rv RNA as the positive control template. In each reaction a RT+ (containing reverse transcriptase) and RT- (containing deionised water as a control) were used to confirm the absence of genomic DNA contaminants in the RNA samples.

2.13 Microarray hybridisation

The microarray slides used for this experiment were printed on GAPS amino-sialine coated slides (Corning, VWR, Lutterworth, UK) by the BμG@S Group at St. George’s Hospital London. The GAPS coating gives the slide a positively charged surface to which the negatively charged DNA can bind. Additionally, the microarray protocol used in this thesis was “the DNA-RNA: labelling and hybridization”, which was kindly provided by the BμG@S Group at St. George’s Hospital London.

To label the reference DNA, 1 µg of genomic DNA from the wild type *M. tuberculosis* H37Rv (Colorado University, USA) was randomly primed by the addition of 3 µg random hexameric primers (Invitrogen, Strathclyde, UK) in 41.5 µl water, heated to 95°C for 5 minutes before cooling on ice. The mixture was then
combined with 5 μl of 10x Klenow polymerase buffer (Promega, Southampton, UK), 1 μl of dNTP mix (5mM dA/G/TTP, 2mM dCTP), 1.5 μl Cy3-labelled dCTP (Amersham, Amersham, UK), and 1 μl of Klenow polymerase (Promega, Southampton, UK) to a final volume of 50 μl, and incubated at 37°C in the dark for 90 minutes.

To label the RNA sample, 3 μg of *M. tuberculosis* RNA was randomly primed by the addition of 3 μg random hexameric primers (Invitrogen, Strathclyde, UK) in 11 μl water, heated to 95°C for 5 minutes before cooling on ice. The mixture was then combined with 5 μl 5x First Strand buffer (Invitrogen, Strathclyde, UK), 2.5 μl 100mM DTT, 2.3 μl dNTP mix (5mM dA/G/TTP, 2mM dCTP), 1.7 μl Cy5-labelled dCTP (Amersham, Amersham, UK), and 500u Superscript III (Invitrogen, Strathclyde, UK) to a final volume of 25 μl. The sample was then incubated at 25°C in the dark for 10 minutes, followed by 42°C in the dark for 60 minutes.

To eliminate unincorporated labelled dCTP, the labelled samples were mixed and purified through a MinElute column (Qiagen, Crawley, UK), following the manufacturer instructions, except for separating the wash step into two washes, one of 500 μl and one of 250 μl. The purified labelled sample was eluted in 15.9 μl de-ionised water.

The microarray slides were prepared for hybridization by incubation in 50 ml pre-hybridization buffer (3.5x SSC, 0.1% SDS, 10mg/ml BSA (Sigma, Shaftesbury, UK)), at 65°C for 20 minutes. After incubation the slides were rinsed with 400 ml water for 1 minute and then 400 ml propan-2-ol for another minute, before
centrifugation in a 50 ml Falcon tube (Becton Dickinson, Plymouth, UK) at room temperature at 2000xg (Sigma, Shaftesbury, UK) for 5 minutes.

The purified labelled sample was combined with 4.6 µl 0.22µm-filtered (Millipore, Stonehouse, UK) 20x SSC, and 3.5 µl 0.22µm-filtered (Millipore, Stonehouse, UK) 2x SDS. The sample was then heated to 95°C for 2 minutes, briefly centrifuged and applied onto the prepared microarray slide. A 22 mm x 22 mm Lifterslip (Erie Scientific, Portsmouth, USA) was placed over the array to evenly distribute the sample after loading.

The loaded array was sealed in an array chamber with two 15 µl water on each end and incubated in a water bath at 65°C for 18-24 hours. After the incubation the array was washed in 400 ml Wash A solution (1 x SSC, 0.05% SDS). It was then washed twice in 400 ml Wash B solution (0.15x SSC), the first at 65°C and the second at room temperature. The array was then spun dry by centrifugation in a 50 ml Falcon tube (Becton Dickinson, Plymouth, UK) at room temperature at 2000xg (Sigma, Shaftesbury, UK) for 5 minutes, and stored in a dark, dust-free box until scanning.

The microarrays were scanned using an Axon 4000A dual-wavelength scanner at 635 nm (for Cy5, in red) and 532 nm (for Cy3, in green). The scanned image was visualized by using the Genepix Pro 5.0 software (Axon Instruments, Foster City, USA). The ratio of the dual wavelength intensity per array was standardized to between 0.9-1.1. The Genepix Pro 5.0 software was also used to locate each spot on the array using the grid-map supplied by Dr. J. Hinds (BµG@S group, St. George’s Hospital). The grid was then modified to fit each array image, the spots with no
hybridisation were then removed and the results were exported in .gpr format to be analyzed by the Genespring 7.2 Software (Agilent Technologies, Santa Clara, USA).

2.14 Microarray analysis

For the Genespring 7.2 (Agilent Technologies, Santa Clara, USA) programme analysis, data generated from the microarray slides were filtered with a per-chip normalisation that divided the value for each spot by the 50th percentile of the measurements on the array. Control sample values below 0.1 were transformed to 0.1; if both the 635 nm and 532 nm values were below 0.1 then the data was excluded from the analysis. The filtered data were then subjected to the 1-way ANOVA comparison and the Bonferroni multiple testing corrections, with the parameters for significance set at a P-value of <0.05. The genes that were present at different ratios between the two strains were copied, along with their P-value and fold change, into Microsoft Excel for subsequent editing.

In Microsoft Excel, the ratio between each array (AsigC:H37Rv) was calculated through a matrix system (each of the six data points from the H37Rv was compared with each of the six data points from the AsigC samples), and hence 36 data points were generated from each gene. A median of the ratios was then calculated from the 36 data points, along with the standard deviation. For the genes with elevated levels of expression, genes were included when the value of the median of ratios was equal or above 2 (2-fold induction) following the subtraction of the standard deviation. For the genes with repressed levels of expression, genes were included when the value of the median of ratios was less than or equal to 0.5 (2-fold reduction). These repressed
genes were then re-calculated with a reversed ratio (H37Rv:ΔsigC) for each of the 36 data points, and genes were included when the value of the median of ratios was equal or above 2 (2-fold reduction) following the subtraction of the standard deviation. This validation method was introduced by Manganelli et al. in their microarray analysis between the wild type *M. tuberculosis* H37Rv and the ΔsigE mutant (Manganelli et al. 2001), and also used by Sherman et al. in their microarray analysis of the genes involved in hypoxic response in *M. tuberculosis* (Sherman et al. 2001).

### 2.15 Primer extension protocol

The oligodeoxyribonucleotide primer ACGTCTTGCTGAGCTGGTGTGTTTGGATAA was labelled at the 5’ terminus with [γ\(^{32}\)P] ATP using T4 polynucleotide kinase as described in the primer extension kit (Promega, Southampton, UK). 0.1-1 pmol of the labelled primer was added to 5 μl of nuclease free water containing 70 μg of total RNA and 5 μl of avian myeloblastosis virus (AMV) primer extension buffer (2X, Promega, Southampton, UK). The annealing step was performed at 50°C for 1 hour. 5μl of the AMV primer extension buffer (2X, Promega, Southampton, UK) 1.4 μl of 40 mM sodium pyrophosphate and 1 unit of AMV reverse Transcriptase (Promega, Southampton, UK) were added to each reaction. The reaction mixture was incubated at 42°C for 1 hour, ethanol precipitated, washed with 70% ethanol and resuspended in 5 μl of loading dye (Promega, Southampton, UK). The extension products were heated at 90°C for 10 minutes, and run on a 8% polyacrylamide-urea gel at 250V, 150mA and 50W for 2 hours, and visualized by autoradiography.
For the sequencing reaction 5 μg of plasmid DNA pCulter2 is required (see Appendix I). The first step was to linearize the plasmid. A total of 0.1 volume of 2 M NaOH-2 M EDTA was added to the sample, and incubated at 37°C for 30 minutes. 0.1 volume of 3M Sodium Acetate (pH 4.8) was then added to the sample before the mixture was ethanol precipitated with 2.5 volume of absolute ethanol. The precipitated sample was then incubated at -70°C for 15 minutes, and centrifuged at 15,000xg (Philip Harris, Shaftesbury, UK) for 15 minutes at room temperature, washed once with 70% ethanol, and centrifuged at 15,000xg (Philip Harris, Shaftesbury, UK) at room temperature for a further 5 minutes. The supernatant was then removed, and the pellet air-dried. The pellet was then resuspended in 7 μl of nuclease-free water. The annealing step was performed on the resuspended sample with the addition of 2 μl sequenase reaction buffer (Promega, Southampton, UK) and 1 μM primer-ACGTCTTGCTGGTGCTTTGATAA. The reaction was incubated at 65°C for 2 minutes, and allowed to cool to room temperature. The reaction was briefly spun and kept on ice. The labelling step involved the addition of 1 μl 0.1 M DTT, 2 μl labelling mix 0.1x, Promega, Southampton, UK), 0.25 μl sequenase and 0.5 μl α-35S dATP to the 10 μl template DNA. The annealing mixture was incubated at room temperature for 2 minutes. 3.2 μl of this mixture was added to 2.5 μl of each of the pre-warmed (37°C for 1 minute) ddGTP, ddCTP, ddTTP and ddATP for the termination step. Each of the four termination reactions were incubated at 37°C for 5 minutes, and the reactions were stopped by the addition of 4 ml Stop solution (Promega, Southampton, UK), and stored at -20°C. Prior to the gel run, the four sequencing samples were incubated at 90°C for 10 minutes, and 3 μl of each sample was loaded in the 8% polyacrylamide-urea gel adjacent to the primer extension reactions.
2.16 Stress studies in *M. tuberculosis*

The bacterial viability following exposure to various stress conditions *in vitro* was determined as follows. 175 ml liquid cultures of each strain were grown to an OD$_{600}$ of 0.3-0.4, and then 40 ml was aliquoted into new roller bottles. Each culture bottle was treated with a single condition (25 mM paraquat, 5 mM hydrogen peroxide, 25 μM cumene hydroperoxide, 3 mM sodium nitrite at pH5.4, or acidic conditions at pH5.4, pH 4, pH 3) for 24 hours, with one sample left untreated. For the sodium nitrite pH 5.4 stresses, 100 ml of culture was harvested at 10,000xg Sorvall, DJB Labcare, Newport Pagnell, UK) at room temperature for 10 minutes, and resuspended in 100 ml modified Dubos at pH 5.4. Two 40 ml aliquots were transferred to roller bottles, to one which 3 mM sodium nitrite was added, while the other was left untreated. For the other low pH conditions, the culture was processed the same way, with a centrifugation step and resuspension in media at pH 4 or pH 3 respectively.

After 24 hours incubation in the rolling incubator (Bellco, New Jersey, USA), samples of the cultures were serially diluted in DMEM (Dulbecco’s Modified Eagle’s Medium-Difco, Plymouth, UK) + 50% FCS (Foetal calf serum, heat inactivated). 50 μl of each culture was added to 450 μl DMEM + FCS solution in a microfuge tube 1/3 filled with 2.5-3.5 mm autoclaved glass beads and vortexed. 10 μl of each serially diluted sample was plated onto ¼ 7H11 agar (Difco, Plymouth, UK) plates. Colonies were counted after 2-3 weeks.
2.17 *M. tuberculosis* mouse infection study

The mouse infection experiment was carried out with the help of Mr. E. Stavropolous and Mr. J. Brennan under containment level 3 conditions. *M. tuberculosis* strains were grown in 7H9 broth containing 0.05% tween and 10% albumin-dextrose complex (ADC) supplement. The strains were allowed to grow to an OD$_{600}$ of 0.3 - 0.4, then the cultures were briefly spun at a slow speed (200xg; Sorvall, DJB Labcare, Newport Pagnell, UK) for 5 minutes at room temperature to separate any bacterial clumps from the liquid cultures. The strains were then diluted in phosphate-buffered saline to an OD$_{600}$ of 0.018 - 0.022. The cultures were briefly vortexed with 2 mm-diameter glass beads (VWR, Lutterworth, UK) to prevent clumping before intravenously (IV) injected into the tail vein of 6 – 8-weeks-old female BALB/c mice with approximately 5 x 10$^5$ bacteria. At different time intervals the lungs and spleens were harvested, individually weighed, homogenised, serially diluted and plated onto 7H11 agar plates. Plates were incubated at 37°C and colonies were counted after 11-15 days.
CHAPTER 3: Functional characterisation of an alternate sigma factor $\sigma^C$

A mutant strain lacking the $\text{sigC}$ gene had been generated previously from wild-type $M.\text{tuberculosis}$ H37Rv via the inverse PCR technique (Gopaul 2002). In this mutant, 410 bp of the 558 bp coding sequence of $\text{sigC}$ is deleted, including the whole of region 4 and most of region 2 (Figure 3.1), such that the functionality of $\text{sigC}$ can reasonably be assumed to be abolished. However, the role of $\sigma^C$ remains to be elucidated. Therefore, investigation of both in vitro and in vivo phenotypes of the $\Delta\text{sigC}$ strain in comparison with the wild-type H37Rv should provide an insight into the role of $\sigma^C$ in $M.\text{tuberculosis}$.

Figure 3.1  (A) Schematic diagram of the $\text{sigC}$ gene replacement construct. 410 bp were deleted from the $\text{sigC}$ coding sequence.  
(B) Schematic diagram of the deletion of regions 2 and 4 ($\Delta$ region) of the $M.\text{tuberculosis} \sigma^C$. Adapted from the NCBI database (NCBI 2006).
3.1 *in vitro* phenotypic analysis of *M. tuberculosis* ΔsigC

3.1.1 Growth of *M. tuberculosis* ΔsigC in aerobic conditions

In previous studies, sigC was found to be highly expressed during the exponential growth phase (Manganelli *et al.* 1999). Therefore, an *in vitro* growth study was conducted to investigate whether ΩC is required for growth of *M. tuberculosis* in aerobic conditions.

