The role of the ECF sigma factor SigG in *Mycobacterium tuberculosis*

Alison Emma Gaudion

March 2011

A thesis submitted in fulfilment of the requirements of University College London for the degree of Doctor of Philosophy

Division of Mycobacterial Research
MRC National Institute for Medical Research
The Ridgeway
Mill Hill
London

Declaration

I Alison Emma Gaudion confirm that the work presented in this thesis is my own. When information has been derived from other sources, I confirm that this has been indicated in the thesis.

A.E. Gaudion

Abstract

Mycobacterium tuberculosis is the causative agent of Tuberculosis (TB). Two billion people are currently infected with *M. tuberculosis* bacilli, one in ten of whom will develop active TB in their lifetime. *M. tuberculosis* is able to survive within macrophages but the exact mechanisms used for intracellular survival are poorly understood.

DNA-dependent RNA polymerase is the enzyme responsible for transcription in all living organisms. In bacteria this enzyme recognises different promoters by binding to sigma factors that recognise those promoters. This study focuses on the role and regulation of the *M. tuberculosis* extracytoplasmic function (ECF) sigma factor, SigG. ECF sigma factors are responsible for upregulating genes necessary for bacterial stress responses. SigG has previously been shown to be upregulated in response to DNA damage and during macrophage infection.

It has been demonstrated that *sigG* is expressed from at least 2 promoters and that only promoter P1 is DNA-damage inducible. *sigG* is co-transcribed with the two downstream genes *Rv0181c* and *Rv0180c*, which were hypothesised to be anti- and anti-anti-sigma factors to SigG. Protein-protein interaction studies showed that SigG and Rv0181c do not interact. Potential anti-sigma factors to SigG were identified, the most promising of which was the thioredoxin family protein Rv1084.
Two potential SigG-dependent genes had previously been identified, *Rv0887c* and *Rv0911*. It has been demonstrated that SigG is able to bind to the promoter regions of these genes but this interaction was not specific.

An *M. tuberculosis sigG-Rv0180c* deletion strain was constructed and complemented with the whole operon as well as with *sigG* alone. The phenotype of the mutant strain was examined *in vitro* as well as *in vivo* to test the hypothesis that SigG has a role during infection. Use of a phenotype microarray revealed an enhanced susceptibility of the mutant strain to oleic acid and its ester, Tween 80, leading to investigation of the sensitivity of the strains to a range of fatty acids.

Acknowledgments

I would like to thank the MRC for providing the funding for this project.

I would like to express my thanks and utmost gratitude towards my supervisors Elaine Davis and Ian Taylor for their guidance and unlimited advice. I would also especially like to thank Katherine Smollett for creating and analysing the SigG overexpression strain, optimising expression conditions for His-tagged SigG, assisting with macrophage and animal experiments and for her help, advice and endless patience throughout this project.

I would like to thank Alan Williams for all his help with protein expression and purification. Kathryn Lougheed for her assistance with the alamar blue and Biolog assays. Thanks to Joanna Dillury for all her help with the macrophage and animal experiments, to Debbie Hunt for providing the yeast-2-hybrid positive control plasmids and the anti-CRP antibody, David Goldstein for the large scale purification of Rv0181c protein and to Steven Howell for performing the MALDI-MS analysis.

Thanks also go to Melanie Stapleton for discussion regarding protein-DNA binding assays and Mark Buttner for providing the positive control and advice for the *in vitro* transcription assays.

Finally I would like to thank all past and present members of the Division of Mycobacterial Research for their advice and discussion.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5` RACE	5` Rapid amplification of cDNA ends
ABC	Ammonium bicarbonate
Amp^R	Ampicillin resistant
ATP	Adenosine-5'-triphosphate
BCG	Bacille Calmette Guerin
BER	Base nucleotide excision repair
BMDM	Bone marrow derived macrophage
bp	base pairs
BSA	Bovine serum albumin
cDNA	complementary deoxyribonucleic acid
CFE	Cell-free extract
CFU	Colony forming units
$Chlor^R$	Chloramphenicol resistant
CIP	Calf intestinal phosphatase
CO_2	Carbon dioxide
CRP	Cyclic AMP Receptor Protein
Da	Daltons
DCO	Double crossover
DE3	λDE3 prophage
DEPC	Diethyl pyrocarbonate
dH_2O	Distilled water
DMF	Dimethylformamide
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonuclaeic acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleoside 5'-triphosphate
DSB	Double-strand break
DTT	Dithiothreitol
ECF	Extracytoplasmic function
ECL	Enhanced chemiluminescence

ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
Gent^{R}	Gentamycin resistant
GSP	Gene-specific primer
GST	Glutathione S-transferase
HBSS	Hank's buffered salt solution
HEPES	4-(2-hydroxyethyl)-1-piperaxineethanesulphonic acid
His	Histidine
HIV	Human immunodeficiency virus
HRP	Horseradish peroxidase
IFN-γ	Interferon-gamma
IPTG	Isopropyl-beta-D-thiogalactopyranoside
IVT	*In vitro* transcription
Kan$^{R/S}$	Kanamycin resistant/sensitive
kb(p)	Kilo base (pairs)
kDa	Kilo daltons
KO	Knockout
LB	Luria-Bertani
LCC	L-cell conditioned medium
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LiAc	Lithium acetate
MALDI-MS	Matrix associated laser desorption ionisation mass spectrometry
MOI	Multiplicity of infection
mRNA	Messenger RNA
N/A	Not applicable
Nanodrop	ND-1000 spectrophotometer
NEB	New England Biolabs
NER	Nucleotide exdision repair
NO	Nitric oxide
OD	Optical density
ONPG	Ortho-nitrophenol-beta-D-galactopyranoside
PAGE	Polyacrylamide gel electrophoresis

PBS	Phosphate buffered saline
PCI	Phenol:chloroform:isoamyl alcohol
PCR	Polymerase chain reaction
PNK	Polynucleotide kinase
PVDF	polyvinylidene diflouride
qRT-PCR	Quantitative reverse transcription PCR
rcf	Relative centrifugal force
RLM-RACE	RNA ligase-mediated rapid amplification of 5' cDNA ends
RNA	Ribonucleic acid
RNAP	DNA-dependent RNA polymerase
RNase	Ribonuclease
RNI	Reactive oxygen intermediates
ROI	Reactive nitrogen intermediates
rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute medium
RPMI+	Supplemented RPMI
RT	Reverse transcription
RT-PCR	Reverse transcription PCR
SCO	Single crossover
SDM	Site-directed Mutagenesis
SDS	Sodium dodecyl sulphafe
SDT	Semi-dry transfer
SSC	Saline-sodium citrate
TAP	Tobacco acid phosphatase
TB	Tuberculosis
TBE	Tris-borate EDTA
TCA	Tricarboxylic acid cycle
TE	Tris EDTA
TEMED	Tetramethylethylenediamine
TOPO	DNA Topoisomerase I
TRAFO	Transformator
U	Unit(s)
UV	Ultraviolet

VLA	Veterinary Laboratories Agency
WHO	World Health Organisation
WT	wild-type
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactoside
YNB	Yeast Nitrogen Base

Work Timeline

1st Year:

Design and creation of the *sigG* operon knockout mutant, ΔsigGoperon

Analysis of oligonucleotide promoter-*lacZ* constructs

Creation of pAG04 and pAG04-mut1 promoter-*lacZ* fusion constructs

Translational start site mapping

Co-transcription analysis of the *sigG* operon by RT-PCR

Yeast-2-hybrid assay to test for interactions between the proteins encoded by the *sigG* operon

2nd Year:

Complementation of ΔsigGoperon with full and partial complementation plasmids

Phenotypic analysis of ΔsigGoperon:

- *in vitro* growth curve
- Alamar blue using Quercetin, Butein and Methylglyoxal
- Macrophage infection model

β-galactosidase assays using CFEs from *M. tuberculosis* WT and ΔsigGoperon transformed with pAG04 and pAG04-mut1 promoter-*lacZ* constructs including exposures to mitomycin C, methylglyoxal, bleomycin, t-butylhydroperoxide and butein.

Purification of His-tagged SigG protein and production of an anti-SigG antibody

Cloning and determination of expression conditions for soluble His-tagged Rv0181c protein.

5'RACE and qRT-PCR for *Rv0654*

Demonstrating successful overexpression of the SigG protein in *M. tuberculosis*

3rd Year

Demonstrating that His-SigG and His-Rv0181c do not interact via pulldown.

Creation by SDM and β-galactosidase assay analysis of all pAG04 mutant constructs as well as 5'RACE on these constructs.

qRT-PCR to determine the number of promoters expressing *sigG*.

DNA-binding assays and *in vitro* transcription to determine whether *Rv0887c* and *Rv0911* expression was regulated by SigG.

Mouse infection model.

The BIOLOG microarray became available and was performed halfway through my final year. All experiments pertaining to fatty acids and Tween were performed subsequent to the BIOLOG results. This included:

- Alamar blue assays of ΔsigGoperon and all complement strains
- Growth curves using fatty acids as sole carbon source
- β-galactosidase assays on *M. tuberculosis* transformed with pAG04 following exposure to fatty acids
- qRT-PCR on RNA from *M. tuberculosis* exposed to fatty acids

Pulldown of interacting proteins from *M. tuberculosis* CFE using His-SigG as bait with subsequent protein identification by mass spectrometry was one of the final experiments to be performed.

1. Introduction

1.1 Mycobacteria and the *Mycobacterium tuberculosis* complex

Mycobacteria are members of the family *Mycobacteriaceae* and genus *Mycobacterium*. The mycobacteria form a large group, which can be split into two depending on how fast the species grows, with slow growers taking at least one week (usually longer) to form colonies on solid medium compared to two to three days for fast growers (Jacobs *et al.*, 1991). In general slow growers tend to be pathogens whereas fast growers are environmental species (Primm *et al.*, 2004). Mycobacteria are characterised by an acid-fast cell wall, containing mycolic acids, and a G/C rich genome. The mycolic acids form a tight, impermeable layer and this confers a degree of resistance to killing by acidic and alkaline compounds, many antibiotics, osmotic lysis and free radicals (Madigan, 2000). Together with *M. tuberculosis* and *M. bovis*, the cause of tuberculosis in cattle, the slow growing mycobacterial pathogens include *M. leprae, M. ulcerans, M. paratuberculosis*, and *M. avium*, which cause Leprosy, Buruli ulcer, Johne's disease (cattle) and disseminated disease particularly in HIV patients respectively (Cosma *et al.*, 2003) and the *M. tuberculosis* complex.

The *M. tuberculosis* complex is made up of the Tuberculosis (TB)-causing mycobacteria in various host species; comprising the well known *M. tuberculosis, M. bovis* and vaccine strain *M. bovis* Bacille Calmette-Guérin (BCG) as well as *M. africanum, M. canetti* and *M. microti* (Zink *et al.*, 2003). The members of the complex differ in host specificity, pathogenicity and a small number of phenotypic and/or genotypic characteristics (Gordon *et al.*, 2009).

1.2 Tuberculosis the Disease

1.2.1 Incidence and therapy

TB is the disease caused by the bacterium *Mycobacterium tuberculosis*. It is often considered to be a disease that is under control but it remains a major worldwide cause of morbidity and mortality and kills millions of people each year (Gupta *et al.*, 2007). Two billion people (1/3 of the world's population) are currently infected with *M. tuberculosis* bacilli and one in ten of these will develop active TB in their lifetime (WHO, 2009). The highest incidences of TB cases are reported in Africa and Asia (WHO, 2009), however, since 1988 the incidence of the disease in developed countries has been increasing and this has been attributed to co-infection with HIV, immigration and social and economic deprivation (Antoine *et al.*, 2006). In the United Kingdom, London has a particularly high incidence rate of TB, with 45 % of reported TB cases in 2004 being in London. In addition some London boroughs e.g. Hackney have been shown to have incidence rates comparable to countries where TB is considered endemic (Antoine *et al.*, 2006).

TB is a worldwide pandemic and coupled with the HIV pandemic it is becoming a growing problem. The weakened immune status of HIV infected individuals facilitates progression to active disease (Corbett *et al.*, 2003). An active *M. tuberculosis* infection will require six to nine months of therapy with a cocktail of antibiotics, including isoniazid, rifampicin, pyrazinamide and ethambutol (Russell, 2001). It is due to the existence of a persistent bacilli population that there is this necessity for such a lengthy treatment process (Cole, 2002, Betts *et al.*, 2002). Recent years have seen the emergence of the additional complication of the development of resistance to anti-TB drugs, including some strains that are resistant

to all of the major drugs. This has led to an urgent need for new therapies. The only available vaccine against TB is Bacille Calmette Guerin (BCG), an attenuated strain of a closely related mycobacterium (*M. bovis*). However, efficacy of BCG against pulmonary TB varies around the world from a maximum of 80 % down to 0 % in India (Gupta *et al.*, 2007). It is clear that new ways of treating TB infections need to be found and if this is to be achieved, a better understanding of the genetics of *M. tuberculosis* and of the *M. tuberculosis*-host interaction is necessary (Smith, 2003).

1.2.2 Pathogenesis and the host response

Infection with *M. tuberculosis* can be classified into three stages: primary infection (symptomatic, active disease), chronic or latent infection (asymptomatic) and post primary infection (reactivation of a latent infection). During the course of infection bacteria are 'walled-off' by the host immune response which forms a granuloma around the site of infection. A low proportion of infected individuals progress to active disease, and become infectious. The infection process must, therefore, be very efficient and the infective dose has been estimated to be as low as one bacterium (Smith, 2003).

The course of infection with *M. tuberculosis* is well defined. Droplets containing infectious bacteria are inhaled into the lungs. Bacteria are phagocytosed by alveolar macrophages leading to a localised proinflammatory response, whereby mononuclear cells are recruited to the site of infection from peripheral blood vessels. A granuloma forms consisting of a central region of infected macrophages surrounded by foamy macrophages and mononuclear phagocytes. The periphery of

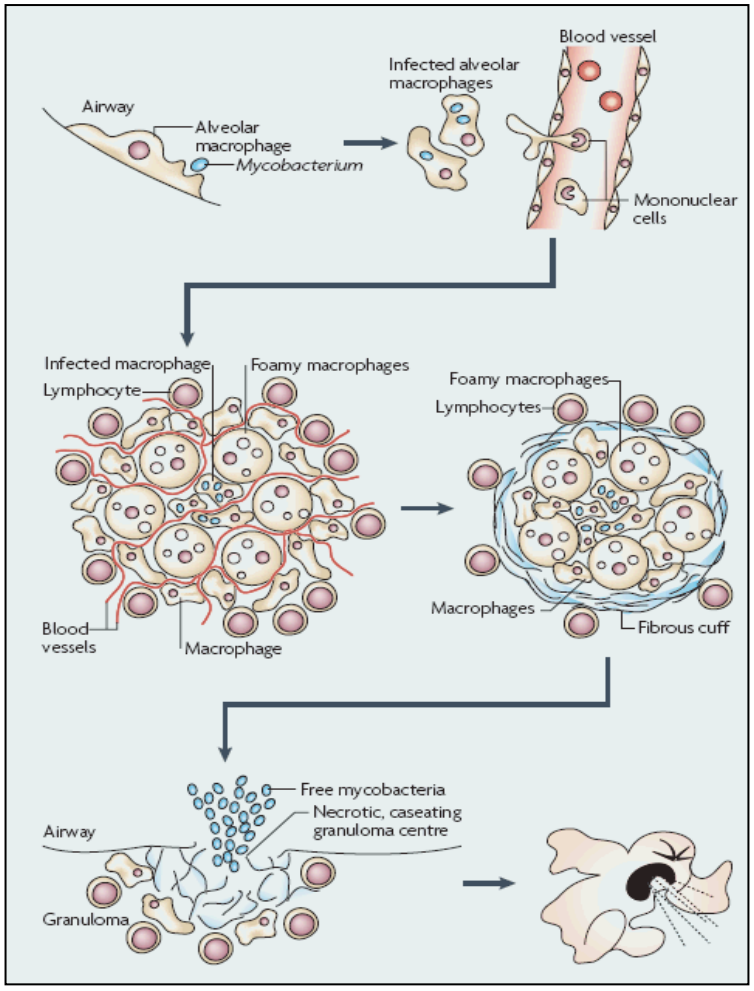


Figure 1.1 - The pathology of the granuloma (Russell, 2007)

Macrophages phagocytose bacteria and release cytokines causing a proinflammatory response. A granuloma forms with a central area of foamy macrophages and mononuclear phagocytes and a periphery of lymphocytes, collagen and extracellular matrix components. At this stage the bacteria are contained. A change in immune status can lead to caseation (necrosis) and rupture of the granuloma and the host develops a productive cough.

the granuloma is made up of lymphocytes, together with collagen and other extracellular matrix components (Figure 1.1). This response contains the infection: the host shows no overt signs of disease and does not transmit infection. Conditions within the granuloma are considered to be hypoxic (low oxygen) and acidic, which would normally lead to bacterial cell death. A weakening of the immune status of the host e.g. old age, HIV infection etc. causes necrotic decay of the granuloma, which ruptures allowing infectious bacteria to enter the airways. The host then develops a productive cough leading to transmission through droplet infection (Russell, 2007).

The success of this pathogen is due to the fact that once inside the macrophage, mycobacteria are able to avoid being killed (Russell, 2001). Inside the macrophage, mycobacteria are contained within a vacuole known as a phagosome, which under normal conditions would fuse with lysosomes to produce the phagolysosome. Phagolysosome fusion exposes the bacteria to acidic pH, proteolytic enzymes, unsaturated fatty acids and reactive oxygen and nitrogen species (Schnappinger *et al.*, 2003). However, pathogenic mycobacteria are able to prevent this process from occurring within naive macrophages which enables the bacteria to survive and replicate within the alveolar macrophages (Hestvik *et al.*, 2005). Once macrophages become activated by the adaptive immune response they are able to undergo phagosome maturation and in order to combat this, *M. tuberculosis* activates additional survival genes (Figure 1.2). The exact mechanisms employed by *M. tuberculosis* for survival inside the hostile macrophage environment are poorly understood (Houben *et al.*, 2006).

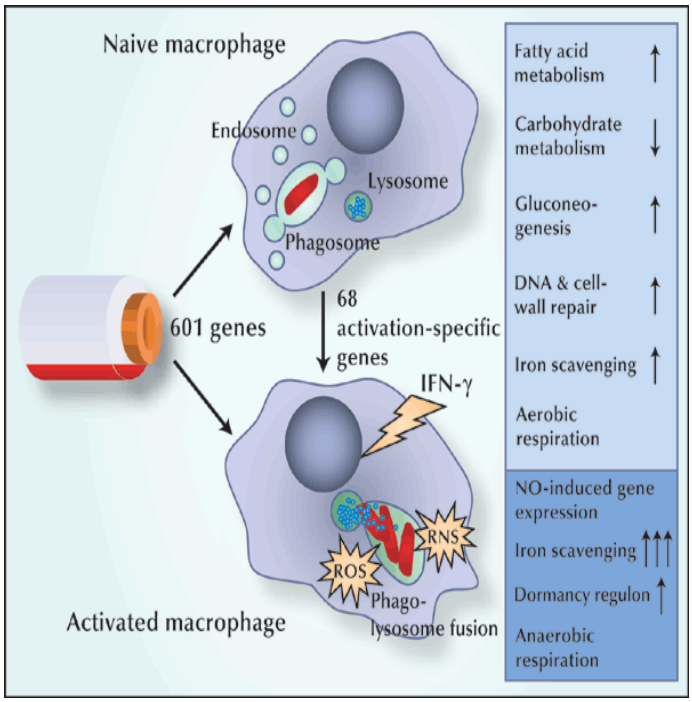


Figure 1.2 - *M. tuberculosis* in the macrophage (McKinney *et al.*, 2003)
In naïve macrophages *M. tuberculosis* survives in a modified phagosome by preventing it from undergoing acidification, maturation or fusion with lysosomes. Macrophages become activated by IFN-γ and this promotes phagosome acidification and phagolysosome fusion causing the activation of additional survival genes in the bacteria.

The environment within a macrophage is not conducive to bacterial growth and causes the induction of stress responses. In the case of *M. tuberculosis*, pathways for fatty acid metabolism are activated, enabling the bacteria to use lipids, instead of carbohydrates, as an energy source. Other genes induced include those for iron scavenging, starvation responses and those needed for the synthesis and modification of DNA and cell wall components (Figure 1.2).

Activated macrophages produce reactive oxygen species and nitric oxide (NO), which are damaging to bacteria. Voskuil *et al* (2004) identified that the 'dormancy' regulon is activated by NO, which causes inhibition of bacterial aerobic respiration, as well as by hypoxia (Park *et al.*, 2003). Activation of this regulon causes a shift to a non-replicating, dormant form that is able to persist within the macrophage. The existence of these persistent bacteria leads to asymptomatic, latent infections that can persist for decades (Russell, 2001).

1.3 The role of fatty acids during *M. tuberculosis* infection of macrophages

1.3.1 Macrophages produce mycobactericidal fatty acids

Unsaturated fatty acids such as oleic and linoleic acids are secreted by activated macrophages both into the phagolysosome (Schnappinger *et al.*, 2003) and outside of the cell (Hemsworth *et al.*, 1978). They form an important part of the mycobactericidal activity of macrophages and a synergistic effect has been demonstrated between production of fatty acids and reactive nitrogen species (Hemsworth *et al.*, 1978, Akaki *et al.*, 2000). Oleic acid is a long-chain (C_{18}) unsaturated fatty acid that has been shown to have mycobactericidal effects.

Unsaturated fatty acids, including oleic acid, were shown to be highly toxic to fast-growing mycobacteria and it was hypothesised that this was due to disturbance of membrane permeability. It was noted in the same study that pathogenic strains showed more resistance to unsaturated fatty acids (Saito *et al.*, 1984). Oleic acid has also been shown to inhibit the enzyme topoisomerase I (topo I), a key enzyme in the cleavage and repair of DNA strands. Topo I is also important for cellular processes such as replication, transcription and recombination and for relaxation of transcription-induced negative supercoiling (Carballeira, 2008, Tse-Dinh, 2009). Inhibition of topo I leads to DNA damage by trapping the topoisomerase cleavage-complex, therefore preventing relaxation of supercoiled DNA during transcription. In mycobacteria a transposon insertion mutant in the *topA* gene could not be isolated suggesting that this gene is essential (Tse-Dinh, 2009).

Whether the mycobactericidal effect of macrophage-produced fatty acids is due to disruption of the mycobacterial membrane or DNA damage caused by inhibition of topo I, these are an important aspect of the host immune response that must be avoided by *M. tuberculosis* in order for it to survive.

1.3.2 Fatty acids as a carbon source for *M. tuberculosis*

Upon infection of a macrophage *M. tuberculosis* upregulates genes involved in fatty acid metabolism, which allow the bacteria to use lipids instead of carbohydrates as an energy source (McKinney *et al.*, 2003). Use of lipids as the primary carbon source *in vivo* is supported by the fact that respiration of *M. tuberculosis* in mouse lungs is strongly stimulated by fatty acids but not carbohydrates (Bloch *et al.*, 1956) and also the fact that several enzymes of the glycolytic pathway appear to be non-essential

for growth and persistence in mice (Sassetti *et al.*, 2003). There are two pathways that are specifically required for growth on fatty acids, the catabolic β-oxidation cycle, which degrades fatty acids to acetyl-CoA and when glycolytic substrates are in low abundance, the glyoxylate shunt (McKinney *et al.*, 2000). The glyoxylate shunt is essential for preventing the loss of carbon molecules as carbon dioxide via the tricarboxylic acid (TCA) cycle during growth on fatty acids (Figure 1.3). Extensive duplication of genes encoding β-oxidation enzymes exists in *M. tuberculosis* (Cole *et al.*, 1998) and this creates an obstacle for studying genes in this system due to functional redundancy between paralogous genes. The glyoxylate shunt however, is limited to three genes encoding malate synthase and the two isocitrate lyase genes (Munoz-Elias *et al.*, 2005).

1.4 Transcription and sigma factors in bacteria

DNA-dependent RNA polymerase (RNAP) is the main enzyme responsible for transcription and gene expression in all living organisms (Madigan, 2000). In bacteria the core RNAP comprises four subunits; an α dimer, β, β' and ω (Manganelli *et al.*, 2004). Core RNAP combines with a dissociable sigma factor to form the holoenzyme, capable of specific promoter recognition and transcription initiation (Figure 1.4) (Borukhov *et al.*, 2003). Sigma factors fall into two broad families: the sigma 70 family and the sigma 54 family (Bashyam *et al.*, 2004). Sigma factors from the sigma 70 family recognise the -10 and -35 promoter regions, whereas those from the sigma 54 family recognise -12 and -24 promoter regions. The names of these regions relate to their positions upstream of the transcriptional start site. Sigma factors are responsible for global regulation of gene

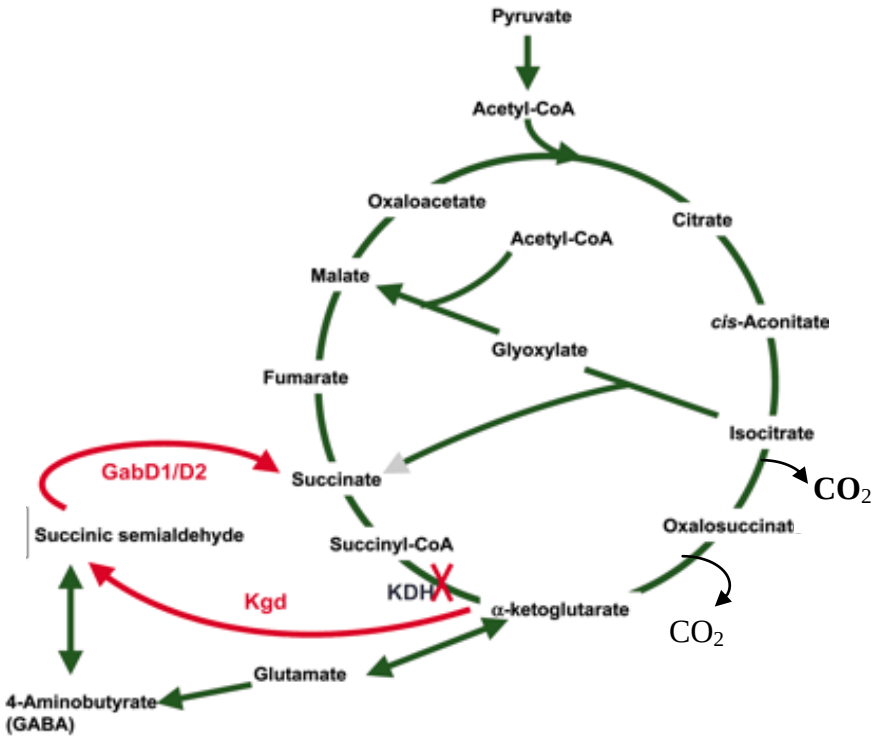


Figure 1.3 – The TCA and glyoxylate pathways of *M. tuberculosis*
During growth on carbohydrate substrates *M. tuberculosis* utilises an alternative version of the TCA cycle (shown in red) converting isocitrate to succinate via ketoglutarate and succinic semialdehyde. The α-ketoglutarate dehydrogenase (KDH) enzyme, which would normally convert α-ketoglutarate to succinyl-CoA is absent in the *M. tuberculosis* genome. Instead α-ketoglutarate is converted to succinate via the activities of α-ketoglutarate dehydrogenase (Kgd) and succinic semialdehyde dehydrogenase (GabD1/D2) enzymes. However, during growth on C_2 substrates such as fatty acids, using the TCA cycle would result in the loss of carbon molecules as carbon dioxide. The glyoxylate pathway allows isocitrate to be converted to succinate and glyoxylate without the loss of carbon molecules. Figure adapted from Tian *et al.* (2005).

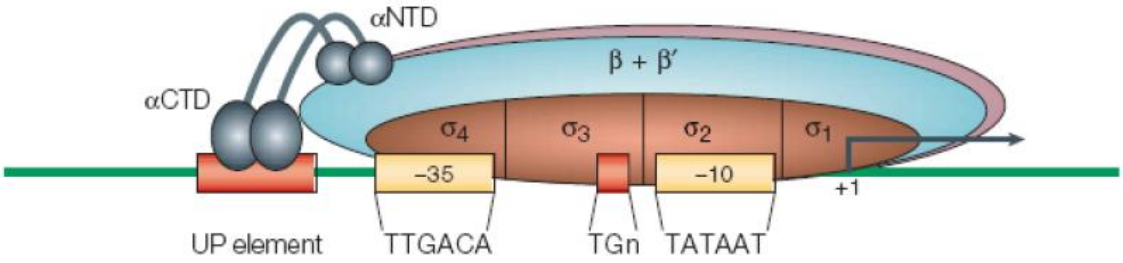


Figure 1.4 - Diagram to show the RNAP/Sigma factor holoenzyme

The sigma factor is responsible for recognition of the -35 and -10 sequences of the promoter and different sigma factors recognise different sequences. The five subunits of the holoenzyme can be seen, including the four domains of the σ factor (σ1-σ4) as well as the N- and C-terminal extensions of the α dimer (αNTD and αCTD, respectively). The -10 and -35 regions are highlighted in yellow, and bind to domain two and domain four of the σ factor, respectively. The extended -10 element is located immediately upstream of the -10 region (TGn) and binds to domain three of the σ factor. The UP element is upstream of the -35 region and binds to the carboxyl-terminal domains of the α dimer. The consensus sequences shown for these elements are those for sigma 70, the primary *E. coli* σ factor. Together, the -10, -35, extended -10 and UP elements specify the initial binding of the RNA polymerase to a promoter, but the relative contribution of each element differs from promoter to promoter as well as by the composition of the sigma factor. Figure adapted from Browning & Busby (2004).

transcription as they are able to bind via domains two and four to the specific promoters of different genes by recognition of two promoter regions (Figure 1.4).

The sigma 70-related family of sigma factors can be divided into four groups depending on the presence or absence of four sigma factor domains or regions. Region 1 is associated with inhibiting free sigma factor from binding with DNA in the absence of RNAP. Region 2 is responsible for recognition of the -10 promoter element. Region 3 has a role in recognising extended -10 promoter elements and finally region 4 is responsible for recognition of the -35 promoter element (Rodrigue *et al.*, 2006). Group 1 consists of principal sigma factors, which contain all four conserved regions and are essential genes. Group 2 sigma factors also contain all four conserved regions but are only found in a limited number of bacterial species (e.g. GC-rich species) and are non-essential for growth under laboratory conditions. Group 3 sigma factors lack region 1 and generally fall into clusters with specific functions, such as heat shock response or sporulation. The final group, group 4, is the largest and most diverse group of sigma factors (Gruber *et al.*, 2003). These sigma factors contain only regions 2 and 4 and are known as extracytoplasmic function (ECF) sigma factors (Figure 1.5).

Under normal growth conditions a bacterial cell constitutively expresses so-called 'housekeeping' genes, which have promoters that are recognised by primary group 1 sigma factors such as σ^{70} (Borukhov *et al.*, 2003). However, bacteria have to continuously adapt to changes in their environment and for pathogenic bacteria, this means having to contend with the immune response of the host. One of the most common bacterial adaptation to stress is to replace the primary sigma factor with an

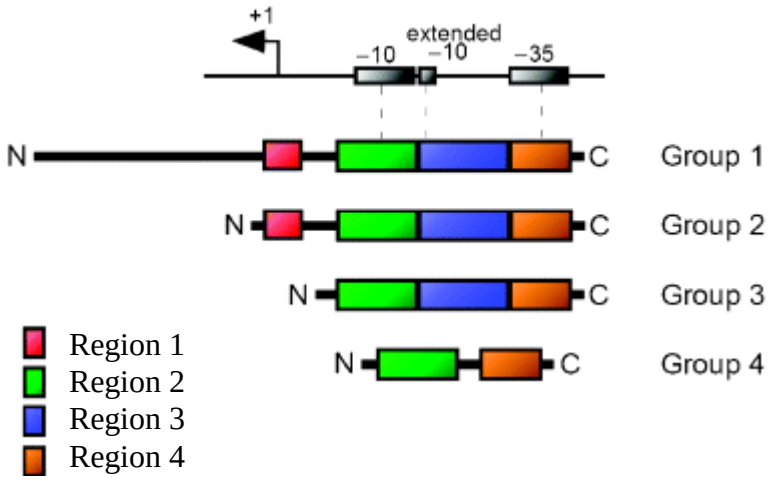


Figure 1.5 – Sigma factor classification and domain organisation

Structural organisation of the sigma factors in groups 1, 2, 3 and 4. The +1 and arrow represent the transcriptional start site and direction of transcription respectively.

Figure adapted from Rodrigue *et al.* (2006).

alternative sigma factor (from groups 2-4), enabling expression of a specific set of genes (Kazmierczak *et al.*, 2005).

1.5 Response to stress – two-component systems and extracytoplasmic function sigma factors

Bacterial stress responses can be categorised into two main types: cytoplasmic and extracytoplasmic depending on whether stress is applied to the cytoplasm or membrane/periplasm respectively. There is a lot of information available relating to cytoplasmic stress responses, which include the heat shock response and the response to DNA and protein damage by toxic radicals (Storz *et al.*, 2000, Walker *et al.*, 2000, Yura *et al.*, 2000). In contrast, the extracytoplasmic stress response has only been the subject of more recent study. Originally the ECF subfamily of sigma factors was named for the fact that they commonly mediate the extracytoplasmic stress response by regulating genes necessary for sensing and responding to changes in the bacterial periplasm and the extracellular environment (Kazmierczak *et al.*, 2005). However, it is possible for ECF sigma factors to respond to intracellular signals e.g. *S. coelicolor* SigR, which responds to oxidative stress in the cytoplasm (Helmann, 2002). As such ECF sigma factors are currently defined by their domain structure; the presence of only regions 2 and 4 (Figure 1.5), rather than the type of stress to which they respond.

ECF sigma factors are highly conserved in both gram-positive and gram-negative bacterial species. The first to be identified was *E. coli* SigE, a second heat shock factor which, although predicted to have a role in protecting extracytoplasmic compartments from stress (Mecsas *et al.*, 1993), was not classified as an ECF sigma

factor until after this group had been defined by Lonetto *et al.* (1994). Although SigE does not play a role in pathogenesis, it has been suggested that ECF sigma factors may contribute to the regulation of virulence genes in pathogenic bacteria such as *Salmonella enterica* serovar Typhimurium, *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis* (Bashyam *et al.*, 2004).

Often ECF sigma factors are co-transcribed with one or more regulators, negative and positive. These can be referred to as anti-sigma factors and anti-anti-sigma factors, respectively. Anti-sigma factors act as negative regulators which bind to the ECF sigma factor preventing it from combining with RNAP to form the holoenzyme. Most commonly, anti-sigma factors are transmembrane proteins consisting of an extracytoplasmic sensory domain and an intracellular domain that binds to the ECF sigma factor to inhibit it (Raivio *et al.*, 2001, Yoshimura *et al.*, 2004). However, this is not always the case as ECF sigma factors can also be involved in responses to intracellular signals such as the oxidative stress response. Regulation of anti-sigma factors often occurs via partner-switching where the anti-sigma factor has a choice of binding two partners in a mutually exclusive manner. These partners are the ECF sigma factor and an anti-anti-sigma factor. Thus, when the anti-anti-sigma factor is present in the correct form and binds the anti-sigma factor, the ECF sigma factor is free to bind RNAP and form the holoenzyme (Helmann, 1999). Other suggested mechanisms include phosphorylation of an anti-anti-sigma factor (Greenstein *et al.*, 2007) and protease degradation of the anti-sigma factor in response to stress signals (Bashyam *et al.*, 2004). Examples of a variety of different anti-sigma factors can be found in Table 1.1.

Table 1.1 – Anti-sigma factors

Organism	Sigma Factors	Anti-sigma Factors	Function	Subcellular location	Mechanism of Action
	σ^F and σ^G	SpoIIAB	Sporulation	Forespore	Partner-switching. The anti-anti-sigma factor SpoIIAA relieves SpoIIAB binding to the sigma factor in a phosphorylation-dependent manner.
	σ^B	RsbW	Stress response	Cytoplasm	Partner-switching. Dephosphorylation of the anti-anti-sigma factor RsbV activates it to bind RsbW and release σ^B
Bacillus subtilis	σ^D	FlgM	Flagellar biosynthesis	Cytoplasm	During assembly of the flagellar hook basal body (HBB) structure FlgM binds to σ^D. Once the HBB is complete FlgM is exported from the cytoplasm allowing σ^D-dependent transcription of flagellin.
Escherichia coli	σ^{32}	DnaK	Heat Shock	Cytoplasm	DnaK binds to the sigma factor when complexed with ADP. The presence of ATP causes dissociation of the complex.
ECF Subfamily					
Bacillus subtilis	σ^{SigX}	RsiX	Cell envelope modifications	Cytoplasmic membrane	Similar to signal transduction by two component histidine kinases. Periplasmic domain senses misfolded proteins and triggers cytoplasmic domain to release the sigma factor
Escherichia coli	σ^E	RseA	Extreme temperature survival	Cytoplasmic membrane	1) Similar to signal transduction by two component histidine kinases. Oxidative or heat shock stress cause a conformational change. 2) Phosphorylation-dependent proteolytic cleavage of RseA can cause release of the sigma factor in response to surface stress.
	σ^E	RseA	Surface stress response		1) Similar to signal transduction by two component histidine kinases. Oxidative or heat shock stress cause a conformational change. 2) PknB phosphorylation-dependent proteolytic cleavage of RseA can cause release of the sigma factor in response to surface stress.
	σ^F	UsfX	Adaptation to stationary phase	Cytoplasm	Partner-switching mechanism. Two anti-anti-sigma factors RsfA and RsfB are able to relieve UsfX inhibition of σ^F under reducing conditions and in an ATP/ADP-dependent manner respectively.
Mycobacterium tuberculosis	σ^H	RshA	Oxidative and heat stress	Cytoplasmic membrane	1) Similar to signal transduction by two component histidine kinases. Oxidative or heat shock stress cause a conformational change. 2) PknB phosphorylation-dependent proteolytic cleavage of RseA can cause release of the sigma factor in response to surface stress.
Pseudomonas aeruginosa	σ^{22}	MucA	Alginate biosynthesis	Cytoplasmic membrane	Similar to signal transduction by two component histidine kinases. Periplasmic domain senses misfolded proteins and triggers cytoplasmic domain to release the sigma factor
Streptomyces coelicolor	σ^R	RsrA	Oxidative stress	Cytoplasm	Binds sigma factor under reducing conditions. Oxidising conditions trigger conformational change and sigma factor release.

(Helmann, 2002, Hughes *et al.*, 1998, Song *et al.*, 2003, Beaucher *et al.*, 2002, Missiakas *et al.*, 1998)

The relationship between the sigma and anti-sigma factors is similar to that of a two-component signal transduction system. These systems exist to link external stimuli to specific adaptive responses and in prokaryotes they usually consist of a sensory histidine protein kinase and a response regulator protein (West *et al.*, 2001). One such example in *M. tuberculosis* is the MprAB system, which is responsible for responding to extracellular stimuli that cause membrane stress (He *et al.*, 2006). In eukaryotes serine/threonine protein kinases (STPKs) are able to sense and transduce environmental signals. *M. tuberculosis* contains 11 eukaryotic-like STPKs, which have been suggested to regulate cell shape, virulence and nitrogen balance. It has also been demonstrated for *Bacillus subtilis* that STPKs are able in some instances to function as anti-sigma factors (Greenstein *et al.*, 2007). It is the interplay of all of these regulatory systems that allow bacteria such as *M. tuberculosis* to constantly adapt to changes in the environment and to evade killing by the host immune system.

1.6 The sigma factors of *M. tuberculosis*

1.6.1 Structure and conserved regions

M. tuberculosis expresses 13 sigma factors, which fall into all four groups of the sigma 70 family (Figure 1.6). SigA, SigB and SigF represent groups 1, 2 and 3 respectively and the remaining 10 sigma factors all belong to the group 4 ECF sigma factors (Manganelli *et al.*, 2004). These ECF sigma factors are thought to contribute to *M. tuberculosis* pathogenesis by enabling it to respond to the stresses faced during infection. Three of the *M. tuberculosis* sigma factors, SigG, SigI and SigJ have an unusual carboxy-terminal extension of approximately 120 amino acids (Figure 1.6), which has been suggested to play a role in interactions between these sigma factors

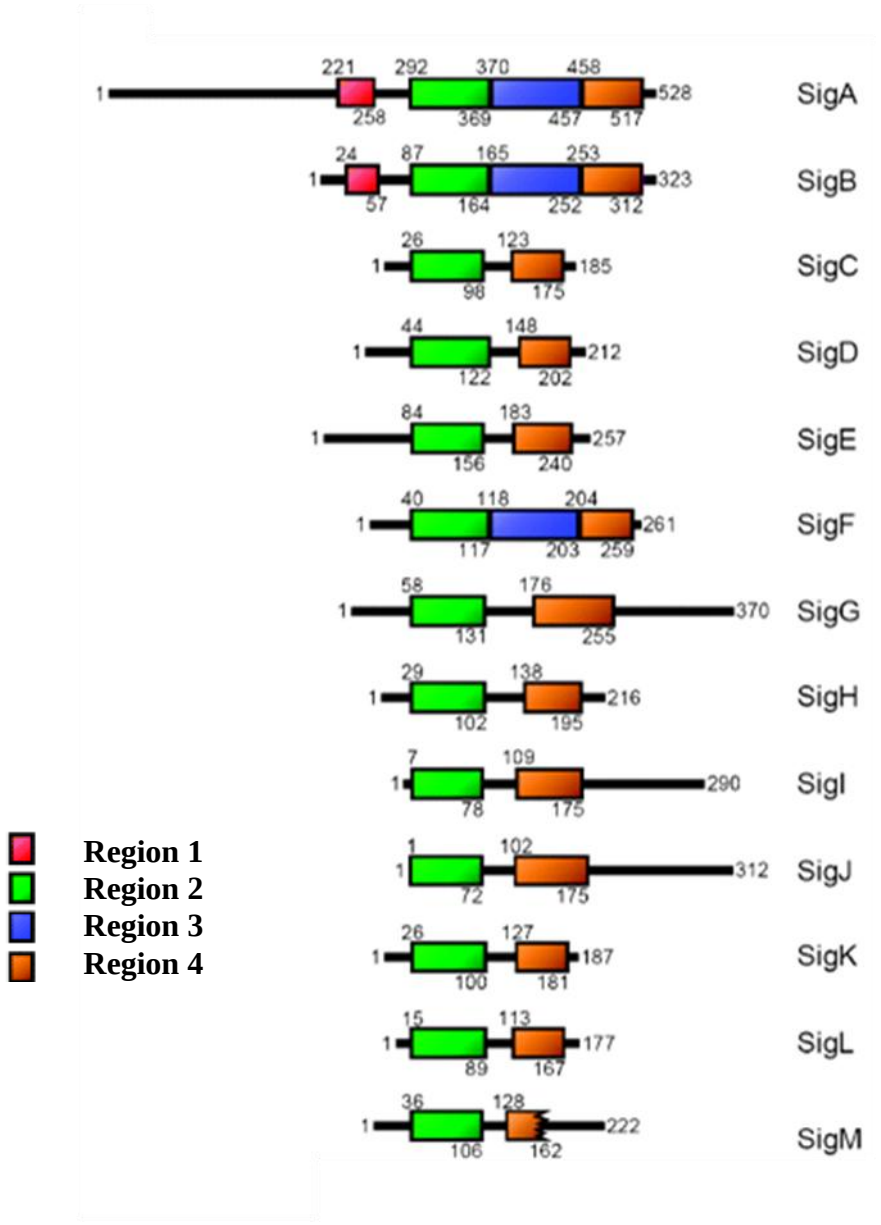


Figure 1.6 – Schematic representation of the *M. tuberculosis* sigma factors
Architecture of the sigma factors expressed by *M. tuberculosis* H37Rv. SigA, SigB and SigF represent groups 1, 2 and 3 respectively while the remaining 10 sigma factors fall into group 4. Figure adapted from Rodrigue *et al.* (2006).

and proteins or other molecules (Rodrigue *et al.*, 2006).

1.6.2 Sigma factor regulation in *M. tuberculosis*

Sigma factors are subject to complex regulation at the transcriptional, translational and post-translational levels. The three best characterised sigma factors in *M. tuberculosis* are SigE, SigF and SigH. SigE has been shown to be upregulated during human macrophage infection (Graham *et al.*, 1999) and in response to stress conditions such as alkaline pH and oxidative stress (Manganelli *et al.*, 1999, He *et al.*, 2006). It is expressed from three different promoters, one that is regulated by the two-component system MprAB in response to surface stress and alkaline pH (He *et al.*, 2006), one that is SigH-dependent, responding to heat shock exposure (Raman *et al.*, 2001) and a third that is less well characterised (Dona *et al.*, 2008). In addition SigE can also be translated from three alternative translational start codons, creating different isoforms depending on the environmental cues sensed by the bacterium (Dona *et al.*, 2008). SigE is subject to post-translational regulation by its cognate anti-sigma factor RseA, the gene for which forms an operon with *sigE* (Rodrigue *et al.*, 2006).

SigH induces the expression of a set of genes that includes heat shock proteins, in response to oxidative and heat stress (Song *et al.*, 2003), diamide stress (Raman *et al.*, 2001) and during macrophage infection (Graham *et al.*, 1999). SigH is thought to regulate expression of both *sigB* and *sigE* as well as its own expression (Figure 1.7).

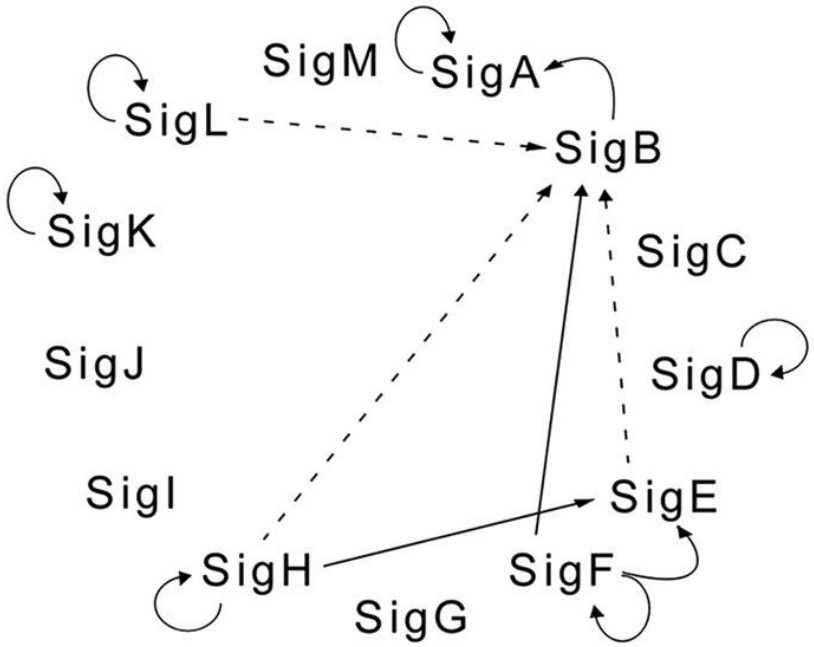


Figure 1.7 – The *M. tuberculosis* sigma factor regulatory network (Rodrigue *et al.*, 2007)

Schematic representation of the sigma factor regulatory network showing the roles of the various *M. tuberculosis* holoenzymes in expression of sigma factor genes.

Dashed lines represent the work of Dainese *et al.* (2006).

sigH is expressed from two promoters that are both active under heat shock conditions (Fernandes *et al.*, 1999). SigH is regulated by its cognate anti-sigma factor RshA, which is encoded downstream of and forms an operon with the *sigH* gene. RshA senses redox potential and in a reducing environment, binds and sequesters SigH, preventing its interaction with RNA polymerase. Upon sensing oxidative stress, RshA undergoes a conformational change leading to the release of SigH, which can then bind RNA polymerase (Song *et al.*, 2003).

Regulation of SigF in *M. tuberculosis* provides the only documented example of regulation by both an anti-sigma and anti-anti-sigma factor in *M. tuberculosis*. *sigF* is co-transcribed with its upstream gene *usfX* (also named *rsbW*), which encodes an anti-sigma factor to SigF (Beaucher *et al.*, 2002). Negative regulation of SigF by UsfX can be relieved by two anti-anti-sigma factors, RsfA under reducing conditions and RsfB possibly due to phosphorylation, which are both able to bind the UsfX-SigF complex causing release of SigF (Beaucher *et al.*, 2002).

Figure 1.7 shows how the *M. tuberculosis* sigma factors form an interaction network, with *sigB* expression predicted to be under the control of SigE, SigF, SigH and SigL as well as several sigma factors appearing to autoregulate their own expression (Rodrigue *et al.*, 2007).

It is clear from these three examples that regulation of the *M. tuberculosis* sigma factors is highly complex (Figure 1.8), with several being expressed from multiple promoters, six of the 13 sigma factors being autoregulatory and several sigma factors

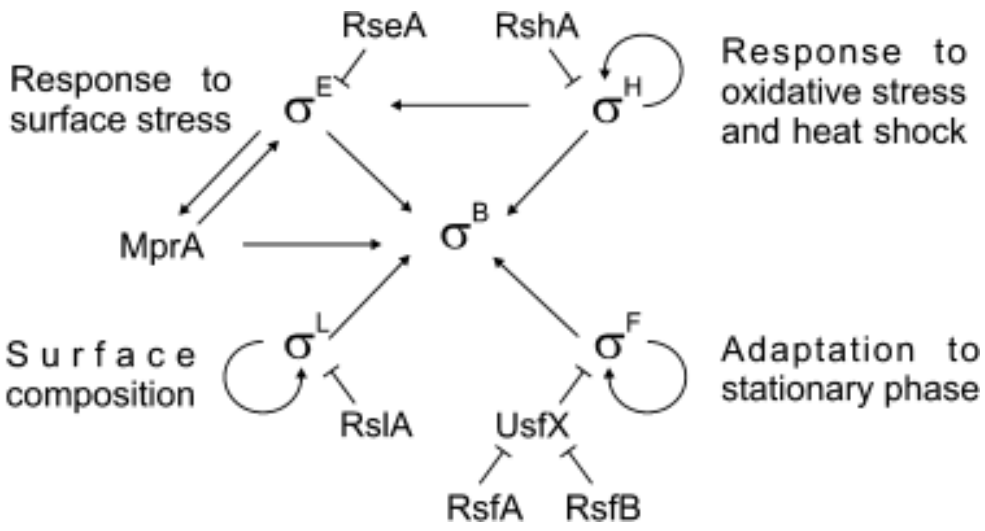


Figure 1.8 – Schematic representation of transcriptional and post-translational regulation of *M. tuberculosis* sigma factors (Rodrigue *et al.*, 2006)

Diagram showing the complex transcriptional and post-translational regulation of *M. tuberculosis* sigma factors. ⟶ represents positive transcriptional regulation; ⟞ represents negative post-translational regulation.

exerting control over expression of other sigma factors. Four confirmed anti-sigma factors are encoded in the *M. tuberculosis* genome; the three discussed above as well as RslA and RskA which act as SigL- and SigK-specific anti-sigma factors respectively (Thakur *et al.*, 2010, Saïd-Salim *et al.*, 2006), as well as two predicted anti-sigma factors to SigD and SigM (Sklar *et al.*, 2010). Add to this the two SigF-specific anti-anti-sigma factors RsfA and RsfB and the picture built up is of a highly flexible system in which, there is likely to be a degree of functional redundancy (Rodrigue *et al.*, 2006).

1.7 *M. tuberculosis* ECF sigma factor, SigG

1.7.1 Expression of *sigG*

To date, very little study has been conducted into the role or regulation of the predicted ECF sigma factor, SigG. During *in vitro* growth, *sigG* has the lowest level of expression of all of the sigma factors in exponentially growing *M. tuberculosis* cultures (Manganelli *et al.*, 1999). SigG has been shown to be the most highly upregulated *M. tuberculosis* sigma factor in response to DNA damage (Rand *et al.*, 2003) as well as being upregulated during stationary phase growth in synthetic medium (Lee *et al.*, 2007) and during human macrophage infection (Cappelli *et al.*, 2006). In addition, SigG has been hypothesised to regulate expression of the repressor LexA, and therefore to have a role in regulating expression of genes involved in the RecA/LexA-dependent SOS response to DNA damage (Lee *et al.*, 2007). This suggested a possible role for SigG in adaptation to intracellular environmental stress.

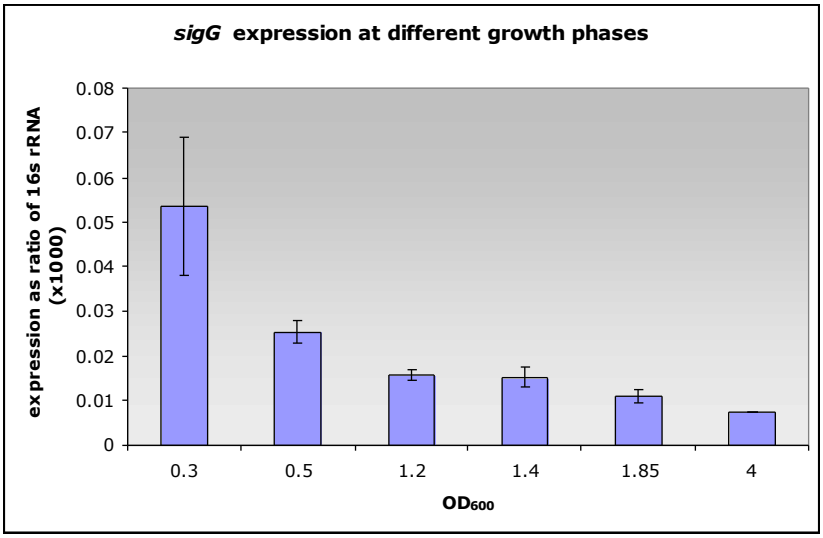


Figure 1.9 - Level of *sigG* expression during different growth phases.
qRT-PCR (SYBR-Green) demonstrating expression of *sigG* during different growth phases. Expression data for each gene was normalised to equivalent 16S ribosomal RNA data and columns and error bars represent the averages and standard deviations of three biological replicates, each assayed in triplicate. The highest expression is observed during early exponential growth (K. Smollett, personal communication).

Contrary to the findings of Lee *et al* (2008) real-time RT-PCR results in this laboratory have shown that *sigG* expression is highest during early exponential phase growth (Figure 1.9; K. Smollett, personal communication). The discrepancy between this work and that of Lee *et al* (2008) can be explained by the fact that in Lee's paper data was normalised to *sigA* expression levels. While *sigA* expression remains constant throughout exponential phase it decreases during stationary phase (Manganelli *et al.,* 1999) making it an unsuitable gene to use for normalising gene expression levels during stationary phase. It is therefore likely that the apparent increase in *sigG* expression observed by Lee *et al.,* in stationary phase was actually due to the decrease in *sigA* expression. It has also been demonstrated that SigG does not regulate the expression of *lexA* and that a *sigG* deletion strain was no more or less susceptible to DNA damage than wild-type. It was therefore concluded that SigG does not regulate expression of genes involved in either the RecA-dependent or RecA-independent DNA-damage SOS responses (Smollett *et al.*, 2011). This again was in direct contradiction to the paper of Lee *et al* (2008). In this paper microarrays were used to analyse a *sigG* deletion strain compared to wild-type *M. tuberculosis* CDC1551. The authors set a 2-fold change as the cut-off for significance however, despite the fact that none of the SOS response genes reached this cut-off, the major conclusion of this paper was that SigG regulated SOS response genes. In addition there was no correlation between genes having identified SOS boxes (signifying LexA regulation) and whether they were up- or down-regulated in the *sigG* deletion strain. Genes in the same operon and therefore co-expressed, were affected differently in the *sigG* deletion strain i.e. in the *rucA*/*B*/*C* operon, *rucC* was up-regulated, while *rucA* and *rucB* were down-regulated (Lee *et al.*, 2008). Microarray comparisons of a *sigG* deletion strain and wild-type *M. tuberculosis* H37Rv under

normal growth and DNA-damaging conditions were performed in this laboratory and showed no difference between the two strains (Smollett *et al.*, 2011). SigG may therefore regulate a set of genes that, although related to the response to DNA damage, are not responsible for DNA repair.

Primer extension experiments performed by Dawson (2005) have indicated the presence of three possible promoter sequences upstream of *sigG*. Promoter P1 appears to be mitomycin C inducible, suggesting a role in the response to DNA damage, and Promoter 3 may be SigG-dependent, indicating an autoregulatory role, which is a common feature of ECF sigma factors. These observations will form the basis for investigation of the control of expression of *sig*G.

1.7.2 The *sigG* operon

There are three genes downstream of *sigG*: *Rv0181c*, *Rv0180c* and *lprO* (Figure 1.10). Pairwise RT-PCR has been used to investigate whether *sig*G is co-transcribed with any of its downstream genes. Results suggested that *sig*G, *Rv0181c* and *Rv0180c* were co-transcribed and that *lpr*O was transcribed separately (Dawson, 2005). As previously stated ECF sigma factors are often co-transcribed with anti- and anti-anti-sigma factors suggesting that these roles could apply to Rv0181c and Rv0180c. Further investigation of the transcription of the *sigG* operon and of the roles of *Rv0181c* and *Rv0180c* will be conducted in this study.

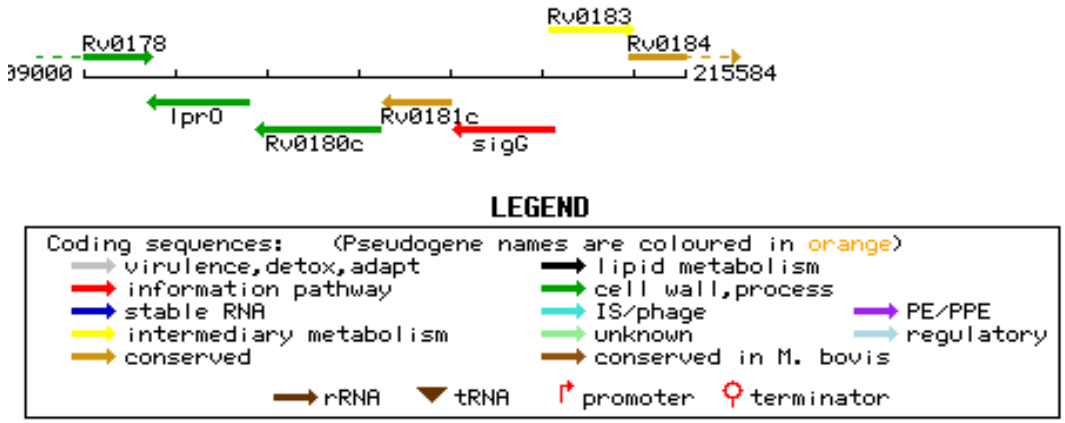


Figure 1.10 - A schematic representation of the location of *sig*G (TubercuList)
A schematic showing the chromosomal locations and predicted functions of the *sigG*, *Rv0181c* and *Rv0180c* genes as well as the downstream gene, *lprO*.

1.8 Identification of a possible SigG-dependent promoter

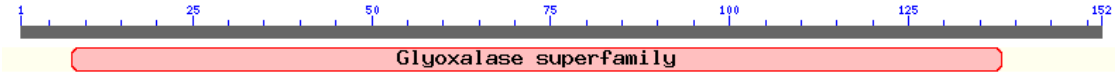
1.8.1 The probable dioxygenase, Rv0654

It has been shown using an *E. coli* two-plasmid system that the expression of the gene *Rv0654* is significantly increased in the presence of SigG (R. Balhana and J Smith, personal communication). This result suggests that *Rv0654* is expressed from a SigG-dependent promoter. However, expression of *Rv0654* was still observed in the absence of SigG, but at a lower level. Possible reasons for this are that there are in fact two promoters, one SigG-dependent and one that is recognised by an *E. coli* sigma factor or that there is just one promoter that is recognised by both *M. tuberculosis* SigG and an *E. coli* sigma factor. These findings will be investigated further in this study.

1.8.2 Expression of *Rv0887c* and *Rv0911* increased in a SigG overexpression strain

Rv0887c and *Rv0911* were the two most up-regulated genes in a SigG overexpression strain when analysed by microarray; 15- and 13-fold respectively (K. Smollett, personal communication). Expression of these two genes has been shown to be induced by DNA damage in wild-type *M. tuberculosis* H37Rv using qRT-PCR. When expression in a *sigG* deletion strain was compared with wild-type, under normal *in vitro* growth conditions, expression was similar to wild-type but neither *Rv0887c* nor *Rv0911* could be induced by DNA damage in this strain (K. Smollett, personal communication).

Rv0887c



Rv0911

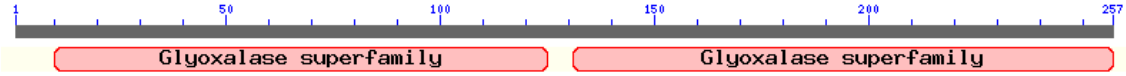


Figure 1.11 – Schematic representation of the domain structures of Rv0887c and Rv0911

Domain structures of Rv0887c and Rv0911 as predicted by protein BLAST analysis. Both proteins are predicted to contain domains belonging to the Glyoxalase (also known as Glo_EDI_BRP_like) domain superfamily, with Rv0887c containing one domain and Rv0911 containing two domains.

It was therefore hypothesised that these two genes were expressed from two promoters; one responsible for their basal expression and one that was regulated by SigG in response to DNA damage. *Rv0887c* and *Rv0911* encode conserved hypothetical proteins (TubercuList). BLAST analysis of the Rv0887c and Rv0911 protein sequences revealed that they belong to the Glo_EDI_BRP_like domain superfamily, with Rv0887c containing one domain and Rv0911 containing two domains (Figure 1.11). This superfamily contains several structurally related metalloproteins, including type I extradiol dioxygenases, glyoxalase I and a group of antibiotic resistance proteins.

1.9 The toxic electrophile methylglyoxal

1.9.1 Methylglyoxal and its detoxification

The glyoxalase I enzyme is the first enzyme in the methylglyoxal detoxification pathway (Figure 1.12). Rv0911 was predicted to be involved in this detoxification pathway based on the fact that it was able to interact with and was therefore a target of, pyrimidine-imidazole compounds, which target the methylglyoxal detoxification pathway (Marrero *et al.*, 2010). Methylglyoxal is an electrophile, which can cause cell death at high concentrations by reacting with the nucleophilic centres of molecules such as DNA, RNA and proteins. However, despite it being toxic, bacteria still produce this metabolite and it is thought that this provides them with the ability to control the rate of carbon flux when moving between environments (Ferguson *et al.*, 1998). In addition the *in vivo* tissue environment, the granuloma especially, has been shown to be rich in methylglyoxal, which is able to induce apoptosis of macrophages (Rachman *et al.*, 2006). While it is unclear whether the methylglyoxal

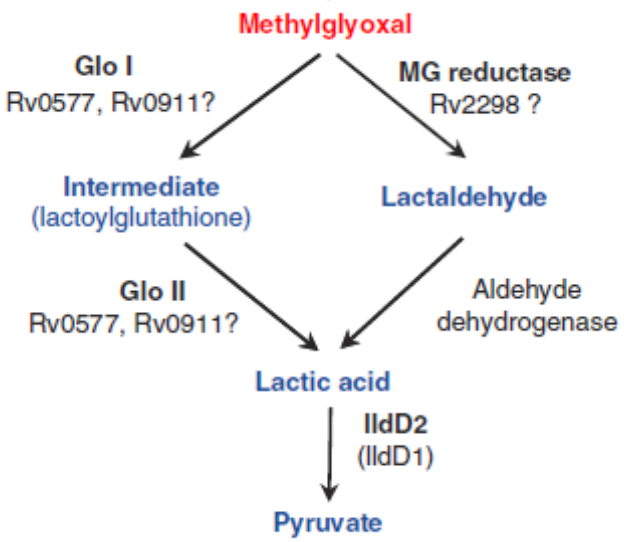


Figure 1.12 – The glyoxalase pathway for detoxification of methylglyoxal
The toxic metabolite methylglyoxal is degraded to produce pyruvate via lactic acid. This can occur via two mechanisms, either glyoxalase I-II enzymes (Glo I and Glo II) or methylglyoxal (MG) reductase and aldehyde dehydrogenase degrade methylglyoxal to lactic acid. Lactate dehydrogenase (lldD2) then converts lactic acid to pyruvate (Marrero *et al.*, 2010).

in the granuloma is entirely produced by the host or also excreted from *M. tuberculosis*, one hypothesis is that the bacterial glyoxalase system may be involved in not only protecting the bacteria from host-derived methylglyoxal but also in preventing methylglyoxal-induced apoptosis of macrophages.

1.9.2 Methylglyoxal and DNA damage

DNA damaging agents can come either from the environment or they can be endogenously produced by the cell's metabolism (Ferguson *et al.*, 1998). Methylglyoxal primarily attacks guanine bases of DNA molecules (Krymkiewicz, 1973), which are then repaired by the base and nucleotide excision repair mechanisms (BER and NER respectively).

BER is a two stage process where initially the damaged bases are recognised by N-glycosylases, which create single strand nicks near the damaged base and the damaged DNA is then repaired by DNA polymerase and DNA ligase (Friedberg, 2003). NER relies on recognition of damaged bases by the enzymatic complex formed by UvrA, UvrB and ATP. UvrA dissociates while UvrB remains bound to the damaged DNA and together with UvrC and UvrD excises a 13 bp region of single-stranded DNA around the damaged base. This is then repaired by DNA polymerase and DNA ligase (Sancar, 1996). A crucial feature of excision repair mechanisms is that while repair of one base is taking place, a section of the DNA becomes single-stranded. If a second damaged base is present on the opposite strand, close to the base being repaired, repair may start on this base before it finishes on the opposite strand and this can cause a double-strand break (DSB). DSBs can be fatal to

the cell and this has been predicted to play a role in cell death due to exposure of *E. coli* cells to methylglyoxal (Karschau *et al.*, 2011).

1.9.3 Methylglyoxal in *E. coli* – a fine line between survival and cell death

Methylglyoxal is produced by many bacteria yet it is highly toxic and therefore the reason for its production and what role(s) it plays has come under scrutiny. The system of methylglyoxal production and detoxification in *E. coli* has been well characterised. The glyoxalase I-II system converts methylglyoxal to lactic acid (Figure 1.12) in a glutathione-dependent manner. During this process two glutathione adducts (hemithiolacetal and S-lactoylglutathione) are produced and these are able to activate the KefB and KefC potassium efflux systems (Booth *et al.*, 2003). In the absence of an electrophile, glutathione negatively regulates the KefB and KefC systems but when an electrophile, such as methylglyoxal, is present the glutathione adducts formed activate these systems causing a rapid loss of intracellular potassium (Ferguson *et al.*, 1998). The loss of potassium is accompanied by an influx of sodium ions and protons leading to acidification of the cytoplasm, which has been shown to protect against the toxic effects of electrophiles (Ferguson *et al.*, 1995). It has been hypothesised that production of methylglyoxal and the resulting acidification of the cytoplasm may create a window of opportunity for the bacteria to adapt to a move from one environment to another (Booth *et al.*, 2003).

While there is very little information regarding the synthesis and detoxification of methylglyoxal in *M. tuberculosis*, this bacterium does express a *kefB* gene, which is predicted to encode a membrane transporter protein related to sodium/hydrogen

exchanger proteins (TubercuList). In addition there are several genes such as *Rv0577* and *Rv0911* that have been hypothesised to encode enzymes that form part of the methylglyoxal detoxification pathway (Figure 1.12).

1.10 Hypotheses & Aims

This study focuses on determining the role of the ECF sigma factor, SigG, in *M. tuberculosis* as well as its regulation, both transcriptional and post-translational. The interest in this sigma factor comes from the observation that it is upregulated in response to DNA damage and during infection of human macrophages.

The main hypotheses to be investigated are:

- The expression of SigG is regulated at the level of transcription
    - SigG is expressed from multiple promoters
    - These promoters are subject to differential regulation
- The activity of SigG is controlled post-translationally
    - SigG is co-transcribed with Rv0181c and Rv0180c
    - Protein interactions occur amongst the gene products of the SigG operon
    - Rv0181c and Rv0180c act as anti-sigma and anti-anti-sigma factors to control the activity of SigG
- *Rv0654*, *Rv0887c* and *Rv0911* are expressed from SigG-dependent promoters
- SigG regulates genes involved in *M. tuberculosis* survival inside macrophages potentially in detoxification of electrophiles

In order to address these hypotheses, the major aims of this study are to:

- Characterise the promoters expressing SigG
- Determine whether Rv0181c and Rv0180c act as anti-sigma and anti-anti-sigma factors respectively
- Identify the stimulus/stimuli to which SigG responds
- Characterise the phenotype of a *sigG* operon deletion strain

2. Materials and Methods

Please see Appendix I for detailed composition of all commonly used media/buffers.

2.1 Bacterial and Yeast Strains & Growth Media

The names and genotypes of all bacterial and yeast strains used in this study are detailed in Table 2.1.

E. coli strains were grown either on LB agar or in LB broth supplemented with the appropriate antibiotic(s) for selection (Table 2.2). Unless otherwise stated, growth was conducted overnight at 37˚C. Liquid cultures were shaken at 250 rpm. *M. tuberculosis* H37Rv and its derivatives and *M. smegmatis* mc^2155 strains were grown at 37˚C on 7H11 agar (Middlebrook) or in modified Dubos medium (Difco) supplemented with the appropriate antibiotics for selection (Table 2.2). When conducting fatty acid susceptibility tests, *M. tuberculosis* was grown in modified Dubos medium where the Tween-80 was replaced with 0.02 % Tyloxapol (Sigma-Aldrich) and without bovine serum albumin. For carbon-use experiments, bacteria were grown in 7H9 broth (Middlebrook) with 0.5% albumin, 0.085% NaCl, 0.05% Tyloxapol and carbon substrate at 0.1% (wt/vol) as per Marrero *et al.*, (Marrero *et al.*, 2010). Small scale liquid cultures were grown statically for approximately one week. They were then sub-cultured into the required final volume and incubated in rolling apparatus (Bellco Biotechnology) at 2 rpm. Yeast was grown on Yeast Nitrogen Base (YNB) agar containing 20 µg/ml histidine but without leucine or tryptophan.

2.2 Plasmids

Plasmids used and created in this study are detailed in Tables 2.3 and 2.4.

Table 2.1 - Bacterial & Yeast strains used in this study

Strain	Genotype
***E. coli* strains**	
α-select silver/gold efficiency (Bioline)	*deoR endA1 recA1 relA1 gyrA96 hsdR17*(rk -mk+) *supE44 thi-*1 Δ(*lacZYA-argFV169*) φ80d*lacZ*ΔM15 F-
XL-1 Blue Supercompetent cells (Stratagene)	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*q*ZΔM15* Tn*10* (Tetr)]
One-shot® TOP10 chemically competent cells (Invitrogen)	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *araD139* Δ(*araleu*)7697 *galU galK rpsL* (StrR) *endA1 nupG*
Tuner (DE3) (Novagen)	F^- *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm lacY1*(DE3)
***M. smegmatis* strains**	
mc^2155	High-efficiency transformation mutant (Snapper *et al.*, 1990)
***M. tuberculosis* strains**	
H37Rv	Wild-type laboratory strain (Kubica *et al.*, 1972)
ΔsigGoperon	H37Rv with the *sigG* operon deleted
Yeast strains	
Y187	MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met–, gal80Δ, MEL1, URA3::GAL1UAS - GAL1TATA-lacZ

Table 2.2 - Media Supplements used in this study

Media Supplement	**Concentration for *E. coli* (µg/ml)**	**Concentration for *M. tuberculosis* (µg/ml)**
Kanamycin	50	25
Ampicillin	100	n/a
Gentamycin	20	15
Hygromycin	250	50
Chloramphenicol	34	n/a
X-Gal	200	50
Sucrose	n/a	20 (mg/ml)

Table 2.3 - Plasmids used in this study

Plasmid	Description	Reference
Commercial plasmids		
pCR4Blunt-TOPO	TOPO-activated vector for cloning blunt-end PCR products, Kan^R, Amp^R	Invitrogen
pCR-BluntII-TOPO	TOPO-activated vector for cloning blunt-end PCR products, Kan^R, Zeo^R	Invitrogen
pET28a	pBR322 derivative, N-terminal His-Tag/thrombin/T7-Tag and C-terminal His-Tag expression vector, Kan^R	Novagen
pGEX-6P-1	N-terminal GST-Tag expression vector, Amp^R	Amersham Biosciences
Laboratory plasmids		
pGAD-C1	Y2H plasmid containing the Gal4 activation domain. Amp^R for *E. coli* selection, LEU2 for yeast selection	(James *et al.*, 1996)
pGBD-C1	Y2H plasmid containing the Gal4 DNA binding domain. Amp^R for *E. coli* selection. TRP1 for yeast selection.	(James *et al.*, 1996)
pDHADWhiB1	pGAD-C1 containing the WhiB1 gene	D. Hunt (personal communication)
pDHBDWhiB1	pGBD-C1 containing the WhiB1 gene	D. Hunt (personal communication)
pEJ414	Integrating mycobacterial *lac*Z transcriptional reporter vector	(Papavinasasundaram *et al.*, 2001)
pCA26	*E. coli lac*Z reporter vector, $Chlor^R$	R. Balhana (personal communication)
pBackbone	pBluescript derivative, *M. tuberculosis* suicide vector, Kan^R, Amp^R	(Gopaul, 2002)
pUC-GM	Contains $Gent^R$ cassette, $Gent^R$, Amp^R	(Schweizer, 1993)
pGoal17	Contains *sac*B/*lac*Z cassette, Amp^R	(Parish *et al.*, 2000)
pBS-Int	Suicide vector containing integrase, Amp^R	(Springer *et al.*, 2001)
pKP186	pMV306 derivative, integrating vector but does not encode integrase, Kan^R	(Rickman *et al.*, 2005)
pEJMyc	Myc tag vector, Kan^R	(Smollett *et al.*, 2009)
pLDlac1	pEJ414 containing an oligonucleotide of SigG putative promoter P1	L. Dawson (personal communication)
pLDlac1-mut	pLDlac1derivative with an A to C mutation in the -10 region	L. Dawson (personal communication)
pLDlac2	pEJ414 containing an oligonucleotide of SigG putative promoter P2	L. Dawson (personal communication)
pLDlac2-mut	pLDlac2derivative with a GTA to TGC mutation in the -10 region	L. Dawson (personal communication)
pLDlac3	pEJ414 containing an oligonucleotide SigG putative promoter P3	L. Dawson (personal communication)
pLDlac3-mut1	pLDlac3derivative with a GTC to TGT mutation in the -10 region	L. Dawson (personal communication)
pLDlac3-mut2	pLDlac3derivative with a GACC to TCTT mutation in the -35 region	L. Dawson (personal communication)
pLDcomp8T	pKP186 containing full *sig*G operon	L. Dawson (personal communication)
pLDcompΔ1	pKP186 containing *sig*G only	L. Dawson (personal communication)

Table 2.3 continued

pLDcompΔ2	pKP186 containing *sig*G and *Rv0180c*	L. Dawson (personal communication)
pLDcompΔ3	pKP186 containing *Rv0181c* and *Rv0180c*	L. Dawson (personal communication)
pLDcompΔ4	pKP186 containing *sig*G and *Rv0181c*	L. Dawson (personal communication)
pJH05	pET28a containing *sig*G	J. Dillury (personal communication)
pRAE2	pCA26 containing 223 bp Rv0654 promoter region, $Chlor^R$	R. Balhana (personal communication)
pKS12	Replicating overexpression vector containing the *sigG* promoters but not the *sigG* gene	K. Smollett (personal communication)
pKS09	Replicating overexpression vector, expressing *sigG* from its own promoters	K. Smollett (personal communication)
pBAC17	Bacterial artificial chromosome containing the 970935-1028020 region of the *M. tuberculosis* H37Rv genome	(Brosch *et al.*, 1998)

Table 2.4 - Plasmids created in this study

Plasmid	Description
Plasmids specific to this study	
pAG01	pEJMyc containing SigG and 488 bp upstream of H37Rv annotated start
pAG01-mut1	pAG01 derivative missing residue 1a; SDM using primers SDMSGF1 and SDMSGR1
pAG01-mut2	pAG01 derivative missing residue 1b; SDM using primers SDMSGF2 and SDMSGR2
pAG01-mut3	pAG01 derivative missing residue 1c; SDM using primers SDMSGF3 and SDMSGR3
pAG02	pEJMyc containing SigG, Rv0181c and 488 bp upstream of H37Rv SigG annotated start
pAG02-mut1	pAG02 derivative missing residue 2a; SDM using primers SDM181F1 and SDM181R1
pAG02-mut2	pAG02 derivative missing residue 2b; SDM using primers SDM181F2 and SDM181R2
pAG03	pEJMyc containing SigG, Rv0181c, Rv0180c and 488bp upstream of H37Rv SigG annotated start
pAG03-mut1	pAG03 derivative missing residue 3a; SDM using primers SDM180F1 and SDM180R1
pAG04	pEJ414 + 335 bp upstream region of SigG containing all 3 putative SigG promoters
pAG04-mut1	pAG04 derivative with an A to C mutation in P1-10; SDM using primers P1SDMF and P1SDMR
pAG04-mut2	pAG04 derivative with a GTA to TGC mutation in P2-10 and GTC to TGT mutation in P3-10; SDM using primers P2SDMF, P2SDMR, P3SDMF1 and P3SDMR1
pAG04-mut3	pAG04 derivative with an A to C mutation in P1-10 and GTC to TGT mutation in P3-10; SDM using primers P1SDMF, P1SDMR, P3SDMF1 AND P3SDMR1
pAG04-mut4	pAG04 derivative with an A to C mutation in P1-10 and CGG to TCT mutation in P3-10; SDM using primers P1SDMF, P1SDMR, P3SDMF2 AND P3SDMR2
pAG04-mut5	pAG04 derivative with an A to C mutation in P1-10 and GTA to TGC mutation in P2-10; SDM using primers P1SDMF, P1SDMR, P2SDMF and P2SDMR
pAG04-mut6	pAG04 derivative with an A to C mutation in P1-10, a GTA to TGC mutation in P2-10 and a GTC to TGT mutation in P3-10; SDM using primers P1SDMF, P1SDMR, P2SDMF, P2SDMR, P3SDMF1 and P3SDMR1
pAG04-mut7	pAG04 derivative with an A to C mutation in P1-10, a GTA to TGC mutation in P2-10 and a CGG to TCT mutation in P3-10; SDM using primers P1SDMF, P1SDMR, P2SDMF, P2SDMR, P3SDMF2 and P3SDMR2
pAG04-mut8	pAG04 derivative with the entire P2-10 and P3-10 regions mutated to just C residues; SDM using primers P2SDMF2, P2SDMR2, P3SDMF3 and P3SDMR3
pAG04-mut9	pAG04 derivative with an A to C mutation in P1-10 and the entire P3-10 region mutated to C residues; SDM using primers P1SDMF, P1SDMR, P3SDMF3 and P3SDMR3
pAG04-mut10	pAG04 derivative with an A to C mutation in P1-10 and the entire P2-10 region mutated to C residues; SDM using primers P1SDMF, P1SDMR, P2SDMF2 and P2SDMR2
pAG04-mut11	pAG04 derivative with an A to C mutation in P1-10 and the entire P2-10 and P3-10 regions mutated to just C residues; SDM using primers P1SDMF, P1SDMR, P2SDMF2, P2SDMR2, P3SDMF3 and P3SDMR3
pAG05	pBackbone containing 1379 bp 5' and 1368 bp 3' regions of SigG operon, $Gent^R$ and *sac*B/*lac*Z
pAG06	pET28a containing *Rv0181c*
pAG07	pGAD-C1 containing full *sig*G sequence
pAG08	pGAD-C1 containing full *Rv0181c* sequence
pAG09	pGAD-C1 containing full *Rv0180c* sequence

Table 2.4 continued

pAG10	pGAD-C1 containing a 510 bp non-transmembrane region of *Rv0180c*
pAG11	pGBD-C1 containing full *sig*G sequence
pAG12	pGBD-C1 containing full *Rv0181c* sequence
pAG13	pGBD-C1 containing full *Rv0180c* sequence
pAG14	pGBD-C1 containing a 510 bp non-transmembrane region of *Rv0180c*

2.3 Recombinant DNA techniques

2.3.1 Polymerase chain reaction (PCR)

PCR reactions where the product was required for downstream cloning or sequencing reactions were conducted using a proofreading DNA polymerase. Reactions were performed in a final volume of 50 µl consisting of 1 x *Pfx* buffer, 1 mM $MgSO_4$ and 1 x PCR_x Enhancer Solution (Invitrogen), 300 µM each dNTP (Promega), 300 nM each primer (Eurogentec/Sigma-Aldrich), 100-150 ng DNA or 5 µl InstaGene Matrix (Bio-Rad) DNA preparation as template, and 1 U Platinum *Pfx* DNA polymerase (Invitrogen). PCR reactions that didn't require a proofreading DNA polymerase such as colony PCR and reverse transcriptase PCR (RT-PCR), were performed in a final volume of 50 µl containing 1 x REDTaq ReadyMix PCR mix, which includes 1.5 mM $MgCl_2$, 200 µM each dNTP and 1.5 U Taq DNA polymerase (Sigma-Aldrich), 5 % DMSO, 300 nM each primer, 75-100 ng DNA/100 ng cDNA or a single colony resuspended in the PCR mix as template. PCR reactions were performed in a DNA Engine Peltier thermal cycler (Bio-Rad) with the lid heated to 110 ^{o}C to prevent evaporation of samples. An initial denaturation step of 95 ^{o}C for four minutes was performed, followed by 25-30 cycles of 95 ^{o}C for 30 seconds, 50-65 ^{o}C annealing temperature for 30 seconds (this was determined according to the melting temperature of the primers used) and 68 ^{o}C (*Pfx* polymerase*)* or 72 ^{o}C (*Taq* polymerase) extension for one minute per kb. Reactions finished with a seven minute extension at the appropriate temperature for the enzyme used (68/72 ^{o}C). PCR products were analysed by agarose gel electrophoresis.

2.3.2 Agarose gel electrophoresis

A 1-2 % (w/v) agarose (Bio-Rad) solution was prepared in 1 x TBE buffer and boiled to ensure the agarose completely dissolved. Ethidium bromide (Bio-Rad) was added to a final concentration of 0.2 μg/ml and the solution was poured into a gel cast with a well comb inserted. Gels were run in 1 x TBE buffer. Samples containing 1 x DNA loading buffer were loaded into the wells alongside an appropriate DNA ladder (100 bp BenchTop Ladder, Promega or Hyperladder I, Bioline). Gels were run at 80-120 V for approximately one hour and the DNA in the gel was visualised and photographed using an ultraviolet (UV) transilluminator (BioDoc It Imaging System, UVP).

2.3.3 Plasmid DNA extraction

Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen) as per the manufacturer's guidelines. 10 ml *E. coli* stationary phase cultures were lysed under alkaline conditions and SDS in the presence of RNase A. SDS solubilised the cell membrane leading to lysis, while the alkaline conditions denatured proteins, chromosomal and plasmid DNA. The lysate was then neutralised and adjusted to high salt conditions, causing denatured proteins, chromosomal DNA, cellular debris, and SDS to precipitate out, leaving the smaller plasmid DNA to renature in solution. The plasmid DNA was adsorbed onto a silica membrane, washed with an ethanol-based solution to remove salts and eluted in salt-free dH_2O. Plasmid DNA was quantified with a ND-1000 spectrophotometer (Nanodrop).

2.3.4 Restriction enzyme digestion of DNA

DNA was digested using FastDigest restriction enzymes (Fermentas) in 1 x FastDigest Buffer. Manufacturer's guidelines were followed when selecting the temperature and the reaction time for optimal digestion. Digested DNA fragments were visualised via agarose gel electrophoresis. When necessary for downstream procedures, restriction enzymes were either inactivated by heat denaturation according to manufacturer's guidelines or separated from the digested DNA by purifying the DNA following agarose gel electrophoresis.

2.3.5 PCR product & digested DNA purification

PCR products and selected digested DNA fragments were purified using the QIAquick Gel Extraction Kit (Qiagen) either straight from the PCR reaction or following agarose gel electrophoresis as per the manufacturer's guidelines. When required the agarose was solubilised to release the DNA into solution. The DNA was adsorbed onto a silica membrane at pH 7.5 under high salt conditions. Following a wash with an ethanol-based solution DNA was eluted in dH_2O. Purified DNA was quantified with a ND-1000 spectrophotometer (Nanodrop).

2.3.6 Dephosphorylation and ligation of DNA

Removal of the 5' phosphate group from linearised vector DNA prevents re-ligation unless an insert is present, thus increasing the proportion of successful ligation reactions. Dephosphorylation and ligation was performed using the Rapid DNA Dephos & Ligation Kit (Roche) as per manufacturer's guidelines. For each cloning reaction the linearised vector plasmid was first dephosphorylated. The maximum

dephosphorylation capacity of the kit; 1 µg DNA per 20 µl reaction, containing 1 x rAPid Antarctic Phosphatase buffer and 1 U rAPid Antarctic Phosphatase and maximum ligation capacity; 200 ng total DNA per 21 µl reaction, containing 1 x T4 DNA Ligation buffer and 5 U T4 DNA Ligase, were never exceeded. Dephosphorylation reactions were incubated at 37 ºC for 30 minutes and the enzyme was inactivated at 75 ºC for 2 minutes. Ligation reactions were performed using a vector to insert ratio of 1:3 and incubated at room temperature for 5 minutes. Reactions were used immediately to transform chemically competent *E. coli* as per Section 2.3.10.

2.3.7 RNA extraction

RNA was extracted from either 50 ml *E. coli* or 30 ml *M. tuberculosis* roller cultures, grown to the appropriate $OD_{600nm,}$ using the FastRNA Pro Blue Kit (Q-Biogene), as per the manufacturer's guidelines. Cells were harvested by centrifugation at 1,500 rcf for 15 minutes, resuspended in 1 ml RNApro solution and transferred to a Lysing Matrix B tube. Samples were lysed in a RiboLyser (Hybaid) at setting 6.5 for 30 seconds. Cell debris and matrix were removed via centrifugation at 15,000 rcf for 5 minutes at 4 ºC. RNA was purified via chloroform extraction (Section 2.4.1) and concentrated via ethanol precipitation (Section 2.3.8). Finally RNA was resuspended in DEPC-treated H_2O and quantified with a ND-1000 spectrophotometer (Nanodrop).

2.3.8 Ethanol precipitation of nucleic acids

Nucleic acids were precipitated from solution by addition of 2.5 volumes cold 100 % ethanol and 0.1 volumes 3 M sodium acetate. The mixture was incubated at -20 ^{o}C overnight and then centrifuged at 15,000 rcf for 15 minutes at + 4 ºC. The pellet was washed with 2.5 volumes cold 70 % ethanol, centrifuged at 15,000 rcf for one minute and air-dried. DNA was resuspended in dH_2O and RNA in DEPC-treated H_2O.

2.3.9 Site-directed Mutagenesis (SDM)

Two complimentary primers containing the desired mutation were designed according to the QuikChange SDM Kit protocol (Stratagene). The SDM PCR final reaction volume was 50 µl, consisting of 1x *Pfx* Buffer, 1 mM $MgSO_4$ and 1 x PCR_x Enhancer Solution (Invitrogen), 300 µM of each dNTP (Promega), 125 ng each primer (Eurogentec/Sigma-Aldrich), 100 ng plasmid DNA Miniprep and 1 U Platinum *Pfx* DNA polymerase (Invitrogen). The following PCR cycle was used for SDM: initial denaturation for 1 minute at 95 ºC followed by 25 cycles of 50 seconds at 95 ºC; 50 seconds at 65 ºC; 1 minute per kb whole template plasmid at 68 ºC and a final extension of 7 minutes at 68 ºC. Following PCR, 30 U FastDigest *Dpn*I restriction enzyme (Fermentas) was added to SDM reactions before incubation at 37 ºC for 15 minutes. A further 20 U *Dpn*I and 15 minute incubation was used to ensure complete digestion of the non-mutated template DNA. SDM reactions were ethanol precipitated (Section 2.3.8) and resuspended in 5 µl dH_2O. 1.25µl was used to transform 25 µl XL-1 Blue Supercompetent cells. A list of all SDM primers can be found in Appendix II.

2.3.10 Transformation of competent *E. coli*

Chemically competent *E. coli* were transformed according to manufacturer's guidelines. Unless otherwise stated 1 µl DNA solution was added to 25 µl chemically competent *E. coli* cells. The mixture was incubated on ice for 30 minutes, heat-shocked at 42 ºC for 45 seconds and placed back on ice for a further 2 minutes. 250 µl LB broth was added to the cells, which were incubated at 37 ºC for 1 hour prior to plating on LB agar containing the appropriate supplements.

2.3.11 *E. coli* frozen stocks

E. coli transformed with plasmid constructs were stored at -80 ºC in LB broth containing 15 % glycerol.

2.3.12 Sequencing of DNA

Plasmids and some PCR products were checked by automated DNA sequencing, carried out by GeneService at UCL, London.

2.4 Mycobacteria-specific Techniques

2.4.1 DNA Extraction

Crude DNA for confirming the presence of a specific plasmid in *M. tuberculosis* transformants by PCR was extracted using the InstaGene Matrix (Bio-Rad) as per the manufacturer's guidelines.

High quality genomic DNA was extracted by resuspending *M. tuberculosis* (approximately ¼ of a plate) in 300 µl sterile TE buffer. Bacteria were heat-killed at 80 ºC for at least 1 hour. Lysozyme (Sigma) and lipase (Sigma) were added at a final concentration of 2 mg/ml together with 2.5 µg of DNase-free RNase (Roche) and the reaction was incubated for 2 hours at 37 ºC. Samples were snap-frozen on dry ice/ethanol and incubated for 10 minutes at 75 ºC. After cooling to room temperature, Proteinase K (Roche) and SDS (Bio-Rad) were added at final concentrations of 0.5 mg/ml and 0.5% respectively. Following an incubation of 1 hour at 50 ºC, DNA was extracted by mixing with an equal volume of phenol:chloroform:isoamyl alcohol (PCI 25:24:1, Sigma) and removing the aqueous layer. This was repeated twice for PCI and once using neat chloroform (Sigma). DNA was ethanol precipitated (Section 2.3.8) and resuspended in sterile TE buffer.

2.4.2 Preparation of competent *Mycobacteria*

M. smegmatis was grown at 37 ºC in a 100 ml culture to OD_{600nm} 0.7-1.0 whereas *M. tuberculosis* was grown at 37 ºC in a 100 ml roller culture to OD_{600nm} 1.0 at which time 10 ml of 2 M glycine (filter sterile) was added and the culture incubated for a further 20-24 hours. Both *M. smegmatis* and *M. tuberculosis* cells were harvested by centrifugation at 16,266 rcf for 20 mins and washed four times in 100 ml sterile 10% glycerol. Cells were resuspended in 10 ml sterile10% glycerol and were used for electroporation.

2.4.3 Electroporation of *Mycobacteria*

400 µl competent cells were electroporated, with 100 ng plasmid miniprep (for transformations) or with 1-2 µg plasmid DNA (for knockout constructs), at either 2.2 kV (*M. smegmatis*) or 2.5 kV (*M. tuberculosis*), 25 µF, 1000 Ω (Bio-Rad gene pulser). *M. smegmatis* cells were incubated for three hours and *M. tuberculosis* cells overnight in static culture at 37 ºC prior to plating. For transformations a serial dilution to ~ 10^{-3} was set up and 100 µl of each dilution was plated onto 7H11 containing appropriate antibiotic. For knockout mutants, cells were harvested by centrifugation, resuspended in ~ 100 µl supernatant and plated onto 7H11 containing the appropriate antibiotic and X-gal.

2.4.4 Preparation of cell-free extract (CFE) from *Mycobacteria*

M. smegmatis or *M. tuberculosis* cells were grown in a 30 ml roller culture to the appropriate OD_{600nm}. Cells were harvested by centrifugation at 1,500 rcf for 15 minutes then washed twice in 10 mls and once in 1 ml cold PBS (1x) or Z buffer (Miller, 1972) without β-mercaptoethanol (referred to as Z* buffer). Cell pellets were re-suspended in 1 ml cold PBS (1x) or Z* buffer and placed in tubes one third full of 150-212 µm glass beads (Sigma-Aldrich). Samples were lysed in a RiboLyser at speed 6.5 for 30 seconds and then centrifuged for 5 mins at 15,000 rcf at 4 ºC. To ensure complete removal of intact cells from *M. tuberculosis* CFE only, the supernatant was filtered through Ultrafree-MC 0.22 µm spin columns (Millipore) following the manufacturer's protocol, prior to removal from the category 3 containment facility.

2.4.5 Mycobacterial frozen stocks

Frozen stocks from agar plates were made by re-suspending colonies in 200 µl freeze mix, whereas liquid cultures were frozen without any supplement due to the fact that the mycobacterial cell wall can withstand freezing.

2.5 Transcriptional analysis

2.5.1 DNA removal from RNA preparations

The TURBO DNA-free kit (Ambion) was used to clean up RNA samples as per the manufacturer's guidelines. Samples were incubated with 4 U TURBO DNase in 1 x TURBO DNase buffer at 37 ºC for 30 minutes. A further 4 U TURBO DNase was added followed by another 30 minute incubation at 37 ºC. The enzyme was inactivated by adding 0.1 volumes DNase Inactivation Reagent and incubating at room temperature for 2 minutes. RNA, in solution, was separated from the DNase Inactivation Reagent by centrifugation at 15,000 rcf for 2 minutes. The supernatant was transferred to a clean tube and RNA was quantified on a ND-1000 spectrophotometer (Nanodrop).

2.5.2 cDNA synthesis

Conversion of RNA to cDNA was performed using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems) as per the manufacturer's guidelines. Reactions were carried out in a final volume of 20 µl, consisting of 1 x Reverse Transcription (RT) buffer, 1 µg DNase-treated RNA, 4 mM each dNTP, 1 x RT Random Primers and 50 U MultiScribe Reverse Transcriptase. For each cDNA reaction (RT+) a control RT-negative (RT-) reaction was also set up where the

reverse transcriptase was replaced with DEPC-treated H_2O. Reactions were incubated in a DNA Engine Peltier thermal cycler (Bio-Rad) for 10 minutes at 25 ºC followed by 2 hours at 37 ºC and finally 5 seconds at 85 ºC to inactivate the enzyme.

2.5.3 Quantitative RT-PCR (qRT-PCR)

qRT-PCR is a method used to quantitatively compare the expression of genes within different bacterial strains or under different conditions. It does this by measuring the amplification reaction as it progresses. SYBR Green qRT-PCR was used in this study. SYBR Green binds to the double stranded DNA produced by the RT-PCR reaction whereby it fluoresces, allowing direct quantification of the amount of product and therefore the amount of initial template in the sample. The RT reaction was carried out as per Section 2.5.2. cDNA was diluted 1 in 100 prior to quantitative analysis. A genomic DNA dilution series was used to produce a standard curve for each primer set. Each 20 µl qRT-PCR reaction consisted of 1 x Fast SYBR Green Master Mix (Applied Biosystems), 900 nM each primer and 5 µl diluted cDNA or 5 µl genomic DNA standard as template or 5 µl dH_2O no template control. qRT-PCR was performed on a 7500 Fast Real-Time PCR System (Applied Biosystems) using the following conditions; 95 ^oC for 20 seconds to activate the enzymes, followed by 40 cycles of 95 ^oC denaturation for 3 seconds and 60 ^oC annealing and extension for 30 seconds. Finally a dissociation step was performed at the end of the qRT-PCR reaction in order to determine the melting temperature (T_M) of the PCR product. This showed whether more than one product had been made and therefore whether the primers were specific or not. Results were only used if the primers were specific. cDNA was quantified using the cycle threshold (Ct) values, which is the PCR cycle where fluorescence became detectable above background levels, using the

appropriate DNA standard curve. RT- reactions were run alongside RT+ reactions to control for any chromosomal DNA contamination. Data was only used if RT- values were less than 5% of RT+ values, the only exception being where the gene of interest had been deleted. Values were averaged, adjusted for chromosomal DNA contamination by subtracting the RT- value from RT+ and finally normalised to the equivalent *sig*A values to obtain a normalised expression level for each gene.

2.5.4 RNA ligase-mediated rapid amplification of 5' cDNA ends (RLM-RACE)

Transcriptional start sites were identified using the GeneRacer kit (Invitrogen) for RLM-RACE as per the manufacturer's guidelines. This technique utilises RNA ligase to cap the RNA with a RNA oligonucleotide of known sequence followed by RT-PCR to amplify strands specific to the gene of interest. Firstly the 5' cap structure was removed from messenger RNA (mRNA) in a final volume of 10 µl, which included 1-5 µg DNase-treated RNA, 1 x tobacco acid phosphatase (TAP) buffer, 40 U RNaseOut and 0.5 U TAP. Samples were incubated at 37 ºC for 1 hour. RNA was then purified via phenol:chloroform extraction followed by precipitation as per Section 2.3.8 but with the addition of 20 µg mussel glycogen to aid precipitation. The de-capped RNA was added to 0.25 µg GeneRacer RNA Oligo and heated at 65 ºC for 5 minutes. The ligation reaction was then performed in 1 x ligase buffer, with 10 nm ATP, 40 U RNaseOut and 5 U T4 RNA ligase in a final volume of 10 µl, samples were incubated at 37 ºC for 1 hour. RNA was extracted and precipitated as described above.

RNA was re-suspended in DEPC-treated H_2O, to which 100 ng random primers and 0.5 mM each dNTP were added, and incubated at 65 ºC for 5 minutes. The RT reaction was carried out in a final volume of 20 µl, consisting of the RNA and primer mixture, 1 x First Strand buffer, 5 mM DTT, 40 U RNaseOut and 200 U SuperScript III RT, initially at 25 ºC for 5 minutes followed by 1 hour at 50 ºC. RT was inactivated at 70 ºC for 15 minutes, 2 U RNase H was added to remove the RNA template and incubated for 20 minutes at 37 ºC. cDNA was amplified using the GeneRacer 5' Primer and a gene-specific reverse primer (GSP), using Platinum *Pfx* DNA polymerase as per Section 2.3.1. The PCR reaction followed the manufacturer's guidelines for GeneRacer where a variation of touchdown PCR is used. This exploits the high annealing temperatures of the GeneRacer 5' Primer and the GSP to increase product specificity. Starting with a high annealing temperature causes only GeneRacer-tagged/gene-specific cDNA to accumulate in the initial cycles. The annealing temperature is then decreased and the remainder of the reaction follows a more conventional reaction cycle of denaturation, annealing and elongation in order for efficient amplification to occur. Cycle parameters were an initial 94 ºC denaturation step, 5 cycles of 94 ºC for 30 seconds and 72 ºC for 1 minute/kb DNA, 5 cycles of 94 ºC for 30 seconds and 70 ºC for 1 minute/kb DNA followed by 25 cycles of 94 ºC for 30 seconds, 65 ºC for 30 seconds and 68 ºC for 1 minute/kb DNA. The cycle finished with an elongation step of 68 ºC for 10 minutes. PCR products were analysed by agarose gel electrophoresis, gel extracted and sequenced directly. Transcriptional start sites were identified at the junction with the GeneRacer RNA Oligo.

2.6 Protein analysis

2.6.1 β-galactosidase assay

The β-galactosidase assay provides a quantitative measure of β-galactosidase enzyme activity. When the gene encoding β-galactosidase is linked to a promoter, this assay can be used to measure promoter activity.

Quantification of protein in β-galactosidase CFEs was conducted using the BCA protein assay kit (Pierce) as per the manufacturer's guidelines. Protein standards of bovine serum albumin (BSA) were prepared using Z* buffer and used at concentrations of 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 and 1.6 mg/ml. Reagents A and B were mixed in the ratio 50:1 and 200 µl of this solution was added to 10 µl standard or CFE in a 96-well PVC microtitre plate. Samples were incubated at 37 ºC for 30 minutes and the OD_{550nm} was measured in a microplate reader (Benchmark Plus Microplate Spectrophotometer, Bio-Rad). A standard curve was produced, from which CFE protein concentrations were calculated.

β-galactosidase assays were performed as described by Miller (1972). Following protein quantification 1.5 µl β-mercaptoethanol was added to the remaining CFE samples. 10-150 µl CFE was made up to 500 µl with Z buffer. Samples were pre-warmed for 5 mins at 28 ºC and 100 µl ONPG (4 mg/ml) was added to start the reaction. The development of a yellow colour over time is representative of the β-galactosidase activity in the sample. The reaction was stopped after development of the yellow colour or after 30 minutes (controls) by addition of 250 µl 1 M sodium carbonate. The intensity of the yellow colour was measured at OD_{405nm}. For each

bacterial clone, three biological replicates were used to make CFE and each CFE sample was assayed in duplicate to provide reliable data.

β-galactosidase activity was calculated as Miller units (MU) using the following formula:

$$MU = \frac{380 \times OD_{405\ nm}}{\text{Time (min)} \times \text{volume of CFE (ml)} \times \text{protein concentration (mg/ml)}}$$

2.6.2 Protein expression and purification of SigG and Rv0181c in *E. coli*

The *E. coli* expression strain Tuner (DE3) was used for expression of tagged proteins. The DNA sequence for the protein of interest was cloned into the desired IPTG-inducible expression vector pET28a. Expression conditions for soluble SigG protein had been determined previously (K. Smollett, personal communication). Initial tests to determine optimum expression conditions for soluble protein were conducted as per standard protocol (Sambrook *et al.*, 2001). Both SigG and Rv0181c were expressed and purified from the *E. coli* Tuner strain.

2.6.2.1 Growth conditions to express soluble SigG and Rv0181c

Liquid media was inoculated with 0.01 volumes of an overnight culture of *E. coli* Tuner containing either pJH05 (SigG expression) or pAG06 (Rv0181c expression). Cultures were incubated at 37 ºC until they reached an OD_{600nm} of 0.5. For both SigG and Rv0181c, protein expression was induced by addition of IPTG to a final concentration of 0.01 mM. Post-induction, SigG cultures were incubated overnight

(~ 15 hrs) at 25 ºC and Rv0181c cultures were incubated for 5 hours at 25 ºC followed by an overnight (~ 15 hr) incubation at 20 ºC. Cells were harvested by centrifugation at 18,592 rcf for 30 minutes.

2.6.2.2 Preparation of soluble SigG and Rv0181c proteins

Cell pellets were resuspended in 0.01 culture volumes 1x binding buffer, containing 1 mg/ml lysozyme (Sigma-Aldrich), 10 U/ml DNase (Promega) and 1 µg/ml RNase (Roche) and incubated on ice for 30 minutes. Cells were lysed by sonication at 20 % amplitude with 9 second pulses for a total of 5 minutes using a Digital Sonifier (Branson). The supernatant, containing the soluble protein fraction was separated from the cell debris and insoluble protein by centrifugation at 18,592 rcf at 4 ºC for 30 minutes. The soluble protein fraction was filtered through a 0.2 µm Ministart syringe filter (Sartorius Stedim Biotech) prior to purification using an Äkta prime (Amersham Biosciences) on a HiTrap Affinity Chelating HP column (GE Healthcare) manually loaded with 100 mM nickel sulphate hexahydrate. The protein sample was loaded onto the column, which was then washed with 1 x binding buffer until the UV trace returned to baseline levels. Loosely bound protein was washed from the column using 6 % 1 x elution buffer (containing ~ 60 mM imidazole). Pure SigG or Rv0181c was eluted from the column using 100 % elution buffer (containing 1 M imidazole). The full recipes for the binding and elution buffers can be found in Appendix I. Fractions collected during purification were analysed on a 12% SDS-PAGE gel. Further purification was conducted through a HiLoad 26/60 Superdex 200 prep grade gel filtration column (GE Healthcare), which was equilibrated using 1 x PBS containing 50 % glycerol prior to loading the protein sample. Pure SigG and Rv0181c were concentrated using a 10 kDa Vivaspin

concentrator (Sartorius Stedim Biotech). Pure SigG was used to prepare an anti-SigG antibody from rabbits (BioServ UK Ltd, Sheffield University).

2.6.3 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Denaturing SDS-PAGE gels were prepared according to Sambrook and Russell (2001) in 1 mm gel cassettes (Invitrogen) using a 37.5:1 acrylamide:bisacrylamide mixture (Bio-Rad). Gels were run in a 1 x Tris/Glycine/SDS buffer (Bio-Rad). 1 x Lane Marker Reducing Sample buffer (Thermo Fisher Scientific) was added to the samples, which were incubated for 5 minutes at 100 ºC prior to being loaded into the wells alongside a Dual Colour Prestained Precision Plus Protein Standard (Bio-Rad). Gels were run at 100-150 V for approximately 1 hour. Gels were removed from the cassette and either placed in Instant Blue Coomassie stain (Expedeon) for 15 minutes prior to destaining with multiple washes with deionised water or silver stained using the method as described by Schevchenko *et al*. (1996). Gels were imaged under white light using a UV transilluminator, then placed in drying solution (45 % methanol and 5 % glycerol) for 5 minutes, sealed between cellophane sheets (Invitrogen) and allowed to dry in a drying frame (Invitrogen). Alternatively gels were used for Western Blot analysis (Section 2.6.4).

2.6.4 Western Blot analysis

Protein samples were run on a SDS-PAGE gel. The gel was placed in semi-dry transfer (SDT) buffer for 5 mins and the PVDF Immobilon-P membrane (Millipore) was prepared by washing in methanol for a few seconds, then dH_2O followed by SDT buffer for 5 minutes. Proteins were electro-blotted onto the prepared membrane

using a semi-dry blotter (Continental Lab Products) at 1 mA per cm^2 for 1 hour as described in the QIAexpress Detection and Assay Handbook (Qiagen). In order to detect specific proteins, the membrane was placed in methanol, followed by water and washed twice in PBS for 10 mins. The membrane was then incubated in blocking buffer either for 1 hour at room temperature or overnight at 4 ºC. To remove excess buffer, the membrane was washed three times (2x PBS-Tween 20 followed by 1x PBS), each for 10 mins. The membrane was then placed in 10 mls blocking buffer containing the primary antibody at the required dilution and incubated for 1 hour. The three wash steps were repeated and if required the membrane was incubated in 10 mls blocking buffer containing a secondary antibody (e.g. goat anti-rabbit-HRP) for 1 hour followed by a further three wash steps. Primary antibodies used in this study were rabbit anti-SigG at a 1:1000 dilution, rabbit anti-Myc at a 1:2000 dilution, mouse anti-His HRP conjugate (Qiagen) at a 1:4000 dilution and rabbit anti-CRP/Rv3676 (D Hunt, personal communication) at a 1:5000 dilution.

The secondary antibody used was goat anti-rabbit HRP conjugate.

Detection was via ECL Western blotting detection reagents (GE Healthcare). Equal volumes of reagents 1 and 2 were mixed and poured onto the protein side of the membrane. After 1 minute excess liquid was removed, the membrane was wrapped in Saranwrap™ and exposed to autoradiograph film (GE Healthcare). Films were developed using an automatic developer (Fujifilm).

2.6.5 Pulldown using magnetic beads

Purified SigG protein at a concentration of 1.2 mg/ml or more was buffer exchanged into 1 x PBS using Zeba Desalt Columns (Pierce) as per the manufacturer's guidelines. 2 x 10^8 magnetic Epoxy Dynabeads (Invitrogen) were equilibrated in 0.1 M sodium phosphate pH 7.4. Dynabeads in 0.1 M sodium phosphate were mixed with 60 µg bait protein or antibody and ammonium sulphate was added to a final concentration of 50 mM. Protein was bound to the beads by incubation with shaking overnight at 37 ºC. The beads with bound bait protein were separated from the buffer using a magnet. The beads were blocked to prevent non-specific interactions by washing four times with 1 x PBS containing 0.5 % (w/v) BSA. Cell-free extract or protein ligand were added to the protein-coated beads together with 0.1 volumes 10 x PBS and incubated for 2 hours at +4 ºC in a rotator. Unbound protein was washed from the beads by washing three times with 1 x PBS. Samples were boiled for 5 minutes at 100 ºC in SDS-PAGE loading dye to release the bound proteins, the beads were removed using a magnet and proteins were analysed via SDS-PAGE. Where the target of the bait protein was unknown, protein bands of interest were excised and sent to the Cambridge Centre for Proteomics for LC-MS/MS analysis.

2.6.6 Trypsin digest for MALDI-MS

Following SDS-PAGE analysis, the protein band of interest was excised from the gel. SDS and Coomassie stain were extracted by incubation in 200 mM ammonium bicarbonate (ABC)/50% acetonitrile buffer for 30 minutes at 37 ºC. Incubation was repeated until the band was no longer blue. The protein was reduced with 20 mM DTT/200 mM ABC/50% acetonitrile buffer for 1 hour at 37 ºC and then washed with 200 mM ABC/50% acetonitrile buffer. Cysteine residues were alkylated by adding 5

mM iodoacetamide/200 mM ABC/50% acetonitrile buffer in the dark for 20 minutes. The band was then washed twice with 20 mM ABC/50% acetonitrile to remove the iodoacetamide. 100% acetonitrile was then added and incubated at room temperature until the band became white. All liquid was removed and the protein band was dried at room temperature for approximately 1 hour.

20 µg dry modified (methylated) trypsin (Promega) was dissolved in 5 mM ABC to a final concentration of 2 µg/ml. 0.1 µg trypsin was added to the dry protein band, which was then incubated at 32 ºC for 24 hours. The protein sample was then analysed by MALDI-MS (S. Howell, NIMR).