For this *in vitro* analysis, wild-type *M. tuberculosis* H37Rv and ΔsigC strains were allowed to grow to exponential phase (OD<sub>600</sub> ~ 0.3 - 0.4) in 175 ml Dubos plus albumin supplement, in roller bottles at 37°C. Each culture was then subcultured into two new roller bottles to give an OD<sub>600</sub> of 0.010. OD<sub>600</sub> readings were taken at 24-hour intervals for 15 days. For absorbance readings greater than 1, the samples were diluted 5-fold prior to the measurement to ensure a more precise value for each sample.

As shown in Figure 3.1.1.1, no differences in the growth rates were observed between ΔsigC mutant strain and the wild-type H37Rv in the *in vitro* aerobic growth conditions. Thus, the deletion of sigC does not affect the *in vitro* growth of *M. tuberculosis*.
Figure 3.1.1.1  The growth of wild-type *M. tuberculosis* H37Rv and ΔsigC strains in aerobic growth conditions. The experiments were carried out in duplicate cultures per strain. The error bars indicate standard deviations.
3.1.2 Survival of *M. tuberculosis* ΔsigC under various stress conditions in exponential growth phase

*M. tuberculosis* encounters hostile conditions within the macrophage, such as RNI, ROI and low pH (McKinney *et al.* 2003). To investigate any role of $\sigma^C$ in surviving such conditions, the viability of the ΔsigC strain was compared with that of the wild-type following exposure to a selection of conditions chosen to mimic different aspects of the environment within the macrophage. Wild-type H37Rv and ΔsigC strains grown to exponential phase ($OD_{600} \sim 0.3 - 0.4$) were used in this study.

The superoxide radical generator paraquat dichloride (Sigma, Shaftesbury, UK) was used at 25 mM concentration (Rand 2004) to investigate the effect of ROI on the *M. tuberculosis* strains *in vitro*. The wild-type strain had a 1.9-fold decrease in viability as compared with the untreated wild-type control after the 24-hour exposure to paraquat, and the ΔsigC mutant strain had also exhibited a 1.9-fold decrease in viability as compared with its untreated control after the paraquat stress (Figure 3.1.2.1). Since the mutant did not show increased sensitivity *in vitro* to superoxide stress, it is difficult to envisage that sigC has a role in the survival of the bacterium to superoxide stress generated *in vivo*. In order to examine the effect of peroxides on *M. tuberculosis* *in vitro*, hydrogen peroxide and cumene hydroperoxide, an organic hydroperoxide, were added to a final concentration of 5 mM and 25 $\mu$M respectively (Rand 2004). The sigC null mutant exhibited a 1.2-fold decrease in viability as compared with a 1.6-fold decrease of the wild-type strain after the hydrogen peroxide stress. A 1.3-fold decrease in viability from the sigC null mutant was observed following the cumene hydroperoxide stress, as compared with a 1.2-fold decrease in
viability of the wild-type strain (Figure 3.1.2.1). Overall, the results from the peroxide stress are very similar between the wild-type and the ΔsigC strains, suggesting that $\sigma^C$ does not play a role in the resistance to ROI stress in vitro.

Similarly, a 1.6-fold decrease in viability from the sigC null mutant was observed following the RNI stress at low pH, as compared with a 1.6-fold decrease in viability of the wild-type strain (Figure 3.1.2.1). This implied that sigC is not involved in the resistance to RNI stress. Surprisingly, both the sigC null mutant and the wild-type strains appeared to be more susceptible to mild acidic pH alone (pH 5.4), than to the acidified nitrite stress, in contrast to previous studies (Rand 2004). The sigC null mutant and the wild-type exhibited 1.3-fold and 1.5-fold decreases in viability respectively in pH 5.4 stress alone as compared with the acidified nitrite stress, although these results were found to be statistically insignificant ($P > 0.18$, by using a two-tailed Student $t$-test for groups of unequal variance). Overall, the sensitivity of the sigC null mutant towards mild acidic conditions is similar to that of the wild-type strain, suggesting that $\sigma^C$ is not involved in the response of mild pH stress conditions.

In order to investigate the role of $\sigma^C$ in the resistance to lower pH stress conditions in vitro, acidic environments of pH 4 and pH 3 were tested. The sigC null mutant exhibited a 1.8-fold decrease in viability as compared with a 2-fold decrease of the wild-type strain after the pH 4 condition, while a 3.6-fold decrease in viability from the sigC null mutant was observed following the pH 3 stress, as compared with a 3.7-fold decrease in viability of the wild-type strain (Figure 3.1.2.1). This suggests that the pathogen’s ability to survive is affected by extreme acidic conditions, but that $\sigma^C$ is not essential for withstanding less severe acidic environments.
Figure 3.1.2.1 Survival of the wild-type *M. tuberculosis* H37Rv and ΔsigC mutant strains under various stress conditions. The experiments were carried out twice independently with duplicate plating. The error bars indicate standard deviations.
3.2 *in vivo* phenotypic analysis of *M. tuberculosis* ΔsigC

Prior to evaluating the role of σ^C^ in the pathogenesis of *M. tuberculosis* using the mouse infection model, it was important to establish a complemented mutant strain in which the expression of a functional sigC is restored. A sigC complementing clone had been constructed as illustrated in Figure 3.1.3.1 (Gopaul 2002). In this clone, the wild-type sigC coding sequence and 164 bp of its upstream sequence, (which contains its native promoter) were isolated following PCR, and cloned into the *EcoRV* restriction site of the integrating vector pKP149 (a gentamycin resistant derivative of pMV306)(see Appendix I for plasmid details). The bacterial attachment site, *attP*, in the vector allowed it to integrate into the *attB* site within the *M. tuberculosis* H37Rv genome by site-specific recombination. This clone was transformed into the sigC null mutant strain, and plated on 7H11 + gentamycin plates for selection.

![Diagram](image)

**Figure 3.2** A schematic diagram illustrating the construction of the sigC-complemented construct, ΔsigCcomp.
3.2.1 Verification of the sigC-complemented strain ΔsigCcomp

RNA was isolated from the sigC-complemented strain (ΔsigCcomp), the ΔsigC mutant and the wild-type H37Rv at the exponential (OD$_{600}$ ~ 0.3 - 0.4) and the stationary (OD$_{600}$ > 2) growth phases. Figure 3.2.1.1 shows the quality of the RNA from all of the samples shown on a 1% agarose gel.

![Image showing RNA purification of wild-type H37Rv, ΔsigC, and ΔsigCcomp strains at exponential (log) and stationary (stat) growth phases. (M): λ.EcoRI/HindIII marker.]

In order to determine whether the ΔsigCcomp strain had restored the expression of sigC, the RT-PCR method was used. PCR primers were selected within the deleted region of sigC, illustrated in Figure 3.2.1.2 and Table 3.2.1.1 (see Appendix II for the complete primer list). A 240 bp PCR product would only occur if a full length sigC coding sequence is present within the genome of the M. tuberculosis strains.
Figure 3.2.1.2 A schematic diagram of the RT-PCR experiment to verify the presence of the \textit{sigC} RNA transcript. A 240 bp PCR product would only occur if a full length \textit{sigC} coding sequence is present within the genome of the \textit{M. tuberculosis} strains.

<table>
<thead>
<tr>
<th>Description</th>
<th>Primer sequences (5'-3')</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{sigC} RT-PCR primers</td>
<td>GCTCCACGCGCCCGGAACCTTG</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td>CTGCTCGGGGCTGTCTATTGCGG</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2.1.1 Sequences of primers used for the RT-PCR studies.
The RNA samples were quantified, and 1 μg of RNA from each sample was used as a template for cDNA synthesis, followed by PCR. The samples were then run on a 1% agarose gel, as illustrated in Figure 3.2.1.3, and compared with the results of PCR reactions performed directly on the RNA without a reverse transcriptase step to control for any DNA contaminations.

PCR products of 240 bp were observed in the wild-type H37Rv cDNA and the complemented strain ΔsigCcomp cDNA samples, in both the exponential and stationary growth phase, when performed with the reverse transcriptase, as well as in the genomic DNA positive control. In addition, the 240 bp PCR product was absent from the cDNA samples of the ΔsigC mutant strain when performed with the reverse transcriptase, thus confirming the absence of a full length coding sequence of sigC in the ΔsigC mutant strain. These results were further confirmed by the absence of the PCR product in the samples performed without the reverse transcriptase, therefore confirming the reverse transcriptase results were not due to DNA contamination. The results from the RT-PCR studies confirmed that sigC expression was restored in the sigC-complemented ΔsigCcomp strain, in both exponential and stationary growth phases. The results also confirmed that the ΔsigC strain did not express sigC RNA.
Figure 3.2.1.3  RT-PCR analysis of the expression of the \textit{M. tuberculosis} \textit{sigC} gene. H37Rv DNA lane was a positive control, using \textit{M. tuberculosis} genomic DNA as a template. The 240 bp PCR product shown in the \textit{M. tuberculosis} H37Rv genomic DNA positive control lane (far left) was comparable with the PCR products generated from the cDNA of the H37Rv and the \textit{AsigCComp} strains, in both exponential (log) and stationary (stat) growth phases. Notably they were absent from both cDNA samples of the \textit{AsigC} strains. (M): 100 bp DNA ladder

RT+ signifies that reverse transcriptase was used, and RT- signifies that the reverse transcriptase was substituted with DEPC-treated water to provide a negative control. C signifies the no template negative control.
3.2.2 Growth of *M. tuberculosis* ΔsigC in a mouse infection model

To investigate whether the deletion of the *sigC* gene affects *M. tuberculosis* virulence, the ΔsigC mutant strain was compared with its parental and complemented strains in intravenously infected BALB/c mice (Medina *et al.* 1998). The mouse infection study was completed with the help of Mr. E. Stavropoulos and Mr. J. Brennan. For the first two time-points (day 1 and 28), three mice from each group were sacrificed, and for the remaining time-points (day 60, 89, 120, 147) four mice from each group were sacrificed. The survival and multiplication of the *M. tuberculosis* strains were determined by enumerating the bacterial colony forming units (CFU) in the lungs and spleens of the infected mice.

In the lungs (Figure 3.2.2.1), although the initial inoculum of all of the strains was approximately the same, on day 28 the CFU counts observed from mice exposed to the ΔsigC strain were higher than those exposed to the wild-type and the complemented strain (> 5-fold increase, *P* < 0.005, by using a two-tailed Student *t*-test for groups of unequal variance). Throughout the remainder of the study, the CFU counts for the ΔsigC mutant strain remained higher than those for the wild-type and complemented strains. The overall *in vivo* growth pattern of the *sigC*-complemented strain, ΔsigC*comp*, was very similar to that of the wild-type H37Rv strain.

In the spleens (Figure 3.2.2.2), the initial inoculum of the ΔsigC strain was slightly lower than those of the wild-type H37Rv and the complemented strains. The growth of all three strains was very similar for the first 28 days, but following this there was a gradual increase in the CFU counts of the ΔsigC strain compared to those of the wild-
type and the complemented strain. The most significant difference was observed on day 147 (~ 5-fold increase, $P < 0.001$, by using a two-tailed Student $t$-test for groups of unequal variance). The overall in vivo growth pattern of the sigC-complemented strain, ΔsigCcomp, was almost identical to that of the wild-type H37Rv strain.
Figure 3.2.2.1 Growth of the wild-type H37Rv (WT), ΔsigC (sigC mutant) and ΔsigCcomp (sigC complement) strains in the lungs of BALB/c mice. Deletion of the sigC gene lead to higher bacilli counts in the lungs of BALB/c mice. Error bars indicate standard deviations.

Figure 3.2.2.2 Growth of the wild-type H37Rv (WT), ΔsigC (sigC mutant) and ΔsigCcomp (sigC complement) strains in the spleens of BALB/c mice. Deletion of the sigC gene lead to higher bacilli counts in the spleens of BALB/c mice. Error bars indicate standard deviations.
3.3 Transcription regulation of \( \text{sigC} \) in \( \text{M. tuberculosis} \)

The regulatory mechanism involved in the regulation of \( \text{sigC} \) is not known in \( \text{M. tuberculosis} \). The identification of the transcription start point(s) of \( \text{sigC} \) by primer extension would provide an indication as to whether \( \sigma^C \) was involved in its own expression, and possibly identify the promoter elements involved in the expression of \( \text{sigC} \) in \( \text{M. tuberculosis} \).

3.3.1 Mapping the transcription start point of \( \text{sigC} \) in \( \text{M. tuberculosis} \)

The position of the transcription start point for \( \text{sigC} \) was identified by primer extension using a primer (ACGTCTTGCTGGGTGGCTTTGATAA), complementary to position 86 - 110 bp of the \( \text{sigC} \) coding sequence. Importantly, this part of the \( \text{sigC} \) coding sequence is intact in the \( \Delta\text{sigC} \) mutant strain, as the deletion begins at base 137 of the coding sequence, thus allowing analysis of this strain as well as the wild-type H37Rv. The primer extension experiment was carried out using 70 \( \mu \text{g} \) of total RNA from both the wild-type \( \text{M. tuberculosis} \) H37Rv and the mutant \( \Delta\text{sigC} \) strains. The transcription start point was found to be at an \( \text{A} \) in the wild-type H37Rv sample (Figure 3.3.1.1), which coincided with the ATG of the predicted translation initiation codon of \( \text{sigC} \) (Tuberculist 2006). For the \( \Delta\text{sigC} \) sample, no extension product was observed (Figure 3.3.1.1), suggesting that \( \sigma^C \) may be involved in regulating its own transcription, although the possibility remained that the lack of a product could be due to instability of the truncated RNA in the \( \Delta\text{sigC} \) mutant strain.
Figure 3.3.1.1 Primer extension analysis of $\textit{sigC}$ in \textit{M. tuberculosis}.

(A) The base corresponding to the deduced transcription start point and the translation initiation colon of $\textit{sigC}$ are highlighted in red and yellow respectively on the sequence. The proposed -10 and -35 regions of $\textit{sigC}$ are highlighted in the light blue and blue boxes, and the position of the primer used for the primer extension is highlighted in purple, and the arrow indicates the direction of the extension.