2.6.7 Yeast-2-hybrid protein analysis

SigG, Rv0181c, full-length Rv0180c and a 510 bp region of Rv0180c encoding the cytoplasmic domain were all amplified via PCR with the Y2H primers (Appendix II) and cloned into the vectors pGAD and pGBD. The *E. coli* α-select strain was used to select and prepare plasmids, which were sequenced before being used for yeast transformation.

The plasmids were co-transformed into the yeast strain Y187 as per the LiAc TRAFO protocol (Schiestl *et al.*, 1989). A positive control consisting of the self-interacting WhiB1 gene (pDHADWhiB1 and pDHBDWhiB1) and a general negative control consisting of the empty pGAD-C1 and pGBD-C1 vectors were also transformed. Specific negative controls were also used with Y187 being co-transformed with each gene in pGAD-C1 together with empty pGBD-C1 and empty pGAD-C1 together with each gene in pGBD-C1.

Transformed yeast were grown on YNB-leucine-tryptophan+histidine agar to select for cells expressing the LEU2 and TRP1 markers. Colonies were restreaked onto new YNB-leucine-tryptophan+histidine plates and assayed for production of β-galactosidase using a colony-lift filter assay as per the Yeast Protocols Handbook (Clontech).

2.6.8 *In vitro* transcription assay

DNA templates for *in vitro* transcription (IVT) were PCR products purified using the QIAquick Gel Extraction kit (Qiagen). Primer sequences can be found in Appendix II and recipes for all IVT mixes can be found in Appendix I.

E. coli core RNA polymerase and purified SigG in PBS/50 % glycerol were mixed in the ratio 1:10 and incubated for 30 minutes at 37 ºC. DNA was mixed with IVT reaction mix and incubated at 37 ºC for 2 minutes. 1 U RNA polymerase was added and the reaction was incubated for a further 5 minutes at 37 ºC followed by addition of 10 µCi [α-^{32}P]UTP and 30 µg each of ATP, GTP and CTP and a further 2 minute incubation. 5 µg heparin was then added followed by a further 5 minute incubation at 37 ºC. The reaction was finished by addition of 30 µg cold UTP and a final incubation for 10 minutes at 37 ºC. The samples were moved onto ice, 0.25 volumes precipitation mix and 1.5 volumes isopropanol were added and RNA was precipitated for a minimum of 10 minutes. RNA was harvested by centrifugation for 15 minutes at 15,000 rcf, the pellet was re-suspended in a final volume of 400 µl, containing 50 % isopropanol and 150 mM sodium acetate (pH 6.0), incubated on ice for at least 10 minutes prior to centrifugation for 15 minutes at 15,000 rcf. The RNA

pellet was washed in 70 % ethanol, dissolved in run-off loading buffer and boiled at 100 ºC for 2 minutes. Samples were analysed on a standard 6 % polyacrylamide/urea sequencing gel alongside Perfect RNA size markers prepared according to the manufacturer's guidelines (Novagen). Gels were exposed and developed using a phosphoimager screen (GE Healthcare).

2.6.9 Protein-DNA binding assay

2.6.9.1 Preparation of the DNA Probe

PCR products used for *in vitro* transcription were cut to approximately 220 bp by restriction digest leaving blunt ends and gel extracted. The resulting DNA fragments were then treated with calf intestinal phosphatase (CIP, Invitrogen) to remove the 5' phosphate in 1 x CIP buffer (Invitrogen), containing 10 U CIP for 30 minutes at 37 ºC. CIP was inactivated by heating to 75 ºC for 10 minutes. DNA probes were labelled with ^{32}P using T4 polynucleotide kinase (PNK, NEB) in 1 x PNK buffer (NEB) containing 1.5 µg BSA, γ ^{32}P-ATP and 20 U T4 PNK; reactions were incubated for 1 hour at 37 ºC. Labelled DNA was separated from proteins and free nucleotide using the QIAquick Gel Extraction Kit (Qiagen) as per Section 2.3.5.

2.6.9.2 Protein-DNA binding assay

All protein-DNA binding reactions were performed in a final volume of 10 µl consisting of IVT master mix, 0.01 U Poly(dI-dC) (Sigma), 1 ng ^{32}P-labelled DNA probe and 0-20 µg purified SigG in 1 x PBS/50 % glycerol. For competition assays the appropriate amount of un-labelled competitor DNA was added to the reaction.

The recombinant SigG protein was always added last and the reaction was incubated at 37 ºC for 30 minutes.

2.6.9.3 Non-denaturing PAGE and ^{32}P detection

Non-denaturing PAGE gels (6%) were prepared in gel casting cassettes and consisted of 0.5 x TBE buffer, 37.5:1 acrylamide:bis-acrylamide mixture, ammonium persulfate and TEMED. Gels were run in 0.5 x TBE buffer and samples containing 1 x DNA loading buffer were loaded into the wells. Gels were run at 30 mA per gel at +4 ºC for approximately 30 minutes, after which the gel was removed from the cassette, placed on a piece of 3 mm filter paper (Whatman), covered in SaranWrap and dried under heat and vacuum on a Slab Gel Dryer 4050 (Savant). The dried gel was then exposed and developed using a phosphoimager screen (GE Healthcare).

2.7 Phenotypic Assays

2.7.1 Biolog Phenotype MicroArray

A list detailing the range of Biolog plates can be found in Appendix III.

The two *M. tuberculosis* strains to be compared were cultured to an OD_{600nm} 0.5 and cells were harvested from 20 ml culture for 20 minutes at 1,500 rcf before being resuspended in an equal volume of sterile 1 x PBS. Cultures were incubated statically overnight at room temperature to starve the cells. Starved cells were harvested by centrifugation for 20 minutes at 1,500 rcf, and then resuspended in 0.5 volumes IF-0a base (BIOLOG). The OD_{600nm} of the samples was adjusted to a calculated value of 0.682 and cells were added to the media appropriate to the

plate(s) being used as per Appendix III to give a final OD_{600nm} of 0.05. 100 µl media/cells mixture was added to each well in the 96-well Biolog plates (BIOLOG) and the plates were incubated for 7 days at 37 ºC. Cells were lysed by addition of a 20 % SDS/50 % dimethylformamide (DMF) to give final concentrations of 10 % SDS and 25 % DMF per well. Plates were fumigated with formaldehyde overnight to ensure all bacteria had been killed prior to removal from the Class I Biosafety Cabinet. The plates were analysed by reading the absorbance at 570nm using a microplate reader (BMG).

2.7.2 Alamar Blue cell viability assay

Cell-titer blue reagent (Promega) comprises a non-toxic, cell permeable compound called resazurin that is blue in colour and non-fluorescent. Upon entering cells resazurin is reduced to the highly fluorescent compound resorufin, which is red in colour. This reaction only occurs in viable cells, and is quantitative with the amount of fluorescence produced directly relating to the number of viable cells in the media. Used in conjunction with various inhibitory compounds the alamar blue assay can be used to compare the susceptibility of different bacterial strains to these compounds e.g. a mutant strain versus wild-type.

M. tuberculosis strains to be compared were cultured to an OD_{600nm} 0.5 to ensure all cells were in a mid-logarithmic phase of growth. Cells were diluted to an OD_{600nm} 0.01 (5×10^3 cells/µl).

A 2-fold serial dilution of the chosen inhibitor was performed in culture media in a 96-well microtitre plate. Each plate contained two inhibitors with three technical

replicates per concentration, three positive controls (minus inhibitor) and three negative controls (minus cells). 5 x 10^4 cells were added per well and the plates were incubated at 37 ºC for 10 days after which, 0.2 volumes Cell-titer blue reagent was added per well and the plates were incubated overnight at 37 ºC. In order to remove the plates from the Class I Biosafety Cabinet they were sealed using aluminium sealing foil (Elkay). Fluorescence was analysed using a microplate reader (BMG).

2.7.3 Murine Bone Marrow Derived Macrophage (BMDM) Infection Model

2.7.3.1 Differentiation and activation of BMDM

All live animal procedures were performed by Joanna Dillury, who holds a personal licence for animal work. Bone marrow was flushed into 10 ml RPMI 1640 medium (Gibco, Invitrogen) supplemented with 1 % foetal calf serum (Gibco, Invitrogen), 20 μM L-glutamine, 10 mM sodium pyruvate, 100 μM HEPES and 50 nM β-mercaptoethanol (RPMI+) from the hind legs of female eight week old BALB/C mice using a 0.5 mm needle. Cells were centrifuged at 250 rcf for five minutes, the RPMI+ was removed and cells were re-suspended in 10 ml 0.83 % ammonium chloride and incubated at 37 ºC for 5 minutes to lyse the red blood cells in the samples. The cells were centrifuged again to remove the ammonium chloride and were washed in 10 ml 1 x PBS (Gibco, Invitrogen), before being re-suspended in 10 ml RPMI+ containing 20 % L-cell conditioned medium (RPMI-LCC-20). L-cell is the supernatant of mouse fibroblast L929 cultures, containing macrophage colony stimulating factor, which stimulates the differentiation of bone marrow derived monocytes into macrophages (Austin *et al.*, 1971). Bone and muscle tissue were removed by passing the cell suspension through a 70 μm nylon filter (BD

Biosciences), and the remaining monocytes were enumerated under the microscope in 50 % trypan blue stain (Fluka) using a Bright-Line hemocytometer (Sigma). Monocytes were diluted to 4 x 10^5 cells/ml in RPMI-LCC-20, 10 ml aliquots were plated into Petri dishes and were incubated at 37 ºC with 5 % CO_2. After four days incubation an additional 10 ml RPMI-LCC-20 was added to the differentiating macrophages, which were then incubated for a further two days. After this time adherent cells are fully differentiated macrophages. Medium, containing non-adherent cells was removed from the Petri dishes and discarded and 5 ml 1 x PBS containing 4 mM EDTA was added to each dish and mixed gently to detach the adherent macrophages. Macrophages were centrifuged at 250 rcf for 5 minutes, were re-suspended in 10 ml RPMI+ containing 5 % L-cell (RPMI-LCC-5), enumerated as before and further diluted to 4 x 10^5 macrophages/ml in RPMI-LCC-5. In preparation for infection, 2 x 10^5 macrophages were added to the appropriate number of wells of 24-well tissue culture plates (Nunc, Thermo Fisher Scientific), and 500 µl RPMI-LCC-5 with or without 20 ng/ml murine IFN-γ (Roche) was added to each well to produce activated or non-activated/naive macrophages, respectively. Tissue culture plates were incubated at 37 ºC with 5 % CO_2 for 18 hours to allow adherence and activation, after which macrophages were infected.

2.7.3.2 Preparation of *M. tuberculosis* strains

M. tuberculosis cultures were grown to an OD_{600nm} 0.5-1.0, centrifuged at 1,500 rcf for 15 minutes, washed and then re-suspended in 1 x PBS containing 0.05 % Tween 80. The cells were then centrifuged at 15 rcf to remove any aggregated cells. The single-cell suspension was transferred to a new tube and the OD_{600nm} was measured. On the basis that an OD_{600nm} of 1.0 is equivalent to 5 x 10^8 cells/ml (J.

Dillury, personal communication), the cells were diluted to 8 x 10^5 cells/ml in 1 x PBS containing 0.05 % Tween 80 and were used to infect prepared macrophages immediately.

2.7.3.3 Infecting BMDM with *M. tuberculosis* and post infection washing

Infection was performed in triplicate for each *M. tuberculosis* strain, under each condition i.e. naive or activated macrophages, for each time point. Macrophages were infected with a multiplicity of infection (MOI) of 0.1:1, with 2 x 10^4 bacteria being added to the 2 x 10^5 adhered macrophages in the 24-well tissue culture plate. Plates were incubated at 37 ºC with 5 % CO_2 for 4 hours to allow the macrophages to phagocytose the *M. tuberculosis*. After this time the media containing extracellular *M. tuberculosis* was removed from the wells, the infected macrophages were washed three times with 500 µl HBSS (Gibco, Invitrogen), and 1 ml of RPMI-LCC-5 with or without 10 ng/ml IFN-γ was added to the appropriate wells. Plates were incubated at 37 ºC with 5 % CO_2.

2.7.3.4 Lysing BMDM and enumeration of *M. tuberculosis*

At each time point, the supernatant was removed from wells. Macrophages were detached from wells and lysed in 500 µl dH_2O containing 0.05 % Tween 80 at room temperature for a minimum of 15 minutes. *M. tuberculosis* was enumerated by plating serial dilutions of the lysate onto 7H11 plates. Colonies were counted and colony forming units (CFU) per ml were calculated after 2-3 weeks incubation at 37 ºC.

2.7.4 Mouse infection model

2.7.4.1 Preparation of *M. tuberculosis* strains

M. tuberculosis cultures were grown to an OD_{600nm} 0.5-1.0, centrifuged at 1,500 rcf for 15 minutes, washed and then re-suspended in 1 x PBS containing 0.05 % Tween 80. The cells were then centrifuged at 15 rcf to remove aggregated cells. The single-cell suspension was transferred to a new tube and the OD_{600nm} was measured. On the basis that an OD_{600nm} of 1.0 is equivalent to 5×10^8 cells/ml (J. Dillury, personal communication), the cells were diluted to 1×10^7 cells/ml in 1 x PBS without Tween 80 and were used to infect BALB/C mice immediately. Four BALB/C mice were used per time point and were infected intravenously with 1×10^6 CFU (J. Dillury).

2.7.4.1 Harvest and enumeration of *M. tuberculosis* from mouse lungs and spleen

Harvesting of lungs and spleens was performed by Joanna Dillury. Briefly, the lungs and spleens were dissected out from the mice and placed in 5 ml dH_2O containing 0.05 % Tween 80 and 7-8 quarter inch ceramic spheres (MP Biomedicals). The organs were homogenised in a Fastprep (MP Biomedicals) at 4 m/s for 20 seconds. *M. tuberculosis* was enumerated by plating serial dilutions of each homogenate onto 7H11 plates. Colonies were counted and colony forming units (CFU) per ml were calculated after 2-3 weeks incubation at 37 ºC.

3. Characterisation of the promoters expressing *sigG*

3.1 Introduction

ECF sigma factors are subject to highly complex regulation, with one facet of this being at the transcriptional level. It is possible for ECF sigma factors to be expressed from multiple promoters, e.g. *sigB* and *sigE,* or from a promoter that is recognised by several sigma factors, e.g. *sigB* (Sachdeva *et al.*, 2009).

Primer extension experiments performed by Dawson (2005) have indicated the presence of three possible promoters for the *sigG* operon. Promoter 1 appears to be mitomycin C inducible, suggesting a role in the response to DNA damage, and Promoter 3 may be SigG-dependent, indicating an autoregulatory role, which is a common feature of ECF sigma factors (Helmann, 2002). These observations will form the basis for investigation of the control of expression of *sigG*.

3.2 Demonstrating co-transcription of *sig*G, *Rv0181c* and *Rv0180c*

Pairwise RT-PCR has been used to demonstrate co-transcription of *sigG* with *Rv0181c* and of *Rv0181c* with *Rv0180c* (Dawson, 2005). The limitation of this approach is that it does not confirm co-transcription of all three genes as it is possible that Rv0181c and Rv0180c could be co-transcribed from an internal promoter.

In order to confirm that *sigG*, *Rv0181c* and *Rv0180c* are co-transcribed, RT-PCR was performed on RNA extracted from *M. tuberculosis* using primers spanning from *sigG* through to *Rv0181c*, *Rv0180c* and also into the downstream gene *lprO* (Figure

3.1). Control reactions using *M. tuberculosis* H37Rv genomic DNA (positive control) and RNA without reverse transcription (negative control) were included. RT-PCR across the two intergenic regions of the operon yielded a positive result (Figure 3.1, product 5); however it was possible that there were internal promoters present so RT-PCR reactions were set up covering all possible scenarios. Reactions extending into the start of *lprO* were initially included as a control and no product was expected. Figure 3.1 shows conclusively that *sigG*, *Rv0181c* and *Rv0180c* are co-transcribed (reactions 1, 5, 6, and 8). However, there was a surprising result with the *lprO* controls which gave positive rather than negative results (reactions 3, 7, 9 and 11). This suggested that *lprO* may be co-transcribed with the three upstream genes and was in direct contradiction to previous results obtained by Dawson (2005) in which, the same primer at the start of the *lprO* gene had been used. To further investigate the involvement of *lprO*, RT-PCR was repeated using the *lprO*R2 reverse primer at the end of the *lprO* gene. No transcript was obtained spanning from the start of *sigG* to the end of *lprO* (reaction 4) or from the end of *Rv0181c* to the end of *lprO* (reaction 10). A positive result was obtained with RT-PCR from the end of *Rv0180c* to the end of *lprO* (reaction 12).

The RT-PCR results suggest the presence of two transcripts; one which encodes *sigG*, *Rv0181c* and *Rv0180c*, but runs into *lprO* before terminating, and one which starts in the *Rv0180c* gene and encodes *lprO*. Co-transcription of *sigG*, *Rv0181c* and *Rv0180c* but not *lprO* is shown by the fact that it was possible to obtain a transcript

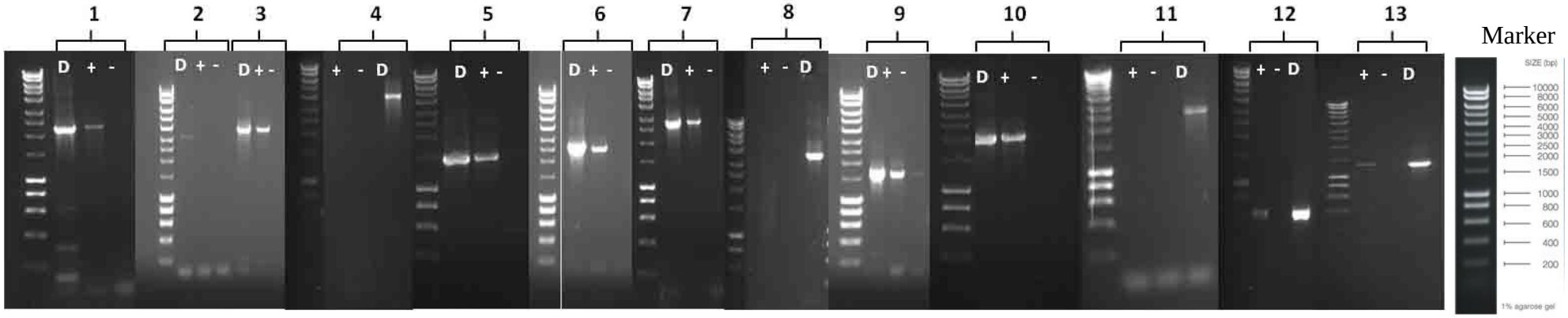


Figure 3.1 – RT-PCR showing co-transcription of *sigG, Rv0181c* and *Rv0180c* and a potential internal promoter expressing *lprO*

RT-PCR results to determine whether *sigG*, *Rv0181c* and *Rv0180c* are co-transcribed and whether *lpr*O is transcribed separately. 1) start of *sig*G to start of *Rv0180c*, 2254 bp; 2) start of *sig*G to end of *Rv0180c*, 3110 bp; 3) start of *sig*G to start of *lpr*O, 3489 bp; 4) start of *sig*G to end of *lpr*O, 4129 bp; 5) end of *sig*G to start of *Rv0180c*, 1464 bp; 6) end of *sig*G to end of *Rv0180c*, 2320 bp; 7) end of *sig*G to start of *lpr*O, 2699 bp; 8) end of *sig*G to end of *lpr*O, 3339 bp; 9) end of *Rv0181c* to end of *Rv0180c*, 1509 bp; 10) end of *Rv0181c* to start of *lpr*O, 1888 bp; 11) end of *Rv0181c* to end of *lpr*O, 2528 bp; 12) end of *Rv0180c* to start of *lpr*O, 649 bp; 13) end of *Rv0180c* to end of *lpr*O, 1289 bp. +) RT positive; -) RT negative and D) genomic DNA control. All reactions shown here were carried out on H37Rv *M. tuberculosis* nucleic acids. Marker is Hyperladder I (Bioline).

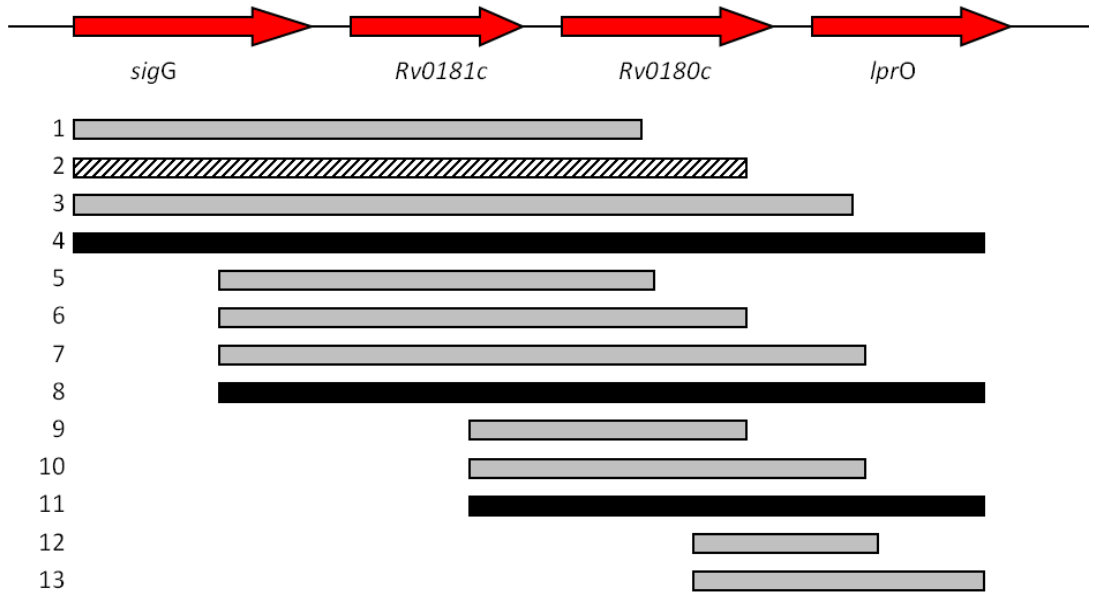


Figure 3.2 – Schematic representation of the locations of transcripts identified by RT-PCR

Grey bars indicate transcripts that were detected by RT-PCR. Black bars indicate transcripts that could not be detected by RT-PCR. Dashed bars indicate reactions where the DNA control was weak, indicating primer incompatibility. 1) start of *sigG* to start of *Rv0180c*; 2) start of *sigG* to end of *Rv0180c*; 3) start of *sigG* to start of *lprO*; 4) start of *sigG* to end of *lprO*; 5) end of *sigG* to start of *Rv0180c*; 6) end of *sigG* to end of *Rv0180c*; 7) end of *sigG* to start of *lprO*; 8) end of *sigG* to end of *lprO*; 9) end of *Rv0181c* to end of *Rv0180c*; 10) end of *Rv0181c* to start of *lprO* ; 11) end of *Rv0181c* to end of *lprO*; 12) end of *Rv0180c* to start of *lprO*; 13) end of *Rv0180c* to end of *lprO*. Transcription was shown from the beginning of the *sigG* gene to the beginning of *lprO* but not to the end of *lprO*, and from the end of *Rv0180c* to the end of *lprO*, suggesting an internal promoter.

that goes from the start of *sigG* to approximately 250 bp into *lprO* (Figure 3.1, band 3) but it is not possible to obtain a transcript going from either the start of *sigG* or the end of *Rv0181c* through to the end of *lprO* (Figure 3.1, reactions 4 and 10 respectively). However, it was possible to obtain transcripts going from the end of *Rv0180c* to both the start and end of *lprO* (Figure 3.1, reactions 11 and 12). The primer used for the RT-PCR reactions for Figure 3.1, reactions 11 and 12 (primer Cf, Appendix II) was 300 bp upstream of the predicted end of the *Rv0180c* coding region. This shows that there may be an internal promoter present >300 bp upstream of the end of the *Rv0180c* coding region, which drives expression of *lprO*.

3.3 Only promoter P1 was active in oligonucleotide promoter-*lacZ* constructs and was mitomycin C inducible

Previous primer extension studies have shown that there are 2 or 3 promoters expressing *sigG* and that promoter P1 was induced by the DNA-damaging agent mitomycin C(Dawson, 2005). In order to confirm that these potential promoters had promoter activity, transcriptional fusions to a *lacZ* reporter gene were used, in which each promoter was contained within an approximately 60bp double-stranded oligonucleotide. The oligonucleotides extended beyond the transcriptional start sites by either 4 or 5 base pairs (Figure 3.3). Wild-type (WT) H37Rv was transformed with the empty vector pEJ414, plasmids pLDlac1, pLDlac2, pLDlac3 (containing promoters P1, P2 and P3 respectively) and their derivatives pLDlac1-mut (A to C mutation in the P1 -10 region), pLDlac2-mut (GTA to TGC mutation in the P2 -10 region), pLDlac3-mut1 (GTC to TGT mutation in P3 -10 region) and pLDlac3-mut2 (GACC to TCTT mutation in P3 -35 region). The mutations made in promoters P2 and P3 were chosen to make a major change in the promoter sequence that did not

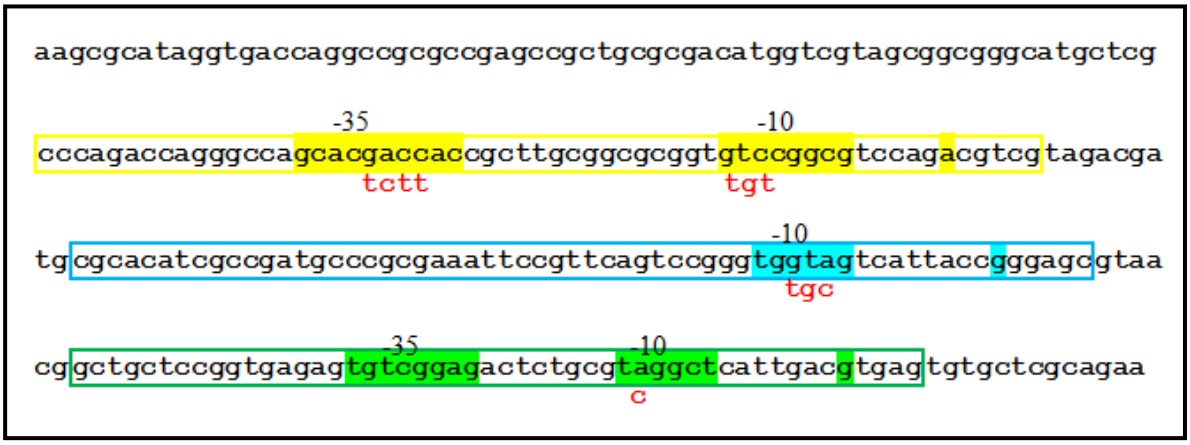


Figure 3.3 – The oligonucleotides and mutations used to produce the pLDlac promoter-lacZ constructs

The positions of the three oligonucleotides used to create the pLDlac oligonucleotide promoter-*lacZ* constructs within the sequence upstream of *sigG*. The sequences of the oligonucleotides for promoters P1, P2 and P3 are boxed in green, blue and yellow respectively, with the mapped transcription start sites and the predicted -10 and -35 motifs highlighted in solid colour. Mutations made in the -10 regions of the promoters, or -35 region for promoter P3, are indicated below the sequence in red.

resemble other known promoter sequences. However, promoter P1 shows significant similarity to the RecA/LexA independent promoter, RecA-NDp, defined by Gamulin (2004). This promoter was shown to be inactivated by mutating an A residue in the -10 region to a C (Gopaul *et al.*, 2003) and so the same mutation has been employed in this study. Oligonucleotide sequences and their mutations are shown in Figure 3.3. *M. tuberculosis* promoter-*lacZ* cultures were grown to early exponential phase (OD_{600nm} 0.3), then split into two with one culture being induced with 20 ng/ml mitomycin C and the other acting as an uninduced control. Cultures were incubated for 24 hours at 37 ºC, after which CFEs were made from all cultures as per Section 2.4.4 and β-galactosidase activity was detected using the β-galactosidase assay (Section 2.6.1). Expression greater than the empty vector control level was detected for promoter P1 alone (Figure 3.4a); this activity increased upon addition of mitomycin C (Figure 3.4b) demonstrating that promoter P1 is mitomycin C inducible.

3.4 A promoter-*lacZ* fusion containing all three promoters still exhibited activity when promoter P1 was inactivated

The discrepancy between the number of promoters shown by primer extension and β-galactosidase assay could be due to the fact that the fragments used to create the oligonucleotide promoter-*lacZ* fusions were too small. Small fragment size can mean that regulatory sequences necessary for activity are not present and this could be why no activity was seen for promoters P2 and P3. To investigate this further, a 335bp fragment containing all 3 promoters was amplified by PCR and cloned into the vector pEJ414 to create pAG04. Figure 3.4a shows that all activity from promoter P1 was lost when the -10 region was mutated. Site-directed mutagenesis (SDM) was

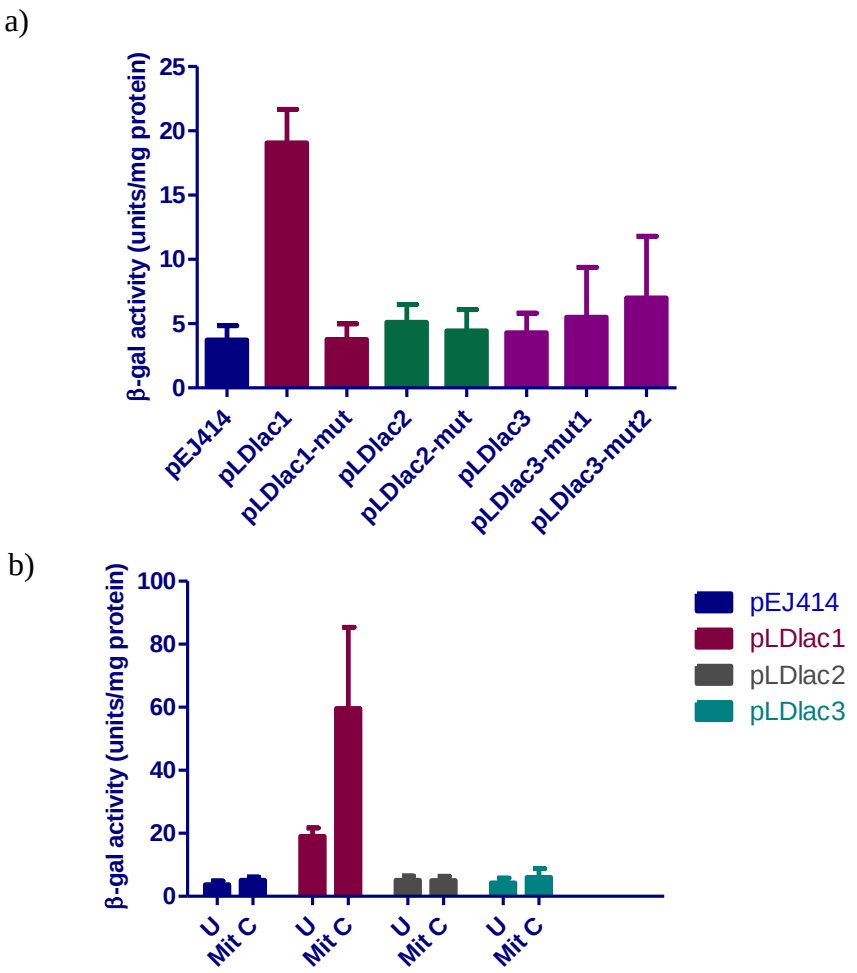


Figure 3.4 – Activity of *sigG* oligonucleotide promoter-*lacZ* fusions and the effect of mitomycin C on expression

β-galactosidase results analysing promoter activity of *sigG* P1, P2 and P3. a) Activity of each promoter construct and its mutant(s) compared to a vector only control (pEJ414). Only construct pLDlac1, containing promoter P1, showed significant activity compared to the vector control $p<0.001$ (Students t-test) and this activity was abolished in pLDlac1-mut by making an A to C mutation in the P1 -10 region. b) Activity of the WT promoter constructs in response to a 24 hr induction with 20 ng/ml mitomycin C at OD_{600nm} 0.3 U) uninduced, Mit C) mitomycin C induced. P1 was significantly induced by mitomycin C $p<0.05$ (Students t-test). Columns represent the means, and the error bars show the standard deviations, of data obtained from three biological replicates.

used to make pAG04-mut1, containing the same A to C mutation in the P1-10 region that was made in the original oligonucleotide P1 promoter-*lacZ* fusion where it eliminated promoter activity. The equivalent change in the related *recA* P1 promoter has also been shown to ablate promoter activity in longer constructs (Gopaul *et al.*, 2003). Should promoter activity be seen in pAG04-mut1, this would support the hypothesis that P1 is not the only promoter expressing *sigG*.

M. tuberculosis H37Rv transformed with pEJ414, pAG04 or pAG04-mut1, was cultured to OD 0.6, CFEs were made as per Section 2.4.4 and were analysed by β-galactosidase assay. The level of activity observed from the pAG04 construct was much higher than that from the oligonucleotide reporter constructs where only promoter P1 showed activity (Figures 3.4a and 3.5). As previously stated, small fragment size such as that in the oligonucleotide reporter constructs can mean that regulatory sequences necessary for activity are not present. The much greater level of activity obtained with the pAG04 construct showed that the larger fragment size had either increased the activity of promoter P1 and/or allowed promoters P2 and P3 to be active. When compared to the empty vector control pAG04-mut1 had promoter activity and the level of activity was one third lower than that of pAG04 due to the inactivation of promoter P1 (Figure 3.5). This shows that there is at least one other promoter besides P1 expressing SigG and that promoter P1 appears to contribute approximately one third of the total activity of the pAG04 construct.

3.5 Expression from pAG04 could only be induced with mitomycin C

3.5.1 Initial exposure to a variety of stress conditions

As shown in Figure 3.4b promoter P1 demonstrated mitomycin C inducibility, suggesting that this promoter is DNA-damage inducible and this is supported by the fact that P1 shows significant homology to the RecA/LexA independent DNA-damage inducible promoter defined by Gamulin (2004). In addition to testing expression from the promoter-*lacZ* construct pAG04 after mitomycin C induction, a variety of additional stresses were examined. *Rv0181c* shows strong homology at the protein level to YhhW pirin family proteins (BLAST). In *E. coli*, YhhW, functions as a quercetinase to break down the antioxidant quercetin and related flavanoid compounds (Adams *et al.*, 2005). Due to the fact that *sigG* and *Rv0181c* are co-transcribed it was hypothesised that the *sigG* promoters may be induced by the presence of an antioxidant flavanoid compound and in this case butein was selected due to its increased solubility compared to the structurally similar quercetin. In addition, microarray analysis of *M. tuberculosis* overexpressing SigG protein identified *Rv0887c* and *Rv0911* as the two most highly upregulated, and therefore, potentially SigG-regulated, genes. These genes contained domains belonging to a family of proteins known as Glo_EDI_BRP_like domain superfamily. This superfamily contains several structurally related metalloproteins, including type I extradiol dioxygenases, glyoxalase I and a group of antibiotic resistance proteins, responsible for aromatic ring cleavage, conversion of methylglyoxal to pyruvate and resistance to antibiotics (e.g. bleomycin) respectively. These proteins and the family they belong to will be discussed further in Chapter 4. However, the fact that they were potentially SigG-regulated, led to the hypothesis that *sigG* expression might be induced by the substrates of *Rv0887c* and *Rv0911*.

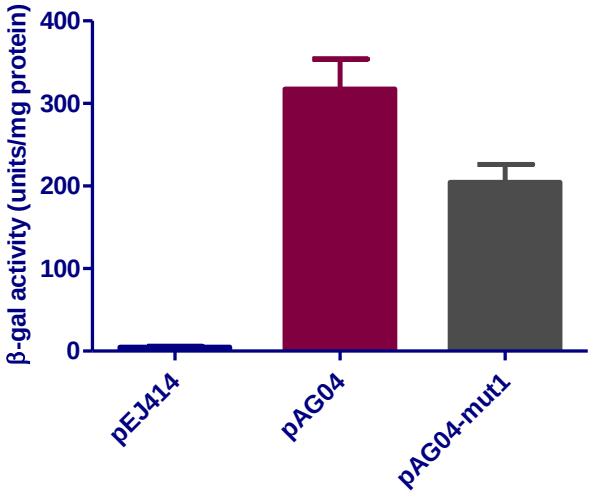


Figure 3.5 - β-galactosidase results showing promoter activity of pAG04 and pAG04-mut1.

Activity of pAG04-mut1 compared to that of pAG04. pAG04 contains a region that includes all three putative promoters and the -mut1 derivative of this vector has a mutation in the P1 promoter. Both pAG04, $p<0.01$ and pAG04-mut1, $p<0.01$ showed a significantly higher level of activity compared to the vector only control (Students t test). Additionally pAG04-mut1 showed a significantly lower level of activity compared to pAG04, $p<0.05$ showing that promoter P1 had been inactivated. Columns represent the means, and the error bars show the standard deviations, of data obtained from three biological replicates.

Two potential substrates were selected to test for induction of the *sigG* promoters: methylglyoxal, a toxic metabolic intermediate that is converted to pyruvate via the glyoxalase system (Inoue *et al.*, 1995), and bleomycin, an antibiotic that causes DNA strand breaks. SigG was also shown to be upregulated during macrophage infection (Cappelli *et al.*, 2006) and to test whether this was due to oxidative stress encountered inside the macrophage, t-butyl hydroperoxide was used to determine whether oxidative stress increased expression from the SigG promoters.

The concentration of mitomycin C and t-butyl hydroperoxide required to cause a log reduction in *M. tuberculosis* viability have been experimentally determined in the laboratory and are 20 ng/ml and 10 µM respectively (J. Houghton, personal communication). Concentrations of methylglyoxal (0.2% w/v), bleomycin (1.5 ng/ml) and butein (25 µg/ml) were selected based on their ability to cause 10 % inhibition of growth of wild-type *M. tuberculosis* H37Rv in an alamar blue cell viability assay (Section 2.7.2). Three biological replicates of *M. tuberculosis* transformed with pAG04 were cultured to early exponential growth phase (OD_{600nm} 0.3) and each culture was split into six. The five stress reagents were added, one per culture, to give the final concentrations stated above, with the remaining sixth culture acting as an uninduced control.

Only mitomycin C significantly induced expression from the SigG promoters ($p < 0.05$, Students t test) and while methylglyoxal appeared to cause a significant decrease in expression ($p < 0.05$, Students t test), it appeared to have had a negative effect on the growth of the induced cultures, which had barely grown in comparison to the uninduced control and it is therefore likely that this ‘reduction’ is actually due

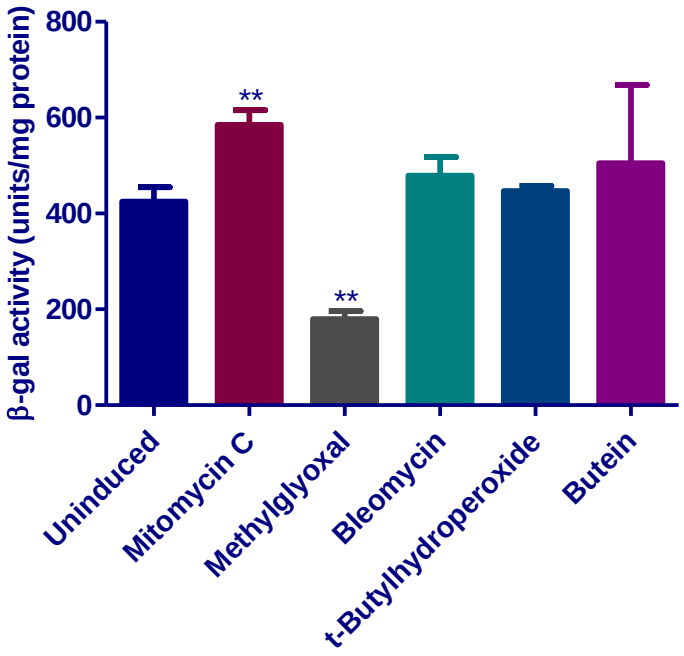


Figure 3.6 – Analysis of β-galactosidase production from pAG04 following exposure to a variety of stress reagents

β-galactosidase activity following exposure of *M. tuberculosis* transformed with pAG04 to 20 ng/ml mitomycin C, 0.2 % (w/v) methylglyoxal, 1.5 µg/ml bleomycin, 10 µM t-butyl hydroperoxide and 25 µg/ml butein for 24 hours at 37 ºC. A significant increase in activity between uninduced and induced samples was observed following exposure to mitomycin C ($p<0.05$, Students t test) and a significant decrease following exposure to methylglyoxal ($p<0.05$, Students t test). No significant difference was observed between the samples exposed to bleomycin, t-butylhydroperoxide and butein when compared to the uninduced control. Columns represent the means, and the error bars show the standard deviations, of data obtained from three biological replicates.

to a wholesale reduction in gene expression due to the toxicity of methylglyoxal (Figure 3.6). The fact that bleomycin was unable to induce promoter activity was surprising as, like mitomycin C, it is a DNA damaging agent and the same concentration as was used here has been shown by qRT-PCR to induce *sigG* expression (Smollett *et al.*, 2011). One explanation could be that bleomycin had more of an effect on growth rate than mitomycin C. Growth rate is known to have an effect on the level of expression of *lacZ* from reporter constructs (Warner *et al.*, 2002).

3.5.2 Investigating the effect on expression of *lacZ* from pAG04 following exposure to fatty acids

Phenotypic analysis of the ΔsigGoperon *M. tuberculosis* strain, discussed in Chapter 5, led to the hypothesis that SigG may play a role during growth on fatty acid carbon sources. It was therefore decided to test whether the presence of a variety of fatty acids could induce expression from the SigG promoters in pAG04. The fatty acids selected and the final concentrations used are shown in Table 3.1 and were chosen to represent both saturated and unsaturated fatty acids as well as different length carbon chains. The concentrations of fatty acid to use were based on previous β-galactosidase experiments conducted in the laboratory (Fivian-Hughes, 2009).

M. tuberculosis transformed with pAG04 was cultured in triplicate to early exponential phase (OD_{600nm} 0.3) and each parent culture was split into seven test cultures. One fatty acid per culture was added to six out of the seven test cultures to give the concentrations shown in Table 3.1. The seventh test culture acted as an uninduced control. None of the fatty acids tested here induced *lacZ* expression from

Table 3.1 - Fatty acids used for β-galactosidase assay of pAG04

Name	Description	Concentration
Acetic Acid	Saturated 2 carbon	500 µM
Valeric Acid	Saturated 5 carbon	500 µM
Hexanoic Acid	Saturated 6 carbon	500 µM
Nonanoic Acid	Saturated 9 carbon	500 µM
Oleic Acid	Unsaturated 18 carbon	500 µM
Linoleic Acid	Unsaturated 18 carbon	500 µM

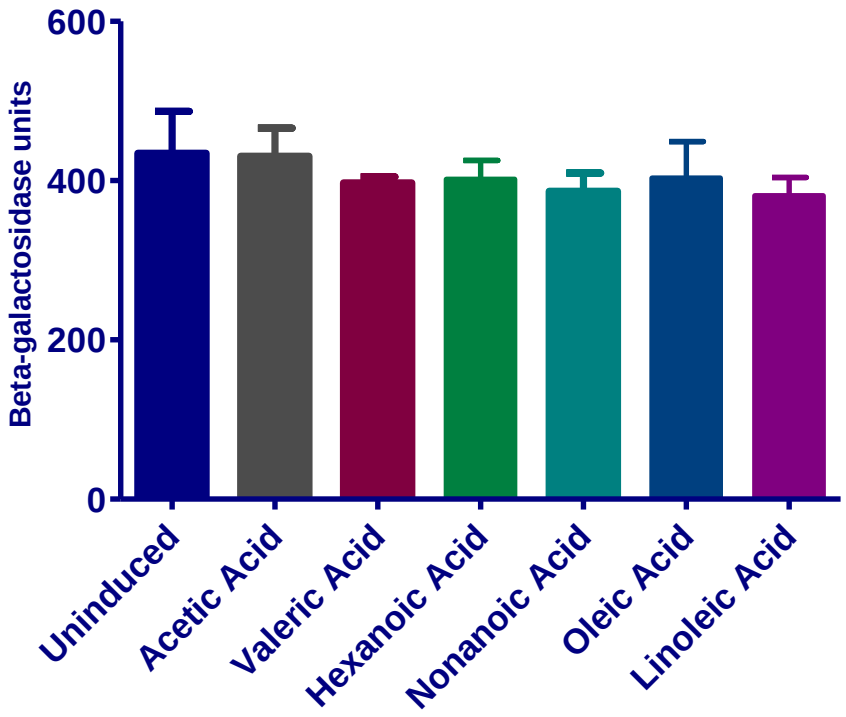


Figure 3.7 – Exposure of *M. tuberculosis* transformed with pAG04 to fatty acids β-galactosidase activity following exposure of *M. tuberculosis* transformed with pAG04 to 500 µM acetic, valeric, hexanoic, nonanoic, oleic or linoleic acid for 24 hours at 37 ºC. No significant difference was observed between any of the fatty acid-treated samples and the untreated control ($p>0.05$, One-way ANOVA with Dunnett multiple comparison test). Columns represent the means, and the error bars show the standard deviations, of data obtained from three biological replicates. Statistical analysis was performed using a One-way ANOVA with Dunnett multiple comparison test to compare each treated sample back to the untreated control.

pAG04 (Figure 3.7) suggesting that the presence of fatty acids in the growth medium does not induce the SigG promoters.

3.6 Expression from pAG04 was not reduced in a *sigG* deletion strain

ECF sigma factors are often autoregulatory (Helmann, 2002) and of the 10 *M. tuberculosis* ECF sigma factors, 4 have been shown to regulate their own transcription (Rodrigue *et al.*, 2007). In addition primer extension studies had suggested that promoter P3 may be SigG-dependent (Dawson, 2005).

In order to determine whether SigG was able to regulate its own expression, the pAG04 promoter-*lacZ* construct was transformed into a *sigG* operon deletion strain (Δ*sigG*operon). Should one of the *sigG* promoters be recognised by SigG it would be expected that LacZ expression in the *sigG* operon deletion strain would be lower than in wild-type *M. tuberculosis.* Wild-type and Δ*sigG*operon *M. tuberculosis* strains transformed with the pAG04 promoter-*lacZ* construct were grown to mid exponential phase (OD_{600nm} 0.6), CFEs were made as per Section 2.4.4 and were analysed by β-galactosidase assay. No significant decrease in β-galactosidase activity, and therefore promoter expression, in the Δ*sigG*operon strain was observed compared to wild-type *M. tuberculosis* (Figure 3.8; $p>0.05$, Students t test). This suggested that none of the *sigG* promoters were SigG-dependent. The -10 region of promoter P3 is similar to the SigM consensus (gtCcgA) predicted by Raman (2006) suggesting that this promoter may actually be SigM-dependent. Another explanation however, is that, with P1 predicted to be the strongest promoter, it was possible that the activity of this promoter was masking any changes that were occurring in the activity of the other two promoters. It was therefore decided to use

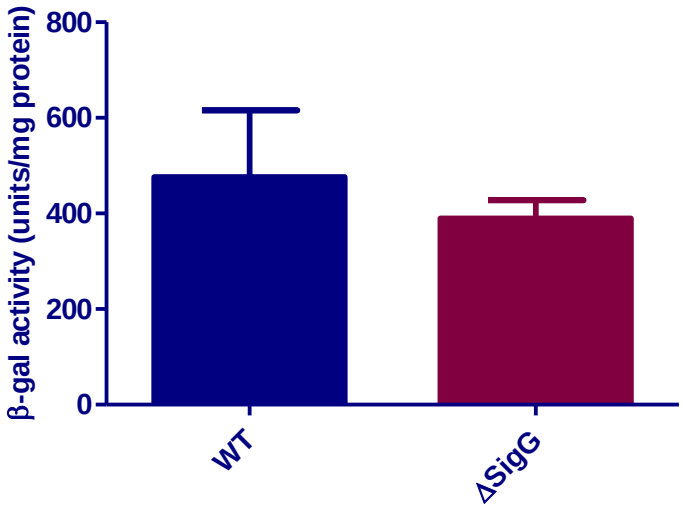


Figure 3.8 – β-galactosidase production from pAG04 in ΔsigGoperon compared to wild-type *M. tuberculosis*

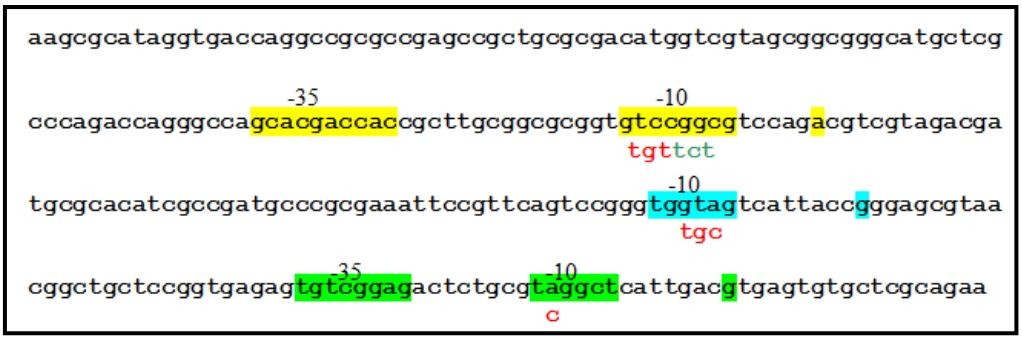
β-galactosidase assay results comparing CFEs from the *sigG* operon deletion strain (ΔsigGoperon) and wild-type *M. tuberculosis* transformed with pAG04. No significant difference was observed in the amount of β-galactosidase activity between the two strains. Columns represent the means, and the error bars show the standard deviations, of data obtained from three biological replicates.

SDM to isolate each promoter within the pAG04 construct by inactivating two promoters at a time.

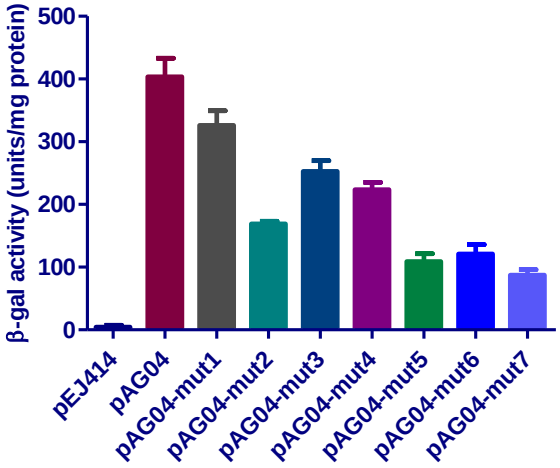
3.7 The original oligonucleotide mutations did not isolate the *sigG* promoters in the pAG04 construct

In order to isolate the three *sigG* promoters individually, SDM was used to mutate the -10 regions of each promoter. The previous experiment using the oligonucleotide promoter-*lacZ* constructs had shown that the A to C mutation in the P1 -10 region was able to abolish promoter activity (Figure 3.4). The same GTA to TGC mutation as used previously was used for the -10 region of promoter P2; however, the -10 region of promoter P3 was less well defined and so in addition to the original GTC to TGT (mutation 1) from the oligonucleotide constructs, a second CGG to TCT mutation (mutation 2) was also tested (Figure 3.9a). The promoter-*lacZ* constructs pAG04-mut2 (P1 wild-type, P2 and P3(1) mutated), -mut3 (P2 wild-type, P1 and P3(1) mutated), -mut4 (P2 wild-type, P1 and P3(2) mutated), -mut5 (P3 wild-type, P1 and P2 mutated), -mut6 (all promoters mutated with P3 mutation 1) and –mut7 (all promoters mutated with P3 mutation 2) were made by making the above mutations via SDM. *M. tuberculosis* was transformed with pEJ414, pAG04 and its derivatives and mid exponential growth phase (OD_{600nm} 0.6) cultures were analysed via β-galactosidase assay as previously described. Both pAG04-mut6 and -mut7, in which all three promoters should have been inactive, showed β-galactosidase production and therefore promoter activity. The amount of activity corresponded to that of pAG04-mut5 where only promoter P3 should be active suggesting that neither of the mutations used to try and inactivate P3 had succeeded. It is also possible that the A to C mutation in the -10 region of promoter P1 did not cause complete

a)



b)



	pAG04	-mut1	-mut2	-mut3	-mut4	-mut5	-mut6	-mut7
P1	✓	X	✓	X	X	X	X	X
P2	✓	✓	X	✓	✓	X	X	X
P3	✓	✓	X_1	X_1	X_2	✓	X_1	X_2

Figure 3.9 – Initial attempt to isolate each promoter within the pAG04 construct

a) Positions of the mutations made by SDM to inactivate the promoters. The mutations from the oligonucleotide constructs were used (shown in red) along with an additional mutation in the P3 -10 region (shown in green). The -35 and -10 regions of promoters P1, P2 and P3 and their transcriptional start sites are represented in green, blue and yellow respectively; b) the β-galactosidase activity of the pAG04 and –mut1-7 constructs. The table shows which promoters are wild-type (✓) or mutated (X) in each construct and the subscript numbers represent the oligonucleotide mutation (1) and new mutation (2) made in promoter P3. The two constructs where all the promoters should have been inactive showed activity indicating that at least one of the mutations had not inactivated the promoter.

inactivation and that the activity seen for the pAG04-mut6 and –mut7 constructs was residual activity from promoter P1 (Figure 3.9b). However, the earlier results using the oligonucleotide constructs showed this mutation abolished activity and there is precedent for this in the literature (Gopaul *et al.*, 2003). Comparison of pAG04-mut1 and pAG04-mut5 where the only difference between the two constructs was a mutation introduced in promoter P2, suggested that this promoter exists as the activity of pAG04-mut5 was approximately one third that of pAG04-mut1. Comparison of the activity of pAG04-mut1 with that of pAG04-mut3 and –mut4, in which the only difference was the mutation in promoter P3, suggested that the P3 promoter has some activity. However, this is less convincing than for promoter P2 as the reduction in activity compared to pAG04-mut1 is approximately one fifth.

3.8 Creation and analysis of constructs where only one SigG promoter should be active

In order to try to completely abolish promoter activity from promoter P3 and also from promoter P2, primers were designed to replace the entire -10 regions of the two promoters with C residues by SDM (the yellow and blue highlighted -10 regions in Figure 3.9a). Plasmids pAG04-mut8 (P2 and P3 mutated), -mut9 (P1 and P3 mutated), -mut10 (P1 and P2 mutated) and -mut11 (all promoters mutated) were created in this way.