(B) The product of primer extension (indicated by the red triangle) using a primer complementary to $\textit{sigC}$ on wild-type H37Rv RNA (Rv), $\Delta$\textit{sigC} RNA ($\textit{AC}$) and No RNA control (c). Sequencing reactions were performed with the same primer (G, A, T, C) and were run adjacent to the extension samples.
3.3.2 Identifying the putative promoter motifs of sigC in *M. tuberculosis*

Following the identification of a putative transcription start point of sigC in *M. tuberculosis*, a high A-T content region (TATGGT) was found at 8 bp upstream of the transcription start point. Additionally, another high A-T content region (TAGACT) was found 17 nucleotides prior to that region. The characteristics of the two high A-T content regions, along with 17 nucleotides between the regions, are very similar to the mycobacterial consensus promoter sequence TTGACG (-35 region) and TATAAT (-10 region) proposed by Mulder et al. (Mulder et al. 1997). The putative -10 region for sigC is identical at four out of six positions with the mycobacterial consensus -10 region (TATggT, identical nucleotides are in upper case). The putative -35 region for sigC is also identical at four out of six positions with the mycobacterial consensus -35 region (TaGACt, identical nucleotides are in upper case). The distance between the -10 and the -35 regions varies between 16 – 18 nucleotides in the mycobacterial consensus motifs, and hence the distance between the putative promoter elements of sigC falls within this region. Therefore, the putative promoter sequence for sigC expression is:

**TAGACT-N_{17}-TATGGT**

The conservation of the -10 and -35 hexamers in sigC was also similar to the *E. coli* σ^{30} consensus promoter (TTGACA and TATAAT) (Mulder et al. 1997), with four out of six positions identical at both the -10 and -35 regions (TaGACt and TATggT, identical nucleotides are in upper case). This suggests that sigC may also be recognized by the principal σ factor in *M. tuberculosis*, σ^{A} and possibly the principal-like σ factor σ^{B}.
3.4 Global gene expression in *M. tuberculosis* ΔsigC during exponential growth phase

A complete genomic microarray analysis comparing gene expression in the ΔsigC mutant with that of the wild-type would allow an insight into which genes were regulated by σC in *M. tuberculosis*. Previous studies by quantitative RT-PCR had shown that sigC expression was influenced by growth state, with a higher expression in the exponential phase than in the stationary phase (Manganelli *et al.* 1999). Therefore, in this study global gene expression in the two strains was compared using exponential phase cultures.

3.4.1 Microarray analysis of *M. tuberculosis* ΔsigC during exponential growth phase compared to wild-type *M. tuberculosis*

For this microarray study, wild-type *M. tuberculosis* genomic DNA (Colorado University, USA) was used as a reference control. RNA samples were isolated from three separate cultures at the exponential growth phase (OD_{600} ~ 0.3 - 0.4) from both the wild-type *M. tuberculosis* H37Rv and the ΔsigC strains, and the quality of the RNA samples was verified using a bioanalyzer (Bioanalyzer 2100, Agilent) prior to use for the microarray hybridisations as described in section 2.13.

In this DNA-RNA microarray hybridization experiment, the Cy3-labelled DNA from the *M. tuberculosis* H37Rv genomic DNA should represent all genes equally, whereas the Cy5-labelled cDNA from an RNA source should contain different levels of the transcript of each gene depending upon its expression level. This should result in a
uniform intensity for each gene in the 532 nm channel (green – from the reference DNA), and a variety of intensities for each gene in the 635 nm channel (red – from the RNA derived cDNA sample). The results were expressed as a cDNA/genomic DNA ratio, and genes that showed a change in the ratio between the wild-type H37Rv cDNA/DNA, and ΔsigC cDNA/DNA groups were believed to be present at different quantities in the two strains.

The microarray experiments were carried out using duplicate slides for each of the six RNA samples (three wild-type H37Rv and three mutant ΔsigC samples). The microarray images were scanned (Figure 3.4.1.1 and Figure 3.4.1.2) and grid-mapped to trace the intensity of each *M. tuberculosis* gene-spot on the arrays. The results of each microarray slide were merged in the Genespring 7.2 software program to generate an overall analysis of the data. The analysis of the data will be discussed in more detail in the next section.
Figure 3.4.1.1  A sample scan image of a microarray slide with the wild-type *M. tuberculosis* H37Rv RNA-derived cDNA isolated at the exponential growth phase versus the wild-type *M. tuberculosis* H37Rv genomic DNA.

Figure 3.4.1.2  A sample scan image of a microarray slide with the *M. tuberculosis* ΔsigC RNA-derived cDNA isolated at the exponential growth phase versus the wild-type *M. tuberculosis* H37Rv genomic DNA.
3.4.2 Genes regulated by $\sigma^C$ detected during the microarray analysis

Following the Genespring software analysis with the 1-way ANOVA comparison and the Bonferroni multiple testing corrections ($P < 0.05$), 101 genes were obtained and further analysed, as described in section 2.14. From this 32 genes were deemed significantly different between the two strains, of which 23 genes showed elevated levels of expression in the $\Delta$sigC strain when compared to the wild-type H37Rv (shown in Table 3.4.2.1), and 9 genes showed reduced levels of expression (shown in Table 3.4.2.2).

Of the 23 genes that showed elevated levels of expression in the $\Delta$sigC strain, 4 clusters were observed to contain genes located adjacent within the *M. tuberculosis* genome (highlighted in yellow, Table 3.4.2.1), suggesting that these genes may belong to operons regulated by $\sigma^C$. Further investigation into the roles of the induced genes from the TubercuList database (TubercuList 2006) have identified that they belong to several specific functional groups in *M. tuberculosis*:

1) Heat shock response (DnaK-GrpE-DnaJ1, GroEL1)

2) Hypoxic response (DosR-Rv3134c, Rv0569, Rv1733c, Rv1813c, Rv1998c, Rv2004c, Rv2005c, FdxA, Rv2623, Rv2629 and Rv2631)

3) Fatty acid degradation (FadA-FadB)

4) Detoxification reactions (HpB)

5) Electron transfer (FprB)
Of the 5 functional groups identified from this analysis, the hypoxic response genes showed the most elevated levels of expression in the ΔsigC strain during exponential growth phase (from 3.1-fold to 46.5-fold induction). In particular, high levels of induced expression were observed from the heat shock response genes (from 5.2-fold to 6.9-fold induction). These results suggest that σC may play a role in the regulation of the heat shock and hypoxic responses, in particular the regulation of the DosR two-component regulator and subsequently the regulation of its downstream genes.

Amongst the 9 genes that showed repressed levels of expression in the ΔsigC strain when compared with the wild-type, one cluster (Rv3402c-Rv3403c) was observed to contain genes located adjacently within the M. tuberculosis genome (highlighted in yellow, Table 3.4.2.2), suggesting that these genes may belong to an operon regulated by σC. Further investigation of the roles of these repressed genes from the TubercuList database (TubercuList 2006) have revealed that the specific functions of these genes are very diverse. The greatest level of repression was observed for DesA3 (5.4-fold reduction), which is involved in lipid metabolism. Other genes that may be induced by σC include: RpsL (2.7-fold reduction), which is involved in the translation initiation step for protein expression in M. tuberculosis, MbtD (3.1-fold reduction), which is involved in mycobactin biogenesis, and PntAB (2.5-fold reduction), which is an integral membrane protein. The functions of the remaining genes have yet to be determined, although they all display significant levels of repression in the ΔsigC strain during exponential growth phase (Rv1233c-4.5-fold reduction, Rv1461c-2.8-fold reduction, Rv3402c-2.6-fold reduction, Rv3403c-2.7-fold reduction and Rv3822-2.9-fold reduction).
Table 3.4.2.1  Table of 23 genes induced in the *M. tuberculosis* *ΔsigC* strain at the exponential growth phase, compared to their parental H37Rv strains. Genes are listed in genomic order and annotated as described by the Institut Pasteur on the TuberculList database (TuberculList 2006). Genes were highlighted yellow if two or more genes were in the same region of the genome. CHP – conserved hypothetical protein; CTMP – conserved transmembrane protein.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Rv No.</th>
<th>Fold Repression (±SD)</th>
<th>P-value</th>
<th>Gene Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>pntAB</td>
<td>Rv0156</td>
<td>2.5 ± 0.5</td>
<td>2.0E-04</td>
<td>part of a probable NAD(P) Transhydrogenase</td>
</tr>
<tr>
<td>rpsL</td>
<td>Rv0682</td>
<td>2.7 ± 0.7</td>
<td>1.9E-03</td>
<td>Probable 30S ribosomal protein S12</td>
</tr>
<tr>
<td>Rv1233c</td>
<td>Rv1233c</td>
<td>4.5 ± 0.9</td>
<td>5.3E-05</td>
<td>CHMP</td>
</tr>
<tr>
<td>Rv1461</td>
<td>Rv1461</td>
<td>2.8 ± 0.7</td>
<td>1.3E-03</td>
<td>CHP</td>
</tr>
<tr>
<td>mbtD</td>
<td>Rv2381c</td>
<td>3.1 ± 0.9</td>
<td>7.5E-04</td>
<td>Polyketide synthetase</td>
</tr>
<tr>
<td>desA3</td>
<td>Rv3229c</td>
<td>5.4 ± 1.4</td>
<td>2.0E-04</td>
<td>Possible linoleyl-CoA desaturase</td>
</tr>
<tr>
<td>Rv3402c</td>
<td>Rv3402c</td>
<td>2.6 ± 0.5</td>
<td>1.8E-02</td>
<td>CHP</td>
</tr>
<tr>
<td>Rv3403c</td>
<td>Rv3403c</td>
<td>2.7 ± 0.5</td>
<td>1.4E-04</td>
<td>HP</td>
</tr>
<tr>
<td>Rv3822</td>
<td>Rv3822</td>
<td>2.9 ± 0.8</td>
<td>6.6E-03</td>
<td>CHP</td>
</tr>
</tbody>
</table>

Table 3.4.2.2  Table of 9 genes repressed in the *M. tuberculosis* *AsigC* strain at the exponential growth phase, compared to their parental H37Rv strains. Genes are listed in genomic order and annotated as described by the Institut Pasteur on the TuberculList database (TuberculList 2006). Genes were highlighted yellow if two or more genes were in the same region of the genome. CHP – conserved hypothetical protein; CHMP – conserved hypothetical membrane protein; HP – hypothetical protein.
3.5 Discussion: defining the role of \( \sigma^C \) in *M. tuberculosis*

The pathogenic *M. tuberculosis* is able to establish long-term persistent infection in its host despite the induction of host-activated inflammatory and antimicrobial responses. This demands that *M. tuberculosis* cells sense and respond to various host-induced stress signals. In order to examine the role of \( \sigma^C \) in mycobacterial virulence, a *M. tuberculosis* mutant generated within the department with an unmarked deletion of the *sigC* gene was used. A *M. tuberculosis* *sigC*-complemented strain, \( \Delta \text{sigC} \text{comp} \), which introduces a functional wild-type *sigC* gene back into the \( \Delta \text{sigC} \) mutant strain via an integrated vector, was isolated, and the expression of *sigC* was verified by RT-PCR during this study. The survival of the \( \Delta \text{sigC} \) strain confirmed that *sigC* was not an essential gene for *in vitro* growth, which is consistent with previous findings (Sassetti *et al.* 2003; Sun *et al.* 2004).

The deletion of the *sigC* gene did not appear to affect the *in vitro* growth properties of the bacterium, although *sigC* is highly expressed during the exponential growth phase (Manganelli *et al.* 1999). Possible explanations for this phenomenon are that the function of \( \sigma^C \) may be compensated by other sigma factors and/or possible transcriptional regulators at a time when exponential growth is required, or the genes that are controlled by \( \sigma^C \) are not required under the growth conditions tested. In addition, the loss of function in *sigC* did not affect the survival of the mutant strain under superoxide, hydroperoxide, nitric oxide, or acidic pH conditions in relation to its parental strain, H37Rv. These findings suggest that \( \sigma^C \) is not likely to be involved in the ROI and RNI responses in *M. tuberculosis*, although other ECF sigma factors \( \sigma^E \) and \( \sigma^H \) were found to be involved in a variety of host-induced stress responses in
*M. tuberculosis*, including heat shock and oxidative stress (Fernandes et al. 1999; Manganelli et al. 2002; Manganelli et al. 2001; Wu et al. 1997).

The mouse infection studies revealed that, compared to its wild-type parental strain H37Rv, the ΔsigC mutant survived to a higher bacillary load in the lungs and spleens of the BALB/c mice, particularly during the late stages of infection. Additionally, the near-identical in vivo growth pattern of the sigC-complemented strain ΔsigCcomp when compared with the parental strain, H37Rv, confirms that the observed increased bacillary load in the ΔsigC mutant-infected mouse organs was the direct result of the deletion of the sigC gene. During the mouse infection studies being investigated for this thesis Sun et al. published the results of a similar study (Sun et al. 2004). Sun et al. showed in their mouse infection models that mice infected with a *M. tuberculosis* ΔsigC mutant resulted in delayed time to death. However, in contrast to the mouse infection data in this thesis, the ability of the mutant to grow in the lungs was the same as that of the wild-type and their sigC-complemented strain through their aerosol infection model. These differences of results could be due to several reasons. Firstly the wild-type strain used for mouse infection study in this thesis is H37Rv, while Sun et al. used strain CDC1551; this latter strain was shown to be less virulent than H37Rv in animal models (Bishai et al. 1999). Also, the mouse strains were different: the mice used for the mouse infection study in this thesis were BALB/C mice, while Sun et al. used the DBA/2 inbred strain. The former are more resistant to *M. tuberculosis* that the latter (Medina et al. 1998). Finally, their comparative data presented in a graph between their ΔsigC mutant and their wild-type and complemented strains were from aerosol infection, a different route of infection than the intravenous injection conducted for the mouse infection study in this thesis. In
their intravenous injection study, Sun et al. did not show the comparative data between their ΔsigC mutant strain and their wild-type. However, they did conclude in their intravenous infection study that their ΔsigC mutant was able to proliferate and survive at high titres in mouse lungs for a prolonged interval.

In spite of these differences, the conclusions from both groups are complementary in the intravenous injection model in that the ΔsigC mutant was able to proliferate and then survive at high bacillary load in the lungs and spleens of the mice. Additionally, Sun et al. have shown that the pattern of attenuation of the ΔsigC mutant belongs to the recently identified immunopathology category (Hingley-Wilson et al. 2003; Kaushal et al. 2002). Mutants of this class grow and persist in mouse lungs, but fail to cause mortality as rapidly as the wild-type. Moreover, the histopathologic analysis identified that the ΔsigC mutant infected mice had reduced inflammatory infiltrates compared to mice infected with the wild-type or the complemented strains. These findings suggested that mortality in mice resulting from the wild-type and complemented strains were due to greater alveolitis in lungs as compared to those infected with the mutant strain. This immunopathological-dysfunction phenotype has also been observed for other ECF sigma factor mutants in M. tuberculosis: ΔsigE (Ando et al. 2003), ΔsigF (Geiman et al. 2004), and ΔsigH (Kaushal et al. 2002). In addition, strains lacking the whiB3 gene, which functions as a transcription regulator (Steyn et al. 2002), had also exhibited the characteristics of this phenotype. The immunopathological phenotype findings suggest that this phenotype may result from several diverse defects in key expression pathways during infection, but are not necessary for bacterial survival in the post-immune host.
The *in vivo* transcription of *sigC*, assessed by primer extension during this study, has identified a putative promoter region and transcriptional start site for *sigC* in *M. tuberculosis*. The absence of a primer extension product in the *ΔsigC* strain suggests that expression of this gene is dependent on RNA polymerase containing $\sigma^C$, therefore it suggests autoregulation. This finding is supported by data published by Sun *et al.* (Sun *et al.* 2004). Based on their quantitative RT-PCR analysis of the expression of *sigC* in their *ΔsigC* strain during early growth phase they have deduced that the expression of *sigC* is largely autoregulated. This autoregulation phenomenon is also observed in another *M. tuberculosis* ECF $\sigma$ factor, $\sigma^H$ (Fernandes *et al.* 1999), which plays a central role in a network that regulates heat and oxidative-stress responses that are likely to be important in *M. tuberculosis* pathogenesis (Raman *et al.* 2001). A review on bacterial ECF sigma factors by John Helmann concluded that ECF $\sigma$ factors often positively regulate their own synthesis (Helmann 2002). However, both the primer extension study and RT-PCR study are based on analysis of mRNA levels, and the possibility remains that such an effect could be a consequence of reduced stability of the truncated *sigC* mRNA in the mutant strain. The use of transcriptional fusions to compare expression in the strains would provide one way of avoiding this limitation.

The coincidental location of the transcription start point of *sigC* identified in this study (A) and the predicted translation initiation codon of *sigC* (ATG), although relatively uncommon in *M. tuberculosis*, is not without precedent. The *M. tuberculosis purC* gene, which is required for purine biosynthesis, shares the same property (Jackson *et al.* 1996). This phenomenon was also reported in other mycobacterial species, for example the P1 promoter of the *M. leprae oxyR* gene
(Dhandayuthapani et al. 1997), and the \textit{M. fortuitum} \textit{blaF} gene, which encodes a β-lactamase (Timm et al. 1994). The latter study found that the fusion of reporter genes to the \textit{M. fortuitum} \textit{blaF} start codon resulted in efficient expression of these genes in the absence of a ribosome binding site. Approximately 35-40 leaderless mRNAs have also been identified in other bacteria (Moll et al. 2002). Studies in \textit{E. coli} have brought to light that the ratio of initiation factors IF2 and IF3 plays a decisive role in translation initiation of leaderless mRNA, indicating that the translational efficiency of this mRNA class can be modulated depending on the availability of components of the translational machinery (Moll et al. 2002). In the \textit{Streptomyces} genus, 11 actinomycete genes were found to display this transcription-translation feature (Strohl 1992). A possible explanation for the rather common occurrence of such a gene structure in streptomycetes, which are also in the same bacterial order as the mycobacteria, i.e. actinomycetales, is the possibility that ribosomes in these organisms may not require extended complementarity of the mRNA template and the 3' end of the 16S rRNA for efficient peptide translation.

The putative promoter sequence identified for \textit{M. tuberculosis} \textit{sigC} in this study, TaGACt-N_{17}-TATggT (identical nucleotides are in upper case), is identical to the \textit{M. tuberculosis} consensus promoter sequence proposed by Mulder et al. (Mulder et al. 1997), TTGACG and TATAAT, with four out of six positions identical at both the -10 and -35 regions. The -10 and -35 hexamers in \textit{sigC} were also similar to the \textit{E. coli} σ^{70} consensus promoter motifs, TTGACA and TATAAT (Mulder et al. 1997), with four out of six positions identical at both the -10 and -35 regions.
The conservation of these promoter elements suggests that the sigC promoter may be recognized by the principal sigma factor $\sigma^A$, or the related principal-like $\sigma$ factor, $\sigma^B$. Alternatively, $\sigma^C$ may also recognize a similar sequence to the *M. tuberculosis* consensus and *E. coli* $\sigma^{70}$ consensus promoter elements. Earlier studies on transcriptional signals of mycobacteria had revealed that strong promoters occur less frequently in the slow growing pathogen *M. tuberculosis* than the fast growing saprophyte *M. smegmatis* (Das Gupta *et al.* 1993). The higher GC content of *M. tuberculosis* promoters, as observed in the -10 region of sigC in this study, may have a bearing on the lower strength of its promoters relative to those of the fast growing *M. smegmatis* (Bashyam *et al.* 1996). Being an intracellular human pathogen, *M. tuberculosis* evades the immune response of the host by remaining dormant in macrophages; hence, weaker promoters may be advantageous to its pathogenic nature.

Whole-genome DNA microarray technology provides a robust tool to assess expression of many genes simultaneously (DeRisi *et al.* 1997). In this microarray study, 32 genes were identified as differentially expressed in the absence of $\sigma^C$ in *M. tuberculosis*. Of the 23 genes which showed elevated levels of expression in the $\Delta$sigC strain, 10 were identified to be involved in hypoxic responses when compared to the genes induced by hypoxia published by Sherman *et al.* (Sherman *et al.* 2001). Additionally, Park *et al.* and Voskuil *et al.* have shown that Rv3133c/dosR, which is induced by at least 7-fold in the $\Delta$sigC strain when compared to the wild-type H37Rv strain in this microarray study, is a transcription factor that mediates the hypoxic response of *M. tuberculosis*. The dormancy regulon, composed of at least 48 genes and controlled by the dormancy survival regulator DosR, is induced during conditions that inhibit aerobic respiration (oxygen deprivation and/or the presence of nitric
oxide) (Park et al. 2003; Voskuil et al. 2003). In this microarray analysis 12 of the genes involved in the dormancy regulon were identified (Rv0569, Rv1733c, Rv1813c, Rv1998c, Rv2004c, Rv2005c, fdxA, Rv2623, Rv2629, Rv2631, dosR and Rv3134c). The functionality of the DosR/DosS/DosT two-component system has been well studied (Roberts et al. 2004), as the dosS gene adjacent to dosR and another gene, dosT (Rv2027c) encode sensor kinases, which can autophosphorylate at a conserved histidine residue and then transfer a phosphoryl group to an aspartate residue of DosR, enhancing the binding affinity of DosR to its cognate DNA binding sequence. Recently the precise structure of the M. tuberculosis DosR and its DNA binding mechanism to induce gene activation during adaptation to hypoxic latency has been determined (Wisedchaisri et al. 2005). Of the remaining genes within the dormancy regulon that are not present in this microarray study, possible explanations could be that the dormancy mechanisms within M. tuberculosis are more complex than the 48-gene dormancy regulon suggested by Voskuil et al. (Voskuil et al. 2003), with the possibility that additional regulators are involved. Moreover, dual regulations of genes are common within the regulative mechanisms of M. tuberculosis. These results suggest that σ^C may play a role in the regulation of the hypoxic response in M. tuberculosis by regulating possible repressor(s) that exert antagonistic effects on the DosR/DosS/DosT two-component system, and subsequently the regulation of its downstream genes. However, microarray analysis of the sigC- complemented strain is required to verify that these observations are attributable to σ^C.

Amongst the other genes exhibiting elevated levels of expression in the ΔsigC strain are genes involved in heat shock response (dnaK, grpE and dnaJ/cpn60.1) and peptide folding mechanism under stress conditions (groEL/cpn60.1). Their gene products,
along with GroES (Cpn10), are collectively known as heat shock proteins. They are
normally induced when the cells are exposed to elevated temperatures or to ethanol or
heavy metal stress conditions. Heat shock proteins have been shown to play important
roles in the appropriate folding, assembly, transport and degradation of other proteins.
*M. tuberculosis* contains two copies of the *groEL* genes (*groEL1/cpn60.1* and
*groEL2/cpn60.2*) which belong to the Hsp60 family, with the *groEL2/cpn60.2* gene
arranged in an operon with the *groES/cpn10* gene, forming a *groESL* operon (Rinke
de Wit *et al.* 1992), which is negatively regulated by HrcA (Stewart *et al.* 2002).
Previous studies have shown that the GroEL1/Cpn60.1, identified in this microarray
study, is a powerful stimulator of pro-inflammatory cytokine production and may play
a role in the inflammatory pathology of tuberculosis (Lewthwaite *et al.* 2001). A
recent unusual finding of the *M. tuberculosis* GroEL1/Cpn60.1 protein is that it can
inhibit allergen-induced pulmonary eosinophilia in the mouse in a therapeutic manner;
hence it may serve as a potential vaccine target for the prevention of asthma in
humans (Riffio-Vasquez *et al.* 2004).

The *dnaK-grpE-dnaJ1* operon belongs to the Hsp70 family, which has been found to
be negatively regulated by HspR, a transcriptional repressor in *M. tuberculosis*
(Stewart *et al.* 2002). In their microarray studies, Stewart *et al.* observed elevated
expression levels of *sigC* along with the genes involved in the Hsp60 and Hsp70 regulons in their *ΔhspR* and *ΔhspRΔhrcA* strains when compared to the wild-type *M.
tuberculosis* H37Rv in exponential growth conditions. These findings complement the
microarray results from this study, suggesting that σ<sup>C</sup> may be involved in the heat
shock response of *M. tuberculosis*, and that σ<sup>C</sup> may be regulated by transcriptional
regulators HspR and HrcA.
Other than heat shock and hypoxic response genes, there were 4 genes (*fadA, fadB, hpx* and *fprB*) with elevated levels of expression in the ΔsigC strain that have been assigned with possible roles in *M. tuberculosis* (TubercuList 2006). The *fadA* and *fadB* genes belong to the group of genes that are involved in fatty acid degradation. These genes have the same orientation for gene expression and only 4 bp separate *fadA* and *fadB* in *M. tuberculosis*, suggesting that the *fadA* and *fadB* genes are expressed from the same promoter upstream of *fadA*. The precise function of *fadA* in *M. tuberculosis* has yet to be determined, although through sequence analysis *fadA* is highly similar to an acetyl-CoA thiolase (T35428) from *S. coelicolor* (TubercuList 2006). FadB, a probable fatty acid oxidation protein, is involved in fatty acid metabolism; it is also highly similar to the probable fatty acid oxidation protein T35429 from *S. coelicolor* (TubercuList 2006). Both *fadA* and *fadB* were found to be non-essential for growth in *M. tuberculosis* (Sassetti *et al.* 2003). Overall, these findings suggest that σ^C^ may be involved in the regulation of the *fadAB* operon, and hence σ^C^ may play a role in the regulation of fatty acid degradation in *M. tuberculosis*. The *Hpx* gene codes for a possible non-heme haloperoxidase in *M. tuberculosis*, which may be involved in detoxification reactions (TubercuList 2006) and was found to be non-essential for growth in *M. tuberculosis* (Sassetti *et al.* 2003); however, the precise role of Hpx has yet to be determined. The *fprB* gene codes for a probable NADPH-ferredoxin reductase that may be involved in the electron transfer mechanism (TubercuList 2006), and was found to be non-essential for growth in *M. tuberculosis* (Sassetti *et al.* 2003); however, the precise role of FprB has yet to be determined.
The specific roles of the remaining genes with elevated levels of expression in the ΔsigC strain (Rv0095c, Rv2033c and Rv2256c) have yet to be determined, though transposon mutagenesis studies in the wild-type H37Rv strain found Rv0095c and Rv2033c to be non-essential genes in *M. tuberculosis* (Sassetti et al. 2003), and the hypothetical protein Rv2256c was found to have conserved regions with hypothetical proteins in the *Streptomyces* family through computational sequence analysis (TuberculList 2006).

Of the 9 genes (*pntAB, rpsL, Rv1233c, Rv1461, mbtD, desA3, Rv3402, Rv3403, and Rv3822*) that exhibited significantly reduced levels of expression in the ΔsigC strain, *rpsL* has been most extensively studied. The *rpsL* gene has been found to encode the ribosomal protein S12, which is involved in the translation initiation step in *M. tuberculosis* and is essential for its survival (Sassetti et al. 2003). The promoter studies conducted by Kenney et al. have identified the putative promoter regions for *rpsL* (TTGACc and TATTgT; identical nucleotides to the *E. coli* σ\(^70\) consensus promoter are in upper cases), and discovered a remarkable similarity with the *E. coli* σ\(^70\) consensus promoter (Kenney et al. 1996; Mulder et al. 1997). These findings are also highly similar to the putative *sigC* promoter regions proposed in this study (TtGACc and TATtGT; identical nucleotides to the putative *sigC* promoter regions are in upper cases), suggesting that σ\(^C\) may be involved in the regulation of *rpsL*, possibly in addition to principal and/or principal-like σ factors.