At the same time as these constructs were analysed for promoter activity they were also tested for promoter induction by a number of stress reagents. The DNA-damaging agents mitomycin C and ofloxacin were used to determine whether it was just promoter P1 that responded to DNA damage, as had been hypothesised from

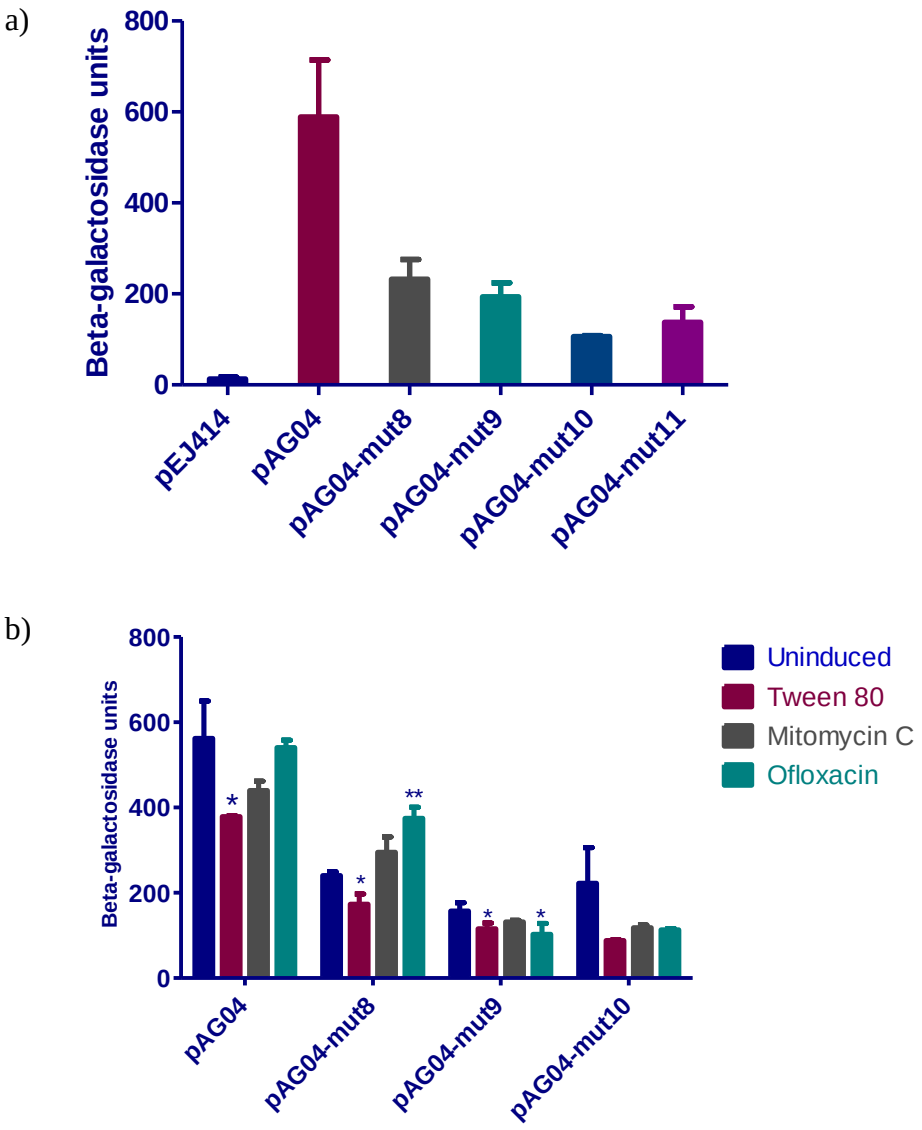


Figure 3.10 – β-galactosidase activity of pAG04 mutant constructs

a) analysis of pAG04 mutant constructs where the -10 regions of promoters P2 and P3 had been mutated into rows of C residues. pAG04-mut8 (P1 only), -mut9 (P2 only), –mut10 (P3 only) and –mut11 (no promoters). pAG04-mut11 demonstrated activity showing that at least one mutation was ineffective; b) β-galactosidase activity produced by pAG04 and –muts8-10 after 24 hour exposure to 5 % Tween 80, 20 ng/ml mitomycin C or 7.5 µg/ml ofloxacin. Only pAG04-mut8 showed significant induction with ofloxacin ($p<0.05$, unpaired t test). Columns represent mean and standard deviation of three biological replicates.

previous primer extension studies (Dawson, 2005). In addition, due to results from Biolog and alamar blue experiments discussed later in Chapter 5, the ΔsigGoperon strain appeared to be more susceptible than wild-type to the presence of Tween 80 in the growth medium. As such, it was hypothesised that Tween 80 would be able to induce expression from one or all of the SigG promoters.

The new pAG04 mutated constructs were transformed into wild-type *M. tuberculosis*. As outlined earlier, cultures were grown to early exponential phase (OD_{600nm} 0.3) and split into four test cultures. Three out of the four test cultures were induced with either 20 ng/ml mitomycin C, 7.5 µg/ml ofloxacin or 5 % w/v Tween 80, with the remaining test culture acting as the uninduced control. Cultures were incubated for a further 24 hours at 37 ºC after which, CFEs were made as per Section 2.4.4 and analysed by β-galactosidase assay.

Despite the apparent complete removal of the -10 regions of promoters P2 and P3, promoter activity was still detected from the pAG04-mut11 construct, in which all of the promoters should have been inactive (Figure 3.10a). This suggests that either one or more of the promoters is still active despite being mutated or possibly that a new promoter has been formed during the creation of the pAG04 plasmid and this will be investigated further in Section 3.9. None of the stress reagents tested were able to induce β-galactosidase production from the pAG04 construct and in the case of Tween 80 and mitomycin C a reduction in β-galactosidase activity was observed compared to the uninduced sample. In contrast the pAG04-mut8 construct, in which only P1 was supposed to be active, demonstrated an increase in activity with the two DNA-damaging agents mitomycin C and ofloxacin, however, only the ofloxacin

result was statistically significant ($p < 0.05$, unpaired t test). Neither pAG04-mut9 or pAG04-mut10 were induced by the DNA-damaging agents and although the fact that there could be an additional promoter present means that these results are not conclusive, this supports the initial hypothesis that only promoter P1 was DNA-damage inducible. Contrary to earlier findings in this study pAG04 did not demonstrate DNA-damage inducibility in this experiment (Figure 3.10b). A reduction in β-galactosidase activity was observed for all of the constructs when induced with Tween 80 compared to the uninduced control and this reduction was significant for pAG04, -mut8 and –mut9 ($p < 0.05$, unpaired t test). While it is tempting to say that Tween 80 represses the activity of the SigG promoters, it is more likely that this is an effect of growth rate on expression level, a known complication when using promoter-*lacZ* reporter constructs (Warner *et al.*, 2002) and this will be discussed in more detail at the end of the chapter.

The results obtained here, coupled with the earlier observation that bleomycin was unable to induce activity in pAG04, suggested that there may be a problem with the pAG04 construct and/or that the mutations made in the promoters had not completely abolished promoter activity. It was decided to investigate which promoters were still active in each of the constructs using 5'RACE.

3.9 5'-RACE was used to try and determine positions of promoters within the pAG04 construct

In order to determine whether the activity observed from the pAG04-mut11 promoter construct was due to a failure to inactivate one or more of the promoters or whether a new promoter had been inadvertently created, RNA samples from *M.*

tuberculosis containing pAG04 and pAG04–mut8 - 11 were analysed by 5'-RACE. Due to the fact that *sigG* has been shown to be most highly expressed during early exponential phase (K. Smollett, personal communication), RNA samples were prepared from cultures at OD_{600nm} 0.3. Samples were DNase treated as per Section 2.5.1 and 5'-RACE was conducted on the clean RNA samples using the GeneRacer kit (Section 2.5.4). Briefly, a commercial oligonucleotide was ligated to the 5' ends of the RNA. Oligo-capped RNA was reverse transcribed into cDNA using random primers. PCR using a forward primer specific to the commercial oligo (GeneRacer 5' Oligo) and a reverse primer specific to the *lac*Z tag (GSPLacR) was used to amplify the *sigG* gene-specific transcript(s) which were subsequently analysed by agarose gel electrophoresis.

Assuming that transcripts originated from the transcriptional start sites identified by Dawson (2005) via primer extension, the 5'-RACE products expected for promoters P1, P2 and P3 were approximately 820, 880 and 950 bp respectively. A product was observed that corresponded to the expected size of the promoter P1 transcript in all samples except pAG04-mut9 despite the fact that this promoter should have been inactive in all but the pAG04 and pAG04-mut8 samples. pAG04-mut9 gave a product just under 800 bp in length, which did not correspond to the expected sizes of any of the *sigG* promoters. No products were observed corresponding to the expected sizes of the promoter P2 and P3 transcripts. However, one additional product that was approximately 500 bp long was observed in all of the samples (Figure 3.11). This could have been due to additional promoters or due to processing of longer transcripts as this 5'-RACE method does not distinguish between full length and processed transcripts. PCR products were excised and sequenced using

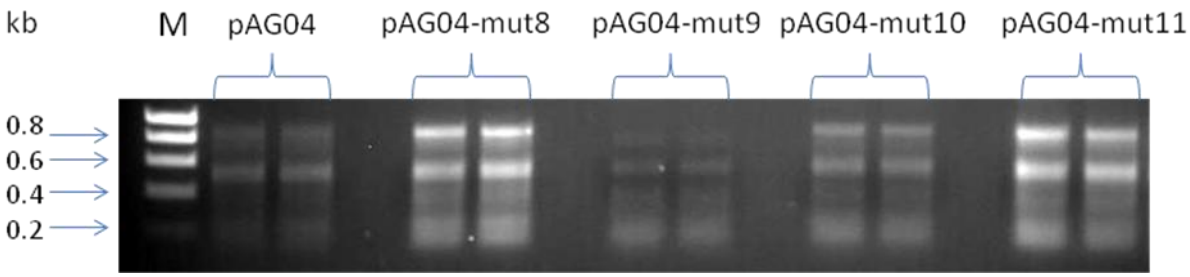


Figure 3.11 – 5'-RACE of *sigG* promoter-*lacZ* constructs

5'RACE was used to determine the transcriptional start sites present within the pAG04 construct and its derivatives pAG04–mut8 (P1 only) –mut9 (P2 only), -mut10 (P3 only) and -mut11 (all promoters mutated), M represents the Hyperladder I size standard. The expected sizes of products produced by P1, P2 and P3 were approximately 820, 880 and 950 bp respectively. All constructs gave a product of approximately 500 bp in size, which does not correspond to any of the promoters. All constructs except pAG04-mut9 gave a product just over 800 bp in size, which corresponded to the P1 product. The pAG04-mut9 construct gave a product slightly smaller than 800 bp, which does not correspond to the size expected from any of the promoters.

the *lacR* primer to attempt to identify the transcriptional start sites, however despite repeated attempts, sequencing failed to yield transcriptional start sites for any of the promoter-*lacZ* fusion constructs. In future work these PCR products could be cloned into a TOPO vector (Table 2.3), which would then be expressed and purified from *E. coli*. This would provide the advantage of increasing the amount of product available to be sequenced and would also enable the use of standard sequencing primers such as M13.

Despite not having obtained transcriptional start site data for the pAG04 construct and its derivatives, it was apparent that none of these constructs were expressing the correct number of promoters and this explains why in earlier experiments these constructs did not behave as expected.

3.10 Promoter activity was assessed using quantitative RT-PCR

3.10.4 Promoters P2 and/or P3 were confirmed to have activity by qRT-PCR

Following on from the inconclusive results of the promoter-*lacZ* fusion constructs it was decided to use quantitative RT-PCR to demonstrate, initially, that a transcript could be detected that started upstream of promoter P1. Two sets of primers were designed such that one amplicon, produced by qRTSigGF/R, would be downstream of promoter P1 and would therefore represent transcripts from all three promoters whereas the second amplicon, produced by qRTSigGP2F/R, would be between the transcriptional start site of promoter P2 and promoter P1 (Figure 3.12a). This second amplicon would detect transcripts initiated at promoters P2 and P3. qRT-PCR was conducted on RNA samples from wild-type *M. tuberculosis* H37Rv and gene

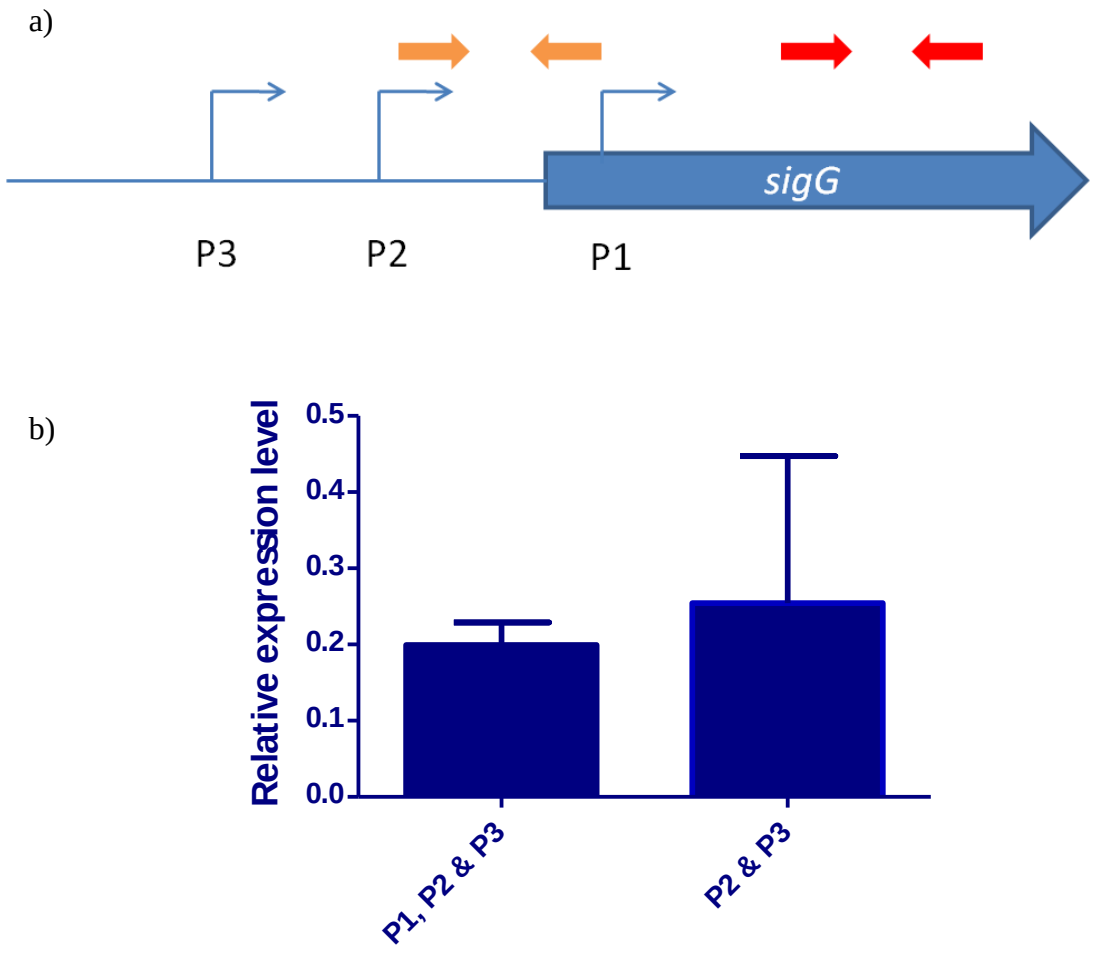


Figure 3.12 – qRT-PCR to show that P1 is not the only promoter expressing *sigG*

a) qRT-PCR primer sets qRTSigGF/R (red) and qRTSigGP2F/R (orange) were designed to detect the presence of transcripts either from all three of the *sigG* promoters or just from promoters P2 and P3; b) qRT-PCR results using primer sets qRTSigGF/R and qRTSigGP2F/R to detect expression from all three promoters or just from promoters P2 and P3 respectively. Transcripts were detected from downstream of P1 and with much higher variability downstream of promoters P2 and P3 showing that transcription does occur from one or both of promoters P2 and P3. Expression was normalised to expression of *sigA.* Columns represent mean and standard deviation of three biological replicates.

expression was normalised to *sigA*. Although the number of transcripts detected upstream of promoter P1 showed much more variation than that detected downstream of promoter P1, transcripts in this region were detected (Figure 3.12b). This confirmed that there is activity from either one of or both promoters P2 and P3.

3.10.5 Only promoter P1 is DNA-damage inducible

Having determined that there was promoter activity other than that from promoter P1 it was next necessary to determine whether the DNA-damage inducibility observed with the promoter-*lacZ* fusions was solely due to promoter P1 or whether promoters P2 and P3 played a role. qRT-PCR was conducted on RNA samples from wild-type *M. tuberculosis* H37Rv that had been induced with either 20 ng/ml mitomycin C or 7.5 µg/ml ofloxacin overnight at 37ºC (samples provided by K. Smollett). A significant increase in gene expression was detected using the primer set downstream of promoter P1 for both mitomycin C and ofloxacin ($p < 0.005$, unpaired t test) but not from the primer set located between promoters P2 and P1. In addition the level of expression detected from promoters P2 and P3 was reduced following DNA damage compared to the uninduced control, suggesting that under DNA-damaging conditions promoter P1 becomes the primary promoter for expression of *sigG* (Figure 3.13).

3.10.6 The promoters were not induced by the presence of fatty acids

Using the promoter-*lacZ* constructs, fatty acids did not cause induction of transcription from the *sigG* promoters but this could have been due to the problems already reported with the promoter-*lacZ* constructs or could also have been due to the growth medium used. When initially defining growth media for Mycobacteria

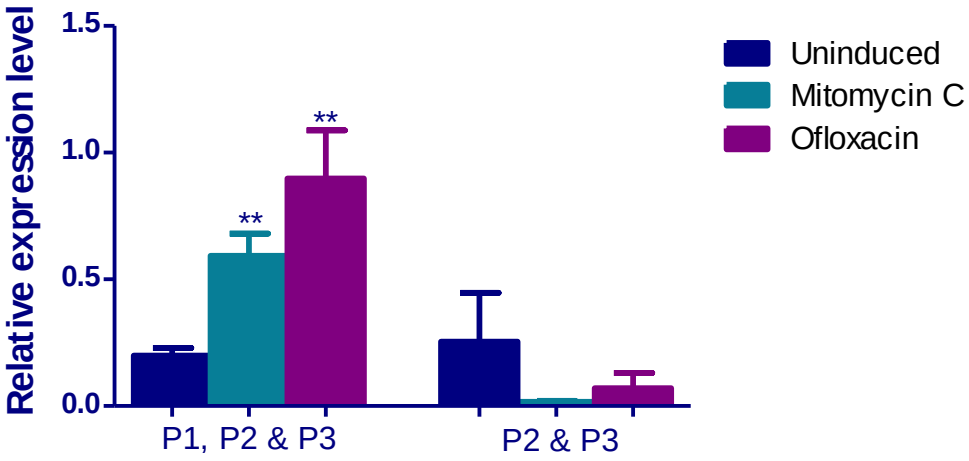


Figure 3.13 – qRT-PCR showing that only promoter P1 is DNA-damage inducible

qRT-PCR results using primer sets qRTSigGF/R and qRTSigGP2F/R to detect expression from all three promoters or just from promoters P2 and P3 respectively following 24 hour exposure to 20 ng/ml mitomycin C or 7.5 µg/ml ofloxacin. A significant increase in expression in response to DNA damage was only observed when promoter P1 was present ($p<0.05$, One-way ANOVA with Dunnett multiple comparison test for both mitomycin C and ofloxacin). This confirmed that only promoter P1 responded to DNA damage. Columns represent mean and standard deviation of three biological replicates.

Dubos and Middlebrook (1947) noted that the bacteria were susceptible to toxic free fatty acid (oleic acid) that contaminated the Tween 80 used in the media and that the presence of albumin was able to protect against this. The exposure of the promoter-*lacZ* constructs to fatty acids was conducted in modified Dubos medium, which contained albumin, and this could have negated the effect of the fatty acids. In order to confirm whether fatty acids could induce expression from the *sigG* promoters, qRT-PCR was conducted on RNA samples made from wild-type *M. tuberculosis* H37Rv cultured in a variation on the modified Dubos medium, in which there was no albumin and the Tween 80 was replaced with the non-hydrolysable surfactant tyloxapol to prevent toxicity (Section 2.1). The fatty acids tested were three that had been shown to have a toxic effect against the ΔsigGoperon strain in an alamar blue assay, arachidonic acid, a fatty acid produced by macrophages (Akaki *et al.*, 2000) and Tween 80, which is an esterified form of oleic acid (results discussed in Chapter 5). Bacteria were cultured to early exponential phase (OD_{600nm} 0.3) and split into six test cultures. Each test culture was induced with either 1 % Tween 80, 500 µM linoleic acid, 500 µM arachidonic acid, 500 µM oleic acid or 5 µM palmitic acid with the sixth test culture acting as an uninduced control. The cultures were incubated for a further 24 hours at 37 ºC prior to harvesting the RNA. qRT-PCR was conducted using primer set qRTSigGF/R and expression was normalised to *sigA*. No significant difference was observed between the fatty acid or Tween 80-induced samples and the uninduced control demonstrating that *sigG* expression is not induced by the presence of fatty acids.

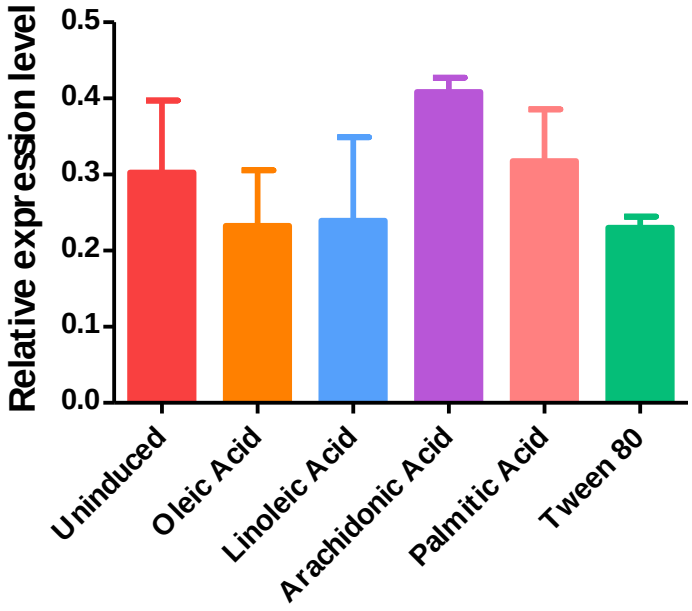


Figure 3.14 – qRT-PCR of *sigG* expression levels using *M. tuberculosis* RNA samples following exposure to fatty acids

qRT-PCR conducted using primer set qRTSigGF/R to compare the expression of *sigG* following 24 hour exposure to 500 µM oleic, linoleic or arachidonic acid, 5 µM palmitic acid or 1 % Tween 80 compared to an untreated control. Cultures were grown in a variant form of modified Dubos albumin containing tyloxapol in place of the standard Tween 80. Columns represent mean and standard deviation of three biological replicates. No significant difference was observed between any of the treated samples and the untreated control ($p>0.05$, One-way ANOVA with Dunnett multiple comparison test).

3.11 Discussion

Previous work characterising the promoters expressing *sigG* had identified that there were two, potentially three, promoters and that these were responsible for the expression of *sigG, Rv0181c* and *Rv0180c* (Dawson, 2005). In addition one of the promoters, P1, showed significant homology to the predicted RecA-NDp promoter motif (Gamulin *et al.*, 2004) and *sigG* expression has been shown to be induced by DNA damage independently of *recA* (Rand *et al.*, 2003). However, it has recently been shown that SigG does not regulate either the RecA-independent or SOS responses (Smollett *et al.*, 2011).

Initially the structure of the *sigG* operon was confirmed by RT-PCR. It was demonstrated that although *sigG, Rv0181c* and *Rv0180c* are co-transcribed, there appears to be some read-through into the downstream *lprO* gene. Transcripts were detected that ran from the start of *sigG* to the very start of the *lprO* gene but not to the end, whereas RT-PCR from the end of the *Rv0180c* gene to the end of the *lprO* gene yielded a positive result. This led to the explanation that although the *sigG* operon consists of just the *sigG, Rv0181c* and *Rv0180c* genes, there is read-through of transcription into the start of *lprO.* In addition, it would appear that there is a promoter internal to *Rv0180c* that transcribes the *lprO* gene and in future work this could be confirmed by 5'RACE using a gene-specific primer set in the *lprO* gene to determine its transcriptional start site coupled with the use of transcriptional fusions.

When investigating the regulation of a gene or operon in bacteria, the method of choice is to use a fusion to the *E. coli lacZ* gene, which encodes β-galactosidase. Transcriptional fusions combine a transcriptional element from the gene being

investigated to a *lacZ* gene lacking its promoter but containing the translational start codon and ribosome binding site (Pessi *et al.*, 2001). The *sigG* promoters were initially investigated by using oligonucleotide promoter-*lacZ* fusions, in which only promoter P1 showed activity. This activity could be abolished by an A to C mutation in the -10 region, and was shown to be induced by the DNA-damaging agent mitomycin C. The fact that promoters P2 and P3 did not demonstrate activity in the oligonucleotide promoter-*lacZ* fusions did not necessarily mean that these promoters were not active. The oligonucleotides used to create these constructs were only 50-60 bp in length and it is therefore possible that one or more regulatory elements could have been missing. A construct containing all three of the potential *sigG* promoters was created and this was used to demonstrate that, when promoter P1 was inactivated, there was a decrease in promoter activity but it was not abolished, confirming the presence of at least one additional promoter.

A variety of stress reagents were tested for their ability to induce expression from the *sigG* promoters in the pAG04 construct. However, only mitomycin C showed a significant increase in β-galactosidase production. Previously qRT-PCR has been used to demonstrate a 4-fold induction of *sigG* expression by mitomycin C (Smollett *et al.*, 2011), which is much higher than that observed with the promoter-*lacZ* fusion. Although exposure to the toxic metabolite methylglyoxal appeared to show a significant decrease in expression, this was coupled with a severe effect on the growth rate of the cells. The growth rate of a culture has been shown to have a significant effect on the level of expression of *lacZ* from reporter constructs (Warner *et al.*, 2002). Often when using stress reagents they will have the effect of decreasing the growth rate of the induced culture compared to the uninduced culture and this

could account for the lower than expected level of induction with mitomycin C and also the apparent repression with methylglyoxal where a severe retardation of growth had occurred.

Promoter P3 had been suggested to be a SigG-dependent promoter (Dawson, 2005), which would make SigG an autoregulatory ECF sigma factor. To try to prove this fact the pAG04 construct was assessed for expression of β-galactosidase in the *sigG* operon deletion strain (ΔsigGoperon) compared to wild-type *M. tuberculosis* H37Rv. The fact that there was no difference in the level of expression between the two strains demonstrated that SigG was not autoregulatory.

One aim of this transcriptional analysis was to confirm whether promoters P2 and P3 were both active promoters. The initial work inactivating promoter P1 in the pAG04 construct demonstrated that there was still activity from at least one additional promoter but in order to confirm whether this was from one promoter or two, it was attempted to make mutant versions of the pAG04 construct in which only one promoter was active. Two attempts were made at creating these individual promoter constructs; however, in both cases it proved impossible to completely abolish all promoter activity from the pAG04 construct. This suggested that either the mutations being made in the promoter -10 regions were not completely abolishing activity, or another promoter had been created during construction of pAG04. It was possible to demonstrate to a certain extent that it was promoter P1 that was DNA-damage inducible as the construct that should contain only P1 demonstrated significant induction with the DNA-damaging agent ofloxacin; however, with the promoter mutations and the number of promoters in the construct in doubt this result was not

conclusive. As previously stated use of stress reagents can have a negative effect on growth rate and therefore, β-galactosidase expression and this could explain why the degree of DNA-damage induction of promoter P1 was not as high as might be expected. These changes in growth rate can complicate interpretation of results due to the fact that it is impossible to say whether the difference observed between uninduced and induced samples is due to the stress reagent added or simply due to the fact that this reagent caused a change in the growth rate of the bacteria in the induced sample. An ideal system for testing stress reagents would therefore be one in which the growth rate of the cells could be kept at a known and constant rate e.g. in a chemostat.

sigG expression has been shown to be highest in early exponential phase (OD_{600nm} 0.3; K. Smollett, personal communication) and as such CFEs for β-galactosidase assays were made from cultures at OD_{600nm} 0.3 when not being exposed to stress reagents. However, due to the fact that some of the stress reagents had a negative effect on growth rate, cultures were exposed to the reagent at OD_{600nm} 0.3 and incubated for a futher 24 hours meaning that the final OD_{600nm} of un-exposed control cultures was 0.6. This was to ensure that a sufficient quantity of protein was able to be harvested. The expression of *sigG* at OD_{600nm} 0.6 is approximately half that at OD_{600nm} 0.3 (Figure 1.9). This accounts for the variation observed in the β-galactosidase units produced from the wild-type pAG04 construct in experiments where no exposure was used and those where stress reagents were used. It is however, important to note that in all experiments the samples within an experiment were grown to the same OD_{600nm} before either CFE production or exposure to a stress reagent.

5'RACE was used to attempt to determine the transcriptional start site(s) present within the pAG04 and –mut8-11 constructs. Each construct was shown to be producing two transcripts. pAG04, pAG04-mut8, pAG04-mut10 and pAG04-mut11 all produced a transcript that appeared to correspond to the size expected from promoter P1. pAG04-mut9 produced a transcript just under 800 bp in length, which did not correspond to any of the *sigG* promoters. All constructs produced a second 500 bp product that could have been due to the presence of an additional promoter or due to processing of longer transcripts as this 5'-RACE method does not distinguish between full length and processed transcripts. Unfortunately direct sequencing of these PCR products failed to yield transcriptional start sites and in future work, the PCR products could be cloned into a suitable TOPO vector to allow amplification of the template (via expression and purification from *E. coli*) and sequencing using standard commercially available primers such as M13. The promoter-*lacZ* reporter assays provided strong evidence that there were at least two promoters active in the pAG04 construct as mutating promoter P2 was also able to cause a decrease in activity from this construct. It was therefore expected that at least two products would be obtained from the pAG04 construct. The conclusion that could be drawn from the 5'RACE results was that, even allowing for the fact that the mutations made may not have completely abolished promoter activity, the promoter-*lacZ* reporter constructs, including the parental pAG04 construct, were clearly not expressing the number of promoters expected.

Promoter-*lacZ* fusions have several pitfalls, not least that the rate of transcription is not the same as the level of expression (Warner *et al.*, 2002) and because stabilising

and destabilising mRNA sites are not taken into account (Pessi *et al.*, 2001). As such further investigation of the *sigG* promoters was conducted using qRT-PCR.

Using two sets of primers, one downstream of promoter P1 designed to measure expression from all three promoters and one from just promoters P2 and P3, it was shown that, transcripts could be detected originating from promoters P2 and P3. However, following induction with the DNA-damaging agents mitomycin C and ofloxacin, an increase in expression was detected with the primers downstream of promoter P1 but not the primers downstream of promoters P2 and P3, demonstrating that promoter P1 is the only promoter that is DNA-damage inducible. Given that this promoter is associated with the RecA-NDp promoter it would seem that the *sigG* operon forms part of the RecA independent regulon and this is supported by recent work conducted by Smollett *et al* (2011).

Investigation of the *sigG* operon deletion strain (ΔsigGoperon) had suggested that SigG may play a role in growth of *M. tuberculosis* on fatty acid carbon sources. The promoter-*lacZ* constructs failed to show induction with fatty acids; however, these experiments were conducted in modified Dubos medium containing albumin. When defining the Mycobacterial growth media Dubos noted that Tween 80 exhibited a toxic effect on the bacteria and that this was due to contamination with the free fatty acid, oleic acid. It was also noted that inclusion of albumin in the medium protected against this toxicity (Dubos *et al.*, 1947). It was therefore possible that the presence of albumin in the medium during exposure of *M. tuberculosis* transformed with the promoter-*lacZ* constructs to fatty acids had prevented an effect from being seen. A new growth medium was designed for testing the effect of fatty acid exposure on

expression from the promoter-*lacZ* constructs, based on modified Dubos medium but with Tween 80 replaced by the non-hydrolysable tyloxapol reagent and with no albumin present. qRT-PCR conducted on RNA samples from wild-type *M. tuberculosis* H37Rv grown in this new media and exposed to various fatty acids and Tween 80 showed no induction of *sigG* expression. It is therefore, possible to conclude that under the conditions tested here *sigG* expression is not regulated by the presence of fatty acids in the external environment.

Importantly, when assessing the effect of exposure to various stress reagents on expression in both the β-galactosidase and qRT-PCR experiments, only one concentration was tested. In future work it would be better to test a range of concentrations as the ones tested here may just have been too low to have an effect.

4. Analysing the SigG protein: interactions & function

4.1 Introduction

Regulation of sigma factors occurs not only at the transcriptional level but also at the post-translational level by proteins known as anti- and anti-anti-sigma factors. Many sigma factors are co-transcribed with their regulatory proteins (Hughes *et al.*, 1998) and in *M. tuberculosis* this includes SigE, SigF and SigH (Rodrigue *et al.*, 2006, Song *et al.*, 2003). SigG is co-transcribed with the two downstream genes *Rv0181c* and *Rv0180c* and this chapter will investigate their potential roles as regulators of SigG.

Very little is known about SigG, especially with regard to the SigG regulon. Previous work in this laboratory using an *E. coli* two-plasmid system has suggested that the *Rv0654* gene may be expressed from a SigG-dependent promoter (R. Balhana and J. Smith, personal communication). Microarray analysis of a *M. tuberculosis* SigG overexpression strain has revealed several genes that are up-regulated, with the *Rv0887c* and *Rv0911* genes being the most highly up-regulated genes in response to SigG overexpression (K. Smollett, personal communication). In addition to investigating the post-translational regulation of the SigG protein this study has also looked at the function of the SigG protein with a view to identifying potential SigG-dependent genes.

4.2 Investigating whether Rv0181c and Rv0180c function as anti-sigma and anti-anti-sigma factors to SigG

4.2.1 Identification of correct translation start sites for SigG, Rv0181c and Rv0180c

In order to express correct proteins for purification, translational start sites for SigG, Rv0181c and Rv0180c were assessed using a technique described by Smollett *et al.* (2009). The technique introduces a *myc* tag at the carboxy-terminus of the protein, through cloning each gene, in-frame, into the vector pEJ-Myc. Single nucleotide deletions were introduced between potential start codons via SDM. A deletion made upstream of the correct translational start site would have no effect on expression of the myc tag whereas a deletion made downstream of the correct start site would cause the protein and tag to be out of frame. Expression of the tagged proteins was from their native promoters. Western analysis was used to detect the presence or absence of Myc-tagged protein in CFE and therefore, whether the protein was in or out of frame. The correct translational start site could therefore be deduced.

Due to the fact that promoter P1 is within the annotated coding region of SigG, it is likely that the actual translational start site for SigG is further downstream than the annotated start site. The technique described above was therefore employed to attempt to define the correct translational start sites of all three genes in the *sigG* operon. For SigG, the SigG gene and its upstream region, containing all three promoters, was amplified by PCR and cloned into pEJMyc producing pAG01. In order to have the genes expressed from their native promoters, it was necessary to clone *Rv0181c* together with *sigG* and its upstream region and *Rv0180c* with *sigG*,

SigG

tagacgATGcgcacatcgccgATGcccgcgaaattccgttcagtccggGTGgtagtcattaccgggagcgtaac

ggctgctccgGTGagaGTGtcggagactctgcgtaggctcattgaCGTGagTGTGcTcgcagaaaactctg
1a 1b 1c

Rv0181c

cgaaattcggcctgcccgattcgctgtgatggcgaacgTgacgtcATGaccgccaccgtcgagaTaaggcgtgcg
2a 2b

gccgaccgggcggtcaccacgacgtcctggctgaaatcccggcattcgttctcattcggcgaccactacgatcccgaca

acacccaccacgggctcctgTTGGTGaacaacgacgaccaaATGgaaccggcttcgggttttgatccccatcca

caccgtgacATGgaaatcGTGacc

Rv0180c

tgagccgtcggaagtactgatctgggagatgcacgcgaagctgggcgattcggcaacatagagtggccacgacagaca

ggagcccgcATGtcTcaagcgcagccgcgccccgccgcacccaaccccaagcggaacgtcaaagcgattcggac
3a

gGTGcgtttctggATGgcaccgatcgccaccacgTTGgccctgATGtcggcgctggccgcgctctatctggg

cggcatcctaaaccccATGaccaatTTGcggca

Figure 4.1 - Locations of alternative start sites and frameshift mutations for translational start site mapping.

Diagram to show the locations of the annotated start sites (yellow highlighted), alternative start sites (red) and the positions of frameshift mutations made by SDM (blue highlighted) in *sigG*, *Rv0181c* and *Rv0180c*. The location of the -35 and -10 regions of the P1 promoter are indicated by the green lines.

Rv0181c and the upstream promoter region to produce pAG02 and pAG03 respectively.

Six alternative start sites were identified downstream of the SigG annotated start site; however, only two of these were downstream of promoter P1 and it was these two potential start sites that were subsequently investigated (Figure 4.1). Single nucleotide deletions were made downstream of the annotated start site and each candidate start codon subsequent to promoter P1 by SDM of pAG01, to give plasmids pAG01-mut1, -mut2 and mut3 where nucleotides 1a, 1b and 1c respectively, were deleted (Figure 4.1). Although Rv0181c and Rv0180c had five and six alternative start codons respectively, these were a significant distance downstream of the annotated start codons. In each case single nucleotide deletions were made solely to determine whether the annotated start site was correct. For Rv0181c deletions were made both upstream and downstream of the annotated start site in pAG02 to give plasmid pAG02-mut1 and –mut2, where nucleotides 2a and 2b respectively were deleted (Figure 4.1). Mutation 2a was made to rule out the possibility of a start codon upstream in the coding region of SigG. In the case of Rv0180c a single deletion was made downstream of the annotated start site in pAG03, to give plasmid pAG03-mut1, where nucleotide 3a was deleted (Figure 4.1).

Originally wild-type and mutated plasmids were electroporated into *M. smegmatis* mc^2155 strain. Expression of myc-tagged proteins was detected via Western analysis using an anti-myc antibody. Myc-tagged SigG was expected to be detected between 37.8 and 42.1 kDa, corresponding to the smallest and largest potential open reading frames respectively. Rv0181c and Rv0180c were expected to be detected at 27.5 and

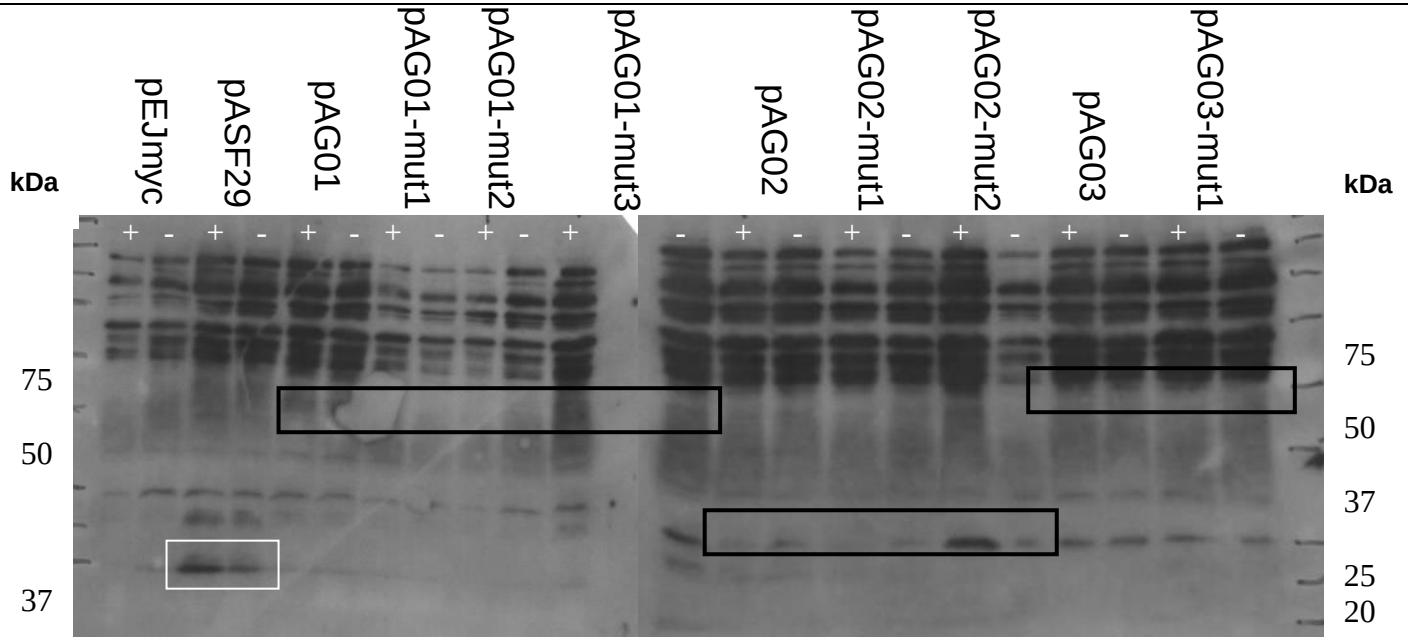


Figure 4.2 -Western analysis of CFEs from H37Rv expressing the myc-tagged constructs.

Western blot analysis (using mouse anti-Myc and rabbit anti-mouse HRP-conjugate antibodies) of Myc-tag constructs in CFEs from H37Rv. No bands can be seen above background for pAG01, pAG02, pAG03 or their derivatives (black boxes indicate positions where protein bands were expected to be visible. In comparison a 15 kDa protein can be seen for the pASF29 mitomycin C positive control (white box). Symbols used represent: +) mitomycin C induced, -) uninduced, WT) wild type, 1a,b,c) frameshift mutations in SigG, 2a,b) frameshift mutations in Rv0181c, 3a) frameshift mutation in Rv0180c.

48.8 kDa respectively. A very high level of background was observed in all extracts from *M. smegmatis,* so much so, that it was impossible to distinguish possible myc-tagged proteins (data not shown). It was suggested that the background level may be lower in *M. tuberculosis* rather than *M. smegmatis* CFEs (A. Fivian-Hughes, personal communication); therefore all plasmids were electroporated into *M. tuberculosis* H37Rv. In addition, due to the fact that transcriptional studies had demonstrated promoter P1 to be DNA-damage inducible, cultures were induced with mitomycin C in an attempt to boost expression of the tagged proteins. A mitomycin C inducible control was included to demonstrate that induction had worked. This control, pASF29, expressed Rv1955-Myc and had been shown to be mitomycin C inducible (Smollett *et al.,* 2009). Cultures were grown to OD_{600} 0.3, split into two and one half culture was induced with 20 ng/ml mitomycin C while the other half was used as an uninduced control. Both cultures were incubated for a further 24 hours prior to preparation of CFE as per Section 2.4.4. Unfortunately, even in the mitomycin C-induced samples it was impossible to detect myc-tagged proteins above background level from pAG01, pAG02, pAG03 or their derivatives (Figure 4.2).

Potential mechanisms for boosting the expression of the myc-tagged proteins above background levels will be discussed at the end of this chapter. For the purpose of expressing the proteins of the operon it was decided to use the first of the two candidate start codons downstream of promoter P1 for SigG and to take the annotated start sites as correct for Rv0181c and Rv0180c.

4.2.2 Yeast-2-hybrid analysis of protein interactions

The yeast-2-hybrid system uses two vectors, pGAD-C1 and pGBD-C1, expressing the activation domain and DNA binding domain, respectively, of the GAL4 transcriptional regulator. GAL4 is responsible for regulating the expression of genes responsible for galactose metabolism. When transformed into a yeast strain such as Y187, containing the *lac*Z reporter gene fused to the GAL4 DNA-binding site, functional GAL4 will activate expression of the *lacZ* reporter to produce β-galactosidase. When two proteins that interact are cloned into the pGAD-C1 and pGBD-C1 vectors and co-transformed into a *lac*Z reporter yeast strain deficient for GAL4, the interaction of the two proteins brings the activation and DNA-binding domains close enough together to produce functional GAL4 (Figure 4.3), which can be detected by analysing β-galactosidase production.

SigG, Rv0181c and Rv0180c were amplified by PCR and cloned individually into the pGAD-C1 and pGBD-C1 vectors to create plasmids pAG07-09 and pAG11-13 respectively. The interaction between the activation and DNA-binding domains takes place in the cell nucleus. For this reason, transmembrane proteins, which don't localise to the nucleus, can interfere with the yeast-2-hybrid system. Therefore, a 510 bp sequence encoding the largest cytoplasmic fragment of Rv0180c was amplified by PCR and cloned into each vector to give plasmids pAG10 and pAG14. Competent Y187 cells were co-transformed with all possible combinations of the test plasmids, empty vector controls and with a positive control, which consisted of the vectors pDHADWhiB1 and pDHBDWhiB1. The WhiB1 protein has previously been shown to interact with itself in the yeast-2-hybrid system (D. Hunt, personal communication).

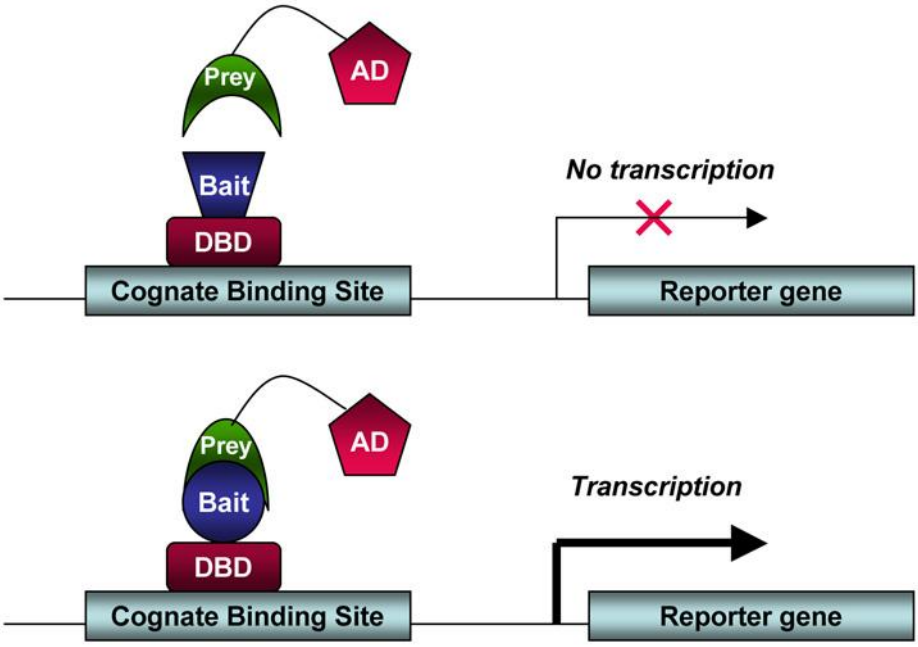


Figure 4.3 - Schematic of the yeast-2-hybrid system (Ratushny *et al.*, 2008). When the bait protein, fused to the DNA-binding domain (DBD) of the transcriptional regulator GAL4, interacts with the prey protein, fused to the activation domain (AD) of GAL4, the two domains are brought into close enough proximity to produce functional GAL4. This leads to expression of GAL4 regulated genes such as the *lac*Z reporter genes.

Transformed yeast were then grown on selective media, YNB containing 20 µg/ml histidine but without leucine or tryptophan (Section 2.1). Growth on this media allowed for selection of yeast transformed with both plasmids because pGAD-C1 encodes the LEU2 gene, which allows growth on leucine-deficient media, and pGBD-C1 encodes the TRP1 gene, which allows growth on tryptophan-deficient media. Transformants were analysed for β-galactosidase activity by colony lift followed by application of X-Gal solution. The development of a blue colour indicated an interaction between the proteins.

It was expected from the initial hypothesis that if Rv0181c and Rv0180c functioned as anti-sigma and anti-anti-sigma factors respectively, SigG co-transformed with Rv0181c and Rv0181c co-transformed with Rv0180c (full or cytoplasmic section only) would demonstrate positive interactions. Figure 4.4 shows the results obtained when SigG was co-transformed with the potential anti-sigma factor Rv0181c and when Rv0181c was co-transformed with the potential anti-anti-sigma factor Rv0180c (full protein). A blue colour was only observed for the positive control suggesting that there is no interaction between SigG and Rv0181c or between Rv0181c and Rv0180c (full). However, a negative result in the yeast-2-hybrid system does not necessarily mean that the proteins do not interact.

The Y2H system is a useful screen for interaction partners but it has several limitations. Expressing the proteins as fusion proteins can cause steric obstruction or incorrect folding, which can prevent protein interactions. In addition, the fusion proteins used in this system have to be transported to the nucleus, where, particularly for membrane-bound proteins (e.g. Rv0180c) there may not be the correct

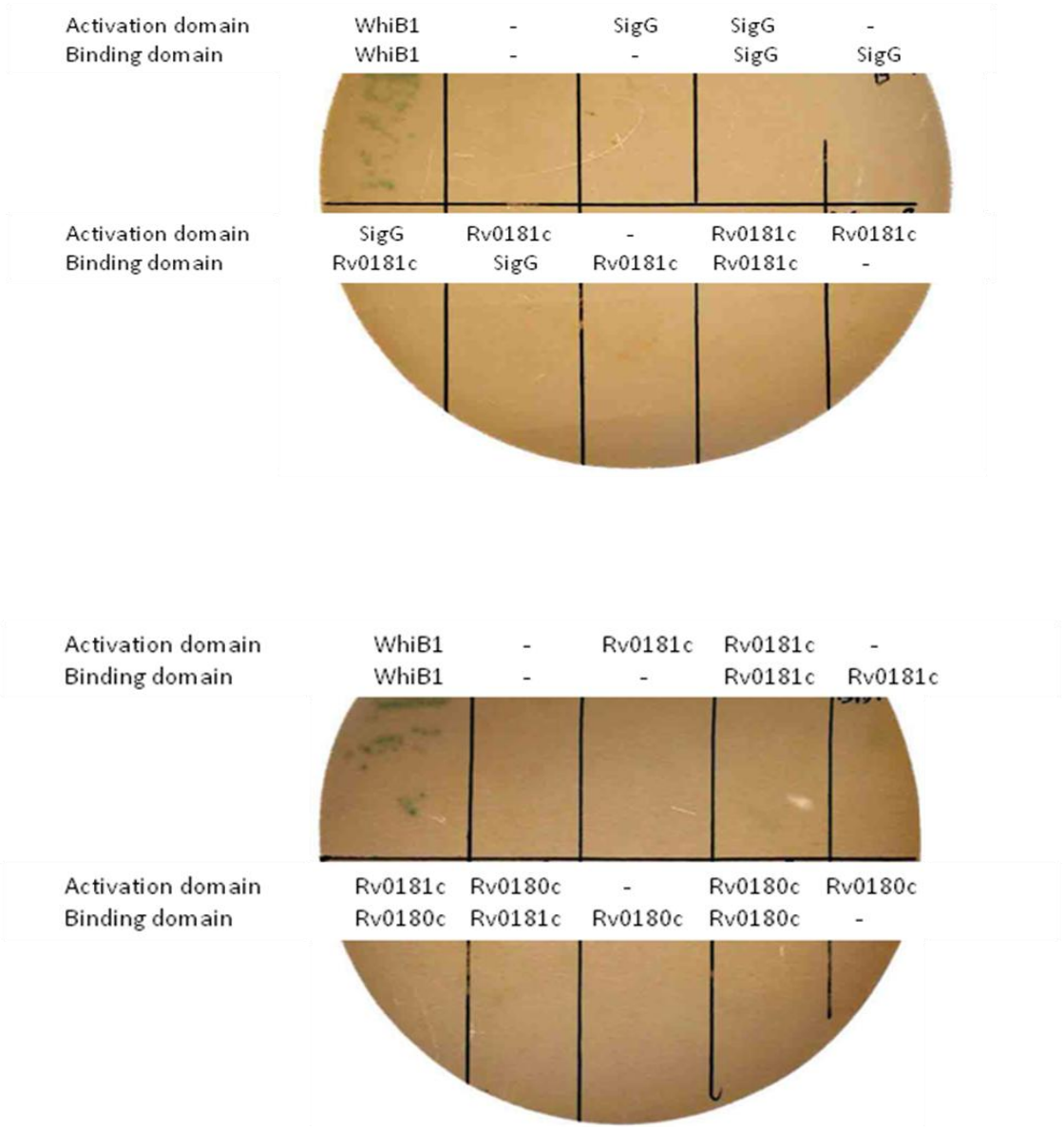


Figure 4.4 –Yeast-2-hybrid analysis of potential protein interactions between SigG, Rv0181c and Rv0180c.

Colony lifts to test whether a) the SigG and Rv0181c proteins and b) the Rv0181c and Rv0180c proteins interact in the yeast-2-hybrid system. The proteins expressed by the pGAD-C1 (Activation domain) and pGBD-C1 (Binding domain) vectors are indicated. – represents an empty vector control. Only the WhiB1 positive control developed a blue colour.

environment for folding, stability and interactions. In view of these factors and the initial results described, it was decided that the Y2H system was not the ideal system to use to investigate interactions between SigG, Rv0181c and Rv0180c.

4.2.3 SigG and Rv0181c do not interact via direct protein pulldown

4.2.3.1 Expression and purification of SigG and Rv0181c proteins

An expression plasmid for SigG had been created previously in the laboratory where the *sigG* gene was amplified by PCR from the first potential start codon downstream of promoter P1 and cloned downstream of the T7*lac* promoter in the vector pET28a to create the expression plasmid pJH05 (J. Houghton, personal communication). This plasmid encodes an amino-terminal His-tagged SigG protein. Similarly in this study the Rv0181c gene was amplified by PCR from its annotated start codon and cloned into the pET28a vector to create plasmid pAG06, which encodes an amino-terminal His-tagged Rv0181c protein. The expected molecular weights of these two proteins were 39.3 kDa and 29.8 kDa respectively.

4.2.3.2 The SigG and Rv0181c proteins were successfully expressed and purified from *E. coli*

pJH05 and pAG06 were transformed into *E. coli* Tuner (DE3) strain, which contains the DE3 prophage integrated into its genome. This prophage encodes T7 RNA polymerase under the control of the *lac*UV5 promoter and this polymerase in turn is required for expression from the T7*lac* promoter. Under normal growth conditions the Lac repressor protein represses expression from both the *lac*UV5 and T7*lac* promoters. Expression of recombinant protein can therefore be induced upon

addition of the lactose analogue IPTG. *E. coli* Tuner cells have a mutation in the lac permease (encoded by *lacY*), which allows uniform entry of IPTG into all cells of a population. This allows for a concentration-dependent, homogeneous level of induction where varying the concentration of IPTG can regulate protein expression from very low levels up to very robust levels. A low level of protein expression can help to enhance protein solubility.

Expression conditions for recombinant SigG had previously been assessed in the laboratory. It had proved difficult to obtain soluble protein, so the option to use a low level of expression in *E. coli* Tuner cells was employed. Cells were induced with 0.01 mM IPTG overnight at 25 ºC (K. Smollett, personal communication); the reduced temperature also helped to solubilise the protein as it slowed the rate of protein production (Sorensen *et al.*, 2005). Figure 4.5a shows how amino-terminal His-tagged SigG was purified initially from cell extract by affinity chromatography using a nickel resin as per Section 2.6.2.2. His-SigG was further purified via gel filtration to produce pure protein (Figure 4.5b). Mass spectrometry showed that pure recombinant protein was *M. tuberculosis* SigG (S. Howell, personal communication).