The probable linoleoyl-CoA desaturase (DesA3) encoded by Rv3229c is thought to be involved in the lipid metabolism of *M. tuberculosis* (TuberculList 2006), and was found to be non-essential for growth (Sassetti et al. 2003). DesA3 belongs to a group
of three aerobic terminal desaturases, encoded by desA1-3, which function as catalysts for the unsaturation of mycolic acids. The reduced level of expression of Rv3229c in the ΔsigC strain suggests that σC may be involved in the downstream lipid metabolic regulation of M. tuberculosis.

Amongst the other genes exhibiting repressed levels of expression in the ΔsigC strain are pntAB and mbtD, which were assigned possible roles in M. tuberculosis. The pntAB gene encodes the membrane integral part of a probable NAD(P) transhydrogenase (TubercuList 2006), and is non-essential for growth in M. tuberculosis (Sassetti et al. 2003). The NAD(P) transhydrogenase, is involved in respiration and ATP hydrolysis and functions as a proton pump across the membrane. The gene mbtD encodes a polyketide synthetase, and is involved in mycobactin biogenesis. The mbt gene cluster in M. tuberculosis was identified by Cole et al. (Cole et al. 1998), which included 8 mbt genes, mbtA to mbtH, all within the same genetic locus. Seven of the genes (mbtB to mbtH) were predicted to be expressed as part of the same operon (Cole et al. 1998), and mbtB was inactivated in another study, where the ΔmbtB mutant was incapable of producing any siderophores and was avirulent in macrophages (De Voss et al. 2000). This gene deletion study suggested that all siderophores are generated from the same biosynthetic pathway, and that siderophores are essential for the in vivo growth of M. tuberculosis. The precise regulative mechanisms of σC towards pntAB and mbtD, have yet to be determined. Further microarray analysis with a sigC-complemented strain would confirm that these observations are attributable to σC.
Of the 5 hypothetical proteins (Rv1233c, Rv1461, Rv3402c, Rv3403c, and Rv3822) that showed significant levels of repression in the ΔsigC strain, Rv1233c and Rv1461 are best understood. The conserved hypothetical membrane protein Rv1233c has been shown to be non-essential for growth in *M. tuberculosis* (Sassetti et al. 2003). It was found to have an increased level of expression in the microarray analysis of the Δ*hspRΔhrcA* strain conducted by Stewart et al. (Stewart et al. 2002), along with *sigC*. This suggests that Rv1233c is negatively regulated by transcriptional repressors, as well as being regulated by *σ^C*. The conserved hypothetical protein Rv1461 is essential for growth in *M. tuberculosis* (Sassetti et al. 2003) and contains an intein, though the mechanism of its post-translational splicing and its precise function has not been explored (Sassetti et al. 2003). The reduced level of expression of Rv1461 in the ΔsigC strain suggests that Rv1461 may be regulated by *σ^C*. For the remaining hypothetical proteins Rv3402c, Rv3403c and Rv3822, their roles in *M. tuberculosis* have yet to be determined, although they were all found to be non-essential for growth in *M. tuberculosis* (Sassetti et al. 2003). The results from this microarray study suggest that Rv3402c and Rv3403c may belong to an operon that is regulated by *σ^C*.

One surprising observation from this microarray analysis was the absence of *sigC* in the list of differentially expressed genes. This may be due to technical limitations during the microarray spot printing process, as the *sigC* spot did not hybridise well with either the Cy3- or Cy5-labelled samples, and the Cy3-labelled samples were of the *M. tuberculosis* H37Rv genomic DNA origin, which should result in a uniform intensity for each gene on the array. Furthermore, the lists of differentially expressed genes generated in this microarray analysis are dissimilar from the data published by Sun et al. (Sun et al. 2004) at the exponential growth phase. These differences of
results could be due to several reasons. Firstly, there may be physiological variations in the RNA samples between the two experiments due to different RNA purification methods. Secondly, the statistical analysis and filtering methods may differ between the two microarray studies; and thirdly, the differences could possibly be due to different strains being used in the two experiments. Sun et al. used the \emph{M. tuberculosis} strain CDC1551 while the microarray study conducted for this thesis used the parental strain H37Rv. Additionally, the \emph{sigC}-complemented microarray analysis was not performed by Sun et al., so it remains to be seen that the genes that were differentially expressed in their microarray analysis were attributable to $\sigma^C$.

Further microarray analyses of the \emph{sigC}-complemented strain, $\Delta$\emph{sigC}comp, when compared with the wild-type H37Rv and the $\Delta$\emph{sigC} mutant strains during exponential growth conditions would confirm that the differential expressions observed in this microarray analysis are due to the mutation of \emph{sigC}. Quantitative-RT-PCR studies involving the genes with differential expression would verify the data obtained from the microarray analysis.

In summary, a hypothetical model of the functional characteristics of $\sigma^C$ has been deduced from this study, shown in Figure 3.5.1. Defining the upstream and downstream players involved in the regulation $\sigma^C$ would have a great impact on the further understanding of the complex role of $\sigma^C$ in \emph{M. tuberculosis}.
Figure 3.5.1 A hypothetical model of the functional characteristics of $\sigma^C$ in *M. tuberculosis.*
CHAPTER 4: Molecular characterisation of the sigB locus; generation of the sigB gene replacement construct

4.1 Generation of *M. tuberculosis* ΔsigB::hyg replacement construct

To determine the role of sigB in *Mycobacterium tuberculosis*, the coding sequence of the gene had to be disrupted. Previous attempts to generate a *M. tuberculosis* sigB null mutant strain with an unmarked sigB replacement construct (Gopaul 2002) were unsuccessful within the department, and hence a different strategy was chosen for this study.

In the previous study (Gopaul 2002), the strategy to disrupt the sigB coding sequence was to delete the majority of the wild type coding sequence via the inverse PCR technique, and to introduce an *EcoRV* restriction site within the deleted region. An illustration of this sigB-gene deletion construct, pBAlmedie2, is shown in Figure 4.1.1 using the primer sequences shown in Table 4.1.1. In this study an antibiotic resistance cassette (Hyg\(^R\)) from the vector pUC-hyg (Mahenthiralingam *et al.* 1998) was introduced within the disrupted sequence, illustrated in Figure 4.1.2. The presence of an antibiotic resistance marker would allow the sigB-disrupted strains to acquire resistance to hygromycin. Hygromycin was preferred as a selectable marker than kanamycin as it was found to be more versatile in transformation and selection (Garbe *et al.* 1994), and thus would provide a more efficient screening step for the single crossover recombinants and mutants.
Figure 4.1.1 A schematic diagram of the strategy previously used to attempt to delete the *M. tuberculosis* `sigB` coding sequence. 794 bp of the coding sequence were deleted by inverse PCR technique (primers shown in red), with 108 bp and 70 bp of coding sequences remaining at the 5' and 3' ends respectively (Gopaul 2002).

<table>
<thead>
<tr>
<th>Description</th>
<th>Primer sequences (5'-3')</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>sigB</code> inversed PCR primers</td>
<td>GATATCCGAGCGCGACGTGAGTAAGC CTTGCCGATGGCCGTTCAGATAGA</td>
<td>7370</td>
</tr>
</tbody>
</table>

Table 4.1.1 Sequences of the primers used for the inverse PCR to make a targeting construct for the deletion of `sigB` (Gopaul 2002).

Figure 4.1.2 A schematic diagram of the strategy adopted in this study in the attempt to generate a `sigB` null mutant in *M. tuberculosis*. The plasmid pBC1 was generated by the insertion of a hygromycin antibiotic resistance marker within the disrupted region of the `sigB` coding sequence, via the *EcoRV* restriction site.
The generation of the sigB gene replacement constructs is illustrated briefly in Figure 4.1.3. The vector pBackbone that generated pBalmedie2 (Gopaul 2002) was derived from a commercial phagemid pBluescript II KS- (Stratagene, La Jolla, USA) (see Appendix I for plasmid details) by the addition of a kanamycin resistance marker from the vector pUC4K (Pharmacia/Pfizer, Tadworth, UK), and the introduction of a unique restriction site for PacI as well as deletion of the lacZ coding sequence within the phagemid (Gopaul 2002). Approximately 1.6 kb of the 5' and 3' sigB flanking regions were retained in the construct generated by inverse PCR. In this study, the hygromycin antibiotic resistance gene was digested with SmaI from pUC-hyg (approximately 1.48 kb), and blunt-end cloned into the EcoRV site between the sigB flanking regions (Figure 4.1.2), thus generating the plasmid pBC1. Subsequent verifications from restriction digestions of pBC1 showed that the insertion of the hygromycin resistance cassette was in the opposite transcriptional orientation to that of the sigB coding sequence. Additionally, the lacZ-sacB selection marker obtained from the plasmid pGOAL17 (Parish et al. 2000a) was introduced into the PacI site of pBC1 to generated pBC2, in order to allow efficient counter selection and screening steps during the mutant isolation. The correct flanking sequences adjacent to sigB in the final sigB gene replacement construct pBC2 were verified by sequencing with the primers shown in Table 4.1.1.
Figure 4.1.3 Summary of the cloning strategy for the construction of the marked \( \text{sigB} \) gene replacement construct. \( \text{Hyg}^R \): hygromycin resistance marker; \( \text{Amp}^R \): ampicillin resistance marker; \( \text{Kan}^R \): kanamycin resistance marker; \( \text{LacZ} \): \( \beta \)-galactopyranosidase marker; \( \text{SacB} \): sucrose sensitive marker. The orientation of the expression of the hygromycin resistance marker was found to be opposite that of the \( \text{sigB} \) flanking region.
4.2 Isolation of single crossover strains for the generation of $\Delta$igB::hyg

In this sigB gene replacement experiment a two-step selection method was used (Pelicic et al. 1996a), illustrated in Figure 4.2.1. This method was advantageous as it overcame the problem of the low recombination efficiency found during a direct, one-step selection by initially selecting for the higher occurrence single recombination event.

The sigB gene replacement construct pBC2 was electroporated into the wild type M. tuberculosis H37Rv competent cells using 3 μg of plasmid per 400 μl of M. tuberculosis competent cells. A total of 4 transformations were performed. The transformed M. tuberculosis cells were plated onto 7H11 agar plates with 50 μg/ml hygromycin and 100 μg/ml X-GAL selections, as described in section 2.2, and incubated at 37°C for 4 weeks. The vector pMV306 (Stover et al. 1991) which carries a kanamycin cassette was used as a measure of transformation efficiency for the electroporation experiment, with an average transformation efficiency of $6.4 \times 10^5$ per μg of DNA being achieved. A total of 85 blue colonies and 10 white colonies were obtained from the 4 transformations.

Twelve blue colonies were selected for further study and allowed to grow on fresh ¼ 7H11 agar plates with 50 μg/ml hygromycin and 100 μg/ml X-GAL selections for 2-3 weeks. Although the blue colonies selected here demonstrated the phenotype of strains with an integrated sigB gene replacement plasmid, this could not exclude the possibility that the plasmid construct had illegitimately recombined into the M.
Figure 4.2.1 A schematic diagram representing the selection process for allelic exchange mutants in *M. tuberculosis*. Selection normally occurs in 2 stages, unless double cross-over mutants have occurred during the first stage of selection. SCO: Single cross-over recombinants; DCO: double cross-over mutants; Hyg\(^R\): hygromycin resistance marker; Kan\(^R\): kanamycin resistance marker; Kan\(^\beta\): kanamycin sensitive; *lacZ*: β-galactopyranosidase marker; *sacB*: sucrose sensitive marker; hyg: hygromycin; Kan; kanamycin; Xgal: 5-bromo-4-chloro-3-indoly-β-D-galactopyranoside; sucrose: 2% sucrose (v/v).
*tuberculosis* genome. Hence, to identify colonies that had undergone legitimate recombination into the wild type *sigB* locus, a PCR-based screening method was used, as illustrated in Figure 4.2.2, using the primer sequences shown in Table 4.2.1.

Primer AS was located within the wild type *M. tuberculosis* genome but external to the 5' *sigB* flanking region of the gene replacement construct pBC2. Primer AT was located within the hygromycin resistance marker gene of pBC2. A 4.2 kb PCR product from primers AS and AT could only be produced if the recombination had occurred within the 5' region of the wild type *sigB* locus. Similarly primer AV was located externally to the 3' *sigB* flanking region of pBC2 with primer AU being located within the hygromycin resistance gene. A PCR product of 5 kb would confirm a legitimate recombination at the 3' end of the wild type *sigB* locus.

The genomic DNA from the twelve blue colonies was extracted by the methods described in section 2.10, and screened using this PCR method. The results are shown in Figures 4.2.3 and 4.2.4. Two single crossover isolates (3 and 8) showed the expected PCR fragment size of 4.2 kb from primers AS and AT, which screened for insertion upstream of the *sigB* locus (Figure 4.2.3). No recombinant isolates were observed to have an insertion downstream of the *sigB* locus, as the expected PCR product of 5 kb was absent from all of the isolates (Figure 4.2.4). Single crossover isolate 3 was then used in the subsequent counter selection steps.
Figure 4.2.2  Schematic diagrams showing the PCR-based screening method for the legitimate recombination of \textit{sigB}. The primers are shown in green and orange, along with the relevant primer names. (A) Primers AS and AT were located within the \textit{M. tuberculosis} genome external to the 5' \textit{sigB} flanking region from pBC2 and within the sequence coding for the hygromycin resistance marker respectively, to screen for 5' single recombinants. (B) Primers AU and AV were located within the sequence coding for the hygromycin resistance marker and within the \textit{M. tuberculosis} genome external to the 3' \textit{sigB} flanking region from pBC2 respectively, to screen for 3' single recombinants. The base pair numbers are the size of the expected PCR products.

<table>
<thead>
<tr>
<th>Description</th>
<th>Primer sequences (5'-3')</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{sigB} 5' crossover primers (AS,AT)</td>
<td>CCTGCGCGACTACCTGGACTGAGA CTCTATTCACAGGGTTACGGGCGGG</td>
<td>4208</td>
</tr>
<tr>
<td>\textit{sigB} 3' crossover primers (AU,AV)</td>
<td>GTCCAGCAGCGGCGGGCGGAGAGGT TGAACACTGGACTACCGCTCGCG</td>
<td>5027</td>
</tr>
</tbody>
</table>

Table 4.2.1  Sequences of primers used for PCR-based screening methods for the legitimate recombination of \textit{sigB}.
Figure 4.2.3  Agarose gel showing the PCR products obtained using the genomic DNA extracted from twelve blue single recombinants from the targeting construct pBC2 as templates. Primers AS and AT were used to screen for insertion upstream of the sigB locus. PCR products of 4.2 kb were observed from isolate 3 and isolate 8. (M)-left: λHindIII marker; (M)-right: λHindIII/EcoRI marker

Figure 4.2.4  Agarose gel showing the PCR products obtained using the genomic DNA extracted from twelve blue single recombinants from the targeting construct pBC2 as templates. Primers AU and AV were used to screen for insertion downstream of the sigB locus. No PCR product of 5 kb was observed. (M): λHindIII marker
4.3 The selection for ΔsigB::hyg from the single recombinants

During the first selection step of the sigB gene replacement strategy, ten white colonies were also observed. The possibility of a recombination event that resulted in the replacement of the wild type sigB coding sequence with the gene replacement construct from the first selection step was highly improbable. Moreover, subsequent selection steps would require an incubation time of 3-4 weeks to observe any phenotypic differences from the colonies; hence there was sufficient time to verify whether the white colonies obtained from the first selection step were ΔsigB mutants. A simple and effective PCR-based screening method was developed to allow potential ΔsigB mutants to be identified from these colonies, as illustrated in Figure 4.3.1 and using the primer sequences shown in Table 4.3.1.

Primers AM and AN were located within the deleted region of the sigB coding sequence in the gene replacement construct pBC2. A 534 bp PCR product from the primers could only be produced if the wild type sigB coding sequence was still present within the genome of the white colonies.

Genomic DNA of the ten white colonies was isolated by using the methods described in section 2.10, and then screened using this PCR method. The results are shown in Figure 4.3.2. The genomic DNA samples from all of the white colonies produced the PCR product of 534 bp, confirming the presence of wild type sigB coding sequences within their genomes. Thus, no ΔsigB mutant was obtained.
Figure 4.3.1 A schematic diagram showing the PCR-based screening method for *ΔsigB* mutant. Primers AM and AN were located within the deleted region of the *sigB* coding sequence (794 bp). A PCR product of 534 bp would only be produced if the wild type *sigB* coding sequence was present.

**Table 4.3.1** Sequences of primers used for PCR-based screening method for *ΔsigB* mutant.

<table>
<thead>
<tr>
<th>Description</th>
<th>Primer sequences (5'-3')</th>
<th>Product size (bp)</th>
</tr>
</thead>
</table>
| *ΔsigB* screening primers (AM,AN) | CGAACTGGCCAAGCGCATAGAAGC  
GAGTCTGGATATGCCGTCGCTC | 534               |

Figure 4.3.2 Agarose gel showing the PCR products obtained using genomic DNA extracted from ten white colonies isolated during the first selection for *ΔsigB::hyg* as templates. Primers AM and AN were used to screen for wild type *sigB* coding sequence. The 534 bp PCR products were present in all of the DNA samples from the white colonies and the wild type sample, but were absent in the gene replacement plasmid sample. Rv represents wild type *M. tuberculosis* H37Rv genomic DNA. P represents gene replacement plasmid construct pBC2. (M): 100 bp DNA ladder.
For the second stage of the generation of a ΔsigB mutant, the single crossover isolate 3 was serially diluted (0, -1, -2) in Dubos broth and plated onto 7H11 agar plates containing hygromycin, XGAL and sucrose. Twenty white colonies were obtained, and they were screened on 7H11 agar plates containing kanamycin to identify colonies with the phenotype expected for a ΔsigB mutant. If a second recombination event had occurred, the kanamycin resistance marker that resides within the vector plasmid sequence would be lost via allelic exchange, and hence the mutant strain would be sensitive to kanamycin exposure. Of the twenty white colonies derived from the single crossover isolate 3, nine were sensitive to kanamycin. The genomic DNA of these nine white colonies was isolated and screened using the PCR-based method shown in Figure 4.3.1. All nine colonies from the second selection step showed the PCR product of 534 bp, as shown in Figure 4.3.3. These results confirmed that the wild type sigB coding sequence was present within their genomes, and therefore a ΔsigB mutant had not been obtained.

The results from this sigB gene replacement experiment suggested that sigB may play an essential role in the survival of *M. tuberculosis*; in which case no gene replacement mutant could be obtained. However, this appeared to be highly improbable as a sigB null mutant was generated in *M. smegmatis* (Mukherjee *et al.* 2005), and the genomic organisation of the sigB locus is conserved in all mycobacterial species (Doukhan *et al.* 1995; Predich *et al.* 1995).
Figure 4.3.3  Agarose gel showing the PCR products obtained using genomic DNA extracted from nine white kanamycin sensitive colonies isolated from the second selection step as templates. The sample numbers given above the lanes correspond to the colony numbers of the second selection. Primers AM and AN were used to screen for wild type $\text{sigB}$ coding sequence. The 534 bp PCR products were present in all of the DNA samples from the white colonies and the wild type sample, but were absent in the gene replacement plasmid sample. Rv represents wild type $\text{M.tuberculosis H37Rv}$ genomic DNA. P represents gene replacement plasmid construct pBC2. (M)-left: 100 bp DNA ladder; (M)-right: $\lambda\text{HindIII/EcoRI}$ marker.
4.4 The influence of the downstream gene, ideR, on the generation of *M. tuberculosis* ΔsigB::hyg

Following the unsuccessful attempts to generate a *M. tuberculosis* ΔsigB mutant, a more detailed investigation within the *M. tuberculosis* sigB gene locus provided new insight into why a conventional gene replacement strategy was unlikely to be successful. An illustration of sigB and its adjacent genes is shown in Figure 4.4.1.

![Figure 4.4.1](image)

Figure 4.4.1 A schematic diagram of the sigB gene locus within the *M. tuberculosis* genome. Adapted from TubercuList (TubercuList 2006)

The two genes adjacent to sigB, Rv2709 and ideR, were found to be transcribed in the same orientation of sigB within the *M. tuberculosis* genome. The ideR gene immediately downstream to sigB is responsible for iron-dependent gene regulation (Dussurget *et al.* 1996). Subsequent to the initial unmarked sigB gene replacement construct being created within the department, ideR was found to be essential for iron-dependent gene expression, iron metabolism and oxidative stress response in *M. tuberculosis* (Rodriguez *et al.* 2002). As the distance between the sequences coding for sigB and ideR was 132 bp, it was generally believed that the genes were independently expressed (Hu *et al.* 1999). However, as no ΔsigB mutant was
generated from this study, the possibility that \textit{sigB} and \textit{ideR} may be co-transcribed was investigated with a view to provide new information for the gene replacement strategy.

A RT-PCR system, based on the reverse transcription of the wild type \textit{M. tuberculosis} H37Rv RNA, was used to investigate the co-transcription of \textit{sigB} and \textit{ideR} within the \textit{M. tuberculosis} genome during exponential growth phase (OD\textsubscript{600} \~ 0.3 – 0.4). The RT-PCR system for \textit{sigB} and \textit{ideR} is illustrated in Figure 4.4.2, and the primers that were used are listed in Table 4.4.1. The RT-PCR method is described in section 2.12.

Primer BK was located 70 bp from the 3' end of the \textit{sigB} coding sequence and primer BL was located 54 bp from the 5' end of the \textit{ideR} coding sequence. A PCR product of 256 bp would only be produced if \textit{sigB} and \textit{ideR} are co-transcribed and part of the same operon.
**Figure 4.4.2** A schematic diagram showing the RT-PCR system used to investigate the expression of *sigB* and *ideR* within the *M. tuberculosis* genome. Primer BK was located at the 3' end of the *sigB* coding sequence and primer BL was located at the 5' end of the *ideR* coding sequence. A PCR product of 256 bp would demonstrate that the two genes are co-transcribed.

**Table 4.4.1** Sequences of primers used for the RT-PCR system used to investigate the co-transcription of *sigB* and *ideR*.

<table>
<thead>
<tr>
<th>Description</th>
<th>Primer sequences (5'-3')</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>sigB-ideR</em> RT-PCR primers (BK,BL)</td>
<td>CGAGCGGCGACGGATGAGTAAG, TACCTGGACCGACCGACCTC</td>
<td>256</td>
</tr>
</tbody>
</table>
A RT-PCR product of 256 bp was observed from RNA that had been reverse transcribed, which was also observed in the genomic DNA positive control (Figure 4.4.3). Importantly, no product was obtained if the reverse transcriptase was omitted (RT-); thus, this negative control result showed that there was no genomic DNA contamination in the RNA samples.

These results suggested that *sigB* and *ideR* are co-transcribed during the exponential growth phase of *M. tuberculosis*. This finding provided valuable information in order to revise the *sigB* gene replacement strategy that was employed during this study. The insertion of the hygromycin resistance marker within the *sigB* coding sequence is likely to have created a polar effect on the expression of the downstream gene *ideR*, which is vital for the survival of *M. tuberculosis*. 
Figure 4.4.3 RT-PCR analysis of the co-transcription of sigB and ideR in *M. tuberculosis*. The 256 bp PCR product shown in the *M. tuberculosis* H37Rv genomic DNA positive control lane (far left) was comparable with the PCR product generated from the wild type *M. tuberculosis* H37Rv RNA that had been reverse transcribed (RT+). No product was observed using RNA where the reverse transcriptase was substituted with DEPC-treated water as a negative control (RT-). (M) 100 bp DNA ladder; (C) No template negative control.
The same RT-PCR system was used to investigate whether the upstream gene, \textit{Rv2709}, which is 175 bp away from the \textit{sigB} coding sequence, is also co-transcribed with \textit{sigB} and \textit{ideR} in \textit{M. tuberculosis}. This is shown in Figure 4.4.4, and the primers that were used are listed in Table 4.4.2.

Primer CE was located 179 bp from the 3' end of the \textit{Rv2709} coding sequence and primer CF was located 33 bp from the 5' end of the \textit{sigB} coding sequence. A PCR product of 387 bp would only be produced if \textit{Rv2709} and \textit{sigB} are co-transcribed and part of the same operon.

No PCR product of 387 bp was observed when the RNA was used as the template for reverse transcription, while the 387 bp PCR product was observed in the genomic DNA positive control (Figure 4.4.5). These results suggested that \textit{Rv2709} and \textit{sigB} are transcribed independently during the exponential growth phase of \textit{M. tuberculosis}, and are not part of the same operon as \textit{sigB} and \textit{ideR}, and hence the promoter region of \textit{sigB} and \textit{ideR} should be located directly upstream of \textit{sigB}.
Figure 4.4.4 A schematic diagram showing the RT-PCR system used to investigate the co-transcription of Rv2709 and sigB within the M. tuberculosis genome. Primer CE was located at the 3' end of the Rv2709 coding sequence and primer CF was located at the 5' end of the sigB coding sequence. A PCR product of 387 bp would demonstrate that the two genes are co-transcribed.

<table>
<thead>
<tr>
<th>Description</th>
<th>Primer sequences (5'-3')</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv2709-sigB RT-PCR primers (CE,CF)</td>
<td>CATGGATGGCCGTCTGAT CACAAGGCCACCACAAGC</td>
<td>387</td>
</tr>
</tbody>
</table>

Table 4.4.2 Sequences of primers used for RT-PCR system for the co-transcription of Rv2709 and sigB.
Figure 4.4.5  RT-PCR analysis of the co-transcription of Rv2709 and sigB in *M. tuberculosis*. The 387 bp PCR product was absent from the wild type *M. tuberculosis* H37Rv RNA that had been reverse transcribed (RT+), but present in the *M. tuberculosis* H37Rv genomic DNA positive control lane (far left). No product was observed using RNA where the reverse transcriptase was substituted with DEPC-treated water as a negative control (RT-). (M) 100 bp DNA ladder; (C) No template negative control.
4.5 Revised strategies for the generation of the *M. tuberculosis* ΔsigB replacement construct

A newly revised *M. tuberculosis* sigB gene replacement strategy was proposed, with an additional aim to preserve the expression of *ideR* downstream of *sigB* while functionally inactivating *sigB*. The new strategy is illustrated in Figure 4.5.1, and the primers that were used are listed in Table 4.5.1.

**Figure 4.5.1** A schematic diagram of the revised strategy for the deletion of *sigB*. The deletion of the *sigB* coding sequence was in-frame (6 bp remained on both ends of the *sigB* coding sequence); hence the expression of the downstream gene *ideR* would not be affected. 1.5 kb flanking lengths adjacent to *sigB* for homologous recombination were designed, and a total of 960 bp of the *sigB* coding sequence were deleted.

<table>
<thead>
<tr>
<th>Description</th>
<th>Primer sequences (5'-3')</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔsigB 5' primers</td>
<td>CCGGAATTCGCCACGACGCCACACGCTACG</td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>CGATCGGGAGGTGCATATGGCCGGATCCCGCG</td>
<td></td>
</tr>
<tr>
<td>ΔsigB 3' primers</td>
<td>CGCGGATCCAGCTGAAGCTGGACATCCTGAGCC</td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>GCCGACGGATCGTGCATAACGCCTCTAGAGCA</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.5.1** Sequences of primers used for the revised deletion strategy of *sigB*.  

117
Primers EO and EP were designed to generate a 5’ *sigB* flanking region of 1.5 kb, which included *EcoRI* and *BamHI* restriction sites on the 5’ and 3’ ends respectively for the cloning into the vector plasmid pBackbone. Primers EQ and ER were designed to generate a 3’ *sigB* flanking region of 1.5 kb, which included *BamHI* and *XbaI* restriction sites on the 5’ and 3’ ends respectively for cloning purposes. The deletion within the *sigB* coding sequence was kept in-frame (960 bp), and a 6 bp insertion containing a *BamHI* site was introduced to ensure directional cloning during the construction of the targeting construct, and to facilitate efficient screening for mutant strains.