Expression conditions for His-Rv0181c were determined by testing a variety of IPTG concentrations (0.01, 0.1 and 1 mM), as well as different temperatures (20, 25 and 37 ˚C) and lengths of time for induction (2, 5 and ~15 hours). Rv0181c was expressed but was sparingly soluble except under the mildest conditions (data not shown). Eventually soluble protein was expressed by induction with 0.01 mM IPTG,

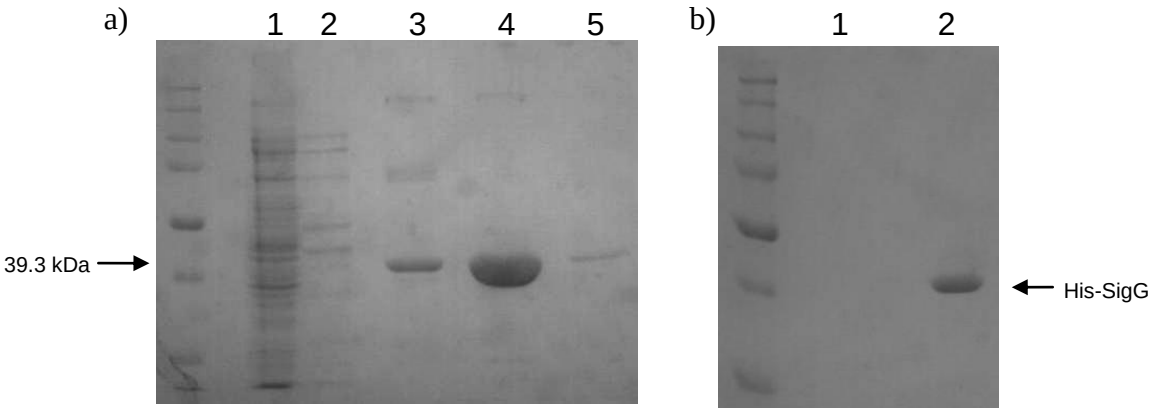


Figure 4.5 – Purification of recombinant His-SigG

SDS-PAGE analysis (12%) showing the purification of recombinant His-SigG by a) affinity chromatography where the five lanes represent: 1) flowthrough fraction, 2) wash fraction, 3-5) elution fractions. Elution fractions 3 and 4 were combined and further purified by b) gel filtration where the two lanes represent two elution peaks. Fraction 2 proved to contain purified His-SigG.

initially for 5 hours at 25 ˚C, followed by an overnight incubation (~ 15 hours) at 20 ˚C. Reducing the temperature ensured that the protein was produced more slowly allowing more time for correct folding and solubility (Figure 4.6). Having determined the optimum expression conditions for His-Rv0181c, large-scale purification was conducted by D. Goldstein (Taylor Group, NIMR) initially by affinity chromatography, followed by gel filtration (as per section 2.6.2.2).

4.2.3.3 Assessing whether there is an interaction between pure His-SigG and His-Rv0181c proteins by direct pulldown

A pulldown uses a bait ligand to identify interaction partners for that protein. In the method used in this study, the bait ligand is bound to magnetic beads (Dynabeads, Invitrogen). The Dynabeads are coated in glycidyl ether (Epoxy) groups, which bind covalently to primary amine or sulphydryl groups in the bait ligand. Ligand-coated beads are then added to a sample containing the target molecule. The use of magnetic beads meant that application of the reaction tube to a magnet (Dynal MPC, Invitrogen) pulled the beads to the side allowing for removal of unbound target without disturbing the beads. Bait protein and any bound target could be eluted by boiling in sample buffer for SDS-PAGE.

Recombinant His-SigG was bound to the Dynabeads overnight as per Section 2.6.5. To prevent false positives due to binding of potential target protein to the beads instead of to the bait protein, BSA was used to block any remaining free Epoxy groups. Pure, recombinant His-Rv0181c (500 µg) was applied to the His-SigG-coated beads and unbound protein was washed from the beads using 1 x PBS.

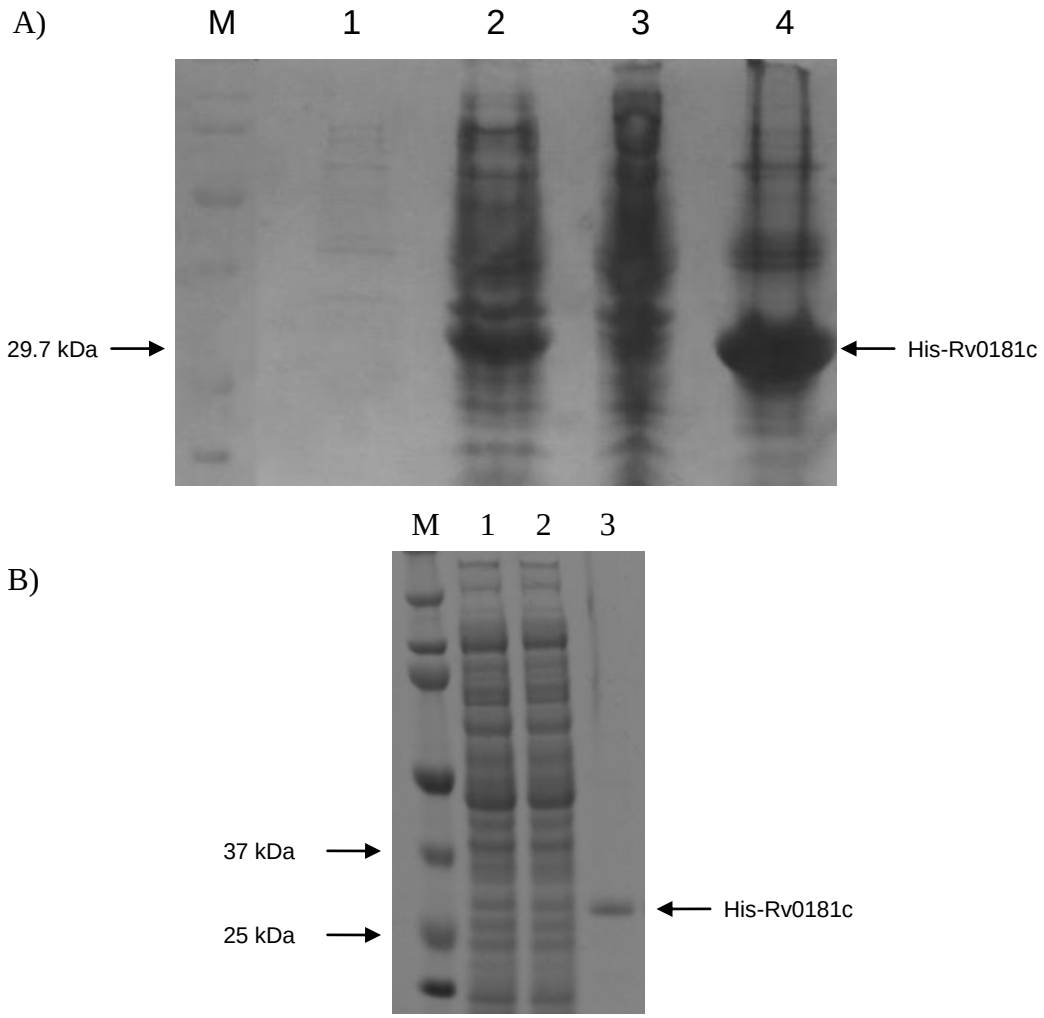


Figure 4.6 – Expression of soluble amino-terminal His-tagged Rv0181c.
SDS-PAGE (15%) analysis showing A) the expression of recombinant His-Rv0181c protein from pAG06 in *E. coli* Tuner cells following induction with 0.01 mM IPTG for 5 hours at 25 ˚C then overnight at 20 ˚C. The four lanes represent: 1) CFE from uninduced *E. coli* Tuner cells, 2) CFE from induced *E. coli* Tuner cells, 3) insoluble fraction, 4) soluble fraction. His-Rv0181c was induced under the conditions described above and was soluble; B) Purification of His-Rv0181c using Ni-NTA resin image supplied by D. Goldstein (Taylor Group, NIMR). The three lanes represent: 1) flowthrough fraction, 2) wash fraction under 40 mM imidazole, 3) elution of His-Rv0181c under 300 mM imidazole. M in both images represents the precision plus dual color size marker (Bio-Rad).

His-SigG and any bound protein was eluted by boiling in SDS-PAGE loading dye and analysed by SDS-PAGE and Western analysis. Due to the fact that both proteins were His-tagged it was possible to detect whether both were present in the elution using the same antibody (mouse anti-His HRP conjugate). Figure 4.7 shows that Rv0181c was not present in the elution from the His-SigG coated beads and therefore does not bind to SigG. The potential roles of Rv0181c and Rv0180c will be discussed at the end of this chapter. However, at this point it was decided that these proteins were unlikely to act as regulators of SigG.

4.3 Identifying potential SigG interaction partners

4.3.1 Potential SigG interaction partners were pulled down from H37Rv CFE

Although Rv0181c and Rv0180c are not regulators of SigG it is likely that there is at least one protein interaction partner. In order to try to identify potential SigG regulators the same pulldown method described above in Section 4.3.2.3 was employed. In this case however, instead of applying a single protein to the bait-covered beads, CFE from *M. tuberculosis* H37Rv was used. This method had a useful 'built-in' control due to the fact that SigG, being a sigma factor, should interact with and therefore, pulldown, subunits of RNA polymerase. Pulldown reactions were analysed via SDS-PAGE and stained with Instant Blue Coomassie stain (Figure 4.8). Clear bands were excised individually and the remainder of the reaction lane was excised from the gel and analysed via LC-MS/MS mass spectrometry by the Cambridge Centre for Proteomics with a view to identifying all the proteins that had been pulled down. It was predicted that the top band pulled

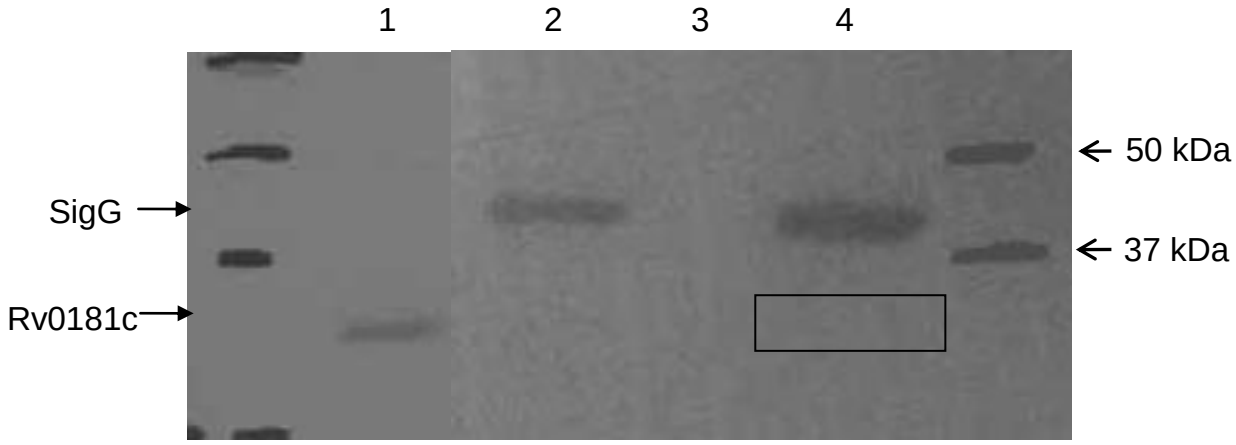


Figure 4.7 – Pure, recombinant His-SigG and His-Rv0181c proteins do not interact.

Western blot analysis (using mouse anti-His HRP conjugate antibody) of the elution from a His-SigG & His-Rv0181c pulldown reaction showing that His-SigG and His-Rv0181c proteins do not interact. Lane 1 represents a positive control to show that the anti-His antibody was able to bind to pure His-Rv0181c protein. The remaining lanes represent elutions from the following pulldown reactions: 2) His-SigG-coated beads only control, 3) Rv0181c applied to BSA-blocked beads in the absence of His-SigG (to confirm the effectiveness of the blocking step), 4) His-SigG-coated beads and His-Rv0181c. The black box represents where Rv0181c should have appeared if it bound to His-SigG; the absence of a band here signifies that the two proteins do not interact.

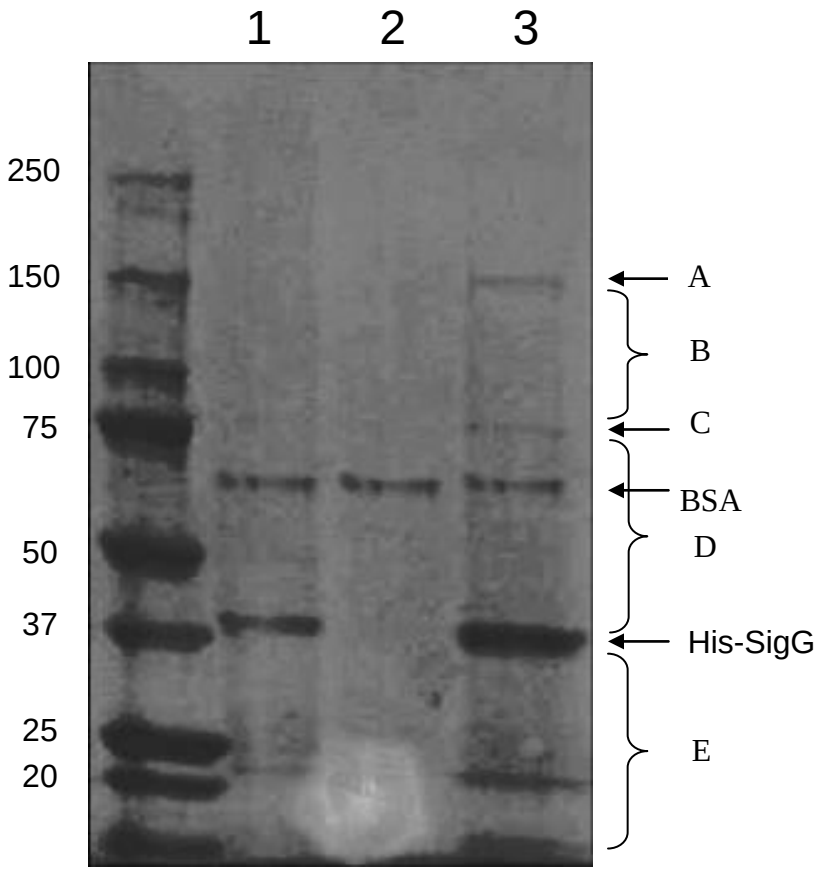


Figure 4.8 – Proteins pulled down from H37Rv CFE by His-SigG
SDS-PAGE (Gradient 4-20%) analysis of the elutions from pulldown reactions using His-SigG as bait and H37Rv CFE as the target sample. The lanes represent: 1) His-SigG-coated beads without CFE, 2) BSA-blocked beads with CFE, 3) His-SigG-coated beads with CFE. BSA (~ 66.7 kDa) was used to block the beads after coupling to His-SigG and is therefore present in all samples. Lane 1 shows that the His-SigG sample did not contain contaminants. Lane 3 shows that two additional clear bands, A and C were pulled down from H37Rv CFE and these were excised along with the His-SigG band for LC-MS/MS analysis. The remainder of the lane was excised as 4 separate sections B, D, and E for LC-MS/MS analysis as further bands could be seen with the naked eye but were below the sensitivity of the imager used to take the gel picture. The BSA band was removed from section D prior to analysis.

down from CFE would be the β' subunit of RNA polymerase, which is approximately 146.7 kDa in size.

4.3.2 Potential SigG interaction partners were identified by LC-MS/MS

The proteins identified by LC-MS/MS analysis of the bands and gel sections are shown in Table 4.1. Band A was identified as the β' subunit of RNA polymerase and the β and α subunits were identified as being present in gel sections B and D respectively (Figure 4.8). The fact that His-SigG was able to interact with and pull down the RNA polymerase subunits from *M. tuberculosis* CFE acted as a control for the experiment to verify that the procedure was working. Anti-sigma factors are often membrane proteins due to the need for them to sense and respond to changes in the environment (Helmann, 1999). Of the proteins identified by LC-MS/MS two were predicted to be membrane proteins; Rv0831c and Rv1084. Rv0831c is a hypothetical protein with no known function, however, Rv1084 is predicted to belong to the thioredoxin-like superfamily. Members of this superfamily contain a redox active CXXC motif that allows them to alter the redox state and therefore the activity of their target proteins. This would make an ideal candidate for further investigation as a potential anti-sigma factor to SigG.

In addition to the two transmembrane proteins identified, the transcription activator (MtrA) of the two component MtrA-MtrB regulatory system was pulled down. Two component regulatory systems typically comprise a response regulator and a membrane-bound histidine kinase sensor. *M. tuberculosis* has been shown to encode 11 complete and 8 unlinked (response and sensor encoded separately in the genome)

Table 4.1 – Proteins pulled down from *M. tuberculosis* CFE

Gel Band/Section	Rv number	Gene name	Size (kDa)	Ions score	Function
A (approx. 150 kDa band)	Rv0668	rpoC	146.7	3689	RNA polymerase β' subunit
B (>75 kDa)	Rv0667	rpoB	129.2	380	RNA polymerase β subunit
C (approx. 75 kDa band)	Rv1133c	metE	81.5	217	Intermediary metabolism - Methionine synthesis
	Rv1821	secA2	88.9	35	Possible preprotein translocase
D (37-75 kDa)	**Rv1084**		**71.0**	**34**	**Transmembrane protein: possible member of thioredoxin superfamily**
	Rv1294	thrA	45.5	75	Homoserine dehydrogenase involved in intermediary metabolism - Threonine, Isoleucine and Methionine biosynthesis
	Rv2245	kasA	43.3	72	Mycolic acid synthesis
	Rv2246	kasB	46.4	71	Mycolic acid synthesis
	Rv3457c	rpoA	37.7	84	RNA polymerase α subunit
E (<37 kDa)	Rv0079		29.5	72	Conserved hypothetical protein
	Rv0728c	serA2	34.4	37	Phosphoglycerate dehydrogenase involved in intermediary metabolism - L-serine biosynthesis
	Rv0831c		**30.2**	**145**	**Conserved hypothetical transmembrane protein**
	Rv1412	ribC	36.8	27	Probable riboflavin synthase alpha chain
	Rv1484	inhA	28.5	47	Mycolic acid synthesis
	Rv1553	frdB	27.2	104	Fumarate reductase iron-sulphur subunit involved in intermediary metabolism - fumarate and succinate interconversion
	Rv2005c		30.9	149	Predicted universal stress protein
	Rv2028c		29.4	38	Predicted universal stress protein
	Rv2161c		30.8	39	Similar to lincomycin production genes involved intermediaty metabolism
	Rv2289	cdh	28.6	66	CDP-diacylglycerol pyrophosphatase involved in phospholipid biosynthesis
	Rv2623		31.6	63	Predicted universal stress protein
	Rv3246c	**mtrA**	**25.3**	**84**	**Transcriptional activator of two component regulatory system**
	Rv3485c		33.2	99	Probable short chain dehydrogenase/reductase involved in intermediary metabolism
	Rv3842c	glpQ1	29.9	28	glycerophophoryl diester phosphodiesterase involved in intermediary metabolism - hydrolysis of phospholipids

Genes of interest as potential SigG regulators are highlighted in bold. Ions score is -10*Log(P), where P is the probability that the observed match is a random event. Individual ions scores > 27 indicate identity or extensive homology ($p<0.05$).

systems and these are thought to play a role in adaptation to different environments (Zahrt *et al.*, 2000). Response regulators are usually DNA-binding proteins that can bind to and either activate or repress transcription from specific promoter sequences (Parish *et al.*, 2003). Should the interaction between SigG and MtrA prove to be true, it is possible that MtrA may assist in recruitment of the SigG holoenzyme to promoter sites.

Multiple universal stress proteins (USP) were pulled down from *M. tuberculosis* CFE. *Rv2005c* and *Rv2623* contain two USP domains, whereas *Rv2028c* only contains one USP domain and a domain of unknown function (O'Toole *et al.*, 2003). All three USPs have been shown to be expressed under hypoxia (Florczyk *et al.*, 2001, Sherman *et al.*, 2001). In addition Rv2623 has been shown to be expressed following phagocytosis (Monahan *et al.*, 2001) and to be required for establishment of persistent infection (Drumm *et al.*, 2009). There is some evidence to suggest that USPs take part in protein-protein interactions in *E. coli*. They have been shown to be able to form homo- and heterodimers with other USPs (Nachin *et al.*, 2008) and to bind to chaperone GroEL (Bochkareva *et al.*, 2002). There has been no report as yet of a USP interacting with a sigma factor.

This is an initial step towards identifying potential regulators of SigG. The pulldown from *M. tuberculosis* CFE should be repeated to determine whether the proteins identified here can routinely be shown to interact with SigG in this type of experiment before candidates can be pursued as potential regulators of SigG.

4.4 Investigating the regulon of the SigG protein

Initially it was hoped that this study would identify an anti-sigma factor to SigG. In order to determine whether a potential anti-sigma factor could inhibit SigG-dependent transcription it was necessary to identify a SigG-regulated gene. This would have enabled an *in vitro* transcription assay to be used to determine the effect of the potential anti-sigma factor on transcription of this gene by the SigG holoenzyme as was performed for the SigF anti-sigma factor UsfX (Beaucher *et al.*, 2002).

In addition, it is not the case that because a gene is expressed the protein is being translated or that once produced the protein is active. In the case of sigma factors it is possible that even though they are being expressed the protein could be sequestered away by an anti-sigma factor preventing it from forming the RNAP holoenzyme. In order to determine when the SigG protein was active, this study attempted to identify a SigG-dependent gene or genes, with the aim of investigating expression of this gene(s) under various conditions.

4.4.1 *Rv0654* is not a SigG-regulated gene

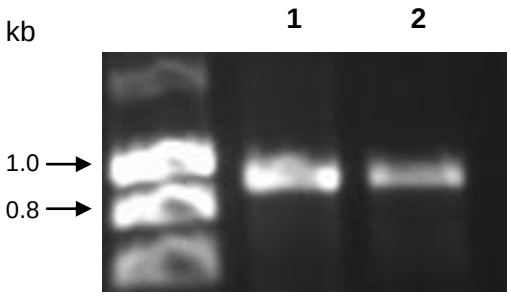
4.4.1.1 Confirmation of the transcriptional start site(s) of *Rv0654* by 5'-RACE

Previous work conducted in this laboratory has shown that expression of the *Rv0654* promoter was significantly increased in the *E. coli* two-plasmid system when in the presence of SigG but that expression was still present in the absence of SigG (R. Balhana and J Smith, personal communication). In this system a protein expression plasmid and a promoter-*lacZ* fusion plasmid are co-transformed into *E. coli* protein

expression cells such as Tuner cells. Promoter activity is assessed via β-galactosidase assay, in the presence and absence of protein expression. If the protein of interest regulates expression from the promoter being tested then a difference in expression is observed between induced strains when a strain expressing the protein of interest is compared to a vector only control. The increase in promoter activity following induction of SigG expression suggested that *Rv0654* may be expressed from two promoters, one that is recognised by an *E. coli* sigma factor and one that is *M. tuberculosis* SigG-dependent. In order to determine whether there were one or two promoters expressing *Rv0654* in *E. coli*, 5'-RACE was used to determine the transcription start site or sites for this gene. *E. coli* Tuner cells co-transformed with pRAE2 (Rv0654 promoter-*lacZ* fusion) and either pET28a (vector control) or pJH05 (His-SigG expressing vector) were induced with IPTG at a final concentration of 0.25 mM for 4 hours. RNA was prepared from each culture and DNase treated as per Sections 2.3.6 and 2.5.1. 5'-RACE was conducted on the clean RNA samples using the GeneRacer kit (Section 2.5.4). Briefly, a commercial oligonucleotide was ligated to the 5' ends of the RNA. Oligo-capped RNA was reverse transcribed into cDNA using random primers. PCR using a forward primer specific to the commercial oligo (GeneRacer 5' Oligo) and a reverse primer specific to the *lacZ* tag (GSPLacR) of the *Rv0654* transcripts was used to amplify the *Rv0654* gene-specific transcript(s) which were analysed by agarose gel electrophoresis.

It was hypothesised that in the absence of SigG (pRAE2 with pET28a), only one promoter would be active so there would only be one transcript and therefore one band would be seen on the agarose gel. In contrast, in the presence of SigG (pRAE2 with pJH05), two promoters would be active, giving two transcripts and therefore

a)



b)

agtccgcgcatcccgccgaaatgggtgtacaccgccatcgtcgaggtgcctgctgcggcggcc
accttgcgggtctgcagcgcgtcgggcccgtgatcgtcgagcagtcgcacgccggcgtgcagc
agctcgtcgcgaacaccggtctgcgaggtcatccttgccatgttctcaccaagggcgtaccgttt
ccaatatcagtgaaataacaatgttataggagatcggcatgaccaccgcacaagccgccgaatc
ccaaaacccatatctcgagggcttcctggcgccggtgagcaccgaggtaactgccaccgacctg
ccggtcaccggccgcattccgg

Figure 4.9 – 5'-RACE in *E. coli* to identify the transcriptional start site of Rv0654 in the presence or absence of *M. tuberculosis* SigG

a) Agarose gel electrophoresis of *Rv0654-lacZ* 5' RACE products obtained from cDNA from *E. coli*, expressing pRAE2 in the presence or absence of SigG, using the GeneRacer RNA 5' primer and the *lac*Z-specific RACE primer. The lanes represent: 1) *Rv0654-lacZ* without SigG, 2) *Rv0654-lacZ* with SigG. A single (800-900 bp) PCR product was obtained in both the presence and absence of SigG. b) Position of *Rv0654* transcriptional start site (highlighted in red) relative to the start of the predicted coding region (highlighted in yellow). Putative -10 and -35 regions are shown in the green boxes.

two bands on the agarose gel. However, Figure 4.9a shows that one transcript (approx. 800-900 bp) was identified in both the presence and absence of SigG.

Sequencing of both transcripts showed that both PCR products mapped to an A residue, 33 bp upstream of the annotated Rv0654 open reading frame (Figure 4.9b). As previously stated, Rv0654 expression has been shown to be increased in the presence of SigG in the *E. coli* two-plasmid system, but a lower level of expression was seen when SigG was absent (R. Balhana and J. Smith, personal communication). The fact that there was only one transcriptional start site for Rv0654 suggested that this gene is expressed from a single promoter and that if it is indeed a SigG-regulated gene this promoter is recognised by both *M. tuberculosis* SigG and an *E. coli* sigma factor.

4.4.1.2 Quantitative RT-PCR of a *sig*G operon deletion strain showed that expression of Rv0654 was not SigG-dependent

Having determined that *Rv0654* was expressed from a single promoter it was decided to investigate expression of this gene in a *sigG* operon deletion strain, ΔsigGoperon in which, the *sigG*, *Rv0181c* and *Rv0180c* genes had been deleted. Creation and analysis of this strain will be discussed in Chapter 5. qRT-PCR was performed on RNA extracts from wild-type H37Rv and Δ*sig*Goperon *M. tuberculosis* strains (OD_{600} 0.6) to determine whether the expression of *Rv0654* was altered in the absence of SigG. It was hypothesized that should *Rv0654* be expressed from a SigG-dependent promoter, expression in the ΔsigGoperon strain would be abolished.

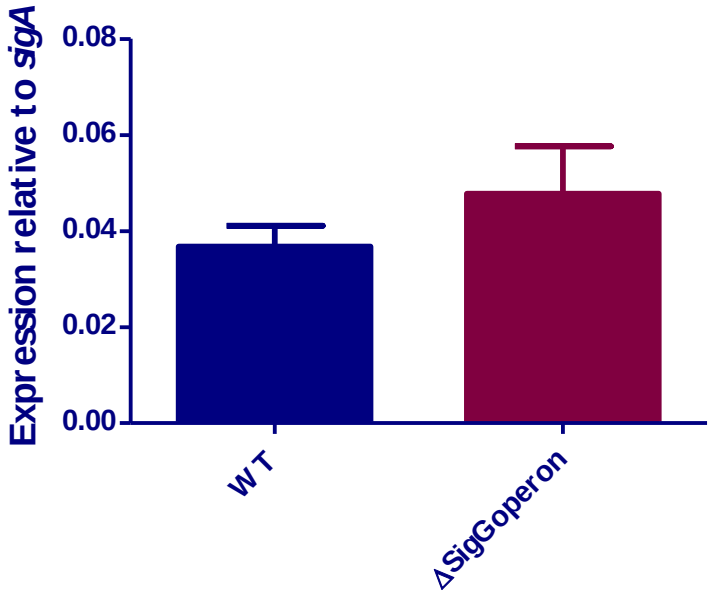


Figure 4.10 – qRT-PCR of *Rv0654* expression in ΔsigGoperon compared to wild-type H37Rv

qRT-PCR (SYBR-Green) of *Rv0654* was performed on RNA extracts of the *M. tuberculosis* H37Rv and ΔsigGoperon strains, and normalised to the corresponding *sigA* data. The columns represent the average and the error bars the standard deviation of three biological replicates, which were assayed in triplicate. The normalised expression level of *Rv0654* was not significantly different between the two strains (Two-tailed T-test, $P > 0.1$).

No significant difference was observed between the expression levels of *Rv0654* in wild-type H37Rv or Δ*sig*Goperon *M. tuberculosis* strains (Figure 4.10; P > 0.1, T-test). Expression in the absence of SigG could be due to the ability of the promoter to be recognised by another sigma factor as well as by SigG, although if this was the case a reduction in gene expression would still be expected. Therefore it was concluded that expression of *Rv0654* is not SigG-dependent in *M. tuberculosis*.

4.4.2 Confirming overexpression of the SigG protein in *M. tuberculosis* transformed with the pKS09 plasmid

A SigG overexpression plasmid, pKS09, had been created previously in the laboratory and *M. tuberculosis* transformed with this plasmid was shown to be overexpressing SigG at the RNA level (K. Smollett, personal communication). However, the fact that a gene is being overexpressed at the RNA level does not necessarily mean that this will translate to the protein level. Prior to using this strain it was, therefore, necessary to determine whether the SigG protein was being overexpressed. His-SigG was used to raise an anti-SigG antibody from rabbits (BioServ UK Ltd, Sheffield University). CFEs from three biological replicates of *M. tuberculosis* H37Rv, transformed either with the empty vector pKS12 or with pKS09 were analysed via Western blot using rabbit anti-SigG and goat anti-rabbit HRP-conjugate antibodies.

The SigG protein was barely detectable in the CFEs from the vector only control (pKS12), whereas in comparison, a much higher level of protein expression was observed in the CFEs from *M. tuberculosis* expressing SigG from the pKS09 plasmid (Figure 4.11). The low level of SigG expression from the vector control

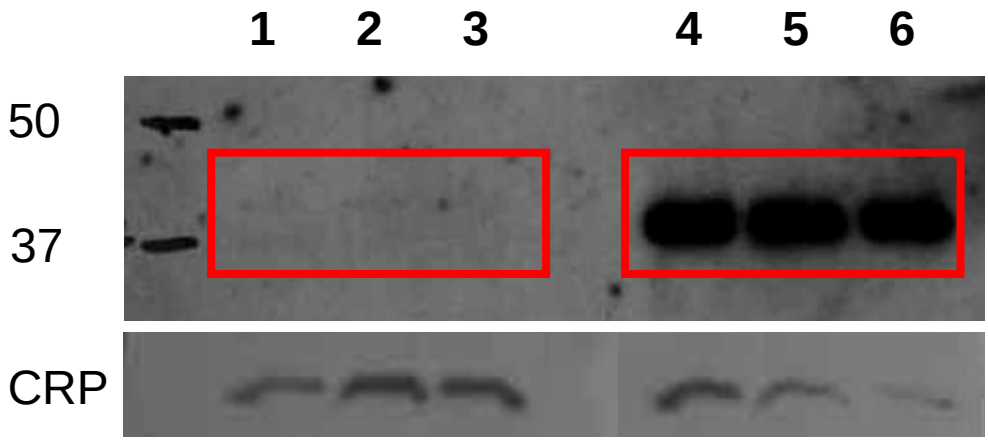


Figure 4.11 – Western blot analysis demonstrating successful overexpression of SigG protein

Western blot analysis (using rabbit anti-SigG, rabbit anti-CRP and goat anti-rabbit HRP-conjugate antibodies) of CFEs from *M. tuberculosis* transformed with either pKS12 or pKS09. The lanes represent 1-3) three biological replicates of pKS12, 4-6) three biological replicates of pKS09. In the vector only control (pKS12) there was only the faintest hint of SigG protein expression whereas in the replicates containing the pKS09 overexpression strain the SigG protein was detected at very high levels in comparison (red boxes denote SigG protein). The CRP protein was used as a control to demonstrate equal loading of protein in the samples.

corresponded with the inability to detect myc-tagged SigG during the translational start site mapping work (Section 4.2.1). These observations suggest that in a wild-type strain under normal growth conditions SigG is expressed at a very low level. This analysis confirmed that SigG was expressed at a highly elevated level from pKS09 at the protein level in addition to the RNA level.

4.4.3 *Rv0887c* and *Rv0911* are potential SigG-dependent genes

4.4.3.1 *Rv0887c* and *Rv0911* were the most up-regulated genes upon SigG overexpression

Microarray analysis conducted in this laboratory on *M. tuberculosis* overexpressing SigG (pKS09) compared to the vector only control (pKS12) showed the two genes *Rv0887c* and *Rv0911* to be highly up-regulated, by 15- and 13-fold respectively (K. Smollett, personal communication). BLAST analysis of the protein sequences of Rv0887c and Rv0911 showed that they belonged to the Glo_EDI_BRP_like domain superfamily, with Rv0887c containing one domain and Rv0911 containing two domains (Figure 1.11). This superfamily contains several structurally related metalloproteins, including type I extradiol dioxygenases, glyoxalase I and a group of antibiotic resistance proteins. Type I extradiol dioxygenases catalyse the cleavage of aromatic rings and are therefore key to the degradation of aromatic compounds (Eltis *et al.*, 1996). Glyoxalase I is the glutathione-dependent enzyme responsible for the first step of the glyoxalase system, whereby the toxic metabolic intermediate, methylglyoxal, is converted to pyruvate (Inoue *et al.*, 1995). Finally, the most common antibiotic resistance protein found in this family is the Bleomycin

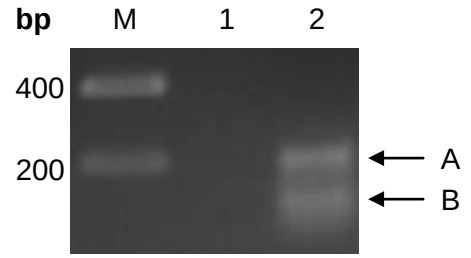
resistance protein, which binds to and sequesters the antibiotic bleomycin (Gatignol *et al.*, 1988).

4.4.3.2 One transcriptional start site for Rv0887c was identified

In order to try to determine the number of promoters expressing *Rv0887c* and *Rv0911* and their position, 5'-RACE was used to determine the transcription start site or sites for these genes. RNA was prepared from a culture of *M. tuberculosis*, which had been transformed with pKS09 and was therefore over-expressing SigG, and DNase treated as per Sections 2.4.2 and 2.5.1. 5' RACE was conducted on the clean RNA samples using the GeneRacer kit (Section 2.5.4). PCR using a forward primer specific to the commercial oligo (GeneRacer 5' Oligo) and a reverse primer specific to the gene of interest (RACE887 and RACE911) was used to amplify the *Rv0887c* and *Rv0911* gene-specific transcripts which were analysed by agarose gel electrophoresis.

Unfortunately no PCR products were obtained for *Rv0911*; however two products of approximately 200 bp and 100 bp were obtained for *Rv0887c* (Figure 4.12a). Sequencing of the two products with the primer Rv0887R only yielded a transcriptional start site for the larger product A, which mapped to a G residue 52 bp upstream of the annotated coding region of Rv0887c. A ClustalW alignment of the upstream regions of *Rv0887c* and *Rv0911* showed that not only was this G residue conserved between the two genes but that it was in a region of significant homology (Figure 4.12b). This suggests that there may be a promoter in this region for Rv0911 as well as for Rv0887c and that they would likely be recognised by the same sigma factor. The smaller PCR product would place that transcriptional start site within the

a)



b)

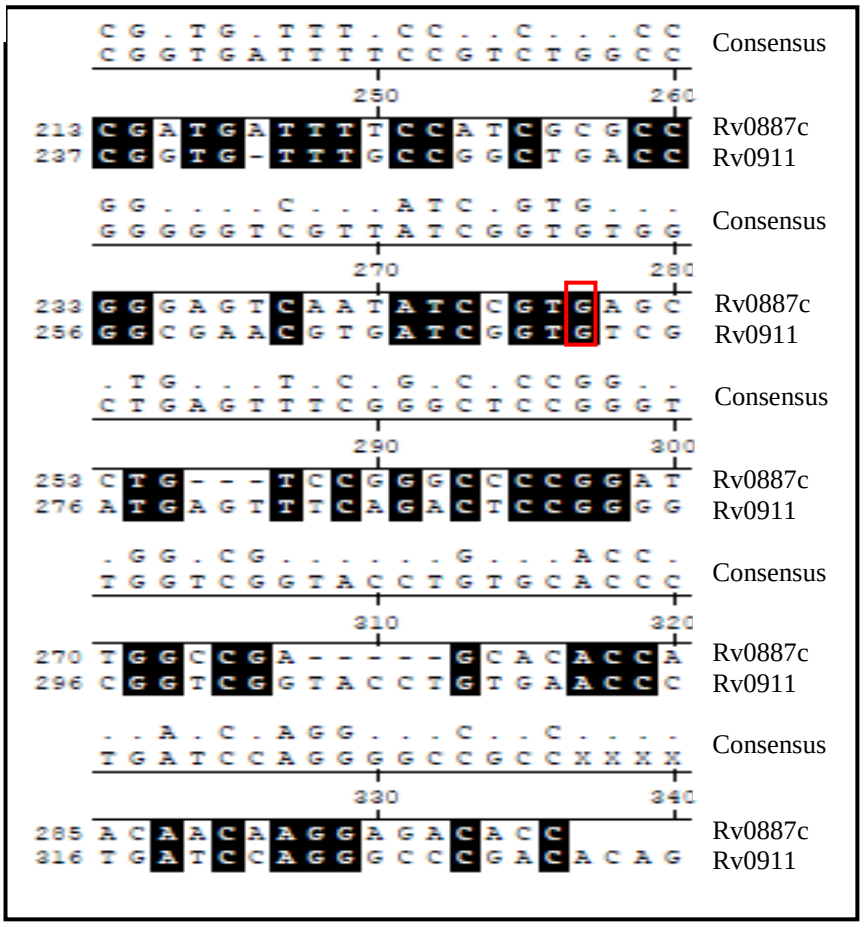


Figure 4.12 – Location of a transcriptional start site for Rv0887c that is in a region conserved with Rv0911

a) 5'RACE of Rv0911 (lane 1) and Rv0887c (lane 2), M corresponds to size marker. Two products (A and B) were obtained for Rv0887c but none were detected for Rv0911. Sequencing yielded a transcriptional start site for PCR product A, which mapped to a G residue 52 bp upstream of the annotated start site for Rv0887c. b) An alignment of the upstream regions of *Rv0887c* and *Rv0911* revealed that this G residue (red box) was conserved between the two genes and that it fell in a region of strong homology.

annotated open reading frame of Rv0887c. This could mean that the annotated ORF is incorrect; however, the GeneRacer kit does not distinguish between full length mRNA and processed RNA molecules. It is therefore possible that the smaller PCR product is merely a processed version of the larger one. The location of the transcriptional start site for Rv0887c was used to design templates for both Rv0887c and Rv0911 for use in *in vitro* transcription and DNA-binding assays.

4.4.3.3 No SigG-specific *in vitro* transcription products were obtained for Rv0887c or Rv0911

In vitro transcription uses a recombinant RNA polymerase, formed from core RNA polymerase and the sigma factor of interest, to determine whether the sigma factor is able to bind to and initiate transcription from a promoter. Transcription is detected by incorporation of a radiolabelled nucleotide into the RNA product(s). Core RNA polymerase on its own is able to bind to the end of a DNA fragment and will transcribe it non-specifically to give an RNA product identical in size to the DNA template. Once a sigma factor has bound to the core RNA polymerase to form the holoenzyme it will greatly reduce, but not eliminate, the non-specific transcription, and most importantly it will enable the RNA polymerase to recognise and initiate transcription from a specific promoter (Borukhov *et al.*, 2003). In an *in vitro* transcription reaction where the sigma factor is able to recognise the promoter, it would therefore be expected that two RNA products would be seen: one corresponding to the non-specific product and the other smaller product being due to transcription from the promoter to the end of the DNA template.

The ability of a SigG RNA polymerase holoenzyme to recognise and initiate transcription from the promoters of Rv0887c and Rv0911 was analysed using His-tagged *M. tuberculosis* SigG and *E. coli* core RNA polymerase. Due to the small size of the PCR products containing the promoter regions of Rv0887c (455 bp) and Rv0911 (557 bp), they were amplified from the bacterial artificial chromosome, pBAC17 containing the relevant region of the *M. tuberculosis* genome, instead of from genomic DNA to increase the specificity of the reaction. DNA fragments were gel purified for use in the *in vitro* transcription assay. In order to confirm that the assay was working a positive control was used, which consisted of the BldN sigma factor from *Streptomyces coelicolor* (*S. coelicolor*) and its template *bldM* kindly provided by M. Buttner, John Innes Centre. The *in vitro* transcription assay was conducted as per Section 2.6.8 with two reactions per template, one with just *E. coli* core RNA polymerase and one with the His-SigG holoenzyme. RNA products were analysed via polyacrylamide gel electrophoresis.

The RNA products expected for the positive control (*bldM*) DNA template were a promoter-specific product of 160 bases and a non-specific one of 240 bases, corresponding to transcription of the whole template. Non-specific RNA products of 455 and 557 bases were expected for *Rv0887c* and *Rv0911* respectively. Specific RNA products expected from the *Rv0887c* and *Rv0911* templates (based on the location of the transcriptional start site identified for *Rv0887c* in Section 4.4.3.2) were 302 and 340 bases respectively.

The *S. coelicolor* BldN positive control demonstrated that the *in vitro* transcription assay was behaving as expected with the production of a single band (~ 240 bases)

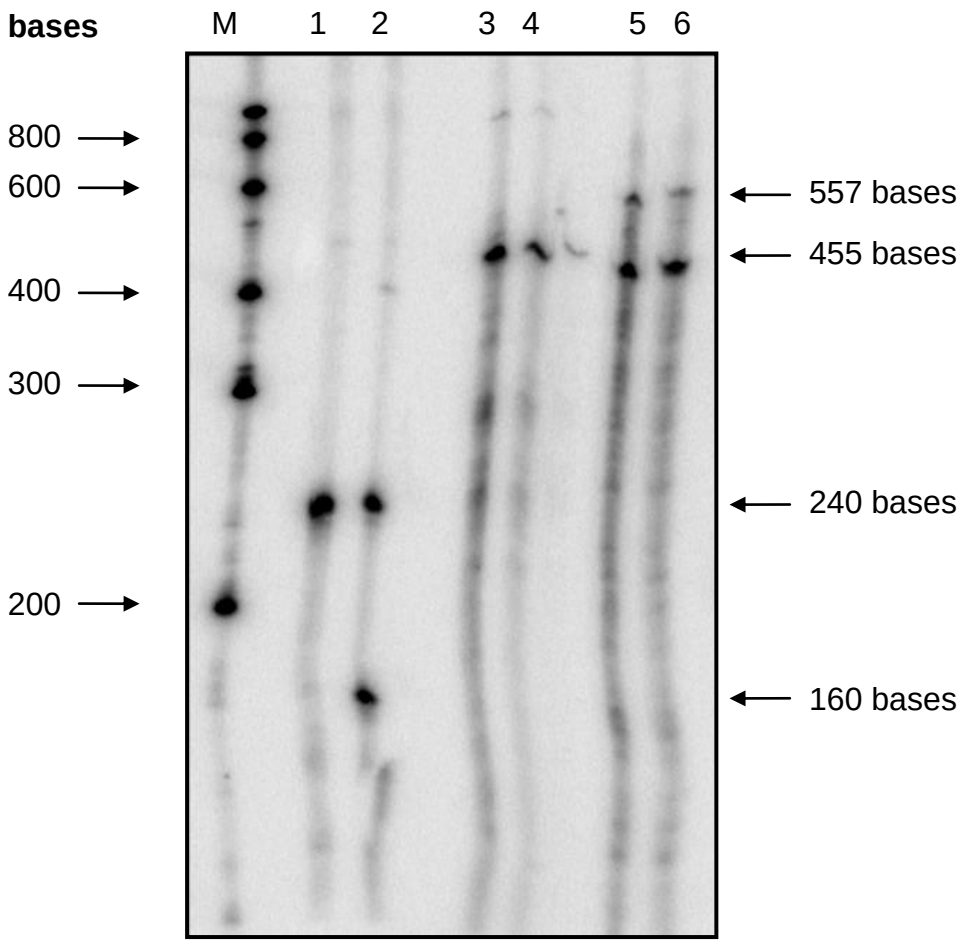


Figure 4.13 – Rv0887c and Rv0911 were not transcribed by His-SigG holoenzyme

Polyacrylamide/urea gel electrophoresis (6%) of the *in vitro* transcription products from *bld*M, *Rv0887c* and *Rv0911* DNA templates transcribed by either *E. coli* core RNA polymerase or either the BldN holoenzyme (for *bld*M) or His-SigG holoenzyme (for *Rv0887c* and *Rv0911*). Each template (0.25 µg) was transcribed by either *E. coli* core RNA polymerase (lanes 1, 3 and 5), reconstituted BldN holoenzyme (lane 2) or reconstituted His-SigG holoenzyme (lanes 4 and 6). The lanes represent: M) Perfect RNA size marker, 1 and 2) *bld*M template, 3 and 4) *Rv0887c* template, 5 and 6) *Rv0911* template. All reactions demonstrated non-specific transcription of the entire template by RNA polymerase but only the *bld*M positive control demonstrated the appearance of a RNA product specific to transcription from a promoter (160 base product in lane 2).

with core RNA polymerase only and the appearance of a 160 base, promoter-specific product when transcribed with reconstituted BldN holoenzyme. In comparison only the non-specific products were observed for both *Rv0887c* (455 bases) and *Rv0911* (557 bases) in both the presence and absence of the His-SigG sigma factor (Figure 4.13). A contaminating band was seen for *Rv0911* around 450 bases, which demonstrates that the DNA template was not pure.

The lack of transcription from the *Rv0887c* and *Rv0911* templates could mean that these genes are not expressed from a SigG-dependent promoter. However, it is possible that the amino-terminal His tag was interfering with either the formation of the holoenzyme or with binding to the promoter. The fact that the His-SigG protein was able to interact with polymerase subunits in the pulldown experiment conducted in Section 4.3.1 suggested that formation of the holoenzyme was not the problem. It was therefore decided to test whether His-SigG was able to specifically bind to DNA fragments containing the promoter regions of *Rv0887c* and *Rv0911*.

4.4.3.4 His-SigG appears to directly bind to 227 bp and 221 bp DNA fragments corresponding to the promoter regions of *Rv0887c* and *Rv0911* respectively

The DNA fragments used for the protein-DNA binding assay were produced by cutting the *Rv0887c* and *Rv0911 in vitro* transcription templates with the restriction enzymes *Bsr*BI (*Rv0887c*) and *Bsr*DI and *Aat*II (*Rv0911*), to retain the promoter region identified for *Rv0887c* in Section 4.4.3.2 but in a DNA fragment more suited in size for a protein-DNA binding assay (*Rv0887c* - 227 bp; *Rv0911* – 221 bp). All protein-DNA binding assays were conducted as per Section 2.6.9. Protein-DNA

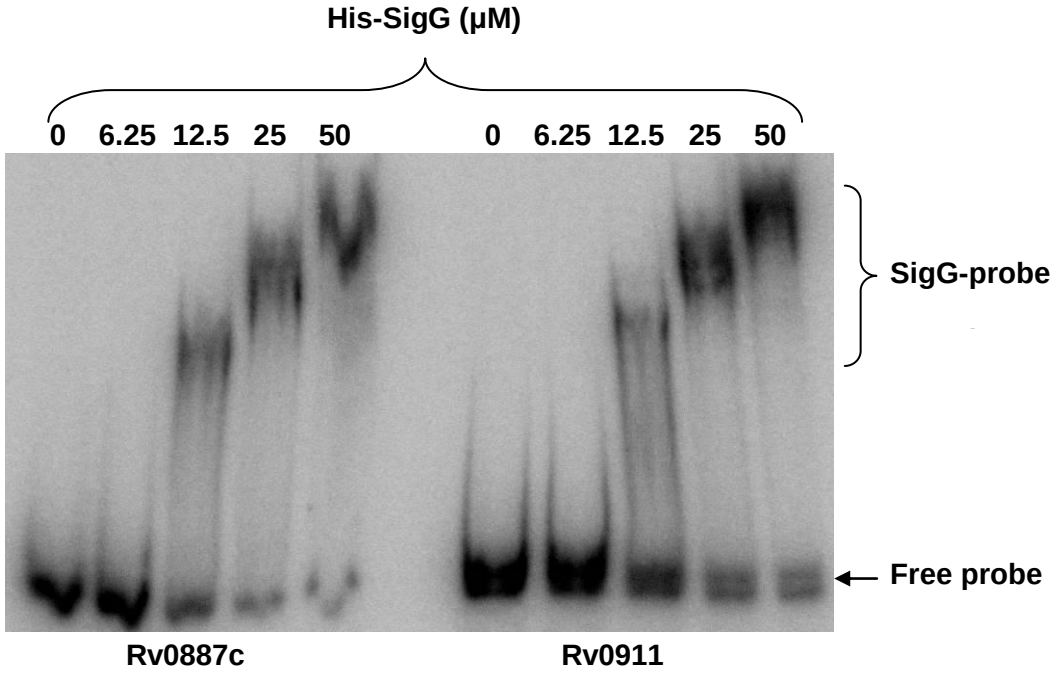


Figure 4.14 – Protein-DNA binding assays between recombinant SigG protein and the *Rv0887c* and *Rv0911* promoter regions

Protein-DNA binding assays were performed between pure His-tagged SigG protein and ^{32}P-labelled *Rv0887c* and *Rv0911* promoter PCR probes (1 fmol; 227 bp and 221 bp respectively) using increasing amounts of protein. Protein-probe complexes were analysed via 6% PAGE. A 'band shift' was observed with 12.5 μM protein but the degree to which the probe was shifted increased with protein concentration, suggesting that the interaction may be non-specific.

interaction is demonstrated by retarded mobility of a labelled DNA probe upon native gel electrophoresis. Potential protein-DNA interactions were observed for both the ^{32}P-labelled *Rv0887c* and *Rv0911* probes (Figure 4.14). However, in both cases there appeared to be a greater degree of retardation as the concentration of His-SigG increased suggesting non-specific binding of the protein to the DNA probe.

4.4.3.5 Binding of His-SigG to the promoter regions of *Rv0887c* and *Rv0911* was non-specific

To test whether the binding of His-SigG to the *Rv0887c* and *Rv0911* probes was specific or non-specific, competition protein-DNA binding assays were performed. The standard reaction was performed as per Section 2.6.9 but with the addition of increasing amounts of either unlabelled specific (the uncut *Rv0887c* and *Rv0911 in vitro* transcription DNA fragments) or non-specific (the promoter region of the *M. smegmatis rrnB* gene, which is a SigA-dependent gene) competitor DNA. Neither the specific, nor the non-specific unlabelled DNA fragments was able to compete when using the highest concentration of His-SigG protein (50 μM, Figure 4.15).

In a complementary approach, it was decided to label the non-specific *rrnB* DNA fragment (~ 250 bp) to determine whether His-SigG was able to bind to this fragment. Labelling of the *rrnB* DNA fragment and the protein-DNA binding assay were carried out as per Section 2.6.9. A 'band shift' was observed with the highest protein concentration used (50 μM) showing that although it required a higher concentration of protein to bind, His-SigG was binding to a non-specific DNA fragment (Figure 4.16). Coupled with the *in vitro* transcription results this suggests either that SigG does not directly regulate the expression of *Rv0887c* and *Rv0911* or

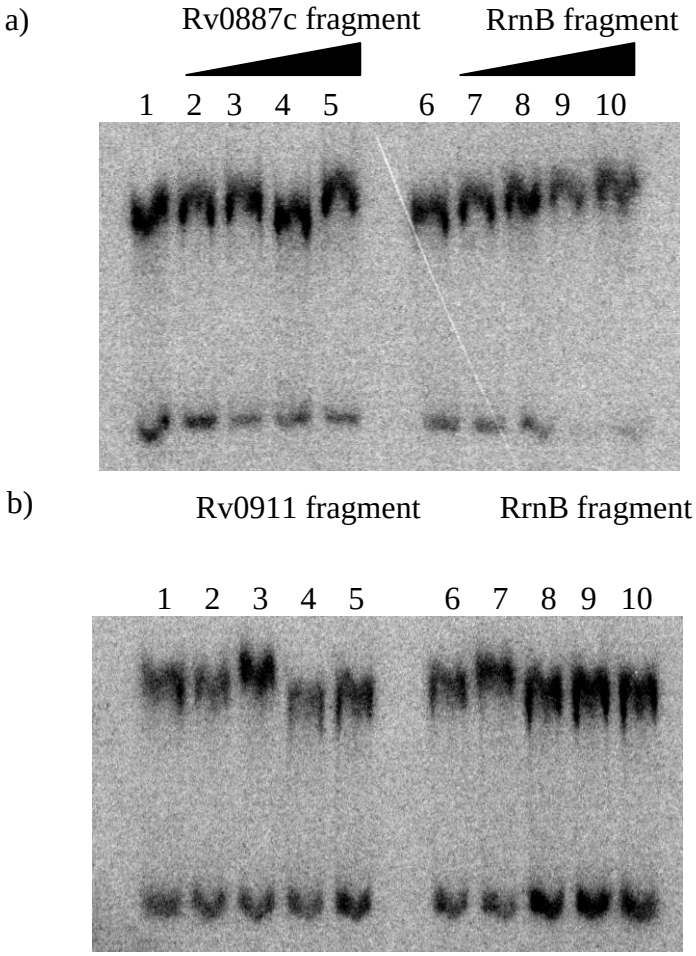


Figure 4.15 – Specific and non-specific competition protein-DNA binding assays
Protein-DNA binding assays were performed between pure His-tagged SigG protein and ^{32}P-labelled a) *Rv0887c* and b) *Rv0911* promoter PCR probes (1 fmol; 227 bp and 221 bp respectively) using increasing amounts of unlabelled a) *Rv0887c* fragment or b) *Rv0911* fragment as specific competitor or a 250 bp unlabelled *M. smegmatis rrnB* fragment as non-specific competitor. Protein-probe complexes were analysed via 6% PAGE. The protein concentration used was 50 µM in all lanes. Specific and non-specific competitor DNA fragments were added at 10-, 25-, 250- and 500-fold molar excess of the labelled probe (lanes 2-5 and 7-10 respectively). Neither the specific nor non-specific unlabelled DNA fragments were able to compete for binding of SigG.