The generation of the revised *sigB* gene replacement construct is illustrated in Figure 4.5.2. The 1.5 kb 5’ *sigB* flanking PCR fragment was digested with *EcoRI* and *BamHI* and inserted into the vector pBackbone. Subsequently the 1.5 kb 3’ *sigB* flanking PCR fragment was digested with *BamHI* and *XbaI* and inserted into the plasmid containing the 5’ *sigB* flanking region, generating the construct pBC19.

The *lacZ-sacB* selection marker from the plasmid pGOAL17 (Parish et al. 2000a) were introduced into the *PacI* region of pBC19 to allow counter-selection during the mutant isolation steps. The final revised *sigB* gene replacement construct pBC20 was verified by restriction digestions and sequencing using primers designed externally to the *sigB* coding sequence (listed in Table 4.5.2.). Unfortunately, due to time constraints, it was not possible to pursue the selection in *M. tuberculosis* using this new targeting construct.
Figure 4.5.2  Summary of the cloning strategy for the construction of an in-frame sigB gene replacement construct. Amp$^R$: ampicillin resistance marker; Kan$^R$: kanamycin resistance marker; LacZ: β-galactopyranosidase marker; SacB: sucrose sensitive marker. The transcriptional orientations of the sigB flanking regions were maintained in the same direction.

<table>
<thead>
<tr>
<th>Description</th>
<th>Primer sequences (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔsigB sequencing primers (DK,DL)</td>
<td>CCGGCTTTTGTGTTAATAACCCCAT</td>
</tr>
<tr>
<td></td>
<td>AGGTAGCAGACGGTATGCCGCCG</td>
</tr>
</tbody>
</table>

Table 4.5.2  Sequences of primers used for the sequencing of the sigB coding region of pBC20.
4.6 Discussion: the generation of the sigB null mutant in *M. tuberculosis*

The *sigB* gene of *M. tuberculosis* encodes a σ\textsuperscript{70}-like σ factor σ\textsuperscript{B}, and is found at the same locus in *M. tuberculosis*, *M. smegmatis*, and *M. leprae* (Doukhan *et al.* 1995; Predich *et al.* 1995), where it is located downstream of *sigA*. The amount of *sigB* transcript in *M. tuberculosis* was found to be increased upon exposure to high temperature, SDS, low aeration and during stationary phase (Hu *et al.* 1999; Manganelli *et al.* 1999). Previous studies have identified that σ\textsuperscript{A} is the principal σ factor essential for *M. smegmatis* (Gomez *et al.* 1998). In the same publication Gomez *et al.* has also stated to have functionally inactivated σ\textsuperscript{B} in *M. smegmatis*, although the experimental evidence of a *M. smegmatis ΔsigB* strain has only been shown in recent publications (Mukherjee *et al.* 2005; Mukherjee *et al.* 2005). Furthermore, Gomez *et al.* were said to have functionally inactivated σ\textsuperscript{B} in *M. tuberculosis* in early 2004, and stated that the mutant strain is sensitive to various environmental stresses, such as SDS-induced surface stress, heat shock, and oxidative stress. Nonetheless, the mutant strain could grow normally in human macrophages and was not attenuated in mice, although the experimental evidence of their *M. tuberculosis ΔsigB* strain has yet to be published (Manganelli *et al.* 2004). However, in contrast to these findings, Sassetti *et al.* have identified both *sigA* and *sigB* as essential genes required for optimal growth in *M. tuberculosis* via transposon mutagenesis studies (Sassetti *et al.* 2003). The essentiality of *sigB* in *M. tuberculosis* is yet to be determined; hence a deletion study of *sigB* was being conducted for this thesis.
A transcription study of the *M. tuberculosis* sigB has identified a transcriptional start site for sigB, 27 bp upstream of its translational start codon (Hu *et al.* 1999). Hu *et al.* have also identified an inverted repeat sequence, 20 to 61 bp downstream of the stop codon, and predicted the formation of a stem-loop structure that possibly acts as a transcriptional terminator; hence they suggested that the sigB gene was transcribed as a monocistronic message.

The initial *M. tuberculosis* sigB deletion studies via allelic replacement methods were based on these previous experimental findings, and an unmarked ΔsigB construct was designed within the department. The lack of a double cross-over mutant identified in previous studies (Gopaul 2002) prompted further enhancements in the selection process of the sigB deletion strategy. A marked sigB deletion construct with a hygromycin resistance marker within the disrupted sigB coding sequence would allow efficient identification of single and double crossover recombinants via additional antibiotic selection, and significantly reduce the number of colonies generated from the secondary selection step to be screened, as previous studies have suggested (Parish *et al.* 1999). An effective PCR-based screening method introduced in this study also allowed rapid identification of two 5' single crossover recombinants, and exclusion of the possibility that colonies chosen for further selections were due to illegitimate recombination. The lack of a ΔsigB mutant from amongst the colonies exhibiting the expected phenotype at the final screening stage suggests that it may not be possible to delete sigB using this approach.

Immediately downstream of sigB is the gene ideR, which encodes an iron-dependent transcriptional repressor. IdeR is found in both pathogenic and non-pathogenic
mycobacteria and is a closely related homologue of \( dtxR \) in \textit{Corynebacterium diptheriae} (Doukhan \textit{et al.} 1995; Manabe \textit{et al.} 2005). \( ideR \) has been found to be essential in \textit{M. tuberculosis}, as it cannot be disrupted unless a second copy of the gene is present or a secondary suppressor mutation occurs (Rodriguez \textit{et al.} 2002). Although Rodriguez \textit{et al.} did not fully understand the nature of the suppressor mutation, their complementation studies of the \( ideR \) mutation with a wild-type gene demonstrated that \( ideR \) is required for the regulation of iron metabolism in \textit{M. tuberculosis}, as their transcriptional studies of the iron-dependent profile of the \( ideR \) mutant compared with the parental and the complemented strains revealed that about one-third of the iron-regulated genes in \textit{M. tuberculosis} are regulated by IdeR.

These findings prompted additional analyses of the \( sigB \) locus in \textit{M. tuberculosis}. There is a distance of 132 bp between the end of the \( sigB \) coding sequence and the translational start codon of \( ideR \). Thus, it was possible that \( sigB \) and \( ideR \) were cotranscribed. Although transcription of \( sigB \) was suggested to be monocistronic by Northern analysis (Hu \textit{et al.} 1999), the accuracy of the sizing of the transcript may have been limited by smearing observed in these experiments. Furthermore, even if the major transcript of \( sigB \) were monocistronic, this does not preclude an additional transcript extending to cover \( ideR \). The putative terminator site proposed by Hu \textit{et al.} (Hu \textit{et al.} 1999) could be responsible for the majority of transcription with some transcription readthrough, which may result in the extension of the mRNA transcript to \( ideR \). Additionally, the inverted repeat sequence downstream of the \( sigB \) stop codon which could create a stem-loop structure was not followed by a poly-A region (4-8 A residues), which should play a crucial role in the weakening and dissociation of the template hybrid leading to the release of the mRNA transcript. Moreover, the
mechanism of bacterial transcription termination is complex and not all symmetrical G-C rich segments are termination sites.

The RT-PCR analysis conducted in this study showed that there is at least some co-transcription between $\text{sigB}$ and $\text{idcR}$ in $M. \text{tuberculosis}$, suggesting that an insertional mutation in $\text{sigB}$ may affect the expression of $\text{idcR}$. Therefore, a revised unmarked $\text{sigB}$-deletion strategy was designed, with the additional objective of creating an in-frame deletion within $\text{sigB}$ such that expression of $\text{idcR}$ should not be altered. The ability to generate unmarked mutations may also be advantageous towards generating new potential vaccine strains, where antibiotic resistance markers cannot be left in the chromosome.

This $\text{sigB}$-deletion study has provided valuable insights into the transcriptional complexity of $\sigma^B$ within $M. \text{tuberculosis}$. As $\text{sigB}$ is dispensable in $M. \text{smegmatis}$, and there are 78% and 91% similarities at the DNA and protein levels between $M. \text{tuberculosis} \text{sigB}$ and $M. \text{smegmatis} \text{sigB}$ respectively (Beggs et al. 1996), it is still logical to postulate that $\text{sigB}$ is dispensable in $M. \text{tuberculosis}$. A revised unmarked in-frame $\Delta\text{sigB}$ construct was produced during this study; further mutant selection studies will provide more information about the essentiality of $\sigma^B$ for the growth and survival of $M. \text{tuberculosis}$.
CHAPTER 5: Conclusion and future work

The success of *M. tuberculosis* as a pathogen is dependent upon its extensive ability to adapt to a wide range of environments, both within and outside the human host. This adaptation is a complex process which involves the detection of changes in the environment and the initiation of responses by switching on or off certain sets of genes. Transcription regulatory proteins, such as σ factors, are likely to play a significant role in mediating the adaptive responses of *M. tuberculosis* to various environmental conditions (Manganelli *et al.* 2004). The finding of multiple σ factors in *M. tuberculosis*, particularly those in the ECF family (Cole *et al.* 1998), suggests that this pathogen is well equipped for adaptation to environmental changes, which may be one of the reasons that *M. tuberculosis* can persist in human hosts for an extensive period of time. Although σ factors are frequently induced under various environmental conditions, their activities are regulated at the transcriptional, translational and post-translational levels (Helmann 1999). A common mechanism at the post-translational level of regulation is the reversible antagonistic mechanism of an anti-σ factor on its cognate σ factor through protein-protein interactions. During the presence of a specific stimulus, the anti-σ factor releases the σ factor, enabling it to bind to the core RNA polymerase, and thus induces its regulon in response to the environmental changes. The ECF family of σ factors are often co-transcribed with their cognate negative σ regulator (Hughes *et al.* 1998), as the gene immediately adjacent to the ECF σ factor often encodes an anti-σ factor. In *M. tuberculosis* several ECF σ factors have exhibited this phenomenon: σ^E^-RseA (Kang *et al.* 1999), σ^H^-RshA (Song *et al.* 2003), and σ^L^-RslA (Hahn *et al.* 2005). Additionally, the
alternative $\sigma$ factor $\sigma^F$ in \textit{M. tuberculosis} is part of a genetic cluster with an anti-$\sigma$ factor \textit{rsbW} directly upstream of \textit{sigF}, with the two genes being co-transcribed (DeMaio \textit{et al.} 1997). Moreover, Beaucher \textit{et al.} have identified two anti-anti $\sigma$ factors for $\sigma^F$ (RsfA and RsfB) in different regions of the \textit{M. tuberculosis} genome to regulate the $\sigma^F$-specific anti-$\sigma$ factor RsbW (Beaucher \textit{et al.} 2002). These findings suggest remote regulations in some of the regulative pathways of \textit{M. tuberculosis}.

The $\textit{sigC}$ gene in \textit{M. tuberculosis} is found to be monocistrionically transcribed through genome sequence analysis (TubercuList 2006); this genomic organisation is highly conserved with the $\textit{sigC}$ gene in another mycobacterial pathogen, \textit{M. leprae} (Leproma 2006). Additionally, $\sigma^C$ is one of 4 $\sigma$ factors ($\sigma^A$, $\sigma^B$, $\sigma^C$, $\sigma^E$) retained in \textit{M. leprae} (Leproma 2006) when compared with \textit{M. tuberculosis}, with 81.2% protein sequence homology in 181 overlapping amino acids (NCBI 2006). Moreover, $\sigma^C$ is also found to be conserved in other actinomycetes, for example, 5 putative $\sigma$ factors in \textit{S. coelicolor} (SCO0864, SCO0866, SCO0942, SCO2954 and SCO3323) have protein sequence homologies ranging between 31% to 49% with \textit{M. tuberculosis} $\sigma^C$; ECF $\sigma$ factors GdnS in \textit{S. hygroscopicus} and AdsA in \textit{S. griseus} have also had 48% and 31% protein sequence homologies with the \textit{M. tuberculosis} $\sigma^C$ respectively (NCBI 2006). However, $\sigma^C$ is distinctively absent in the non-pathogenic \textit{M. smegmatis} (TIGR 2006), suggesting $\sigma^C$ may play a role in the pathogenesis of \textit{M. tuberculosis}. One possible explanation for the monocistronic nature of $\textit{sigC}$ in \textit{M. tuberculosis} may be that regulation of $\textit{sigC}$ occurs by remote transcription regulators. The identification of putative promoter elements for $\textit{sigC}$ in this study and their similarities to the $\sigma^{70}$ consensus promoter motifs suggest that $\textit{sigC}$ may be regulated by the principal $\sigma$
factor $\sigma^A$ and/or the principal-like $\sigma$ factor $\sigma^B$ in *M. tuberculosis*. Alternatively, *sigC* may be autoregulated as suggested by the lack of a primer extension product from the $\Delta$*sigC* strain observed in this study. Autoregulation has also been observed for another ECF $\sigma$ factor in *M. tuberculosis*, $\sigma^H$ (Fernandes et al. 1999). In order to further investigate the hypothesis that *sigC* is autoregulated, it is necessary to determine that the difference observed was not due to *sigC*-mRNA instability in the $\Delta$*sigC* mutant strain. One approach to address this is to conduct transcriptional fusion studies of *sigC* with a promoter-less $\beta$-galactopyranosidase reporter gene (*lacZ*) in the wild-type and the $\Delta$*sigC* mutant strains; expression would be expected in the wild-type strain but not in the mutant. Furthermore, if this is the case, mutational studies of the constructs using site-directed mutagenesis within the promoter motifs of *sigC* should identify the nucleotides that are crucial for promoter recognition and binding of $\sigma^C$.