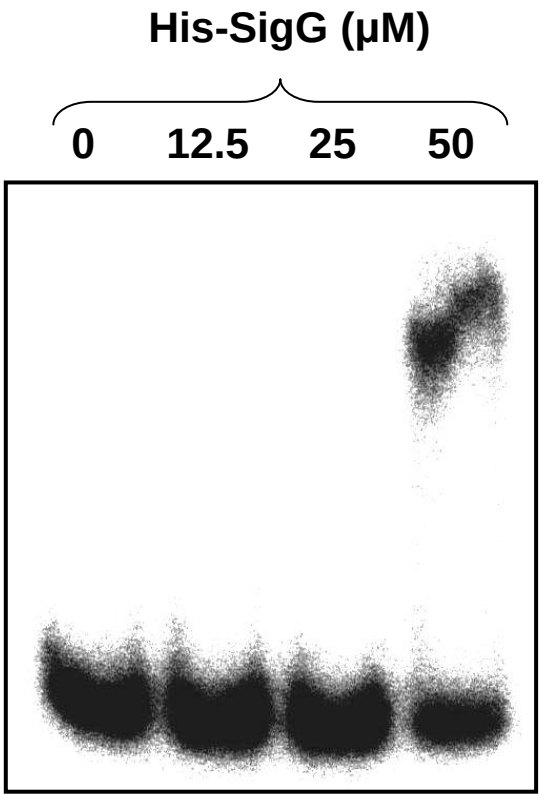


Figure 4.16 – Demonstrating non-specific binding of recombinant SigG protein to the 250 bp *M. smegmatis rrnB* promoter fragment

Protein-DNA binding assays were performed between pure His-tagged SigG protein and ^{32}P-labelled *M. smegmatis rrnB* promoter PCR probe (1 fmol; 250 bp) using increasing amounts of protein. Protein-probe complexes were analysed via 6% PAGE. A 'band shift' was observed only with the highest protein concentration (50 μM) but this showed that His-SigG was non-specifically binding.

that the presence of the His tag is interfering with specific DNA binding. In order to determine whether the His tag is interfering, a new recombinant could be expressed and purified in which, the tag can be cleaved so as to leave the native protein.

4.5 Discussion

This chapter has investigated two different aspects of the SigG protein: its interaction with other proteins, which may be regulators of SigG, and its interaction with promoter regions of genes that could be SigG-dependent. Initially it was hypothesised that Rv0181c and Rv0180c could act as regulators of SigG activity. This hypothesis was based on the fact that ECF sigma factors are often co-transcribed with and regulated by their cognate anti-sigma factors (Hughes *et al.*, 1998).

Prior to investigating the SigG protein it was first decided to try to determine the correct translational start site. Primer extension studies had shown that one of the three promoters expressing *sigG*, promoter P1, was within the annotated coding region suggesting that the annotated start site was not correct (Dawson, 2005). A system for determining translational start sites using detection of a Myc tag had been developed in this laboratory and used to successfully determine the start sites of several proteins (Smollett *et al.*, 2009). However, when it came to the proteins encoded by the *sigG* operon, the level of expression was so low that it was impossible to distinguish the tagged proteins from the high levels of background obtained with the anti-myc antibody. For the purposes of this study in all future experiments the first potential translational start site downstream of promoter P1 was used for SigG and the annotated start sites were used for Rv0181c and Rv0180c.

A potential method that could be used in future, for determining the correct translational start site of SigG could be to use the plasmid pKS09, used to successfully overexpress the SigG protein, and insert either the Myc tag or another tag that reacts with an antibody that produces less background from mycobacterial extracts. The much higher level of expression of the SigG protein in the overexpression strain should make the protein detectable above background and therefore enable the correct translational start site to be determined.

An initial yeast-2-hybrid screen failed to demonstrate interactions between SigG and Rv0181c or between Rv0181c and Rv0180c. This system has its limitations and the fact that no interaction was observed did not necessarily mean that these proteins did not interact. Interactions can be prevented by expression of the proteins as fusion proteins due to steric obstruction or incorrect folding and also due to the lack of post-translational modifications that would occur in the bacterial cell but cannot occur in yeast (Van Criekinge *et al.*, 1999). In addition, the yeast-2-hybrid system relies on translocation of the fusion proteins to the nucleus where there may not be the correct environment for folding, stability and interactions, in particular for membrane proteins such as Rv0180c. SigG and Rv0181c proteins were then expressed and purified with an amino-terminal His tag. A direct pulldown reaction using His-SigG as bait to bind the pure His-Rv0181c protein failed to yield an interaction between the two proteins.

It has been demonstrated that where there is conservation of gene order, this can be used as an indicator of proteins that physically interact (Dandekar *et al.*, 1998). In the case of *sigG* and *Rv0181c*, the gene order is not conserved in all species of

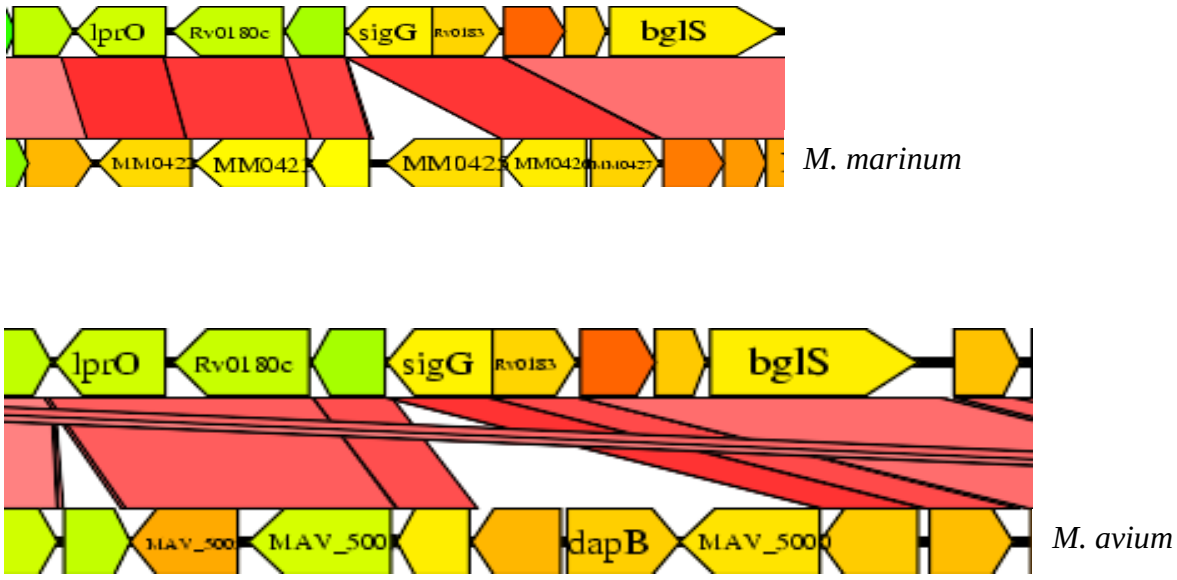


Figure 4.17 – Comparison of the gene order of the *sigG* operon in *M. tuberculosis* with four mycobacterial species

Comparison of the gene order of the *sigG* operon in *M. tuberculosis* with that in *M. marinum* and *M, avium*. In all cases the top genome represents *M. tuberculosis* and the bottom genome labelled with the species name. Red sections indicate the positions of homologous genes in the two genomes being compared. In *M. marinum* there is an extra gene, *MM_0425*, inserted between *sigG* and the *Rv0181c* homologue *MM_0424* and in *M. avium* there are three genes between *sigG* and the downstream *Rv0181c* homologue. This shows that the gene order of the *sigG* operon is not conserved across mycobacterial species. Images were obtained from MycoDB (http://myco.bham.ac.uk).

mycobacteria e.g. in *M. marinum* there is an additional gene *MMAR_0425* encoded between *sigG* and the *Rv0181c* homologue *MMAR_0424* (Figure 4.17) and this is also the case for *M. ulcerans.* In *M. avium* there is an insertion of three genes between *sigG* and the *Rv0181c* homologue (Figure 4.17). This lack of conservation of gene order in the *sigG* operon adds to the evidence that the Rv0181c protein is unlikely to interact with, and therefore regulate, SigG.

At this point a BLAST analysis of the Rv0181c protein sequence revealed that it had significant homology to the *E. coli* YhhW pirin protein. YhhW has been shown to break down the antioxidant, quercetin, and is therefore a detoxifying enzyme known as a quercetinase (Adams *et al.*, 2005). Due to the homology between YhhW and Rv0181c it is likely that Rv0181c is actually an enzyme, probably with a detoxifying role. Given this potential enzymatic role and the lack of an interaction between His-SigG and His-Rv0181c, it is unlikely that Rv0181c has a role as an anti-sigma factor to SigG. Furthermore, in a recently produced global protein-protein interaction network for *M. tuberculosis* H37Rv (Wang *et al.*, 2010) SigG and Rv0181c both appeared as having interaction partners, but noticeably no interaction was observed between the proteins encoded by the genes of the *sigG* operon.

In a recent publication by Cáceres *et al* (2011) Rv0180c was shown to be involved in host cell invasion. In addition, in *Mycobacterium leprae,* the *sigG* and *Rv0181c* homologues are pseudogenes whereas the *Rv0180c* homologue is predicted to still be functional (Mycobrowser). This suggests that Rv0180c does not have a role in the regulation of SigG. Taken together the results of this study and findings of recent publications indicate that Rv0181c and Rv0180c are not regulators of SigG.

In order to try to identify potential regulators of SigG, the recombinant His-SigG protein was used as bait in a pulldown experiment to pull down interacting proteins from *M. tuberculosis* CFE. Three proteins of interest were identified: the potential transmembrane proteins Rv0831c and Rv1084 and the transcriptional regulator MtrA. Of these Rv1084 is an ideal candidate for an anti-sigma factor. It is a predicted thioredoxin-like superfamily protein, meaning it contains a CXXC motif. There is a family of anti-sigma factors known as the zinc-associated anti-sigma (ZAS) factors that are characterised by the presence of a conserved HXXXCXXC motif (Paget *et al.*, 2001). In *M. tuberculosis* the anti-sigma factors RseA, RshA and RslA all belong to this family (Thakur *et al.*, 2010). Analysis of the protein sequence of Rv1084 shows it has the motif GXXXCXXC, with the histidine replaced by a glycine. Although this suggests this protein would not be a member of the ZAS family the precedent exists for *M. tuberculosis* anti-sigma factors containing a CXXC motif. This protein should therefore, be the first protein investigated as a potential anti-sigma factor to SigG in future work.

In order to confirm that the interactions observed here are true, candidate regulators of SigG could be investigated using purified protein to demonstrate a direct interaction with SigG and then further in *in vitro* transcription assays to determine whether they negatively or positively regulate transcription by the SigG holoenzyme. In addition a deletion strain of the candidate SigG regulator e.g. Rv1084 could be created and used to determine the effect this deletion has on expression of a SigG-dependent gene.

Although an anti-sigma factor was not identified, one of the initial aims of this study was to identify and confirm activity of an anti-sigma factor to SigG. One method for confirming whether a protein acts as a negative or positive regulator of a sigma factor is to use an *in vitro* transcription assay. However, this technique relies on the availability of a promoter, whose expression is dependent on the sigma factor of interest. Due to the fact that there has been very little investigation into the role of SigG previously, there were no confirmed SigG-regulated genes. In addition to requiring a SigG-dependent promoter for testing anti-sigma factor function, it was also desired for testing when the SigG protein was active i.e. under what conditions was SigG able to bind core RNAP and direct transcription from a SigG-dependent promoter.

Previous work in this group using the *E. coli* two-plasmid system had suggested that *Rv0654* was a SigG-dependent gene (R. Balhana and J. Smith, personal communication). 5' RACE determined that *Rv0654* was expressed from a single promoter in *E. coli*, suggesting that SigG and an *E. coli* sigma factor must both be able to recognise the same promoter. However, the expression level of *Rv0654* in a *M. tuberculosis* strain in which the *sigG* operon had been deleted (ΔsigGoperon) did not differ from that in the wild-type, showing that this promoter was not SigG-dependent.

Microarray analysis of a SigG overexpression strain highlighted two genes that were particularly highly upregulated compared to a vector only control: *Rv0887c* and *Rv0911*. 5' RACE gave two products for *Rv0887c* but none for *Rv0911*; however sequencing only identified one transcriptional start site for *Rv0887c.* The size of the

second product suggested that if there was a second transcriptional start site, it was in the coding region of Rv0887c. However, 5' RACE does not distinguish between full length mRNA and processed RNA molecules and it is therefore likely that the smaller product for *Rv0887c* is actually a processed form of the longer transcript. Despite not being able to obtain RACE products for *Rv0911*, alignment of the DNA sequence upstream of this gene to that upstream of *Rv0887c* revealed that the *Rv0887c* transcriptional start site fell in a region of high sequence homology between the two genes. It was therefore decided to use this same region as the promoter region for *Rv0911*. Initially *in vitro* transcription was used to determine whether these genes were transcribed from a SigG-dependent promoter. Only non-specific template read-through was obtained for the two genes. However, this could have been due to SigG either not binding to the polymerase or not binding to the promoter due to interference from the His tag. The earlier pulldown experiments using His-SigG as bait had pulled down RNA polymerase subunits from *M. tuberculosis* CFE suggesting that the His tag was not interfering with sigma factor binding to the polymerase. A protein-DNA binding assay was used to determine whether His-SigG was able to bind to the promoter regions of *Rv0887c* and *Rv0911*. Although a 'band shift' was observed with both the *Rv0887c* and *Rv0911* promoter fragments, the shift increased with increasing protein concentration suggesting a non-specific interaction and this was confirmed by the fact that His-SigG was able to bind to the non-specific DNA fragment of the *M. smegmatis rrn*B gene. It has been demonstrated that region 1.1 of *E. coli* sigma 70 inhibits sigma factor binding to DNA by preventing region 4 binding to the -35 promoter element and this inhibition is alleviated when sigma 70 is part of the RNAP holoenzyme. Deletion of region 1 allows sigma 70 to bind DNA in the absence of core RNAP (Camarero *et al.*, 2002, Dombroski *et al.*, 1992). ECF

sigma factors, such as SigG, do not contain region 1 (Figure 1.5) and it was therefore expected that these sigma factors would be able to bind DNA in the absence of RNAP. In this study SigG has been shown to bind the promoter regions of *Rv0887c* and *Rv0911* but this interaction was not specific. This could be due to the fact that SigG was used on its own and not as part of the RNAP holoenzyme. SigA of *Bacillus subtilis,* which contains region 1.1, has been shown to bind DNA in the absence of core RNAP, but with lower specificity than when core RNAP is present (Yeh *et al.*, 2011). In light of this, future work on the ability of SigG to bind the promoter regions of *Rv0887c* and *Rv0911* could use the SigG holoenzyme in order to determine whether SigG is able to specifically bind these promoter regions.

It is also possible that the N-terminal His tag was interfering with the specificity of DNA binding. Therefore, in future work new SigG protein could be purified, in which the tag can be cleaved to leave native protein. These genes were the most highly up-regulated genes in the SigG overexpression strain. They have been shown using qRT-PCR to be DNA damage-inducible and in a *sigG* deletion strain they are still expressed at near wild-type levels but are no longer induced by DNA damage (K. Smollett, personal communication). This corroborates the observation of the 5'RACE experiments conducted in this study, where two products were identified for *Rv0887c* and *Rv0911* suggesting transcription from two promoters. It suggests that basal level expression of *Rv0887c* and *Rv0911* is controlled by another *M. tuberculosis* sigma factor and that their expression is increased in a SigG-dependent manner following DNA damage.

With the advances in experimental techniques a new avenue that could be pursued in future would be to employ ChIP-Seq to identify potential SigG-dependent genes. This technique operates on the same principle as a pulldown experiment except that instead of pulling out interacting proteins, it is interacting DNA fragments that are obtained, amplified and then sequenced to determine the sites in the genome that the protein of interest is able to bind to (Fields, 2007).

5. Creation and phenotypic analysis of a whole operon deletion strain

5.1 Introduction

In bacteria DNA-dependent RNA polymerase combines with a dissociable sigma factor to form the holoenzyme, capable of specific promoter recognition and transcription initiation (Borukhov *et al.*, 2003). Sigma factors are responsible for global regulation of gene transcription as they are able to bind to the specific promoters of different genes. The most common bacterial adaptation to stress is to replace the primary sigma factor with an alternative sigma factor, enabling expression of a specific set of genes (Kazmierczak *et al.*, 2005). *M. tuberculosis* has thirteen σ^{70} class sigma factors, ten of which belong to the ECF family (Manganelli *et al.*, 2001). ECF sigma factors are thought to contribute to *M. tuberculosis* pathogenesis by enabling it to respond to the stresses faced during macrophage infection. An example is SigH, which induces the expression of a set of genes, including heat shock proteins, in response to oxidative and heat stress (Song *et al.*, 2003).

SigG has been shown to be the most upregulated sigma factor in response to DNA damage (Rand *et al.*, 2003) and it was originally thought to have a role in the regulation of the DNA damage response (Lee *et al.*, 2007). However, a recent study conducted in this laboratory has shown that this is not the case and that SigG and the genes of its operon are likely to form part of the RecA-independent regulon (Smollett *et al.*, 2011).

ECF sigma factors are often co-transcribed with one or more regulators, positive and negative (Hughes *et al.*, 1998) and in *M. tuberculosis* this is the case for SigE, SigF, SigH and SigL (Rodrigue *et al.*, 2006, Song *et al.*, 2003, Mecsas *et al.*, 1993). SigG is co-transcribed with the two downstream genes *Rv0181c* and *Rv0180c*.

In this study a deletion strain of the whole *sigG* operon was created with the initial aim of investigating the phenotype of this mutant but also by way of partial complementation, to assess the roles of *Rv0181c* and *Rv0180c* as potential regulators of *sigG*.

5.2 Creating a whole operon deletion strain

This study set out to create a deletion strain of the whole *sigG* operon in which the three genes were replaced with a gentamycin resistance cassette. Creation of a deletion strain relies on allelic exchange by which homologous recombination causes the chromosomal copy of a gene or genes to be replaced with a mutated one. The mutation is introduced using a non-replicating suicide vector (Smith, 2003). The strategy employed in this study uses a two-step process, whereby a single cross-over recombinant is isolated from an initial electroporation, followed by screening and counter-selection to identify cells in which the second recombination event has occurred. The inclusion of multiple markers such as *sacB* and *lacZ* on the knockout plasmid construct makes this screening and counter-selection possible and greatly increases the efficiency of creating deletion mutations in *M. tuberculosis* (Parish *et al.*, 2000).

5.2.1 Design and creation of the deletion plasmid

In this study the whole *SigG-Rv0180c* operon has been deleted and replaced with a gentamycin resistance cassette using a knockout construct made in *E. coli* via sequential cloning steps. In order for homologous recombination to occur between the deletion plasmid and the bacterial chromosome the plasmid must contain the 5' and 3' flanking regions of the gene or operon of interest. When designing the 5' flanking region, the first 146 bp of the *sigG* gene (from the annotated start site) were included to ensure that the promoter region of the divergent gene *Rv0183* was not affected. Likewise in the 3' flanking region the last 380 bp of the *Rv0180c* gene were included to try to avoid downstream effects on the expression of the *lprO* gene.

Primers were designed to amplify a 1390 bp 5' flanking region of *sigG* and a 1384 bp 3' flanking region of *Rv0180c* and in the process incorporate appropriate restriction enzyme sites into the ends of the PCR products.. PCR products were sequentially cloned into the *Sac*I/*Xba*I (5' region) and *Xba*I/*Bam*HI (3' region) sites of the pBackbone suicide plasmid (Gopaul, 2002). The 900 bp gentamycin resistance ($Gent^R$) cassette from the plasmid pUC-GM (Schweizer, 1993) was inserted into the *Xba*I site between the two flanking regions. Finally, the 6.4 kb *sacB*/*lacZ* cassette from pGoal17 (Parish *et al.*, 2000) was cloned into the *Pac*I site in the pBackbone part of the construct. This cassette allows selection of potential mutants via blue/white screening for the *lacZ* gene (encoding β-galactosidase) and via sucrose counter-selection through the *sacB* sucrose sensitivity gene (encoding levansucrase). The resulting deletion plasmid was named pAG05 and contained three selectable markers, $Gent^R$, kanamycin resistance (Kan^R) and *sacB* and the *lacZ* marker, which allows for blue/white screening of colonies.

5.2.2 Selection process for identifying single and double crossover strains

Selection of single (SCO) and double (DCO) crossovers relies on the presence or absence of the various markers in the targeting construct. SCOs have undergone a single recombination event. This means that they contain all of the markers present in the targeting construct and will therefore contain $Gent^R$, Kan^R, *sacB* and *lacZ*. DCOs have undergone a second recombination event and will therefore have lost the backbone of the targeting construct. These bacteria will contain $Gent^R$ only as a marker.

Initial electroporations of the pAG05 plasmid into *M. tuberculosis* H37Rv were grown on 7H11 agar containing gentamycin (gent) and the synthetic lactose analogue, X-Gal, which can be broken down by β-galactosidase to produce a blue colour. Following incubation blue colonies were potential SCOs or random integration strains and white colonies were potential DCOs or spontaneous gentamycin mutants. Blue colonies were analysed by PCR (as described below) to determine whether they were single crossovers. One confirmed single crossover was plated onto 7H11 agar containing gent to allow the second recombination event to take place. Colonies were serially diluted in modified Dubos medium and plated onto 7H11 agar containing gent, X-Gal and sucrose to select for loss of the pAG05 plasmid. The presence of the *sacB* gene in SCOs means that they are sensitive to the presence of sucrose in the media whereas DCOs are able to survive and remain white due to the loss of the *lacZ* marker. Potential DCOs from both the initial electroporation and the serial dilutions were replica plated onto 7H11 agar containing kanamycin (kan) and 7H11 agar containing gentamycin, XGal and

sucrose to confirm their phenotypes. White, gent resistant, sucrose resistant and kan sensitive colonies were used for further analysis by PCR.

Three PCR reactions were designed so that wild-type *M. tuberculosis* H37Rv, single crossover strains of either orientation and double crossover strains could be distinguished (Figure 5.1). Random integration strains would demonstrate a single crossover phenotype on the selection media but due to the fact that recombination had occurred in a random location, PCR of the operon of interest would show a wild-type genotype.

5.2.3 The *sigG* operon deletion strain was successfully created

Seven blue colonies, potential single crossovers, were obtained from the initial electroporation of pAG05 into *M. tuberculosis* H37Rv. InstaGene preparations of these colonies were analysed by PCR and five colonies were confirmed as single crossovers, one of which was used for serial dilution and replica plating. This yielded two white gent resistant, sucrose resistant and kanamycin sensitive colonies. InstaGene preparations of these two colonies were analysed by PCR, which confirmed that both were double crossover strains (Figure 5.2).

5.3 Complementing the ΔsigGoperon strain

5.3.1 The complementation plasmids, their genotypes and intended roles

The initial aim of this study as well as to characterise the phenotype of the ΔsigGoperon strain, was to assess the roles of the *Rv0181c* and *Rv0180c* genes as potential regulators of *sigG*. In order to confirm that any phenotypic changes

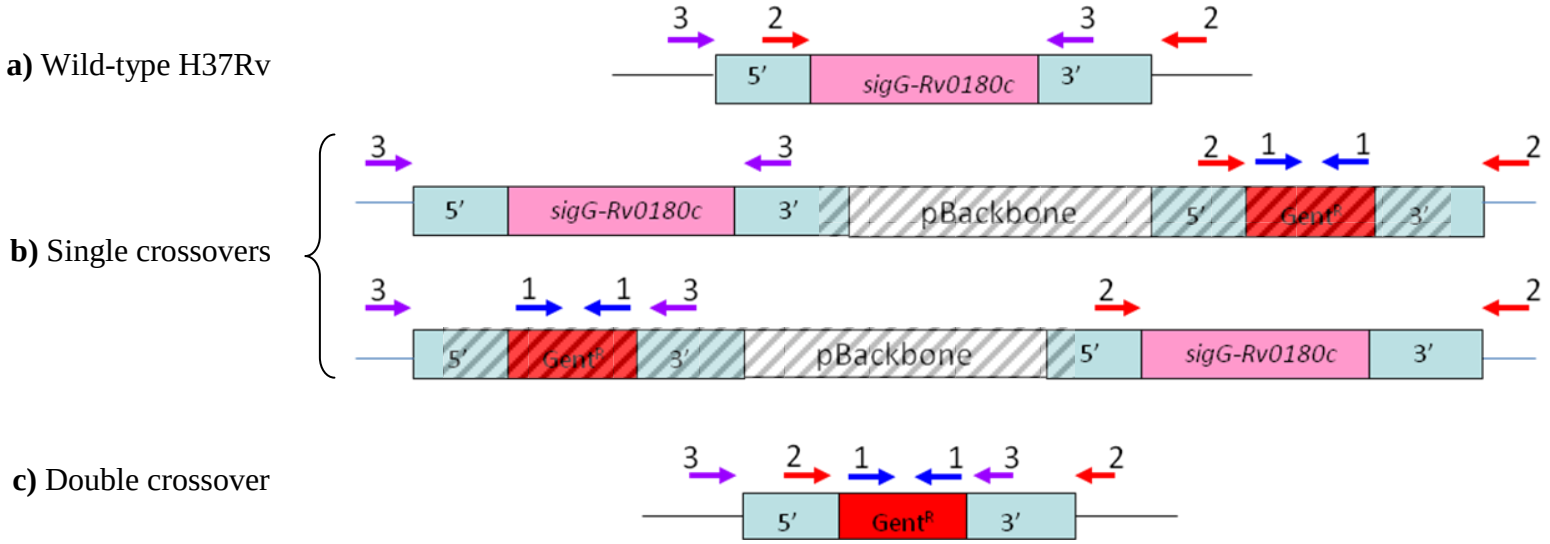
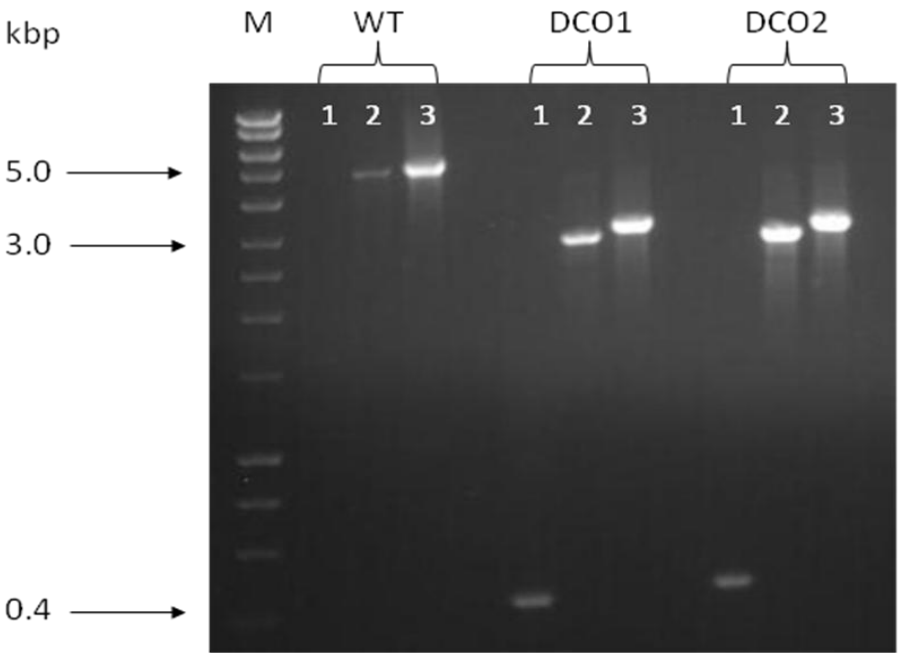


Figure 5.1 – Schematic representation of the PCR reactions used to screen *M. tuberculosis* single and double crossovers

Schematic diagrams illustrating the three PCR reactions used to distinguish between a) wild-type *M. tuberculosis* H37Rv, b) single crossover strains of either orientation and c) double crossover strains. The numbers 1-3 represent the primer pairs KO1F/R, KO2F/R and KO3F/R and the hashed sections represent the plasmid within single crossover strains. The region of *sig*G-*Rv0180c* knocked out was 2723 bp long and this was replaced by the 926 bp $Gent^R$ cassette so PCR reactions 2 and 3 would yield products ~1.8kb smaller if the mutant allele is present. The chromosomal location is specific as one primer for both PCR 2 and 3 was designed to anneal outside of the flanking regions. PCR 1 amplifies the gent cassette.



Expected Sizes (bp):

Wild-type	1	n/a	**DCO**	1	434
	2	4684		2	2887
	3	4891		3	3094

Figure 5.2 – PCR screening for the ΔsigGoperon deletion strain

PCR screening of potential *sigG-Rv0180c* double cross-over colonies 1 and 2 from the serial selection using PCR reactions 1-3 (see Figure 5.1) Lanes 1, 2 and 3 represent PCR reactions 1, 2 and 3 respectively. PCR 1 shows that both colonies contained the gent resistance cassette. For PCR 2 and 3 the size of the products expected if the mutant allele was present was ~1.8 kbp smaller than that for wild-type H37Rv. PCR results indicated that both colonies were double cross-over strains. (WT= wild-type H37Rv genomic DNA positive control; DCO = double crossover; M = Bioline Hyperladder I marker).

observed with the ΔsigGoperon deletion strain are due to the proteins being absent and not due to a secondary mutation elsewhere in the genome, it was necessary to construct a strain in which wild-type copies of all three genes have been inserted back into the genome. This strain should demonstrate the same phenotype as wild-type *M. tuberculosis* H37Rv. In addition, in order to investigate the roles of *Rv0181c* and *Rv0180c*, partial complements were designed, in which the genes of the operon were complemented in pairs or the *sigG* gene alone was complemented. Complementation plasmids had previously been constructed in the laboratory using the stable integrating vector pKP186 (Rickman *et al.*, 2005) to make a series of full and partial complementation plasmids, the genotypes of which are outlined in Table 5.1 (Dawson, 2005). In all complementation plasmids gene expression was driven by the native *sigG* promoters.

The creation of the whole operon (pLDcomp8T) and *sigG* only (pLDcompΔ1) plasmids would allow for the characterisation of phenotypic changes that were solely due to the absence of SigG. Phenotypic changes complemented by both plasmids would be due to the absence of SigG but if the SigG only complement failed to complement then this would demonstrate that the phenotype was due to the absence of either Rv0181c or Rv0180c. The partial complements pLDcompΔ2, 3 and 4 were originally created to assess the effect that removing the potential regulators of SigG had on expression of a SigG-dependent gene. The initial hypothesis was that should Rv0181c be an anti-sigma factor to SigG, a pLDcompΔ2 strain in which this negative regulator was missing, would demonstrate an increase in expression of a SigG-dependent gene. Should Rv0180c act as an anti-anti-sigma factor to SigG, a pLDcompΔ4 strain, lacking this positive regulator, would be expected to show

Table 5.1 – Complementing the ΔsigGoperon strain

Plasmid	Complement name	SigG	Rv0181c	Rv0180c	Deletion size
pLDcomp8T	Whole operon complement	✓	✓	✓	none
pLDcompΔ1	SigG only complement	✓	X	X	2094
pLDcompΔ2	ΔsigGoperon:SigG and Rv0180c	✓	X	✓	690
pLDcompΔ3	ΔsigGoperon:Rv0181c and Rv0180c	X	✓	✓	899
pLDcompΔ4	ΔsigGoperon:SigG and Rv0181c	✓	✓	X	1404

Table showing the genes complemented (✓) and still missing (X) in the full and partial complementation plasmids used to complement the ΔsigGoperon strain together with the names of the resulting complements. Expression was driven from the native *sigG* promoters.

decreased expression of a SigG-dependent gene. The pLDcompΔ3 complementation plasmid, which lacks the *sigG* gene, was created as a control to confirm the SigG-dependency of the gene being used to assess the anti- and anti-anti-sigma factor roles of Rv0181c and Rv0180c.

5.3.2 The whole operon and SigG only complements were successfully created

All five complementation constructs were electroporated into the *M. tuberculosis* ΔsigGoperon strain together with the plasmid pBS-Int (Springer *et al.*, 2001), which encodes the integrase enzyme required for integration of the complementation vector into the bacterial chromosome. At this time the protein-interaction results discussed in Chapter 4 and the findings of recent publications indicated that Rv0181c and Rv0180c were not regulators of SigG and therefore initially only the whole operon (ΔsigGoperon containing pLDcomp8T) and SigG only (ΔsigGoperon containing pLDcompΔ1) complement strains were carried forward.

To confirm that these plasmids complemented gene expression, RNA extracts from wild-type *M. tuberculosis* H37Rv, ΔsigGoperon, the whole operon and SigG only complement strains were analysed using qRT-PCR of the *sigG-Rv0180c* genes. All strains were also analysed for expression of *lprO* in order to determine if there were any downstream effects on expression of this gene caused by deleting the *sigG* operon. qRT-PCR reactions were conducted as per Section 2.5.3. Reverse transcriptase negative reactions (RT-) were run alongside test qRT-PCR reactions (RT+), cDNA values were adjusted for presence of contaminating genomic DNA by subtraction of RT- values from RT+ values and the expression values were then normalised to the equivalent *sigA* cDNA values to give a relative expression level for

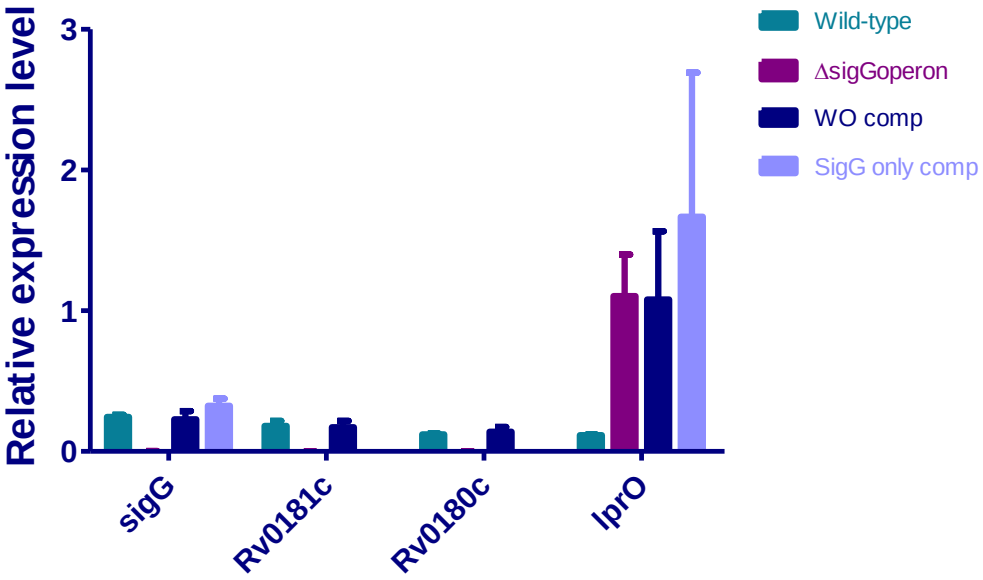


Figure 5.3 – qRT-PCR to assess complementation of the ΔsigGoperon strain
qRT-PCR (SYBR-Green) of the *sigG-Rv0180c* and *lprO* genes was performed on RNA extracts of wild-type *M. tuberculosis* H37Rv, the *M. tuberculosis* ΔsigGoperon deletion strain, the whole operon complementation strain (ΔsigGoperon containing pLDcomp8T) and the partial SigG only complementation strain (ΔsigGoperon containing pLDcompΔ1). Expression data for each gene was normalised to equivalent *sigA* data and columns and error bars represent the averages and standard deviations of three biological replicates, each assayed in triplicate. *sigG-Rv0180c* transcription was undetectable within the ΔsigGoperon deletion strain. No significant difference was observed between the relative expression level of the *sigG-Rv0180c* genes in the wild-type and full complementation strains (Unpaired t-test, $p>0.05$) or between the relative expression level of *sigG* in the wild-type and SigG only complement strain (Unpaired t-test, $p>0.05$) (Unpaired t-tests, $p>0.05$). Normalised *lprO* transcription levels were significantly increased compared to the wild-type strain in the ΔsigGoperon strain and both complementation strains (Unpaired t-test, $p<0.05$).

each gene. Normalisation allows for comparison of genes from different RNA preparations, while taking into account differences in the efficiency of the first strand cDNA synthesis. During exponential growth of *M. tuberculosis* the expression level of the *sigA* transcript has been shown to remain relatively constant and it is therefore accepted as a suitable gene for normalisation (Manganelli *et al.*, 1999).

qRT-PCR demonstrated that the three genes of the *sigG* operon were undetectable in the ΔsigGoperon strain, confirming the deletion of these genes (Figure 5.3). No significant difference was observed between the relative expression level of the *sigG-Rv0180c* genes in the wild-type and full complementation strains (Figure 5.3; Unpaired t-test, $p>0.05$). There was also no significant difference between the relative expression level of *sigG* in the wild-type and SigG only complement strain (Figure 5.3; Unpaired t-test, $p>0.05$). The relative expression level of the *lprO* gene was analysed in all strains to determine whether deletion of the *sigG* operon had downstream effects on this gene. Analysis showed that in ΔsigGoperon and both complement strains expression of this gene was significantly increased compared to wild-type (Figure 5.3; Unpaired t-tests, $p<0.05$). This suggests that either deletion of the *sigG* operon has removed a negative regulatory element of *lprO* or read-through from the gentamycin resistance cassette could be causing increased levels of transcription. Despite the fact that *lprO* expression has been affected by deletion of the *sigG* operon, this effect was still present in the complement strains. Therefore, if a phenotypic change was observed that was due to the overexpression of *lprO* rather than deletion of the *sigG* operon, this phenotype would not be complemented by either of the complement strains.

5.4 Initial attempts to characterise the phenotype of the ΔsigGoperon strain

Initial phenotypic work was conducted using wild-type *M. tuberculosis* H37Rv, ΔsigGoperon and the whole operon complement strain to first attempt to identify a phenotype.

Microarray analysis is often employed to screen for expression differences between a deletion strain and wild-type. However, microarray analysis comparing a ΔsigG strain to wild-type *M. tuberculosis* H37Rv has previously been conducted in this laboratory under exponential growth phase and DNA-damaging conditions. No differences in gene expression were observed between the two strains under either of these growth conditions (Smollett *et. al.,* 2011). As such it was decided that microarray analysis of the ΔsigGoperon strain was unlikely to yield a phenotype for SigG and alternative methods were used.

5.4.1 *In vitro* phenotype analysis of ΔsigGoperon compared to wild-type

5.4.1.1 Deletion of the *sigG* operon had no effect on the *in vitro* growth of *M. tuberculosis*

The effect of the deletion of the *sigG* operon on *in vitro* growth was assessed by comparing the growth of the ΔsigGoperon strain to wild-type *M. tuberculosis* H37Rv and the whole operon complement strain. Cultures were grown to mid-exponential phase (OD_{600nm} ~0.5) and were used to inoculate fresh modified Dubos media (refer to Appendix I) to a calculated OD_{600nm} 0.005. The OD_{600nm} was then monitored over a period of 15 days. No significant difference was observed between the doubling times of the three strains (Figure 5.4; One-way ANOVA with Dunnett post test, $p>0.05$) showing that deletion of the *sigG* operon has no effect on the ability of *M. tuberculosis* to grow *in vitro*.

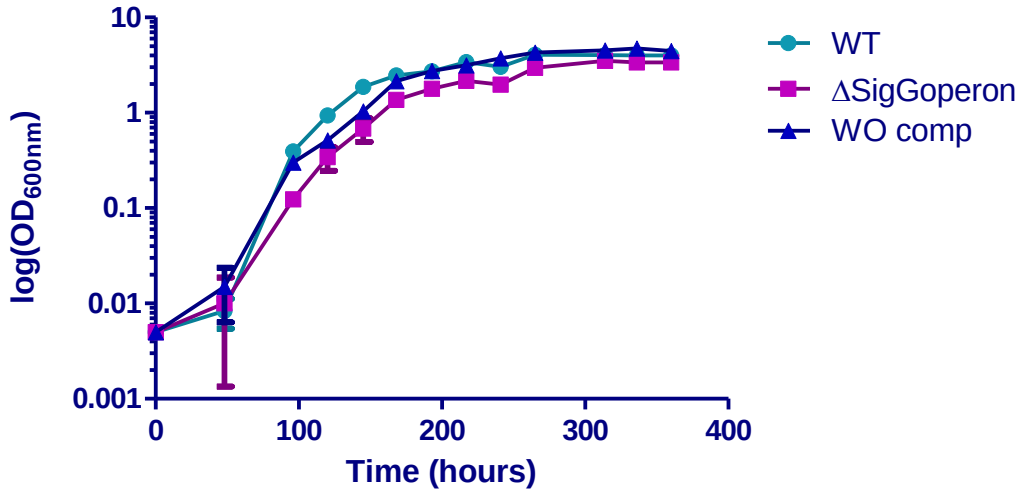


Figure 5.4 – *In vitro* growth curve of the ΔsigGoperon deletion strain

In vitro growth curves of wild-type *M. tuberculosis* H37Rv, the *M. tuberculosis* ΔsigGoperon deletion strain and the whole operon complementation strain. Strains were grown in modified Dubos medium (Appendix I). The data represents the averages and standard deviations of three biological replicates. There were no significant differences between the doubling times of the ΔsigGoperon and whole operon complement strains (17.6 hrs and 15.9 hrs respectively) when compared with the wild-type strain (12.3 hrs; One-way ANOVA with Dunnett post test, $p>0.05$).

5.4.1.2 Deletion of the *sigG* operon did not affect susceptibility of *M. tuberculosis* to a variety of stresses in an alamar blue cell viability assay

The alamar blue assay provides a homogeneous, fluorescent method for monitoring cell viability following exposure to stress reagents. The assay is based on the ability of living cells to convert a redox dye (resazurin) into a fluorescent end product (resorufin). Nonviable cells rapidly lose metabolic capacity and thus do not generate a fluorescent signal. The procedure involves adding the dye directly to cells in culture medium and after an incubation step, data are recorded using a plate-reading fluorometer. The alamar blue assay has been shown to be comparable to alternative methods such as BACTEC 460, for high-throughput screening of compounds against *M. tuberculosis* (Collins *et al.*, 1997).

Recent work in this laboratory has investigated the susceptibility of a ΔsigG deletion strain to three DNA-damaging agents, mitomycin C, bleomycin and ofloxacin, using the alamar blue assay. No significant difference was observed in the MIC_{90} (defined as the lowest drug concentration that caused a 90 % inhibition of growth) and IC_{50} (defined as the midpoint of a dose response curve) values of wild-type *M. tuberculosis* and ΔsigG strains (Smollett *et al.*, 2011).

This study aimed to identify stress reagents to which, the ΔsigGoperon strain showed increased inhibition compared to wild-type *M. tuberculosis* and to then confirm using the whole operon and SigG only complements that this was due to the absence of the *sigG* gene. Initially, selection of stress reagents to test was based on the hypothetical roles of the *Rv0181c* gene and the potential SigG-dependent genes *Rv0887c* and *Rv0911*.

Bioinformatic analysis of the Rv0181c amino acid sequence revealed that it had 37.3% sequence homology to the *E. coli* YhhW pirin protein. YhhW has been shown to be a 2,3-dioxygenase or quercetinase that degrades the antioxidant flavanoid compound quercetin and related compounds (Adams *et al.*, 2005). Due to the fact that *Rv0181c* is part of the *sigG* operon it was hypothesised that, should *Rv0181c* act as a quercetinase, the ΔsigGoperon strain may show an increased susceptibility to flavanoid compounds. Two flavanoids, quercetin and butein, were selected for use in the alamar blue assay.

Microarray analysis of a SigG overexpression strain had identified two genes that were the most up-regulated compared to wild-type *M. tuberculosis*, *Rv0887c* and *Rv0911* (K. Smollett, personal communication). As discussed in Chapter 4, bioinformatic analysis of the Rv0887c and Rv0911 amino acid sequences revealed that these proteins contained domains belonging to the diverse glyoxalase I, bleomycin resistance and Type I extradiol dioxygenase protein family. As bleomycin had previously been tested against a ΔsigG deletion strain of *M. tuberculosis*, with no effect; this study focused on the glyoxalase and dioxygenase properties of this diverse protein family. Type I extradiol dioxygenases catalyse the cleavage of aromatic rings and are therefore key to the degradation of aromatic compounds (Eltis *et al.*, 1996). Interestingly quercetinase also falls into this family of proteins (Iacazio, 2005) adding to the decision to include the flavanoid compounds in the alamar blue assay. Glyoxalase I is the glutathione-dependent enzyme responsible for the first step of the glyoxalase system, whereby the toxic metabolic intermediate, methylglyoxal, is converted to pyruvate (Inoue *et al.*, 1995). Should *sigG* play a role in regulation of the glyoxalase system, it follows that the ΔsigGoperon strain may show increased

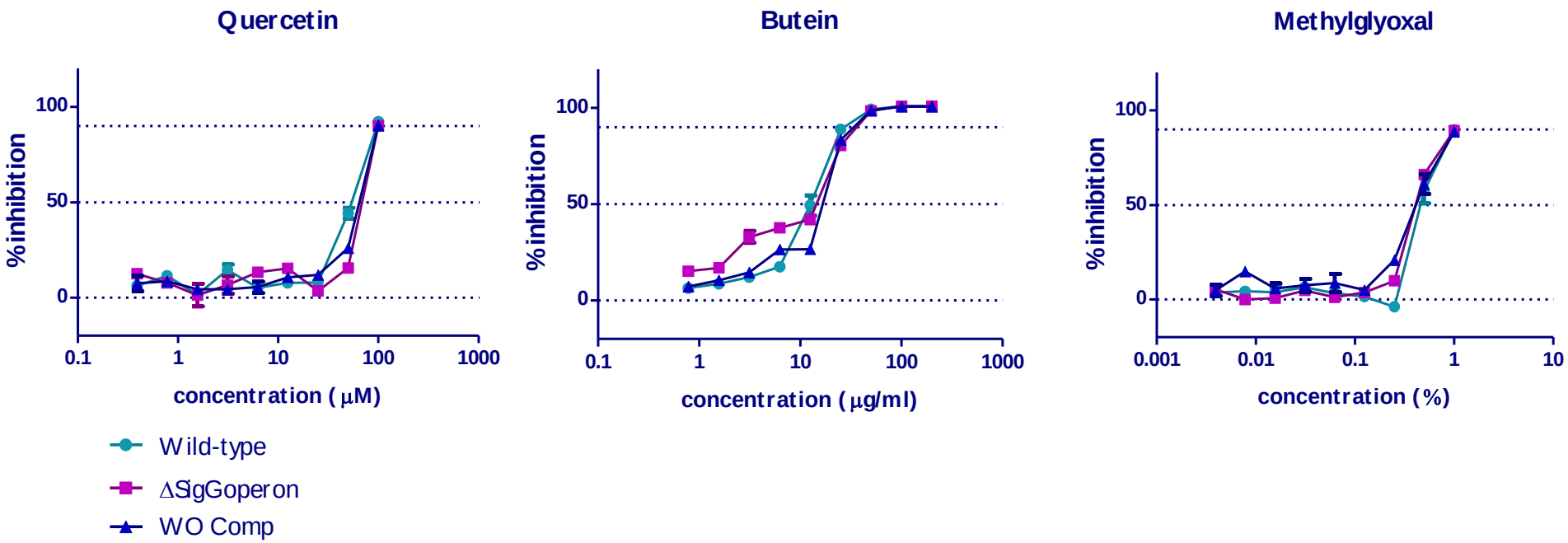


Figure 5.5 – Assessing the susceptibility of the ΔsigGoperon deletion strain to quercetin, butein and methylglyoxal using the alamar blue cell-viability assay

Wild-type *M. tuberculosis* H37Rv, ΔsigGoperon and whole operon complement strains were analysed for their susceptibility to increasing concentrations of quercetin, butein and methylglyoxal in the alamar blue cell-viability assay. No difference was observed between the dose response curves of the three strains in the presence of any of the three compounds. No difference was observed between the IC_{50} and MIC_{90} values of the three strains in response to any of the compounds (see Table 5.2).

Table 5.2 – MIC_{90} and IC_{50} values determined via alamar blue cell-viability assay

	Wild-type		ΔsigGoperon		Whole Operon comp	
Drug	MIC	IC_{50}	MIC	IC_{50}	MIC	IC_{50}
Quercetin (μM)	100	55.67	100	~62.0	100	61.83
Butein (μg/ml)	50	13.13	50	13.20	50	16.29
Methylglyoxal (% w/v)	>1.0	0.50	>1.0	0.44	>1.0	0.45

MIC_{90} was defined as the lowest drug concentration to cause at least 90% inhibition. IC_{50} was defined as the midpoint of a dose response curve and was determined using GraphPad Prism 5.00 except in the case of quercetin for the ΔsigGoperon strain, where the program was unable to assign an IC_{50} and so this value was approximated directly from the dose response curve (see Figure 5.4).

susceptibility to methylglyoxal due to an inability to convert this metabolite to pyruvate.

Viability of wild-type *M. tuberculosis* H37Rv, ΔsigGoperon and the whole operon complement strains in response to different concentrations of quercetin, butein or methylglyoxal was examined using the alamar blue microplate assay. No difference was observed in the dose response curves produced (Figure 5.4) or between the MIC_{90} and 50% inhibitory concentration (IC_{50}) for the three strains (Table 5.2) with any of these reagents.

5.4.2 *In vivo* analysis of ΔsigGoperon compared to wild-type *M. tuberculosis*

The fact that no *in vitro* phenotype was discovered lead to a shift to investigating the *in vivo* phenotype. *M. tuberculosis* is an intracellular pathogen and the mechanisms by which it evades the host immune response are not completely understood. It stands to reason that genes that might not be required for growth *in vitro* in synthetic media, may play an important role during infection. SigG has been shown to be upregulated during macrophage infection (Cappelli *et al.*, 2006), suggesting that it may have a role during infection.

5.4.2.1 Deletion of the *sigG* operon had no effect on *M. tuberculosis* survival *in vivo* within naive or activated macrophages

To investigate the role of *sigG* during infection, wild-type *M. tuberculosis* H37Rv, the ΔsigGoperon deletion strain and the whole operon complement strain were used

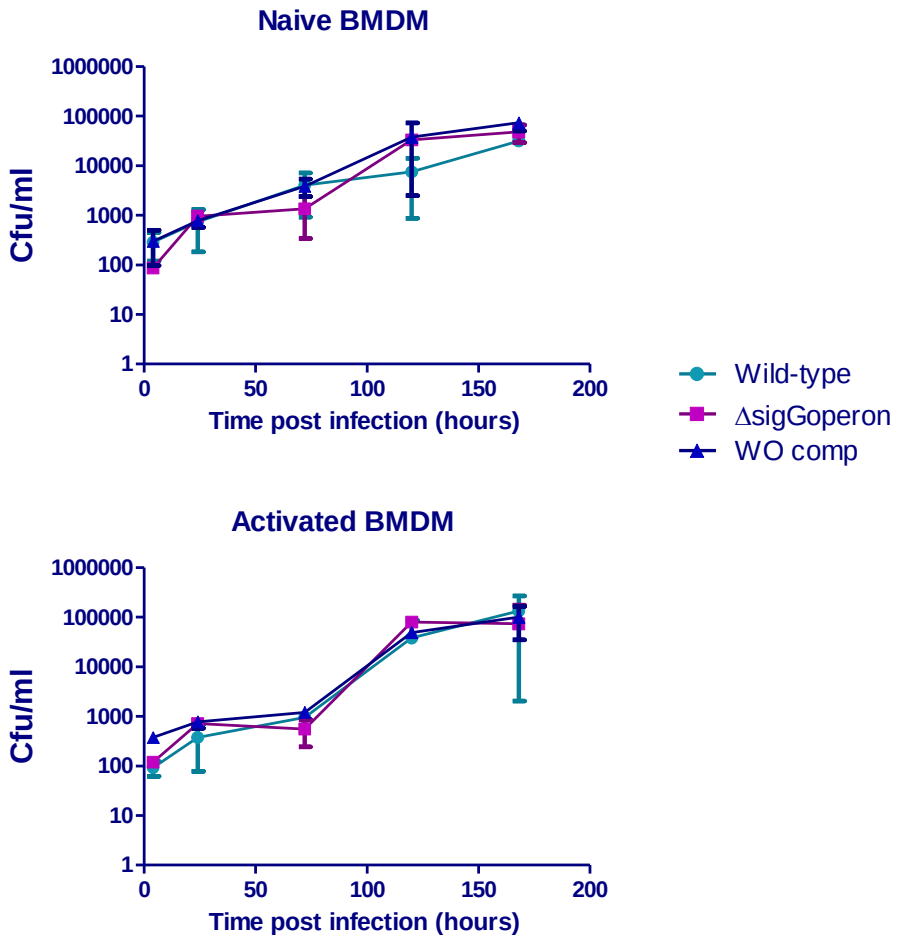


Figure 5.6 – Survival of ΔsigGoperon in the macrophage infection model
Survival of wild-type *M. tuberculosis* H37Rv, ΔsigGoperon and whole operon complement strains within naive and activated BMDM over a time-course of infection. There were no significant differences between the survival of either the deletion or complement strains within naive or activated macrophages when compared with the wild-type strain (One-way ANOVA with Dunnett post test, $P>0.05$). The data points and error bars represent the averages and standard deviations of triplicate infections.

to infect naive and activated murine BMDM (as per Section 2.7.3). Naïve macrophages and macrophages activated with 20 ng/ml IFN-γ were infected with a MOI of 0.1:1 (bacteria:macrophages). Infection was allowed to proceed for 7 days, during which time, the quantity of *M. tuberculosis* surviving within the macrophages was enumerated at five time points. Enumeration of the bacteria surviving within naive and activated BMDM macrophages showed that there was no significant difference between the ΔsigGoperon strain or the whole operon complement strain and wild-type *M. tuberculosis* H37Rv (Figure 5.6; One-way ANOVA with Dunnett post test, $p>0.05$). The fact that the ΔsigGoperon strain was not attenuated in naive or activated macrophages compared to wild-type suggests that it does not have a role in pathogenesis during infection. It is however, important to remember that the macrophage does not fully replicate the conditions encountered during infection.