In this study, *sigC* was found to be a regulator of genes involved in numerous functions in *M. tuberculosis*, for example: heat shock, hypoxic response, fatty acid degradation, detoxification reactions, electron transfer, translation initiation, mycobactin biogenesis and ATP hydrolysis, as well as other unknown functions. Additionally, the identification of putative transcriptional regulators HspR and HrcA for $\sigma^C$ (Stewart et al. 2001) have further suggested that $\sigma^C$ is involved in complex regulatory mechanisms for environmental responses in *M. tuberculosis*. A hypothetical model was constructed in an attempt to understand possible roles of $\sigma^C$ using the results of this study (Figure 3.5.1). To confirm that the differentially expressed genes observed from the $\Delta$*sigC* strain are legitimately controlled by $\sigma^C$, microarray analysis of the *sigC*-complemented strain should reveal restoration of wild-type-like levels of expression of these genes.
The analyses of *in vitro* growth and survival of the *M. tuberculosis ΔsigC* strain revealed that the loss of $\sigma^C$ in the mutant strain did not affect growth in aerobic conditions or the adaptability of *M. tuberculosis* to the selected stress conditions *in vitro* - superoxide, hydroperoxide, nitric oxide or acidic pH. The data acquired from the microarray analysis indicate that *in vitro* stress studies under heat shock and hypoxic conditions might provide evidence that $\sigma^C$ is involved in these environmental responses. Additionally, in the mouse infection study, *ΔsigC* mutant was found to be able to proliferate and survive at high bacillary load in the lungs and spleens of the mice, when compared to the wild-type and complemented strains. An immunopathological phenotype was observed by Sun *et al.* from the histopathological study of a *M. tuberculosis ΔsigC* mutant (Sun *et al.* 2004). Therefore, further histopathological analyses of lung and spleen tissues infected with the strains used in this mouse infection study would be of interest to ascertain whether a similar phenomenon is observed with the *ΔsigC* mutant used in this study.

$\sigma^B$ is conserved in all mycobacterial species as well as other bacterial species due to its close resemblance with the principal $\sigma$ factor, $\sigma^A$ (NCBI 2006). In *S. coelicolor*, the principal $\sigma$ factor $\sigma^{hrdB}$ shares 61% of protein sequence homology with the *M. tuberculosis* $\sigma^B$, and the 3 principal-like $\sigma$ factors in *S. coelicolor*, $\sigma^{hrdA}$, $\sigma^{hrdC}$, $\sigma^{hrdD}$, have protein sequence homologies of 53%, 51% and 51% with the *M. tuberculosis* $\sigma^B$ respectively (NCBI 2006). In order to address the role of $\sigma^B$ in *M. tuberculosis*, the *sigB* gene was targeted for mutagenesis via homologous recombination.
The initial $\text{sig}B$ gene deletion strategy with the insertion of a hygromycin antibiotic resistance cassette into the deleted region of $\text{sig}B$ resulted in the lack of a $\Delta \text{sig}B$ mutant from amongst the colonies exhibiting the expected phenotype at the final screening stage. This suggested that it may not be possible to delete $\text{sig}B$ using this approach. The RT-PCR analysis conducted in this study showed evidence of co-transcription between $\text{sig}B$ and $\text{ide}R$ in $M. \text{tuberculosis}$, suggesting that an insertional mutation in $\text{sig}B$ may affect the downstream expression of $\text{ide}R$, which encodes an essential regulator for iron dependent gene expression, iron metabolism and oxidative stress response in $M. \text{tuberculosis}$ (Rodriguez et al. 2002). Therefore, a revised unmarked $\text{sig}B$-deletion strategy was designed, with the additional objective of creating an in-frame deletion within $\text{sig}B$ such that expression of $\text{ide}R$ should not be altered.

Overall, this study has provided additional information on the role of $\sigma^C$ in the environmental responses and pathogenesis of $M. \text{tuberculosis}$. The foundation for future construction of a $\text{sig}B$ deletion mutant in $M. \text{tuberculosis}$ has also been accomplished. Further understanding of the transcriptional regulatory network involved in the environmental adaptation and virulence of $M. \text{tuberculosis}$ will ultimately help to design new strategies with which to eradicate tuberculosis.
References


Mycobacterium tuberculosis mutant lacking the SigF alternate sigma factor and identification of SigF-dependent genes by microarray analysis." Infect Immun 72: 1733-1745.


http://genolist.pasteur.fr/Leproma/.


NCBI (2006). NCBI-Basic Local Alignment Search Tool (BLAST)


Park, H. D., Guinn, K. M., Harrell, M. I., Liao, R., Voskuil, M. I., Tompa, M.,
Schoolnik, G. K. and Sherman, D. R. (2003). "Rv3133c/dosR is a transcription
factor that mediates the hypoxic response of Mycobacterium tuberculosis."
sacB gene confers sucrose sensitivity on mycobacteria." J Bacteriol 178: 1197-
1199.
polymerase and two sigma-factor genes from Mycobacterium smegmatis."
regulation by the Mycobacterium tuberculosis alternative sigma factor SigD
Raman, S., Song, T., Puyang, X., Bardarov, S., Jacobs, W. R., Jr. and Husson, R. N.
(2001). "The alternative sigma factor SigH regulates major components of
oxidative and heat stress responses in Mycobacterium tuberculosis." J
tuberculosis pathogenesis." Division of Mycobacterial Research, NIMR.
Riffo-Vasquez, Y., Spina, D., Page, C., Tormay, P., Singh, M., Henderson, B. and
bronchial eosinophilia and hyper-responsiveness in a murine model of allergic


TIGR (2006). Mycobacterium smegmatis MC2 genome list, TIGR


http://genolist.pasteur.fr/TuberculList/


*Arch Biochem Biophys* 298: 452-457.
**Appendix I: List of plasmids used in this study**

A list of vector plasmids used in this study:

<table>
<thead>
<tr>
<th><strong>Vector plasmids</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>pBluescript II KS-</td>
<td>A commercial vector possessing the <em>lacZ</em> gene and ampicillin resistance gene (Stratagene, La Jolla, USA)</td>
</tr>
<tr>
<td>pBackbone</td>
<td>Modified pBluescript II KS- with 159 bp deletion of the <em>lacZ</em> gene and the introduction of a unique PacI site within the deleted region; A 1256 bp Kan$^R$ cassette from pUC4K (Pfizer, Tadworth, UK) was also introduced into the Sall restriction site of the construct (Gopaul 2002).</td>
</tr>
<tr>
<td>pKP149</td>
<td>An <em>Xbal</em>-digested fragment of pUC-GM (Schweizer 1993) cloned into the <em>NheI</em>-SpeI site of pMV306 (Hess <em>et al</em>. 1998), replacing Kan$^R$ with Gm$^R$</td>
</tr>
<tr>
<td>pGOAL17</td>
<td>Based on pBluescript II but carrying a 6.367kbp double marked <em>lacZ</em>- <em>sacB</em> cassette which could be removed by PacI digestion (Parish <em>et al</em>. 2000a).</td>
</tr>
<tr>
<td>pUC-hyg</td>
<td>Vector carrying a 1596 bp Hyg$^R$ gene cassette (Mahenthiralingam <em>et al</em>. 1998)</td>
</tr>
</tbody>
</table>
A list of plasmids used and constructed in this study:

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Vector</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCulter2</td>
<td>pBackbone</td>
<td>A sigC deletion construct with 410 bp of sigC CDS deleted via inverse PCR (Gopaul 2002).</td>
</tr>
<tr>
<td>ΔsigCcomp</td>
<td>pKP149</td>
<td>A sigC CDS fragment with 164 bp of its upstream sequences ligated into the EcoRV site of pKP149 (Gopaul 2002).</td>
</tr>
<tr>
<td>pBalmedie2</td>
<td>pBackbone</td>
<td>A sigB deletion construct with 794 bp of sigB CDS deleted via inverse PCR, and the introduction of an EcoRV site within the deleted region (Gopaul 2002).</td>
</tr>
<tr>
<td>pBC1</td>
<td>pBalmedie2</td>
<td>A SmaI-digested 1.48 kb Hyg^R cassette from pUC-hyg was cloned into the EcoRV site of pBalmedie2.</td>
</tr>
<tr>
<td>pBC2</td>
<td>pBC1</td>
<td>A PacI-digested 6.37 kb double marked lacZ-sacB cassette from pGOAL17 was cloned into the PacI site of pBC1.</td>
</tr>
<tr>
<td>pBC19</td>
<td>pBackbone</td>
<td>1.5 kb PCR products of both the 5' and 3' sigB flanking regions were cloned into pBackbone via the EcoRI-BamHI-XbaI restriction sites.</td>
</tr>
<tr>
<td>pBC20</td>
<td>pBC19</td>
<td>A PacI-digested 6.37 kb double marked lacZ-sacB cassette from pGOAL17 was cloned into the PacI site of pBC19.</td>
</tr>
</tbody>
</table>
## Appendix II: List of primers used in this study

A list of primers used and constructed in this study:

<table>
<thead>
<tr>
<th>Description</th>
<th>Primer Sequences (5'-3')</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sigC RT-PCR primers</td>
<td>GCTCCAGCGCCCGAACTTG</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td>CTGCTCGGCTGTCCCTATGCCG</td>
<td></td>
</tr>
<tr>
<td>sigC primer for primer extension</td>
<td>ACGTCTTGCTGGGTTGCTTTTGATAA</td>
<td></td>
</tr>
<tr>
<td>sigB inversed PCR primers</td>
<td>GATATCCGAGCGCGACGTGATGAGTAAGC</td>
<td>7370</td>
</tr>
<tr>
<td></td>
<td>CTGCGGATGCCCCGTACGATAGA</td>
<td></td>
</tr>
<tr>
<td>sigB 5’ crossover primers (AS,AT)</td>
<td>CCTGCGGACTACCTGGACTGAGA</td>
<td>4208</td>
</tr>
<tr>
<td></td>
<td>CTCTATTCACAGGGTACGGGCGGGGGG</td>
<td></td>
</tr>
<tr>
<td>sigB 3’ crossover primers (AU,AV)</td>
<td>GTCCAGAGCCCGGGGCGAGAGT</td>
<td>5027</td>
</tr>
<tr>
<td></td>
<td>TGAACCTACTGGACTACCGCTCGCG</td>
<td></td>
</tr>
<tr>
<td>AsigB screening primers (AM,AN)</td>
<td>CGAACGCGGACGTGATGAGTAAG</td>
<td>534</td>
</tr>
<tr>
<td></td>
<td>GAGTCTGAGATGCTGCAGGTCCTC</td>
<td></td>
</tr>
<tr>
<td>sigB-ideR RT-PCR primers (BK,BL)</td>
<td>CGAGCGCGGACGTGATGAGTAAG</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>TACCTCGGACCATCTACGACCCT</td>
<td></td>
</tr>
<tr>
<td>Rv2709-sigB RT-PCR primers (CE,CF)</td>
<td>CATGAGTCGGCCGTTCTGAT</td>
<td>387</td>
</tr>
<tr>
<td></td>
<td>CACAAAGGGCCCCACCAAGC</td>
<td></td>
</tr>
<tr>
<td>AsigB 5’ primers (EO,EP)</td>
<td>CCGGAATTCCGACAGCCCCAGACACCGCTACG</td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>CGATCGGAGTGCTGCTATGCGCCGGATCCCGG</td>
<td></td>
</tr>
<tr>
<td>AsigB 3’ primers (EO,ER)</td>
<td>CGCGGATCCAGCCTGTACCGACATCTCAGCC</td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>GCCGCAGGATCGTCGATAACGCCTCTAGAGCA</td>
<td></td>
</tr>
<tr>
<td>AsigB sequencing primers (DK,DL)</td>
<td>CCGGCTTTTGTCGTAAACCCCCCAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGGTAGCAGACCGGTATGCCCCGCGG</td>
<td></td>
</tr>
</tbody>
</table>
Appendix III: Media, solutions and gel

Media

Luria-Bertani broth (L-broth)

- Tryptone (Difco, Plymouth, UK) 10 g/l
- Yeast extract 5 g/l
- NaCl 10 g/l

Adjust pH to 7.5 (NaOH), autoclave at 121°C for 15 minutes.

Luria-Bertani agar (L-agar)

- Tryptone (Difco, Plymouth, UK) 10 g/l
- Yeast extract 5 g/l
- NaCl 10 g/l
- Agar (Difco, Plymouth, UK) 15 g/l

Autoclave at 121°C for 15 minutes.

Dubos broth

- K$_2$HPO$_4$ 1 g/l
- Na$_2$HPO$_4$.12H$_2$O 6.25 g/l
- Na$_3$ citrate 1.25 g/l
- MgSO$_4$.7H$_2$O 0.6 g/l
- Asparagine 2 g/l
- Casamino acids (Difco, Plymouth, UK) 2 g/l
- 10% Tween 80 5 ml/l
- 10% Glycerol 20 ml/l
Adjust pH to 7.2 (NaOH), autoclave at 121°C for 15 minutes; once cool, add 40 ml albumin supplement (Difco, Plymouth, UK).

**Middlebrook 7H11 agar**

- Glycerol 5 ml/l
- Tween 80 0.05% (w/v)/l
- 7H11 medium powder (Difco, Plymouth, UK) 21 g/l

Autoclave at 121°C for 15 minutes; once cooled, add 100 ml OADC (oleic acid, albumin, dextrose and catalase) supplement (Difco, Plymouth, UK).

**Solutions**

**TAE Buffer (50X) (Tris-Acetate-EDTA)**

- Tris base 242 g/l
- Acetic acid 57.1 ml/l
- 0.5M EDTA 100 ml/l

Adjust pH to 8.5.

**TBE buffer (10x) (Tris-Borate-EDTA)**

- Tris base 121 g/l
- Boric acid 61.83 g/l
- EDTA 18.61 g/l

**TE buffer**

- Tris-HCl, pH7.5 10mM
- EDTA 1mM
Gel

8% polyacrylamide-urea gel

Urea 21g
TBE (10X) 5ml
Long Ranger (Cambrex, New Jersey, USA) 8ml
ddH2O 18ml

The ingredients are thoroughly mixed to dissolve the urea and 0.22μm-filtered (Millipore, Stonehouse, UK) before adding the following solutions:

APS (10%) 250μl
TEMED 25 μl