5.4.2.2 Deletion of the *sigG* operon had no effect on the ability of *M. tuberculosis* to survive during a mouse model of infection

In order to provide a more complete model of *M. tuberculosis* infection, wild-type *M. tuberculosis* H37Rv, ΔsigGoperon, the whole operon complement and in this case the SigG only complement were used to infect BALB/C mice intravenously (as per Section 2.7.4). Briefly, mice were infected with 1×10^6 cells and infection was allowed to progress for 13 weeks, during which time, lungs and spleens were harvested for enumeration of *M. tuberculosis* at five different time points. No significant difference was observed between numbers of bacteria in the lungs and spleen for the ΔsigGoperon and wild-type *M. tuberculosis* H37Rv strains at any time point (Figure 5.7; One-way ANOVA with Dunnett post test, $p>0.05$). In the lungs differences between the two complement strains and wild-type were observed at the

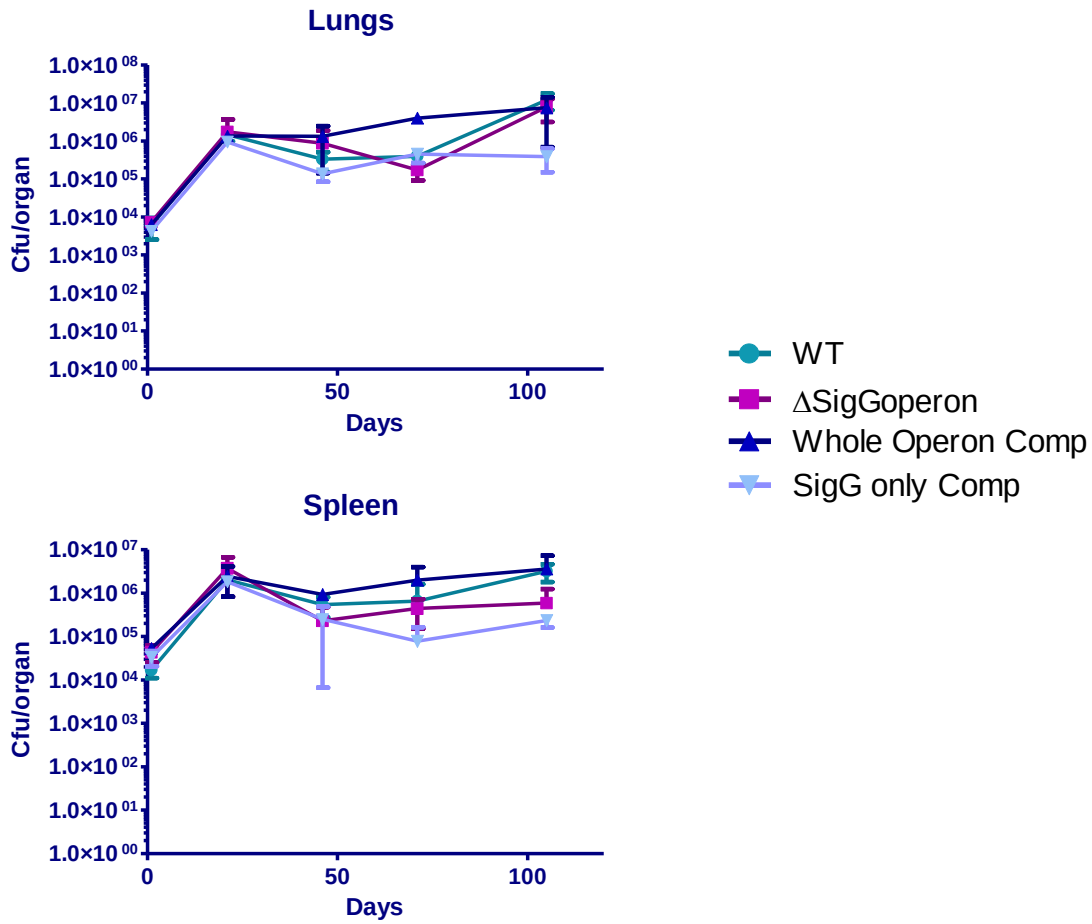


Figure 5.7 – Survival of ΔsigGoperon in the murine infection model

Survival of wild-type *M. tuberculosis* H37Rv, ΔsigGoperon and the whole operon and SigG only complement strains within the lungs and spleens of BALB/C mice over a time-course of infection. There were no significant differences between the survival of either the deletion or complement strains within the lungs and spleens of BALB/C mice when compared with the wild-type strain (One-way ANOVA with Dunnett post test, $p>0.05$). The data points and error bars represent the averages and standard deviations from four mice per timepoint.

later time points, however, this was not observed in the spleen. Due to the fact that no difference was observed with the ΔsigGoperon deletion strain it is likely that errors in enumerating the bacteria may have caused the differences with the complements. The result of the mouse infection suggests that SigG does not play a role during infection and pathogenesis. It is important to note however, that the mouse model does not fully represent the conditions encountered by *M. tuberculosis* during a human infection due to the inability of the mouse to form granulomas. It is therefore still possible that SigG has a function during granuloma formation or bacterial survival within granulomas that we were unable to model in this study.

5.5 A phenotype microarray identified a potential phenotype for ΔsigGoperon

The development of a microbial phenotype microarray (Biolog) provided a platform by which the ΔsigGoperon strain could be tested for nearly 2000 cellular phenotypes. Biolog preconfigures a wide range of phenotypic tests into sets of arrays. Each well of the array is designed to test a different phenotype. A single strain cell suspension is inoculated into the wells of the microarray, thereby testing thousands of phenotypes at once. Phenotype microarrays use redox chemistry (Biolog patented), employing cell respiration as a universal reporter. If the phenotype is strongly positive in a well, the cells respire actively, reducing a tetrazolium dye and forming a strong purple colour. If it is weakly positive or negative, respiration is slowed or stopped, and less or no colour is formed. The endpoint colour for each well after incubation can then be measured by reading the absorbance at 570 nm using a microplate reader.

Table 5.3 – Phenotype microarray carbon source data comparing the ΔsigGoperon deletion strain to wild-type *M. tuberculosis* H37Rv

PM	Compound	WT			ΔsigGoperon			Fold Change				p-value
		1	2	3	1	2	3	1	2	3	Average	
1	**D5 Tween 40**	**0.509**	**0.530**	**0.371**	**0.197**	**0.182**	**0.174**	**0.386**	**0.343**	**0.470**	**0.400**	**0.028**
1	**E5 Tween 80**	**0.450**	**0.470**	**0.517**	**0.194**	**0.283**	**0.279**	**0.430**	**0.602**	**0.539**	**0.524**	**0.005**
4	E8 Methylene Diphosphonic Acid	0.171	0.199	n/a	0.059	0.054	n/a	0.345	0.272	n/a	0.308	0.061
4	C7 6-Phospho-Gluconic Acid	0.100	0.150	n/a	0.077	0.050	n/a	0.774	0.332	n/a	0.553	0.203
1	G12 L-Malic Acid	0.093	0.104	0.095	0.105	0.173	0.170	1.134	1.656	1.783	1.525	0.141
1	B12 L-Glutamic Acid	0.082	0.078	0.049	0.109	0.081	0.119	1.328	1.035	2.414	1.592	0.100
2	F11 D-Tartaric Acid	0.120	0.027	n/a	0.114	0.068	n/a	0.955	2.539	n/a	1.747	0.774
2	F4 Oxalic Acid	0.088	0.114	n/a	0.103	0.275	n/a	1.177	2.418	n/a	1.798	0.488
2	D11 δ-Amino Valeric Acid	0.022	0.015	n/a	0.068	0.047	n/a	3.144	3.174	n/a	3.159	0.143

Values represent absorbance at 570nm normalised to a negative control well and Fold Change was calculated by dividing the normalised absorbance for the ΔsigGoperon deletion strain by that of wild-type *M. tuberculosis* H37Rv (WT). Columns 1-3 represent replicate experiments. p-values were calculated using a Student's t-test to compare the absorbance values of the WT strain and ΔsigGoperon deletion strain. Compounds with a p value <0.05 are marked in bold.

The ΔsigGoperon strain was tested against phenotype microarrays for carbon, phosphorous and sulphur source utilisation (PMs 1, 2 and 4, Appendix III) compared to wild-type *M. tuberculosis* H37Rv (as per Section 2.7.1). The fold change for each substrate was calculated by dividing the normalised absorbance for ΔsigGoperon by that of the wild-type strain tested in parallel. Therefore, a fold change of <1.0 represents phenotypes where the mutant strain demonstrated a lower level of respiration than the wild-type whereas a fold change >1.0 represents phenotypes where the wild-type strain demonstrated a lower level of respiration than the mutant strain. A fold change was considered significant if its value was <0.6 or >1.5 (full results in Appendix IV). This technique proved highly variable in the results produced over three (PM 1) or two (PM 2 and PM4) replicate experiments, however, two compounds were observed to produce a significant phenotype in all experiments: Tween 40 and Tween 80 (Table 5.3; Students t-test, $p<0.05$). In addition several organic acids were identified in the microarray but not as having a significant phenotype. When tested for phosphorous source use (PM4) the acids methylene diphosphonic acid and 6-phosphogluconic acid, on average, were better utilised by the wild-type strain than the ΔsigGoperon deletion strain. In comparison when tested for carbon source use (PM1 and PM2), the ΔsigGoperon deletion strain was, on average, better able to utilise several carboxylic acids e.g. malic and oxalic acids (Table 5.3). In particular 6-phosphogluconic acid and malic acid can act as intermediates in the pentose phosphate and TCA cycle respectively and are therefore intermediates of central metabolism. The results appeared to be leading towards a role for SigG in coping with carboxylic/fatty acids, especially as the two compounds that gave a significant phenotype, Tween 40 and Tween 80, are esters of the fatty acids palmitic and oleic acid respectively. In order to determine whether the

Table 5.4 – Phenotype microarray carbon source data comparing ΔsigGoperon to the SigG only complement for Tween 40 and Tween 80

ΔsigGoperon/SigG only complement fold change		
Plate	**Compound**	**Fold Change**
1	D5 Tween 40	0.52
1	E5 Tween 80	0.97

This data set is from one experiment

phenotype observed with Tween 40 and Tween 80 was due to SigG, the carbon source phenotype microarrays were repeated comparing the ΔsigGoperon strain to the SigG only complement strain. Should the phenotype be due to the deletion of *sigG*, the SigG only complement would behave like wild-type and a fold change would be observed representing attenuation of ΔsigGoperon. However, if the phenotype was due to the absence of either *Rv0181c* or *Rv0180c*, no difference would be observed between the two strains and the fold change would be ~1.0. ΔsigGoperon still showed attenuation with Tween 40 compared to the SigG only complement strain but not with Tween 80 suggesting that the phenotype with Tween 80 may be due to one of the other genes in the *sigG* operon (Table 5.4).

There is strong evidence to suggest that pathogenic Mycobacteria primarily use fatty acids rather than carbohydrates as carbon substrates during infection (Munoz-Elias *et al.*, 2005). In addition the toxicity of free fatty acids has been documented as far back as when the culture media for Mycobacteria were being defined (Dubos *et al.*, 1947).

Although the phenotype microarrays used here were designed for testing carbon source utilisation there were two possible explanations for the results seen. Either SigG could be involved during growth on fatty acids as a carbon source or in coping with the toxic effects of free fatty acids.

5.6 Deletion of the *sigG* operon does not affect growth on fatty acids as the sole carbon source

During infection it is thought that mycobacteria use fatty acids as their primary substrate. This is supported by the fact that respiration of *M. tuberculosis* in mouse lungs is strongly stimulated by fatty acids but not carbohydrates (Bloch *et al.*, 1956) and also the fact that several enzymes of the glycolytic pathway appear to be non-essential for growth and persistence in mice (Sassetti *et al.*, 2003). There are two pathways that are specifically required for growth on fatty acids, the catabolic β-oxidation cycle, which degrades fatty acids to acetyl-CoA and when glycolytic substrates are in low abundance, the glyoxylate shunt (McKinney *et al.*, 2000). The glyoxylate shunt is essential for preventing the loss of carbon molecules via the tricarboxylic acid (TCA) cycle during growth on fatty acids (Figure 1.3). Extensive duplication of genes encoding β-oxidation enzymes exists in *M. tuberculosis* (Cole *et al.*, 1998) and this creates an obstacle for studying genes in this system due to functional redundancy between paralogous genes. If SigG is involved in regulating β-oxidation genes the functions of these genes are highly likely to be duplicated, meaning that no phenotype would be observed for the ΔsigGoperon strain. The glyoxylate shunt however, is limited to three genes encoding malate synthase and the two isocitrate lyase genes (Munoz-Elias *et al.*, 2005). A phenotype for ΔsigGoperon growth on fatty acids is therefore more likely to be observed should SigG have a role in regulating the glyoxylate shunt. In order to investigate the role of SigG during growth on fatty acids, *M. tuberculosis* wild-type, ΔsigGoperon, whole operon and SigG only complement strains were cultured on four different carbon sources; glycerol as a control substrate, two short-chain fatty acids, acetate (C2), which forces the use of the glyoxalate shunt and α-ketoglutarate (C5), which is an intermediate of

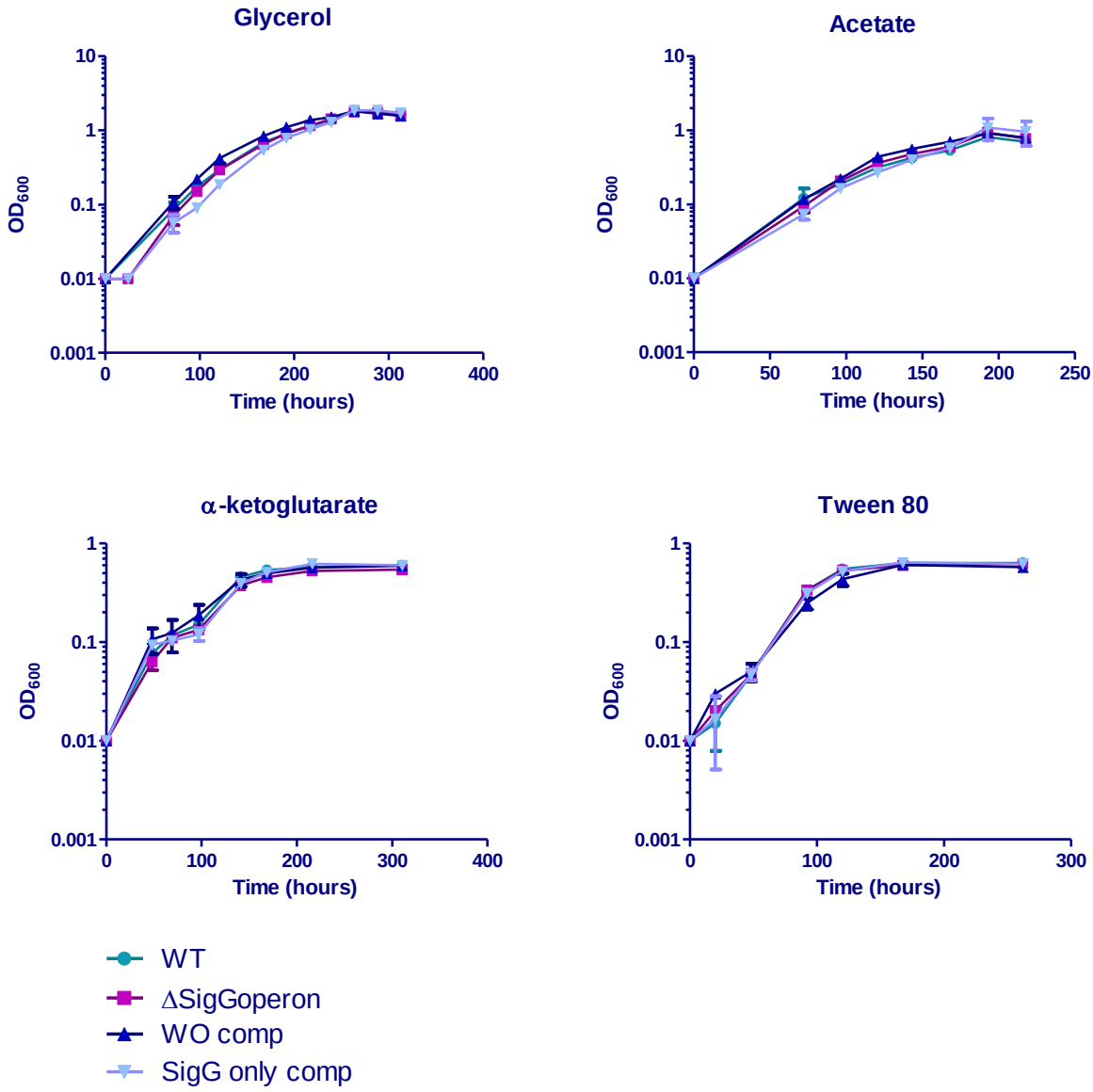


Figure 5.8 – Growth of ΔsigGoperon on fatty acid carbon sources

In vitro growth curves of wild-type *M. tuberculosis* H37Rv, the *M. tuberculosis* ΔsigGoperon deletion strain, the whole operon and SigG only complementation strains using sole fatty acid carbon substrates. Strains were grown in 7H9 medium supplemented with glycerol (control), acetate, α-ketoglutarate or Tween 80 (Appendix I). The data represents the averages and standard deviations of three biological replicates. There was no significant difference between the growth of the four strains on any of the carbon sources tested.

the TCA cycle and Tween 80 (C64). In order to prevent interference from Tween 80, which is a constituent of modified Dubos medium that can be hydrolysed by *M. tuberculosis*, 7H9 medium containing the non-hydrolysable Tween 80 substitute, tyloxapol was used as Marrero (2010).

The deletion of the *sigG* operon had no effect on bacterial replication kinetics in media containing glycerol (control), the short-chain fatty acids acetate and α-ketoglutarate, or the long-chain fatty acid ester, Tween 80 (Figure 5.9). α-ketoglutarate is an intermediate of the part of the TCA cycle that would normally be bypassed by the glyoxylate shunt (Figure 1.3).

The fact that the ΔsigGoperon strain was not attenuated compared to wild-type *M. tuberculosis* H37Rv during growth on fatty acids as the sole carbon source does not necessarily mean that SigG plays no role during fatty acid metabolism. These data suggest that it is unlikely that SigG has any role in regulating the glyoxylate shunt; however, it could be involved in regulation of β-oxidation genes. As previously stated the extensive duplication of these genes in the *M. tuberculosis* genome means that a high degree of functional redundancy exists within this system whereby the deletion of SigG and the resulting down-regulation of its regulon would be compensated for. An alternative explanation was that the results obtained using the phenotype microarrays were not due to the inability of ΔsigGoperon to utilise Tween and fatty acid compounds as carbon sources but that this strain was more susceptible to the toxic effects that free fatty acids exert on the cell and this was investigated further.

5.7 Assessment of the susceptibility of the ΔsigGoperon deletion strain to the presence of fatty acids using the alamar blue cell viability assay

5.7.1 The ΔsigGoperon deletion strain was more susceptible to the presence of oleic and linoleic acids in modified Dubos medium

Initial tests to determine the susceptibility of the ΔsigGoperon strain to fatty acids were conducted using the alamar blue cell viability assay using the standard modified Dubos medium, which contained both bovine serum albumin and Tween 80 (Appendix I). The presence of Tween 80 in the medium meant that in this initial screen, Tween 80 toxicity was not tested due to the potential for interference from the Tween 80 present in the medium.

Wild-type *M. tuberculosis* H37Rv, ΔsigGoperon and whole operon complement strains were assessed for their susceptibility to the short-chain saturated fatty acids acetic (C2), valeric (C5), hexanoic (C6) and nonanoic (C9) acid and the unsaturated long-chain fatty acids palmitic (C16), linoleic (C18) and oleic (C18) acid (Figure 5.10). The three strains were incubated at 37 ºC for seven days in the presence of different concentrations of fatty acid and then analysed for their ability to reduce the cell-titer blue dye by measuring the fluorescence. Dose response curves were plotted for each fatty acid and used to determine the MIC_{90} and IC_{50} values for each strain. A difference in the dose response of the ΔsigGoperon deletion strain compared to wild-type *M. tuberculosis* that was complemented in the whole operon complement strain was only observed with the two long-chain fatty acids, linoleic and oleic acid (Figure 5.10).

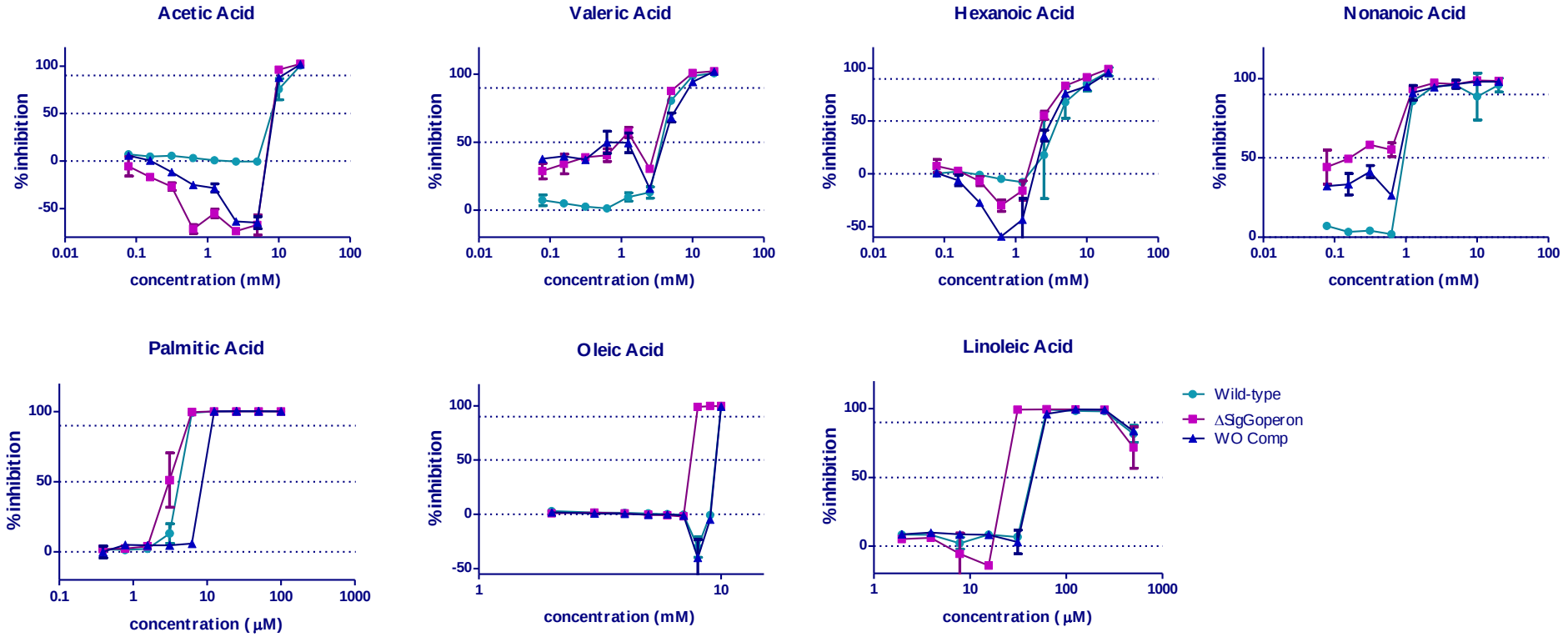


Figure 5.9 – Susceptibility of ΔsigGoperon to fatty acids compared to wild-type *M. tuberculosis* and the whole operon complement in modified Dubos medium

Alamar blue cell-viability assay results assessing the susceptibility of the ΔsigGoperon strain to a variety of fatty acids compared to wild-type *M. tuberculosis* and the whole operon complement strain in modified Dubos medium. Only the unsaturated fatty acids linoleic and oleic acid demonstrated a difference in susceptibility between ΔsigGoperon and wild-type that was complemented in the whole operon complement.

Table 5.5 – MIC_{90} and IC_{50} values for fatty acid susceptibility as determined by alamar blue cell-viability assay

	Wild-type		ΔsigGoperon		Whole Operon complement	
Drug	MIC_{90}	IC_{50}	MIC_{90}	IC_{50}	MIC_{90}	IC_{50}
Acetic acid (mM)	20	9.552	10	9.498	20	9.479
Valeric acid (mM)	10	3.861	10	1.909	10	2.671
Hexanoic acid (mM)	20	4.308	10	2.483	20	3.927
Nonanoic acid (mM)	2.5	1.046	1.25	0.2170	1.25	0.5409
Palmitic acid (mM)	6.25	4.100	6.25	3.102	12.5	8.358
Oleic acid (mM)	10	9.746	8	7.787	10	9.779
Linoleic acid (μM)	62.5	23.4	31.25	46.8	62.5	23.4

MIC_{90} was defined as the lowest drug concentration to cause at least 90% inhibition. IC_{50} was defined as the midpoint of a dose response curve and was determined using GraphPad Prism 5.00 except in the case of linoleic acid, where the program was unable to assign an IC_{50} and so this value was approximated directly from the dose response curve (see Figure 5.10). There was no significant difference between the MIC_{90} and IC_{50} values or the three strains (F test, $p>0.05$).

However, the MIC_{90} and IC_{50} values of the three strains for linoleic and oleic acid were not significantly different ($p>0.05$, F test). An extensive literature search provided a possible explanation. When the media for growth of tubercle bacilli were being defined several independent researchers noted that Tween 80 exerted a bacteriostatic effect on the growth of *M. tuberculosis*, which was attributed to both the presence of contaminating free oleic acid in the Tween 80 preparation and to the fact that during incubation at 37 ºC and in the presence of bacterial enzymes, Tween 80 hydrolyses to release free oleic acid (Dubos *et al.*, 1947, Sattler *et al.*, 1948). This bacteriostatic effect of free oleic acid was shown to be alleviated by the presence of bovine serum albumin in the medium (Davis *et al.*, 1947). Modified Dubos medium contains bovine serum albumin and therefore this could have been preventing the fatty acids tested from exerting their full effect. A new medium was defined for testing the susceptibility of *M. tuberculosis* strains to fatty acids using the alamar blue cell-viability assay. The new fatty acid susceptibility medium was based on modified Dubos medium, in which the bovine serum albumin was removed and Tween 80 was replaced by the non-hydrolysable tyloxapol reagent (Section 2.1 and Appendix I).

5.7.2 ΔsigGoperon was more susceptible to the presence of oleic acid and Tween 80 but was not complemented in the SigG only complement strain

The susceptibility of ΔsigGoperon to the presence of oleic and linoleic acid as well as to Tween 80 and Tween 40 compared to *M. tuberculosis* wild-type H37Rv was assessed using the new fatty acid susceptibility medium. In addition both the whole

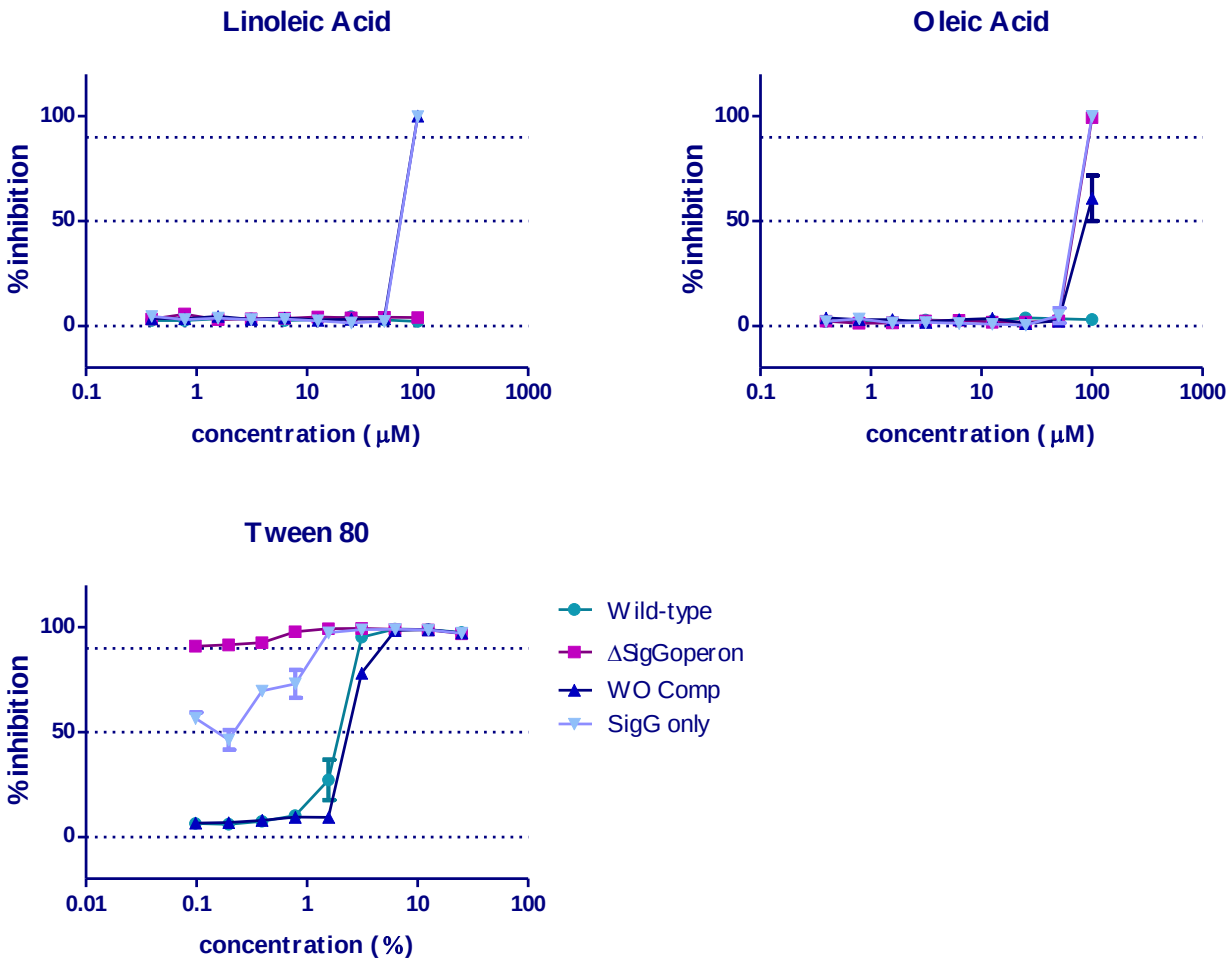


Figure 5.10 – Determining the SigG-dependency of the oleic acid, linoleic acid and Tween 80 phenotypes

Wild-type *M. tuberculosis* H37Rv, ΔsigGoperon, the whole operon and SigG only complement strains were assessed for their susceptibility to oleic acid, linoleic acid and Tween 80 in the fatty acid susceptibility medium (Appendix I). ΔsigGoperon was more susceptible to oleic acid and Tween 80 than wild-type and these phenotypes were either fully (Tween 80) or partially (oleic acid) complemented in the whole operon complement but not in the SigG only complement. ΔsigGoperon behaved as wild-type in the presence of linoleic acid whereas the two complement strains showed greater susceptibility than wild-type *M. tuberculosis* H37Rv.

operon and SigG only complement strains were tested in order to determine whether any phenotype detected was due to the deletion of SigG.

In all experiments conducted using Tween 40, this ester of palmitic acid produced complete inhibition of growth of all strains at all concentrations tested (data not shown) and will therefore not be discussed further under the scope of this project. ΔsigGoperon was more susceptible than wild-type *M. tuberculosis* H37Rv to the presence of oleic acid and its ester Tween 80 but not to the presence of linoleic acid in the fatty acid susceptibility medium (Figure 5.11). The two complements appeared to be susceptible to the highest concentration of linoleic acid tested (100 μM), however as this was not the case with the ΔsigGoperon or wild-type strains, this was not due to the *sigG* operon. As with the previous alamar blue assay conducted in modified Dubos medium, there appeared to be an all or nothing response to oleic acid, with the inhibition of ΔsigGoperon jumping from no inhibition to almost complete inhibition of growth over a very small concentration range (Figure 5.11). The inhibition caused by oleic acid was not complemented in the SigG only complement and only partially restored with the whole operon complement (Figure 5.11) suggesting that it is not SigG that is responsible for the phenotype but potentially one of the other genes in the *sigG* operon coupled with the overexpression of *lprO*, which is overexpressed in the ΔsigGoperon and both complement strains (Figure 5.3). The most dramatic phenotype was observed with Tween 80, which caused complete inhibition of growth of the ΔsigGoperon strain over the concentration range tested, whereas no significant difference was observed between the IC_{50} and MIC_{90} values of the wild-type and whole operon strains (IC_{50} 1.963 & 2.487, MIC_{90} 3.125 & 6.250 respectively). This phenotype was only

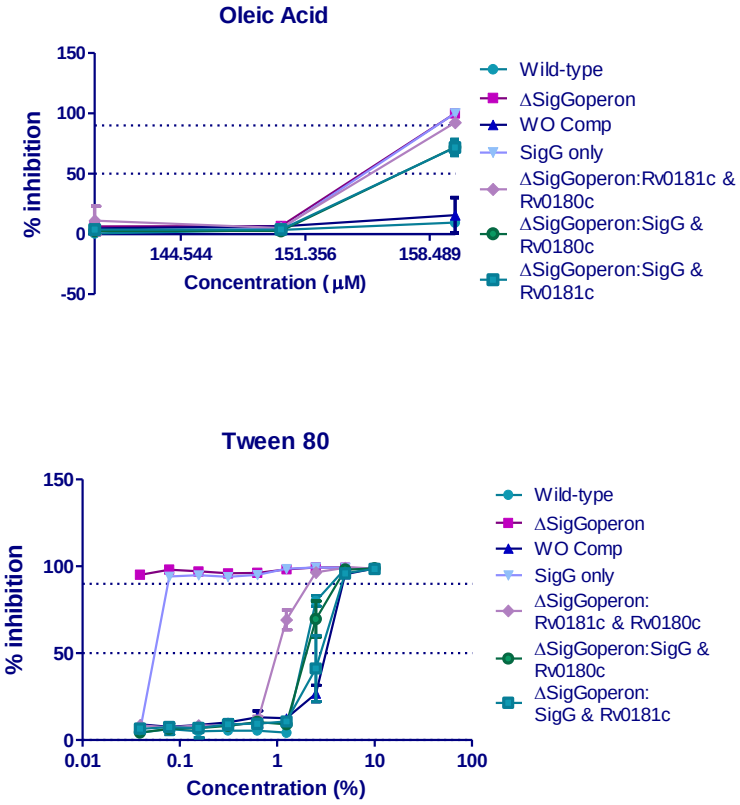
partially complemented in the SigG only strain (Figure 5.11) suggesting that at least one of the other genes in the *sigG* operon may have a role in the response to Tween 80.

5.7.3 Alamar blue assay results to determine the roles of the genes of the *sigG* operon in susceptibility to oleic acid and Tween 80 were inconclusive

In order to investigate the roles of the other genes of the *sigG* operon during fatty acid inhibition of *M. tuberculosis* growth, it was decided to use the partial complementation strains created in Section 5.3.2. In these strains the genes of the operon had been complemented pair-wise to create strains that were missing only one of the genes of the *sigG* operon. These are referred to as ΔsigGoperon:Rv0181c & Rv0180c (ΔsigGoperon containing the pLDcompΔ3 plasmid), ΔsigGoperon:SigG & Rv0180c (ΔsigGoperon containing the pLDcompΔ2 plasmid) and ΔsigGoperon:SigG & Rv0181c (ΔsigGoperon containing the pLDcompΔ4 plasmid).

Linoleic acid was not tested using the new complementation strains due to the fact that no difference was observed between the ΔsigGoperon and wild-type *M. tuberculosis* H37Rv strains in the fatty acid susceptibility medium (Figure 5.11). Therefore, the susceptibilities of wild-type *M. tuberculosis* and ΔsigGoperon to oleic acid and Tween 80 were assessed compared to all five complementation strains using the alamar blue cell-viability assay.

Previously the inhibition of oleic acid had only occurred at the highest concentration tested so this concentration was raised to 160 µM and instead of a two-fold serial dilution, the concentration of oleic acid decreased by 10 µM across the microplate. However, inhibition of growth of ΔsigGoperon again only occurred at the highest



IC50

Wild-type	ΔSigGoperon	WO Comp	SigG only	ΔSigGoperon: Rv0181c & Rv0180c	ΔSigGoperon: SigG & Rv0180c	ΔSigGoperon: SigG & Rv0181c
2.072	< lowest tested value	3.096	0.05879	1.053	2.107	2.776

Figure 5.11 – Determining the gene(s) responsible for the oleic acid and Tween 80 phenotypes

ΔsigGoperon and all five complement strains were assessed for their susceptibility to oleic acid and Tween 80 compared to wild-type *M. tuberculosis* H37Rv in the fatty acid susceptibility medium (Appendix I). The oleic acid phenotype was only complemented by the whole operon complement, with no significant difference observed between the remaining four complements and ΔsigGoperon (F test, $p>0.05$). The Tween 80 phenotype was complemented to wild-type by all complement strains except ΔsigGoperon:Rv0181c & Rv0180c, which partially complemented, and the SigG only complement, which could barely complement the ΔsigGoperon phenotype.

concentration tested compared to no inhibition of wild-type *M. tuberculosis* H37Rv (Figure 5.12). This phenotype was fully complemented by the whole operon complement; however, none of the other complementation strains were able to complement the oleic acid inhibition phenotype. This suggests that all three genes of the *sigG* operon are involved in preventing inhibition of growth due to the presence of oleic acid (Figure 5.12).

As with the previous result, ΔsigGoperon displayed complete inhibition of growth in the presence of Tween 80 over the concentrations tested compared to wild-type *M. tuberculosis* H37Rv, which had an IC_{50} of 2.072. The whole operon, ΔRv0181c and ΔRv0180c complementation strains showed no significant difference in their IC_{50} values compared to wild-type *M. tuberculosis* H37Rv ($p>0.05$, F test) and were therefore able to fully complement the ΔsigGoperon phenotype (Figure 5.12). In comparison both the SigG only and ΔsigGoperon:Rv0181c & Rv0180c complementation strains showed a significant difference between their IC_{50} values and that of wild-type *M. tuberculosis* H37Rv ($p<0.05$, F test) although the SigG only strain showed a much more severe inhibition of growth than the ΔsigGoperon:Rv0181c & Rv0180c complementation strain (Figure 5.12).

The severe inhibition of growth of the SigG only complementation strain in the presence of Tween 80 suggested that *Rv0181c*, *Rv0180c* or both genes were responsible for the phenotype observed with the ΔsigGoperon strain. However, if this was the case then it would be expected that one or both of the ΔsigGoperon:SigG & Rv0180c and ΔsigGoperon:SigG & Rv0181c complementation strains would also demonstrate inhibition of growth in the alamar

blue cell-viability assay, yet this was not the case (Figure 5.12). It was therefore, impossible to determine from these results, which gene or genes of the *sigG* operon were responsible for the phenotype observed with Tween 80. However, when the complementation strains were created, only the whole operon and SigG only complements were analysed for gene expression by qRT-PCR. It was therefore decided to analyse all of the complements to determine whether the inconclusive alamar blue results could be explained by examining the *sigG* operon gene expression profiles of all of the strains.

5.7.4 The ΔRv0181c strain was still expressing *Rv0181c*

To confirm that these plasmids complemented gene expression, RNA extracts from wild-type *M. tuberculosis* H37Rv, ΔsigGoperon and all five complement strains were analysed using qRT-PCR of the *sigG-Rv0180c* genes as well as for expression of *lprO*. qRT-PCR reactions were conducted as per Section 2.5.3. Reverse transcriptase negative reactions (RT-) were run alongside test qRT-PCR reactions (RT+), cDNA values were adjusted for presence of contaminating genomic DNA by subtraction of RT- values from RT+ values and the expression values were then normalised to the equivalent *sigA* cDNA values to give a relative expression level for each gene.

qRT-PCR demonstrated that expression of *Rv0181c* and *Rv0180c* in the ΔsigGoperon:Rv0181c & Rv0180c strain was only half that of wild-type (Figure 5.13), which could explain the reduced level of complementation observed with this plasmid. All the complement strains were complemented for expression of the correct genes except in one instance. The ΔsigGoperon:SigG & Rv0180c strain

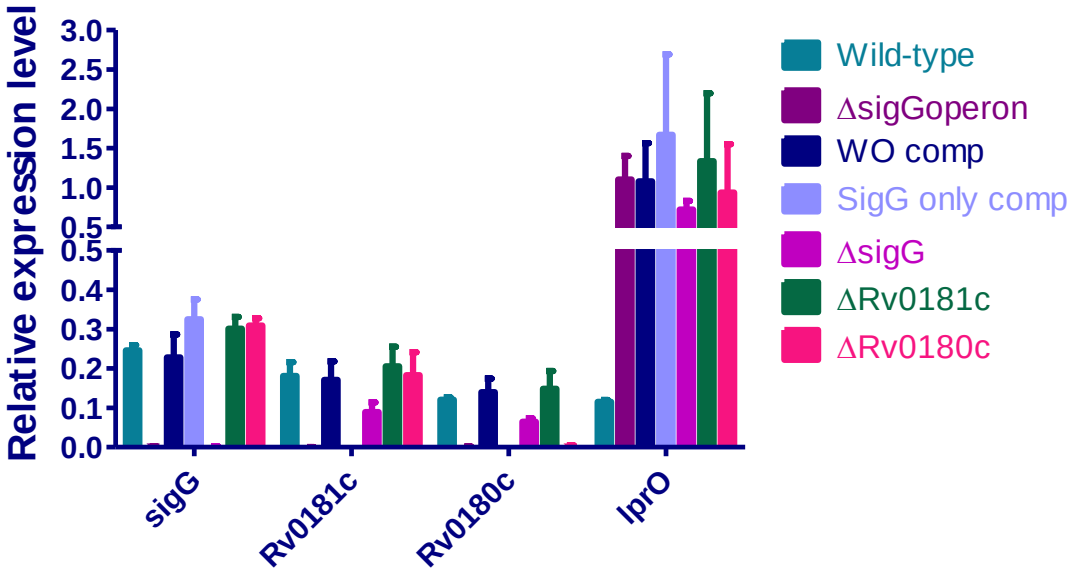


Figure 5.12 - qRT-PCR to assess the partial complementation of the ΔsigGoperon strain

qRT-PCR (SYBR-Green) of the *sigG-Rv0180c* and *lprO* genes was performed on RNA extracts of wild-type *M. tuberculosis* H37Rv, the *M. tuberculosis* ΔsigGoperon deletion strain and all five complementation strains (ΔsigGoperon containing pLDcomp8T, pLDcompΔ1, pLDcompΔ2, pLDcompΔ3 and pLDcompΔ4). Expression data for each gene was normalised to equivalent *sigA* data and columns and error bars represent the averages and standard deviations of three biological replicates, each assayed in triplicate. The SigG only and ΔsigGoperon:SigG & Rv0181c partial complementation strains demonstrated expression levels that were returned to wild-type and the ΔsigGoperon:Rv0181c & Rv0180c strain was expressing *Rv0181c* and *Rv0180c* at a lower level than wild-type. The ΔsigGoperon:SigG & Rv0180c strain demonstrated expression of all three genes of the *sigG* operon at wild-type levels, despite the fact that the *Rv0181c* gene should have been deleted. Normalised *lprO* transcription levels were significantly increased compared to the wild-type strain in the ΔsigGoperon strain and all of the complementation strains (Unpaired t-tests, $p<0.05$).

should only have been expressing *sigG* and *Rv0180c* from the *sigG* operon but wild-type levels of expression for *Rv0181c* were detected (Figure 5.13). The relative expression level of the *lprO* gene was analysed in all strains to determine whether the increase in expression of *lprO*, observed in the ΔsigGoperon, whole operon and SigG only complement strains, was observed in all of the complements.

qRT-PCR analysis showed that expression of *lprO* was significantly induced compared to wild-type in ΔsigGoperon and all complement strains (Figure 5.13; Unpaired t-tests, $p<0.05$). Despite the fact that *lprO* expression has been affected by deletion of the *sigG* operon, the fact that this effect was present in all of the complement strains meant that if a phenotypic change was observed that was due to the overexpression of *lprO* rather than deletion of the *sigG* operon, this phenotype would not be complemented by any of the complement strains.

Examination of the pLDcompΔ2 complementation plasmid revealed that the *Rv0181c* gene was still present. At the time of writing this report, the plasmid has been remade and electroporated into the ΔsigGoperon *M. tuberculosis* strain. The fact that the ΔsigGoperon:SigG & Rv0180c strain was actually expressing all three genes of the *sigG* operon meant that it should have been behaving like the whole operon complement strain. In the case of the Tween 80 alamar blue results, this suggests that the phenotype observed was due to deletion of *Rv0181c* as the two strains where this gene was missing (ΔsigGoperon and the SigG only complement) demonstrated Tween 80-dependent inhibition of growth (Figure 5.12). However, the 'ΔsigGoperon:SigG & Rv0180c' strain did not behave like the whole operon complement in response to oleic acid (Figure 5.12). It would appear that all three

genes of the *sigG* operon are necessary to protect against oleic acid toxicity but this will need to be confirmed using the new ΔsigGoperon:SigG & Rv0180c strain and higher concentrations of oleic acid to obtain more than one point at which inhibition of growth occurs.

5.8 Discussion

The *sigG* operon marked deletion strain designed in this study was successfully created using the method described by Parish & Stoker (2000). This strain was shown to be complemented by the integration of the *sigG-Rv0180c* genes, under the control of the three promoters discussed in Chapter 3. Partial complementation strains in which, *sigG* only, *sigG-Rv0181c* and *Rv0181c-Rv0180c* were integrated into the genome were also successfully created. However, a strain where only the *sigG* and *Rv0180c* genes should have been complemented proved to still be expressing the *Rv0181c* gene and this strain had to be recreated towards the end of this study. The partial complement strains were initially created to assess the roles of Rv0181c and Rv0180c as regulators of SigG; however they became more useful for determining which gene(s) of the operon were responsible for the phenotype observed in the ΔsigGoperon strain. Expression of the downstream gene, *lprO,* increased due to deletion of the *sigG* operon suggesting that either a negative regulatory element for this gene had been deleted, or there was transcriptional read-through from the gentamycin resistance cassette into the *lprO* gene. The increase in expression of *lprO* was also observed in all of the complementation strains and it was therefore possible to continue with phenotypic analysis of the ΔsigGoperon strain, as a phenotype that was due to overexpression of *lprO* would not be complemented by any of the complementation strains.

Initial analysis of *in vitro* growth of *M. tuberculosis* wild-type H37Rv, ΔsigGoperon and whole operon complement strains demonstrated that the deletion strain behaved as wild-type *M. tuberculosis*. Likewise, *in vivo* studies analysing survival within naive and activated macrophages and also during a mouse model of infection showed that the *M. tuberculosis* ΔsigGoperon strain behaved just as wild-type *M. tuberculosis* H37Rv. These results suggest that the *sigG* operon is not required to maintain infection; however, it is important to note that neither the macrophage infection nor mouse infection models completely model the conditions encountered by *M. tuberculosis* during human infection. In particular, the granuloma is not modelled (Russell, 2001) and it is therefore still possible that the *sigG* operon may have a role in granuloma formation or maintenance.

In addition the ΔsigGoperon strain was analysed using the alamar blue cell-viability assay to determine whether it was more or less susceptible to inhibition by the compounds quercetin, butein and methylglyoxal. These compounds were selected based on the potential role of Rv0181c as a pirin protein and the roles of the potential SigG-dependent genes *Rv0887c* and *Rv0911* as glyoxalase enzymes. However, ΔsigGoperon behaved as wild-type *M. tuberculosis* H37Rv in the presence of all three compounds.

There have been very few studies to characterise pirin orthologues in prokaryotes, and they can have a variety of roles. The *E. coli* YhhW pirin protein has been shown to function as a quercetinase enzyme that breaks down the flavonoid compound quercetin and its derivatives (Adams *et al.*, 2005). Flavonoid compounds have been shown to have antimycobacterial properties by inhibiting the fatty acid synthase II system, which is responsible for producing the fatty acid components of the

mycobacterial cell wall, mycolic acids (Brown *et al.*, 2007). Due to the fact that Rv0181c shows strong homology to the *E. coli* YhhW quercetinase protein and forms part of the *sigG* operon it was hypothesised that Rv0181c could function as a quercetinase enzyme and therefore, that the *sigG* operon may have a role in detoxification of flavonoid compounds. However, the ΔsigGoperon strain behaved as wild-type in the presence of the flavonoids quercetin and butein, demonstrating that the *sigG* operon is unlikely to be involved in detoxification of flavonoid compounds.

Interestingly a recent study identified a pirin-orthologue, with strong homology to *E.coli* YhhW, in *Serratia marcescens* that was able to interact with pyruvate dehydrogenase to modulate its activity and therefore was predicted to have a role in central carbohydrate metabolism of this organism (Soo *et al.*, 2007). This may be a worthwhile avenue for future study of the function of Rv0181c.

The two genes, *Rv0887c* and *Rv0911* were shown to be most upregulated in a strain overexpressing SigG (K. Smollett, personal communication). Both of the proteins encoded by these genes belong to the Glo_EDI_BRP_like domain superfamily, with Rv0887c containing one domain and Rv0911 containing two domains (Figure 1.11). This superfamily contains several structurally related metalloproteins, including type I extradiol dioxygenases, glyoxalase I and a group of antibiotic resistance proteins. The glyoxalase I enzyme is the first enzyme in the methylglyoxal detoxification pathway (Figure 1.12). Rv0911 was predicted to be involved in this detoxification pathway based on the fact that it was able to interact with, and was therefore a target of, pyrimidine-imidazole compounds, which target the methylglyoxal detoxification pathway (Marrero *et al.*, 2010). Methylglyoxal is an electrophile, which can cause

cell death at high concentrations by reacting with the nucleophilic centres of molecules such as DNA, RNA and proteins. However, despite it being toxic, bacteria still produce this metabolite and it is thought that this provides them with the ability to control the rate of carbon flux when moving between environments (Ferguson *et al.*, 1998). In addition, the *in vivo* tissue environment, the granuloma especially, has been shown to be rich in methylglyoxal, which is able to induce apoptosis of macrophages (Rachman *et al.*, 2006). While it is unclear whether the methylglyoxal in the granuloma is entirely produced by the host or also excreted from *M. tuberculosis*, one hypothesis is that the bacterial glyoxalase system may be involved in not only protecting the bacteria from host-derived methylglyoxal but also in preventing methylglyoxal-induced apoptosis of macrophages.

In the presence of increasing concentrations of methylglyoxal, ΔsigGoperon behaved as wild-type *M. tuberculosis* H37Rv. However, due to the toxic properties of this metabolite and the importance of the methylglyoxal detoxification pathway, there are likely to be several enzymes involved at each stage of detoxification and as shown in Figure 1.12 this is predicted to be the case (Marrero *et al.*, 2010). Should SigG regulate the expression of genes encoding enzymes involved in the glyoxalase system, its deletion and therefore the loss of production of these enzymes would most likely be compensated for by the presence of alternative glyoxalase system genes, such as *Rv0577* (Figure 1.12) in the *M. tuberculosis* genome.

A novel phenotype microarray system was used to analyse the ΔsigGoperon strain for its ability to use individual carbon substrates compared to wild-type *M. tuberculosis* H37Rv. This highlighted the fatty acid esters Tween 40 and Tween 80 as carbon sources, on which, ΔsigGoperon was less able to grow. There were two

possible explanations for the results obtained with the phenotype microarrays: 1) as the microarrays were designed to test the ability of a bacterial strain to utilise the substrates as a carbon source it was possible that ΔsigGoperon was less able to utilise the Tween substrates or 2) ΔsigGoperon growth was being inhibited by the presence of Tween 40 and 80. Tween 80 (the ester of oleic acid) in particular has been shown to exert an inhibitory effect on mycobacterial growth, which was shown to be due to contaminating free oleic acid as well as the release of oleic acid via hydrolysis over time (Dubos *et al.*, 1947, Sattler *et al.*, 1948).

There are two pathways that are specifically required for growth on fatty acids, the catabolic β-oxidation cycle, which degrades fatty acids to acetyl-CoA, and when glycolytic substrates are in low abundance, the glyoxylate shunt (McKinney *et al.*, 2000). The glyoxylate shunt prevents the loss of carbon molecules via the tricarboxylic acid (TCA) cycle, which would otherwise occur during growth on fatty acids. Extensive duplication of genes encoding β-oxidation enzymes exists in *M. tuberculosis* (Cole *et al.*, 1998) making the study of genes in this system difficult due to functional redundancy. The glyoxylate shunt however, is limited to three genes encoding malate synthase and the two isocitrate lyase genes (Munoz-Elias *et al.*, 2005) and it is therefore more likely that a phenotype will be obtained if regulation of the glyoxylate shunt is disrupted. ΔsigGoperon was able to grow as wild-type *M. tuberculosis* H37Rv on the fatty acid substrates acetate, α-ketoglutarate and Tween 80 as sole carbon sources. This suggests that SigG is not involved in regulation of genes required for growth on fatty acids. However, while it is unlikely that SigG regulates genes involved in the glyoxylate shunt, the functional redundancy in the β-

oxidation cycle means that SigG could regulate genes in this pathway without its deletion causing a noticeable phenotype.

The susceptibility of the ΔsigGoperon strain to a variety of fatty acids as well as to Tween 40 and 80 was analysed compared to wild-type *M. tuberculosis* H37Rv. Due to the fact that modified Dubos medium contains Tween 80 and also bovine serum albumin, which has been shown to protect the bacteria from free fatty acids (Davis *et al.*, 1947), a new medium was defined for assessing fatty acid susceptibility.

The ΔsigGoperon strain proved to be more susceptible than wild-type *M. tuberculosis* H37Rv to the presence of Tween 80 and oleic acid in the growth medium. Susceptibility to oleic acid inhibition was only restored to wild-type levels in the whole operon complement suggesting that all three genes of the *sigG* operon have a role in resisting toxicity of oleic acid. Contrastingly, susceptibility to Tween 80, was restored to wild-type levels in all but the SigG only complement strain, which only very partially complemented the ΔsigGoperon phenotype. At this point qRT-PCR analysis of the expression of the *sigG* operon genes in the partial complements revealed that the ΔsigGoperon:SigG & Rv0180c partial complement was in fact still expressing all three genes of the operon. Applying this knowledge to the Tween 80 alamar blue assay results enabled this phenotype to be attributed to the absence of the *Rv0181c* gene as only the two strains where this gene was missing (ΔsigGoperon and the SigG only complement) demonstrated marked inhibition of growth compared to wild-type *M. tuberculosis* H37Rv. At the time of writing the original partial complementation plasmid, missing just the *Rv0181c* gene, had been remade together with a new partial complementation plasmid expressing only the

Rv0181c gene. Should Rv0181c be responsible for the phenotype observed with Tween 80 it is expected that the new ΔsigGoperon:SigG & Rv0180c complement will demonstrate the same phenotype as the ΔsigGoperon strain, whereas the Rv0181c only complement should restore the phenotype to that of wild-type *M. tuberculosis* H37Rv. Although SigG was not required for resistance to Tween 80, the *sigG* gene is co-transcribed with *Rv0181c* suggesting that the response pathway triggered by Tween 80 may also require genes regulated by *sigG*. Due to the large number of sigma factors and duplication of functionality between genes in the *M. tuberculosis* genome it is possible that the function(s) of the genes regulated by *sigG* are compensated for elsewhere whereas the function of Rv0181c is not.

Tween 80 is a non-ionic surfactant that has been shown to have an effect on the mycobacterial cell wall e.g. in *Mycobacterium avium* Tween 80 reduces the number of glycolipids, causing the cell wall to be more permeable to antitubercular drugs (Masaki *et al.*, 1991). Higher levels of resistance to Tween 80 have been observed in strains with the drug resistance-associated W-Beijing genotype compared to non-W-Beijing. It was hypothesised by the authors of this study that different resistances to Tween 80 in mycobacterial strains may be related to the susceptibility of these strains to pulmonary surfactants, which form an important part of the host defence mechanism (von Groll *et al.*, 2010).

Oleic acid is a long-chain (C_{18}) unsaturated fatty acid that has been shown to have mycobactericidal effects. Unsaturated fatty acids, including oleic acid, were shown to be highly toxic to the fast-growing group IV mycobacteria and it was hypothesised that this was due to disturbance of membrane permeability. It was

noted in the same study that pathogenic strains showed more resistance to unsaturated fatty acids (Saito *et al.*, 1984). Unsaturated fatty acids such as oleic and linoleic acids are secreted by activated macrophages and form an important part of the mycobactericidal activity of macrophages (Hemsworth *et al.*, 1978). Oleic acid has been shown to inhibit the enzyme topoisomerase I (topo I), a key enzyme in the cleavage and repair of DNA strands. Topo I is also important for cellular processes such as replication, transcription and recombination and for relaxation of transcription-induced negative supercoiling (Carballeira, 2008, Tse-Dinh, 2009). Inhibition of topo I leads to DNA damage by trapping the topoisomerase cleavage-complex, therefore preventing relaxation of supercoiled DNA during transcription. In mycobacteria a transposon insertion mutant in the *topA* gene could not be isolated suggesting that this gene is essential (Tse-Dinh, 2009).

Oleic acid inhibition of the ΔsigGoperon deletion strain could therefore be occurring by one of two mechanisms: by changing the membrane permeability of the bacterial cells or by inhibiting the topoisomerase I enzyme.

Recent work published in this laboratory has determined that SigG forms part of the RecA-independent response to DNA-damage but does not regulate DNA repair genes (Smollett *et al.*, 2011). Taken together with the evidence presented here regarding the role of Rv0181c and potentially SigG in resistance to Tween 80 and the whole *sigG* operon in resistance to oleic acid toxicity, it is possible that the *sigG* operon is involved in regulation of genes for detoxifying compounds that cause DNA damage.

6. Discussion

SigG has been shown to be the most highly upregulated *M. tuberculosis* sigma factor in response to DNA damage (Rand *et al.*, 2003) and during human macrophage infection (Cappelli *et al.*, 2006). There is very little literature regarding the genes regulated by SigG or its role and what information there is available would appear to be contradictory. Lee *et al.* (2007) hypothesised that SigG was able to regulate expression of the repressor LexA, and therefore that it has a role in regulating expression of genes involved in the RecA/LexA-dependent SOS response to DNA damage. However, recent investigations in this laboratory have demonstrated that SigG does not regulate the expression of *lexA* and that a *sigG* deletion strain was no more or less susceptible to DNA damage than wild-type. It was therefore concluded that SigG does not regulate expression of genes involved in either the RecA-dependent or RecA-independent DNA-damage responses, although it appears to be part of the RecA-independent regulon (Smollett *et al.*, 2011). SigG may therefore regulate a set of genes that, although related to the response to DNA damage, are not responsible for DNA repair.

The aim of this work was to investigate the function and regulation of the little-studied *M. tuberculosis* ECF sigma factor, SigG. ECF sigma factors can be subject to complex regulation both at the transcriptional level and the post-translational level (Sachdeva *et al.*, 2009). This study set out to investigate the type(s) of regulation that SigG was subject to as well as to determine its role, if any, during *M. tuberculosis* pathogenesis.

Primer extension studies had suggested the presence of three promoters expressing *sigG*, one (P1) which showed significant similarity to the RecA/LexA independent promoter, RecA-NDp (Gamulin *et al.*, 2004) and another (P3) that had a -10 sequence similar to the SigM consensus (Raman *et al.*, 2006). The number and regulation of the promoters expressing *sigG* were investigated initially using promoter-*lacZ* fusions, and then further using qRT-PCR. Promoter-*lacZ* fusion studies supported the presence and activity of promoters P1 and P2 and although there was evidence to support the fact that promoter P3 was active, this was less convincing due to the fact that at least one of the other promoters appeared not to have been completely inactivated. qRT-PCR was used to confirm that there were at least two promoters expressing *sigG*. Future work to characterise the number of promoters expressing *sigG* could include further qRT-PCR experiments using primers designed to produce an amplicon in the region between promoters P2 and P3, as well as upstream of promoter P3. These two primer sets would detect transcripts produced from promoter P3 only and from any as yet unidentified promoters upstream of promoter P3 respectively.

It has been demonstrated that SigG is not an auto-regulatory sigma factor due to the fact that expression from the promoter-*lacZ* construct containing all three promoters was the same in both the wild-type and ΔsigGoperon deletion strains.

The *sigG* promoters were assessed for their ability to be induced by a variety of agents. Promoter P1 was shown to be the only DNA-damage inducible promoter, in keeping with its similarity to the RecA-NDp promoter (Gamulin *et al.*, 2004). Expression of *sigG* was not induced by the presence of the flavanoid compounds

quercetin and butein, by the toxic electrophile methylglyoxal or by the presence of fatty acids in the medium. As discussed previously there is a highly complex regulatory network, with the expression of some sigma factors being regulated by several different holoenzymes e.g. *sigB* regulation by the SigE, SigF, SigH and SigL holoenzymes (Rodrigue *et al.*, 2007). Future work to determine the differential regulation of the *sigG* promoters could include investigating *sigG* expression in *M. tuberculosis* sigma factor deletion strains, of which, only *sigA* and *sigI* deletion strains are unavailable. Although a *sigB* deletion strain has been created (Fontan *et al.*, 2009), this is a marked mutant, where the *sigB* gene has been replaced with an antibiotic resistance cassette and there are concerns that this may have affected expression of the essential downstream gene *ideR*. Determining which holoenzymes regulate *sigG* expression could lead to discovering what stresses, besides DNA damage, lead to its induction as the conditions when some of the better characterised sigma factors are active are well known e.g. SigH during oxidative stress (Song *et al.*, 2003).

One of the initial hypotheses of this study was that the Rv0181c and Rv0180c proteins act as anti-sigma and anti-anti-sigma factors to SigG due to the fact that the *sigG, Rv0181c* and *Rv0180c* genes are co-transcribed. The potential roles of the Rv0181c and Rv0180c proteins as regulators of SigG were investigated using the yeast-2-hybrid system and then, for Rv0181c, using a direct protein pulldown. Protein BLAST analysis of the Rv0181c protein revealed that it showed significant homology to the *E. coli* YhhW protein, which has been shown to function as a quercetinase enzyme (Adams *et al.*, 2005). A recent publication has also shown the Rv0180c protein to function in host cell invasion (Caceres *et al.*, 2011). It was

concluded from the protein-interaction studies and due to the recent literature that Rv0181c and Rv0180c do not function as regulators of SigG.

SigG interacted with several proteins in a pulldown assay using recombinant SigG to pulldown proteins from *M. tuberculosis* CFE, which were identified by mass spectrometry. Anti-sigma factors are often transmembrane proteins as they sense and respond to changes in the environment (Helmann, 1999). Of the potential interaction partners identified the most promising candidate is the predicted transmembrane protein Rv1084, which belongs to the thioredoxin-like superfamily. Members of this superfamily contain a redox active CXXC motif that allows them to alter the redox state and therefore the activity of their target proteins. Further work to validate the interactions between SigG and the proteins pulled down from *M. tuberculosis* CFE, especially Rv1084, could include expressing these proteins in a yeast-2-hybrid screen, or performing the reciprocal pulldown using purified e.g. Rv1084 protein to determine whether it can pulldown SigG from *M. tuberculosis* CFE. Confirmed interaction partners would then need to be characterised as to whether they were regulators of SigG function. This could be performed using an *in vitro* transcription assay to determine whether the presence of an interacting protein was able to increase or decrease transcription from a SigG-dependent promoter.

Just because a gene is expressed does not mean that the protein is being translated or that once produced the protein is active. In the case of sigma factors it is possible that even though they are being expressed the protein could be sequestered away preventing it from forming the RNAP holoenzyme. In order to determine when the SigG protein was active, this study attempted to identify a SigG-dependent gene or

genes, with the aim of investigating expression of this gene(s) under various conditions. Three genes were analysed as potential SigG-dependent genes. The *Rv0654* gene had previously been predicted to be expressed from a SigG-dependent promoter; however, expression of this gene in the ΔsigGoperon deletion strain was unaffected, showing that its expression was not SigG-dependent. The SigG-dependency of the two genes that were most highly up-regulated in the SigG overexpression strain, *Rv0887c* and *Rv0911* was assessed using *in vitro* transcription and protein-DNA binding assays. A SigG-specific transcript could not be obtained for either gene via *in vitro* transcription and although SigG bound to the *Rv0887c* and *Rv0911* promoter DNA fragments, this interaction was not specific. However, this protein was His-tagged and it is possible that the tag was interfering with either binding to RNA polymerase or to the DNA fragments. It was therefore, not possible to draw a definitive conclusion as to whether expression of *Rv0887c* and *Rv0911* was SigG-dependent or not.

Phenotypic analysis of a ΔsigGoperon deletion strain revealed that this strain was not attenuated compared to wild-type *M. tuberculosis* H37Rv for *in vitro* growth or in the *in vivo* macrophage and mouse models of infection. A ΔsigG deletion strain had previously been shown to have the same susceptibility to the DNA damaging agents mitomycin C, bleomycin and ofloxacin as wild-type in an alamar blue microplate assay (Smollett *et al.*, 2011). Subsequently in this study it has been demonstrated that the ΔsigGoperon deletion strain shows the same susceptibility as the wild-type strain to the flavanoid compounds butein and quercetin, the toxic electrophile methylglyoxal and to saturated fatty acids but that it was more susceptible to the presence of the unsaturated fatty acid, oleic acid, and its ester Tween 80 in an alamar

blue microplate assay. Analysis of full and partial complements of the ΔsigGoperon deletion strain revealed that the susceptibility of this strain to oleic acid appeared to be due to the absence of all three genes of the *sigG* operon, whereas its susceptibility to Tween 80 could be attributed to the absence of the *Rv0181c* gene. Due to the fact that the ΔRv0181c partial complement strain was expressing all three genes of the *sigG* operon, further work is needed to confirm that the Tween 80 phenotype is due to deletion of *Rv0181c* using a strain lacking *Rv0181c* alone and another that is only expressing *Rv0181c* of this locus, which should have the same phenotype as the ΔsigGoperon deletion and wild-type *M. tuberculosis* H37Rv strains respectively.

Tween 80 is a non-ionic surfactant that has been shown to have an effect on the mycobacterial cell wall e.g. in *Mycobacterium avium* Tween 80 reduces the number of glycolipids, causing the cell wall to be more permeable to antitubercular drugs (Masaki *et al.*, 1991).

Oleic acid is a long-chain (C_{18}) unsaturated fatty acid that has been shown to have mycobactericidal effects. Unsaturated fatty acids such as oleic and linoleic acids are secreted by activated macrophages and form an important part of the mycobactericidal activity of macrophages (Hemsworth *et al.*, 1978). Oleic acid has been shown to inhibit the enzyme topoisomerase I (topo I), a key enzyme in the cleavage and repair of DNA strands, leading to DNA damage (Carballeira, 2008).

In conclusion, expression of *sigG* has been shown to be induced in response to DNA damage (Rand *et al.*, 2003, Smollett *et al.*, 2011); however, SigG does not regulate the expression of DNA repair genes (Smollett *et al.*, 2011). A *sigG* operon deletion

strain showed increased susceptibility to the presence of the unsaturated fatty acid, oleic acid, and Tween 80 in the medium. SigG is also thought to regulate the expression of two genes predicted to encode glyoxalase enzymes, responsible for the detoxification of the toxic electrophile methylglyoxal, which can cause DNA damage (Ferguson *et al.*, 1998); however direct regulation of these genes could not be conclusively proven in this study. Unsaturated fatty acids and methylglyoxal are produced by macrophages as part of the host immune response (Hemsworth *et al.*, 1978, Rachman *et al.*, 2006). The current hypothesis for the role of SigG and its operon is therefore, that it is involved in regulation of genes for detoxifying compounds that cause DNA damage explaining why it forms part of the RecA-independent DNA damage response regulon.

References

Adams, M. and Jia, Z. (2005). Structural and biochemical analysis reveal pirins to possess quercetinase activity. *J Biol Chem* **280,** 28675-28682.

Akaki, T., Tomioka, H., Shimizu, T., Dekio, S. and Sato, K. (2000). Comparative roles of free fatty acids with reactive nitrogen intermediates and reactive oxygen intermediates in expression of the anti-microbial activity of macrophages against *Mycobacterium tuberculosis*. *Clinical & Experimental Immunology* **121,** 302-310.

Antoine, D., Maguire, H. and Story, A. (2006). Epidemiology and response to the growing problem of tuberculosis in London. *Euro Surveill* **11,** 25-28.

Austin, P.E., McCulloch, E.A. and Till, J.E. (1971). Characterization of the factor in L-cell conditioned medium capable of stimulating colony formation by mouse marrow cells in culture. *J Cell Physiol* **77,** 121-134.

Bashyam, M.D. and Hasnain, S.E. (2004). The extracytoplasmic function sigma factors: role in bacterial pathogenesis. *Infection, Genetics and Evolution* **4,** 301-308.

Beaucher, J., Rodrigue, S., Jacques, P.-E., Smith, I., Brzezinski, R. and Gaudreau, L. (2002). Novel *Mycobacterium tuberculosis* anti-sigma factor antagonists control SigF activity by distinct mechanisms. *Molecular Microbiology* **45,** 1527-1540.

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

BLAST, N. http://blast.ncbi.nlm.nih.gov/Blast.cgi. [WWW document]. URL

Bloch, H. and Segal, W. (1956). Biochemical differentiation of *Mycobacterium tuberculosis* grown *in vivo* and *in vitro. J Bacteriol* **72,** 132-141.

Bochkareva, E.S., Girshovich, A.S. and Bibi, E. (2002). Identification and characterization of the *Escherichia coli* stress protein UP12, a putative in vivo substrate of GroEL. *European Journal of Biochemistry* **269,** 3032-3040.

Booth, I.R., Ferguson, G.P., Miller, S., Li, C., Gunasekera, B. and Kinghorn, S. (2003). Bacterial production of methylglyoxal: a survival strategy or death by misadventure? *Biochem Soc Trans* **31,** 1406-1408.

Borukhov, S. and Nudler, E. (2003). RNA polymerase holoenzyme: structure, function and biological implications. *Current Opinion in Microbiology* **6,** 93-100.

Brosch, R., Gordon, S.V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., *et al.* (1998). Use of a *Mycobacterium tuberculosis* H37Rv Bacterial Artificial Chromosome Library for Genome Mapping, Sequencing, and Comparative Genomics. *Infect. Immun.* **66,** 2221-2229.

Brown, A.K., Papaemmanouil, A., Bhowruth, V., Bhatt, A., Dover, L.G. and Besra, G.S. (2007). Flavonoid inhibitors as novel antimycobacterial agents targeting Rv0636, a putative dehydratase enzyme involved in *Mycobacterium tuberculosis* fatty acid synthase II. *Microbiology* **153,** 3314-3322.

Browning, D.F. and Busby, S.J.W. (2004). The regulation of bacterial transcription initiation. *Nat Rev Micro* **2,** 57-65.

Caceres, S.M., Ocampo, M., Arevalo-Pinzon, G., Jimenez, R.A., Patarroyo, M.E. and Patarroyo, M.A. (2011). The *Mycobacterium tuberculosis* membrane protein Rv0180c: Evaluation of peptide sequences implicated in mycobacterial invasion of two human cell lines. *Peptides* **32,** 1-10.

Camarero, J.A., Shekhtman, A., Campbell, E.A., Chlenov, M., Gruber, T.M., Bryant, D.A., *et al.* (2002). Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc Natl Acad Sci U S A* **99,** 8536-8541.

Cappelli, G., Volpe, E., Grassi, M., Liseo, B., Colizzi, V. and Mariani, F. (2006). Profiling of *Mycobacterium tuberculosis* gene expression during human macrophage infection: Upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function. *Research in Microbiology* **157,** 445-455.

Carballeira, N.M. (2008). New advances in fatty acids as antimalarial, antimycobacterial and antifungal agents. *Prog Lipid Res* **47,** 50-61.

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

Cole, S.T. and Barrell, B.G. (1998). Analysis of the Genome of *Mycobacterium tuberculosis* H37Rv. *Novartis Found Symp* **217,** 160-172; discussion 172-167.

Collins, L. and Franzblau, S.G. (1997). Microplate alamar blue assay versus BACTEC 460 system for high-throughput screening of compounds against *Mycobacterium tuberculosis* and *Mycobacterium avium*. *Antimicrob Agents Chemother* **41,** 1004-1009.

Corbett, E.L., Watt, C.J., Walker, N., Maher, D., Williams, B.G., Raviglione, M.C. and Dye, C. (2003). The Growing Burden of Tuberculosis: Global Trends and Interactions With the HIV Epidemic. *Arch Intern Med* **163,** 1009-1021.

Cosma, C.L., Sherman, D.R. and Ramakrishnan, L. (2003). The secret lives of the pathogenic Mycobacteria. *Annual Review of Microbiology* **57,** 641-676.

Dainese, E., Rodrigue, S., Delogu, G., Provvedi, R., Laflamme, L., Brzezinski, R., *et al.* (2006). Posttranslational regulation of *Mycobacterium tuberculosis* extracytoplasmic-function sigma factor sigma L and roles in virulence and in global regulation of gene expression. *Infect Immun* **74,** 2457-2461.

Dandekar, T., Snel, B., Huynen, M. and Bork, P. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* **23,** 324-328.

Davis, B.D. and Dubos, R.J. (1947). The Binding of Fatty Acids by Serum Albumin, a Protective Growth Factor in Bacteriological Media. *J Exp Med* **86,** 215-228.

Dawson, L.F. (2005) A novel mechanism regulating DNA-damage repair in *Mycobacterium tuberculosis*. University College London.

Dombroski, A.J., Walter, W.A., Record, M.T., Jr., Siegele, D.A. and Gross, C.A. (1992). Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. *Cell* **70,** 501-512.

Dona, V., Rodrigue, S., Dainese, E., Palu, G., Gaudreau, L., Manganelli, R. and Provvedi, R. (2008). Evidence of complex transcriptional, translational, and posttranslational regulation of the extracytoplasmic function sigma factor sigmaE in *Mycobacterium tuberculosis*. *J Bacteriol* **190,** 5963-5971.

Drumm, J.E., Mi, K., Bilder, P., Sun, M., Lim, J., Bielefeldt-Ohmann, H., *et al.* (2009). *Mycobacterium tuberculosis* Universal Stress Protein Rv2623 Regulates Bacillary Growth by ATP-Binding: Requirement for Establishing Chronic Persistent Infection. *PLoS Pathog* **5,** e1000460.

Dubos, R.J. and Middlebrook, G. (1947). Media for tubercle bacilli. *Am Rev Tuberc* **56,** 334-345.

Eltis, L.D. and Bolin, J.T. (1996). Evolutionary relationships among extradiol dioxygenases. *J Bacteriol* **178,** 5930-5937.

Ferguson, G.P., McLaggan, D. and Booth, I.R. (1995). Potassium channel activation by glutathione-S-conjugates in *Escherichia coli*: protection against methylglyoxal is mediated by cytoplasmic acidification. *Mol Microbiol* **17,** 1025-1033.

Ferguson, G.P., Tötemeyer, S., MacLean, M.J. and Booth, I.R. (1998). Methylglyoxal production in bacteria: suicide or survival? *Archives of Microbiology* **170,** 209-218.

Fernandes, N.D., Wu, Q.L., Kong, D., Puyang, X., Garg, S. and Husson, R.N. (1999). A mycobacterial extracytoplasmic sigma factor involved in survival following heat shock and oxidative stress. *J Bacteriol* **181,** 4266-4274.

Fields, S. (2007). Molecular biology. Site-seeing by sequencing. *Science* **316,** 1441-1442.

Fivian-Hughes, A.S. (2009) The Role of the Predicted Transcriptional Regulator Rv1956 in *Mycobacterium tuberculosis*. University College London.

Florczyk, M.A., McCue, L.A., Stack, R.F., Hauer, C.R. and McDonough, K.A. (2001). Identification and Characterization of Mycobacterial Proteins Differentially Expressed under Standing and Shaking Culture Conditions, Including Rv2623 from a Novel Class of Putative ATP-Binding Proteins. *Infect. Immun.* **69,** 5777-5785.

Fontan, P.A., Voskuil, M.I., Gomez, M., Tan, D., Pardini, M., Manganelli, R., *et al.* (2009). The *Mycobacterium tuberculosis* sigma factor sigmaB is required for full response to cell envelope stress and hypoxia *in vitro*, but it is dispensable for *in vivo* growth. *J Bacteriol* **191,** 5628-5633.

Friedberg, E.C. (2003). DNA damage and repair. *Nature* **421,** 436-440.

Gamulin, V., Cetkovic, H. and Ahel, I. (2004). Identification of a promoter motif regulating the major DNA damage response mechanism of *Mycobacterium tuberculosis*. *FEMS Microbiology Letters* **238,** 57-63.

Gatignol, A., Durand, H. and Tiraby, G. (1988). Bleomycin resistance conferred by a drug-binding protein. *FEBS Lett* **230,** 171-175.

Gopaul, K.K. (2002) Transcription of the *Mycobacterium tuberculosis recA* gene. University College London.

Gopaul, K.K., Brooks, P.C., Prost, J.-F. and Davis, E.O. (2003). Characterization of the Two *Mycobacterium tuberculosis* recA Promoters. *J. Bacteriol.* **185,** 6005-6015.

Gordon, S.V., Bottai, D., Simeone, R., Stinear, T.P. and Brosch, R. (2009). Pathogenicity in the tubercle bacillus: molecular and evolutionary determinants. *BioEssays* **31,** 378-388.

Graham, J.E. and Clark-Curtiss, J.E. (1999). Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proc Natl Acad Sci U S A* **96,** 11554-11559.

Greenstein, A.E., MacGurn, J.A., Baer, C.E., Falick, A.M., Cox, J.S. and Alber, T. (2007). *M. tuberculosis* Ser/Thr Protein Kinase D Phosphorylates an Anti-Anti-Sigma Factor Homolog. *PLoS Pathogens* **3,** e49.

Gruber, T.M. and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual Review of Microbiology* **57,** 441-466.

Gupta, U.D., Katoch, V.M. and McMurray, D.N. (2007). Current status of TB vaccines. *Vaccine* **25,** 3742-3751.

He, H., Hovey, R., Kane, J., Singh, V. and Zahrt, T.C. (2006). MprAB is a stress-responsive two-component system that directly regulates expression of sigma factors SigB and SigE in *Mycobacterium tuberculosis*. *J Bacteriol* **188,** 2134-2143.

Helmann, J.D. (1999). Anti-sigma factors. *Current Opinion in Microbiology* **2,** 135-141.

Helmann, J.D. (2002) The extracytoplasmic function (ECF) sigma factors. In *Advances in Microbial Physiology*. Academic Press, pp. 47-110.

Hemsworth, G.R. and Kochan, I. (1978). Secretion of Antimycobacterial Fatty Acids by Normal and Activated Macrophages. *Infect. Immun.* **19,** 170-177.

Hestvik, A.L., Hmama, Z. and Av-Gay, Y. (2005). Mycobacterial manipulation of the host cell. *FEMS Microbiol Rev* **29,** 1041-1050.

Houben, E.N.G., Nguyen, L. and Pieters, J. (2006). Interaction of pathogenic mycobacteria with the host immune system. *Current Opinion in Microbiology* **9,** 76-85.

Hughes, K.T. and Mathee, K. (1998). The anti-sigma factors. *Annual Review of Microbiology* **52,** 231-286.

Iacazio, G. (2005). Increased quercetinase production by *Penicillium olsonii* using fractional factorial design. *Process Biochemistry* **40,** 379-384.

Inoue, Y. and Kimura, A. (1995). Methylglyoxal and regulation of its metabolism in microorganisms. *Adv Microb Physiol* **37,** 177-227.

Jacobs, W.R., Kalpana, G.V., Cirillo, J.D., Pascopella, L., Snapper, S.B., Udani, R.A., *et al.* (1991) Genetic systems for mycobacteria. In *Methods in Enzymology*, H.M. Jeffrey (ed.). Academic Press, pp. 537-555.

James, P., Halladay, J. and Craig, E.A. (1996). Genomic Libraries and a Host Strain Designed for Highly Efficient Two-Hybrid Selection in Yeast. *Genetics* **144,** 1425-1436.

Karschau, J., de Almeida, C., Richard, M.C., Miller, S., Booth, I.R., Grebogi, C. and de Moura, A.P. (2011). A matter of life or death: modeling DNA damage and repair in bacteria. *Biophys J* **100,** 814-821.

Kazmierczak, M.J., Wiedmann, M. and Boor, K.J. (2005). Alternative Sigma Factors and Their Roles in Bacterial Virulence. *Microbiol. Mol. Biol. Rev.* **69,** 527-543.

Krymkiewicz, N. (1973). Reactions of methylglyoxal with nucleic acids. *FEBS Lett* **29,** 51-54.

Kubica, G.P., Kim, T.H. and Dunbar, F.P. (1972). Designation of Strain H37Rv as the Neotype of *Mycobacterium tuberculosis*. *Int J Syst Bacteriol* **22,** 99-106.

Lee, J.-H., Geiman, D.E. and Bishai, W.R. (2007). Role of stress response sigma factor, SigG, in *Mycobacterium tuberculosis*. *J. Bacteriol.***,** JB.00511-00507.

Lonetto, M.A., Brown, K.L., Rudd, K.E. and Buttner, M.J. (1994). Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci U S A* **91,** 7573-7577.

Madigan, M.T., Martinko, J. M. and Parker, J. (2000) Brock: Biology of Microorganisms 9th Edition New Jersey, Prentice Hall.

Manganelli, R., Dubnau, E., Tyagi, S., Kramer, F.R. and Smith, I. (1999). Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Molecular Microbiology* **31,** 715-724.

Manganelli, R., Proveddi, R., Rodrigue, S., Beaucher, J., Gaudreau, L. and Smith, I. (2004). σ Factors and Global Gene Regulation in *Mycobacterium tuberculosis*. *J. Bacteriol.* **186,** 895-902.

Manganelli, R., Voskuil, M.I., Schoolnik, G.K. and Smith, I. (2001). The *Mycobacterium tuberculosis* ECF sigma factor σE: role in global gene expression and survival in macrophages†. *Molecular Microbiology* **41,** 423-437.

Marrero, J., Rhee, K.Y., Schnappinger, D., Pethe, K. and Ehrt, S. (2010). Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for *Mycobacterium tuberculosis* to establish and maintain infection. *Proc Natl Acad Sci U S A.*

Masaki, S., Sugimori, G., Okamoto, A., Imose, J. and Hayashi, Y. (1991). Effect of Tween 80 on formation of the superficial L1 layer of the *Mycobacterium avium-Mycobacterium intracellulare* complex. *J. Clin. Microbiol.* **29,** 1453-1456.

McKinney, J.D. and Gomez, J.E. (2003). Life on the inside for *Mycobacterium tuberculosis*. *Nature Medicine* **9,** 1356-1357.

McKinney, J.D., Honer zu Bentrup, K., Munoz-Elias, E.J., Miczak, A., Chen, B., Chan, W.T., *et al.* (2000). Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406,** 735-738.

Mecsas, J., Rouviere, P.E., Erickson, J.W., Donohue, T.J. and Gross, C.A. (1993). The activity of sigma E, an *Escherichia coli* heat-inducible sigma-factor, is modulated by expression of outer membrane proteins. *Genes Dev* **7,** 2618-2628.

Miller, J.H. (1972) Experiments in molecular genetics, [Cold Spring Harbor, N.Y.] Cold Spring Harbor Laboratory.

Missiakas, D. and Raina, S. (1998) The extracytoplasmic function sigma factors: role and regulation. pp. 1059-1066.

Monahan, I.M., Betts, J., Banerjee, D.K. and Butcher, P.D. (2001). Differential expression of mycobacterial proteins following phagocytosis by macrophages. *Microbiology* **147,** 459-471.

Munoz-Elias, E.J. and McKinney, J.D. (2005). *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nat Med* **11,** 638-644.

Mycobrowser. http://mycobrowser.epfl.ch/. [WWW document]. URL

Nachin, L., Brive, L., Persson, K.-C., Svensson, P. and Nyström, T. (2008). Heterodimer Formation within Universal Stress Protein Classes Revealed By an In Silico and Experimental Approach. *Journal of Molecular Biology* **380,** 340-350.

O'Toole, R. and Williams, H.D. (2003). Universal stress proteins and *Mycobacterium tuberculosis*. *Res Microbiol* **154,** 387-392.

Paget, M.S.B., Bae, J.-B., Hahn, M.-Y., Li, W., Kleanthous, C., Roe, J.-H. and Buttner, M.J. (2001). Mutational analysis of RsrA, a zinc-binding anti-sigma factor with a thiol–disulphide redox switch. *Molecular Microbiology* **39,** 1036-1047.

Papavinasasundaram, K.G., Anderson, C., Brooks, P.C., Thomas, N.A., Movahedzadeh, F., Jenner, P.J., *et al.* (2001). Slow induction of RecA by DNA damage in *Mycobacterium tuberculosis*. *Microbiology* **147,** 3271-3279.

Parish, T., Smith, D.A., Kendall, S., Casali, N., Bancroft, G.J. and Stoker, N.G. (2003). Deletion of Two-Component Regulatory Systems Increases the Virulence of *Mycobacterium tuberculosis*. *Infect. Immun.* **71,** 1134-1140.

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

Park, H.D., Guinn, K.M., Harrell, M.I., Liao, R., Voskuil, M.I., Tompa, M., *et al.* (2003). Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol Microbiol* **48,** 833-843.

Pessi, G., Blumer, C. and Haas, D. (2001). lacZ fusions report gene expression, don't they? *Microbiology* **147,** 1993-1995.

Primm, T.P., Lucero, C.A. and Falkinham, J.O., III (2004). Health Impacts of Environmental Mycobacteria. *Clin. Microbiol. Rev.* **17,** 98-106.

Rachman, H., Kim, N., Ulrichs, T., Baumann, S., Pradl, L., Eddine, A.N., *et al.* (2006). Critical Role of Methylglyoxal and AGE in Mycobacteria-Induced Macrophage Apoptosis and Activation. *PLoS ONE* **1,** e29.

Raivio, T.L. and Silhavy, T.J. (2001). Periplasmic stress and ECF sigma factors. *Annual Review of Microbiology* **55,** 591-624.

Raman, S., Puyang, X., Cheng, T.-Y., Young, D.C., Moody, D.B. and Husson, R.N. (2006). *Mycobacterium tuberculosis* SigM Positively Regulates Esx Secreted Protein and Nonribosomal Peptide Synthetase Genes and Down Regulates Virulence-Associated Surface Lipid Synthesis. *J. Bacteriol.* **188,** 8460-8468.

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 Bacteriol* **183,** 6119-6125.

Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R.S. and Davis, E.O. (2003). The majority of inducible DNA repair genes in *Mycobacterium tuberculosis* are induced independently of RecA. *Molecular Microbiology* **50,** 1031-1042.

Ratushny, V. and Golemis, E. (2008). Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system. *Biotechniques* **44,** 655-662.

Rickman, L., Scott, C., Hunt, D.M., Hutchinson, T., Menéndez, M.C., Whalan, R., *et al.* (2005). A member of the cAMP receptor protein family of transcription regulators in *Mycobacterium tuberculosis* is required for virulence in mice and controls transcription of the *rpfA* gene coding for a resuscitation promoting factor. *Molecular Microbiology* **56,** 1274-1286.

Rodrigue, S., Brodeur, J., Jacques, P.-E., Gervais, A.L., Brzezinski, R. and Gaudreau, L. (2007). Identification of Mycobacterial σ Factor Binding Sites by Chromatin Immunoprecipitation Assays. *J. Bacteriol.* **189,** 1505-1513.

Rodrigue, S., Provvedi, R., Jacques, P.E., Gaudreau, L. and Manganelli, R. (2006). The sigma factors of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev* **30,** 926-941.

Russell, D.G. (2001). *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nat Rev Mol Cell Biol* **2,** 569-586.

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

Sachdeva, P., Misra, R., Tyagi, A.K. and Singh, Y. (2009). The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *FEBS Journal* **9999**.

Saïd-Salim, B., Mostowy, S., Kristof, A.S. and Behr, M.A. (2006). Mutations in *Mycobacterium tuberculosis Rv0444c*, the gene encoding anti-SigK, explain high level expression of MPB70 and MPB83 in Mycobacterium bovis. *Molecular Microbiology* **62,** 1251-1263.

Saito, H., Tomioka, H. and Yoneyama, T. (1984). Growth of group IV mycobacteria on medium containing various saturated and unsaturated fatty acids. *Antimicrob. Agents Chemother.* **26,** 164-169.

Sambrook, J. and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual New York, Cold Spring Harbor Laboratory Press.

Sancar, A. (1996). DNA excision repair. *Annu Rev Biochem* **65,** 43-81.

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

Sattler, T.H. and Youmans, G.P. (1948). The effect of tween 80, bovine albumin, glycerol, and glucose on the growth of *Mycobacterium tuberculosis* var. hominis (H37Rv). *J Bacteriol* **56,** 235-243.

Schiestl, R.H. and Gietz, R.D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet* **16,** 339-346.

Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., *et al.* (2003). Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages. *The Journal of Experimental Medicine* **198,** 693-704.

Schweizer, H.D. (1993). Small broad-host-range gentamycin resistance gene cassettes for site-specific insertion and deletion mutagenesis. *BioTechniques* **15,** 831-834.

Sherman, D.R., Voskuil, M., Schnappinger, D., Liao, R., Harrell, M.I. and Schoolnik, G.K. (2001). Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding α-crystallin. *Proc Natl Acad Sci U S A* **98,** 7534-7539.

Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68,** 850-858.

Sklar, J.G., Makinoshima, H., Schneider, J.S. and Glickman, M.S. (2010). *M. tuberculosis* intramembrane protease Rip1 controls transcription through three anti-sigma factor substrates. *Molecular Microbiology* **77,** 605-617.

Smith, I. (2003) *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. pp. 463-496.

Smollett, K.L., Dawson, L.F. and Davis, E.O. (2011). SigG Does Not Control Gene Expression in Response to DNA Damage in *Mycobacterium tuberculosis* H37Rv. *J Bacteriol* **193,** 1007-1011.

Smollett, K.L., Fivian-Hughes, A.S., Smith, J.E., Chang, A., Rao, T. and Davis, E.O. (2009). Experimental determination of translational start sites resolves uncertainties in genomic open reading frame predictions - application to *Mycobacterium tuberculosis*. *Microbiology* **155,** 186-197.

Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T. and Jacobs, W.R., Jr. (1990). Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4,** 1911-1919.

Song, T., Dove, S.L., Lee, K.H. and Husson, R.N. (2003). RshA, an anti-sigma factor that regulates the activity of the mycobacterial stress response sigma factor SigH. *Molecular Microbiology* **50,** 949-959.

Soo, P.C., Horng, Y.T., Lai, M.J., Wei, J.R., Hsieh, S.C., Chang, Y.L., *et al.* (2007). Pirin regulates pyruvate catabolism by interacting with the pyruvate dehydrogenase E1 subunit and modulating pyruvate dehydrogenase activity. *J Bacteriol* **189,** 109-118.

Sorensen, H. and Mortensen, K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial Cell Factories* **4,** 1.

Springer, B., Sander, P., Sedlacek, L., Ellrott, K. and Bottger, E.C. (2001). Instability and site-specific excision of integration-proficient mycobacteriophage L5 plasmids: development of stably maintained integrative vectors. *International Journal of Medical Microbiology* **290,** 669-675.

Storz, G. and Zheng, M. (2000) Oxidative stress. In *Bacterial Stress Responses*, G. Storz, R. Hengge-Aronis (eds.). Washington, DC, ASM, pp. 47-59.

Thakur, K.G., Praveena, T. and Gopal, B. (2010). Structural and biochemical bases for the redox sensitivity of *Mycobacterium tuberculosis* RslA. *J Mol Biol* **397,** 1199-1208.

Tian, J., Bryk, R., Itoh, M., Suematsu, M. and Nathan, C. (2005). Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: Identification of α-ketoglutarate decarboxylase. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 10670-10675.

Tse-Dinh, Y.C. (2009). Bacterial topoisomerase I as a target for discovery of antibacterial compounds. *Nucleic Acids Res* **37,** 731-737.

TubercuList. http://tuberculist.epfl.ch/. [WWW document]. URL

Van Criekinge, W. and Beyaert, R. (1999). Yeast Two-Hybrid: State of the Art. *Biol Proced Online* **2,** 1-38.

von Groll, A., Martin, A., Stehr, M., Singh, M., Portaels, F., da Silva, P.E.A. and Palomino, J.C. (2010). Fitness of *Mycobacterium tuberculosis* Strains of the W-Beijing and Non-W-Beijing Genotype. *PLoS ONE* **5,** e10191.

Voskuil, M.I., Visconti, K.C. and Schoolnik, G.K. (2004). *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis (Edinburgh, Scotland)* **84,** 218-227.

Walker, G.C., Smith, B.T. and Sutton, M.D. (2000) The SOS response to DNA damage. In *Bacterial Stress Responses*, G. Storz, R. Hengge-Aronis (eds.). Washington, DC, ASM, pp. 131-144.

Wang, Y., Cui, T., Zhang, C., Yang, M., Huang, Y., Li, W., *et al.* (2010). Global protein-protein interaction network in the human pathogen *Mycobacterium tuberculosis* H37Rv. *J Proteome Res* **9,** 6665-6677.

Warner, J.B. and Lolkema, J.S. (2002). LacZ-promoter fusions: the effect of growth. *Microbiology* **148,** 1241-1243.

West, A.H. and Stock, A.M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26,** 369-376.

WHO (2009) Global tuberculosis control - epidemiology, strategy, financing.

Yeh, H.-Y., Chen, T.-C., Liou, K.-M., Hsu, H.-T., Chung, K.-M., Hsu, L.-L. and Chang, B.-Y. (2011). The core-independent promoter-specific interaction of primary sigma factor. *Nucleic Acids Research* **39,** 913-925.

Yoshimura, M., Asai, K., Sadaie, Y. and Yoshikawa, H. (2004). Interaction of *Bacillus subtilis* extracytoplasmic function (ECF) sigma factors with the N-terminal regions of their potential anti-sigma factors. *Microbiology* **150,** 591-599.

Yura, T., Kanemori, M. and Morita, M.T. (2000) The heat shock response; regulation and function. In *Bacterial Stress Responses*, G. Storz, R. Hengge-Aronis (eds.). Washington, DC, ASM, pp. 3-18.

Zahrt, T.C. and Deretic, V. (2000). An Essential Two-Component Signal Transduction System in *Mycobacterium tuberculosis*. *J. Bacteriol.* **182,** 3832-3838.

Zink, A.R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H. and Nerlich, A.G. (2003). Characterization of *Mycobacterium tuberculosis* Complex DNAs from Egyptian Mummies by Spoligotyping. *J. Clin. Microbiol.* **41,** 359-367.

Appendix I – Media/Buffer Composition

Luria-Bertani Broth/Agar

10 g Tryptone

5 g Yeast Extract

10 g NaCl

15 g Agar (if required)

dH_2O to 1 litre

Autoclave at 121 ºC for 15 mins

Mycobacteria 7H11 Agar

21 g 7H11 Medium Powder (Middlebrook)

5 ml Glycerol

dH_2O to 960 ml

Autoclave at 121 ºC for 15 mins

40 ml Dubos Medium Albumin (Difco)

Modified Dubos Medium

1 g KH_2PO_4

6.25 g $Na_2HPO_4.12H_2O$

1.25 g Na_3 Citrate

0.6 g $MgSO_4.7H_2O$

2 g Asparagine

2 g Casamino Acids

5 ml 10% Tween 80

2 ml Glycerol

dH_2O to 960 mls

Adjust to pH 7.2

Autoclave at 121 ºC for 15 mins

40 ml Dubos Medium Albumin (Difco)

Fatty Acid Susceptibility Medium

1 g	KH_2PO_4
6.25 g	$Na_2HPO_4.12H_2O$
1.25 g	Na_3 Citrate
0.6 g	$MgSO_4.7H_2O$
2 g	Asparagine
2 g	Casamino Acids
2 ml	10 % Tyloxapol (Sigma)
2 ml	Glycerol

dH_2O to 1 litre

Adjust to pH 7.2

Autoclave at 121 ºC for 15 mins

Carbon Source Medium

4.7 g	7H9 Broth Powder (Difco, Middlebrook)
0.85 g	NaCl
1 g	Carbon Source
5 ml	10 % Tyloxapol (Sigma)

dH_2O to 995 ml

Autoclave at 121 ºC for 10 mins

5 ml Dubos Medium Albumin (Difco)

Yeast Nitrogen Base-leucine-tryptophan

6.7 g	Yeast nitrogen base without amino acids
20 g	Agar
850 ml	dH_2O

Autoclave at 121 ºC for 15 mins

100 ml	10X sterile –leucine-tryptophan dropout medium (Clontech)
40 ml	50% sterile Glucose

dH_2O to 1 litre if necessary

Freeze Mix

12.6 g	K_2HPO_4
0.9 g	Na_3 Citrate
0.18 g	$MgSO_4.7H_2O$
1.8 g	$(NH_4)_2SO_4$
3.6 g	KH_2PO_4
96 g	Glycerol

dH_2O to 1 litre

Autoclave at 121 ºC for 15 mins

TBE Buffer (10X)

121 g	Tris Base
61.83 g	Boric acid
18.612 g	EDTA

dH_2O to 1 litre

Adjust to pH 8.0

PBS Buffer/PBS-Tween

10 g	NaCl
0.25 g	KCl
1.437 g	Na_2HPO_4
0.25 g	KH_2PO_4
0.5 ml	Tween 20 if required

dH_2O to 1 litre

His-Binding Buffer (1X)

0.5 M	NaCl
20 mM	Tris-HCl
5 mM	Imidazole

Adjust pH to 7.9

His-Elution Buffer (1X)

0.5 M	NaCl
20 mM	Tris-HCl
1 M	Imidazole

Adjust pH to 7.9

Z-Buffer

0.06 M	Na_2HPO_4
0.04 M	NaH_2PO_4
0.01 M	KCl
1 mM	$MgSO_4$
0.05 M	β-mercaptoethanol

Adjust to pH 7.0

Blocking Solution

2.5 g	Milk Powder (Marvel)
50 ml	PBS-Tween

Semi-dry Transfer Buffer

2.9 g	Glycine
5.8g	Tris Base
1.85 ml	20% SDS
200ml	Methanol

dH_2O to 1 litre

Adjust to pH 8.3

In vitro Transcription Master Mix (2X)

1.21 g	Tris Base
9.3 mg	Na_2EDTA
487.2 mg	$MgCl_2.6H_2O$

dH_2O to 35 ml

4 mM Potassium Phosphate to 60 ml

Adjust pH to 7.9

dH_2O to 70 ml

In vitro Transcription Reaction Mix

115 µl	*In vitro* Transcription Master Mix (2X)
300 µg	BSA (RNase-free)
2.4 mM	DTT
0.8 mM	EDTA
80 µl	Sterile 100 % Glycerol

DEPC-H_2O to 250 µl

In vitro Transcription Precipitation Mix

100 µg	tRNA
2M	Sodium Acetate

dH_2O to 80 µl

Run-off Loading Buffer

9.5 ml	Formamide
20 mM	EDTA
0.005 mg	Bromophenol Blue
0.005 mg	Xylene Cyanol

DEPC-H_2O to 10 ml

Appendix II

PCR Primers

Cloning of knockout vector			**Product Size (bp)**
KO5'F	5' flanking region introducing *Sac*I site	TTGAGCTCGAGGTAGGCGTCGGATGC	
KO5'R	5' flanking region introducing *Xba*I site	GGGTCTAGATCGCGGCCAGAGTTTTC	1390
KO3'F	3' flanking region introducing *Xba*I site	GGTCTAGAGCTCGATGGGCCTGTTGGT	
KO3'R	3' flanking region introducing *Bam*HI site	TGGGATCCGTCCGGGGTGTAGTTGC	1384
Cloning of translational start site vectors			
TlnF	Introducing *Eco*RV site	GGGATATCTAAGCCTGGACCACCTCAG	
TlnRSG	Introducing *Xba*I site	GTTCTAGACCCAGCGAATCGGGCA	1598
TlnR181	Introducing *Xba*I site	GGTCTAGATGTTGCCGAATCGCCC	2349
TlnR180	Introducing *Xba*I site	CGTCTAGACGGTTGTTCGCTTGAGC	3734
Cloning of promoter region for pAG04			
SigGP1-3F	Promoter region introducing *Xba*I site	TTTCTAGAGTCAGCGGTGTACTCGGAG	
SigGP1-3R	Promoter region introducing *Hin*dIII site	GGAAGCTTAATGAGCCTACGCAGAGTC	335
Cloning of yeast 2 hybrid vectors			
Y2HSGF	Y2H SigG PCR primers	TTTGGATCCGTGAGTGTGCTCGCAG	
Y2HSGR		GCCAGATCTTCACAGCGAATCGGG	1113
Y2H181F	Y2H Rv0181c PCR primers	TTTGGATCCATGACCGCCACCG	
Y2H181R		GCGAGATCTCTATGTTGCCGAATCGC	735
Y2H180F1	Y2H Full Rv0180c PCR primers	CGTGGATCCATGTCTCAAGCGCAG	
Y2H180R1		GCGAGATCTTTACGGTTGTTCGCTTG	1359
Y2H180F2	Y2H Cytoplasmic Rv0180c PCR primers	GGCGGATCCATCCTAAACCCCATG	
Y2H180R2		TTTAGATCTCGACAGCCCGTTGCC	510
Cloning into protein expression vectors			
SigGF	Introducing *Bam*HI site	GAGGATCCGTGAGTGTGCTCGCAG	
SigGR	Introducing *Xho*I site	CCCTCGAGCTATCACAGCGAATCG	1018
Rv0181cF	Introducing *Bam*HI site	AAGGATCCATGACCGCCACCG	
Rv0181cR	Introducing *Xho*I site	CACTCGAGCTATGTTGCCGAATCG	751
***In vitro* transcription and protein-DNA binding assay templates**			
Rv0887cF	IVT template	GCAGCAATCACGTTGTGACGG	455
Rv0887cR	IVT template	GAATGGTGACCGGTGTTCCG	
Rv0911F	IVT template	CTCAACGGCGACGGTGTG	557
Rv0911R	IVT template	CCACGGTCGCACCGATGTCG	
rrnBF	Protein-DNA binding assay probe	GGATTTGACTCCCAGTTTCCAAGGAC	249
rrnBR	Protein-DNA binding assay probe	TAAACGGGAAAAAGAGGCGGACAAA	

PCR confirmation of knockout mutant

PCR confirmation of KO mutant		
KO1F	Double XO PCR primers	GTGGCGGTTTTCATGGCTTGTTAT
KO1R	Double XO PCR primers	AGATCTCACTACGCGGCTGCTCA
KO2F	Double XO PCR primers	GTTCAGTCCGGGTGGTAGTCATTA
KO2R	Double XO PCR primers	AAACAAACAGGCCAAGATTACCAA
KO3F	Double XO PCR primers	ACGCCAGCTGGAAGTTGAGTCG
KO3R	Double XO PCR primers	TGGCCGCCGTTTGTTCATCTC

5’RACE primers

5'RACE primers		
GeneRacer RNA 5' oligo	5'RACE primers	CGACTGGAGCACGAGGACACTGA
GSPLacR	5'RACE primers	AGGTAGTCACGCAACTCGCCGCACATCTGAA
RACE887	5'RACE primers	GCCCCAGTTCGACGGCGTCAAAG
RACE911	5'RACE primers	GACGAACCGCCCGCGGCAGCCAACTT

Sequencing Primers

Sequencing primers		
pBbF	pBackbone sequencing primers	TCGCCCTGATAGACGGTTTT
pBbR	pBackbone sequencing primers	CACTTTCTGGCTGGATGATGG
pMINT2	General sequencing primers	CAATCTGGTGTGAATGCCCCTCGT
KS306	General sequencing primers	GCCTTTGAGTGAGCTGATACCGCT
306-4	General sequencing primers	CGTAAAAAGGCCGCGTTGCTGG
Af	SigG operon sequencing primers	ACCAACAGTGCCCCGCCTATT
Ar	SigG operon sequencing primers	GCCGCACGCCTTATCTCG
Bf	SigG operon sequencing primers	CCTACGGCCGCCTCACTCTGGA
Br	SigG operon sequencing primers	GCGTCTGGCCGGCGATGCTA
Cf	SigG operon sequencing primers	AGCCGTACCAGCGTTCTTCC
pGEX 5'	pGEX-6P-1 commercial primer	GGGCTGGCAAGCCACGTTTGGTG
pGEX 3'	pGEX-6P-1 commercial primer	CCGGGAGCTGCATGTGTCAGAGG
T7 promoter (F)	pET28a sequencing primer	TAATACGACTCACTATAGGG
T7 terminator (R)	pET28a sequencing primer	GCTAGTTATTGCTCAGCGG
lacR	lacZ-specific sequencing primer	TTCCCAGTCACGACGTTGTAAA

RT-PCR primers

Af	Co-transcription primers	ACCAACAGTGCCCCGCCTATT
Ar	Co-transcription primers	GCCGCACGCCTTATCTCG
Bf	Co-transcription primers	CCTACGGCCGCCTCACTCTGGA
Br	Co-transcription primers	GCGTCTGGCCGGCGATGCTA
Cf	Co-transcription primers	AGCCGTACCAGCGTTCTTCC
RT-PCRSGF	Co-transcription primers	TTTGGATCCGTGAGTGTGCTCGCAG
*Rv0180c*R	Co-transcription primers	CTTTGGCGAGCTTGGTATTGAG
*lpr*OR1	Co-transcription primers	TGGCCGCCGTTGTTCATCTC
*lpr*OR2	Co-transcription primers	TGGGATCCGTCCGGGGTGTAGTTGC

qRT-PCR primers

qRT-PCR primers	
qRTSigAF	TCGGTTCGCGCCTACCT
qRTSigAR	TGGCTAGCTCGACCTCTTCCT
qRTSigGF	TGAACTGCTCGCACACTGCTA
qRTSigGR	AGCGTCTCCTGAACAAGGTCTT
qRTSigGP2F	GGAGACTCTGCGTAGGCTCATT
qRTSigGP2R	TCGCGGCCAGAGTTTTCT
qRTRv0181F	CTGGCCCAACGCATGTC
qRTRv0181R	TCGGTAGCAGAATCGTTCATTTC
qRTRv0180F	TGGCCGCGCTCTATCTG
qRTRv0180R	AACGCGATTGGGAAATGC
qRTlprOF	TGCTCGCCCAAGCGATAG
qRTlprOR	CAGTGACCACCTGCGTTGAG
qRT0654F	CGTCGGACGGTTGATATCG
qRT0654R	GGTCGTAGATCACCACGTAGTTGT

SDM primers

SDM primers	
Translational start sites	
SDMSGF1	GCGTAGGCTCATTGAGTGAGTGTGCTCGCAGAAAACTCTGGC
SDMSGR1	GCCAGAGTTTTCTGCGAGCACACTCACTCAATGAGCCTACGC
SDMSGF2	GGCTCATTGACGTGAGGTGCTCGCAGAAAACTCTGGCC
SDMSGR2	GGCCAGAGTTTTCTGCGAGCACCTCACGTCAATGAGCC
SDMSGF3	GCTCATTGACGTGAGTGTGCCGCAGAAAACTCTGGCC
SDMSGR3	GGCCAGAGTTTTCTGCGGCACACTCACGTCAATGAGC
SDM181F1	GCTGTGATGGCGAACGGACGTCATGACCGCC
SDM181R1	GGCGGTCATGACGTCCGTTCGCCATCACAGC
SDM181F2	CGCCACCGTCGAGAAAGGCGTGCGGCCG
SDM181R2	CGGCCGCACGCCTTTCTCGACGGTGGCG
SDM180F1	GGAGCCCGCATGTCCAAGCGCAGCCGC
SDM180R1	GCGGCTGCGCTTGGACATGCGGGCTCC
pAG04 Promoter mutations	
P1SDMF	CGGAGACTCTGCGTCGGCTCATTAAGC
P1SDMR	GCTTAATGAGCCGACGCAGAGTCTCCG
P2SDMF1	CCGTTCAGTCCGGGTGTGCGTCATTACCGGGAGCG
P2SDMR1	CGCTCCCGGTAATGACGCACACCCGGACTGAACGG
P2SDMF2	CCGTTCAGTCCGGGCCCCCCCCATTACCGGGAGCG
P2SDMR2	CGCTCCCGGTAATGGGGGGGGCCCGGACTGAACGG
P3SDMF1	GCTTGCGGCGCGGTTGTCGGCGTCCAGACG
P3SDMR1	CGTCTGGACGCCGACAACCGCGCCGCAAGC
P3SDMF2	GCTTGCGGCGCGGTGTCTCTCGTCCAGACGTCG
P3SDMR2	CGACGTCTGGACGAGAGACACCGCGCCGCAAGC
P3SDMF3	CTTGCGGCGCGGTCCCCCCCCCTCCAGACGTCGTAG
P3SDMR3	CTACGACGTCTGGAGGGGGGGGACCGCGCCGCAAG

Appendix III – Biolog Phenotype Microarray Plates

Phenotype MicroArrays™

PM1 MicroPlate™ Carbon Sources

A1 Negative Control	A2 L-Arabinose	A3 N-Acetyl-D-Glucosamine	A4 D-Saccharic Acid	A5 Succinic Acid	A6 D-Galactose	A7 L-Aspartic Acid	A8 L-Proline	A9 D-Alanine	A10 D-Trehalose	A11 D-Mannose	A12 Dulcitol
B1 D-Serine	B2 D-Sorbitol	B3 Glycerol	B4 L-Fucose	B5 D-Glucuronic Acid	B6 D-Gluconic Acid	B7 D,L-α-Glycerol-Phosphate	B8 D-Xylose	B9 L-Lactic Acid	B10 Formic Acid	B11 D-Mannitol	B12 L-Glutamic Acid
C1 D-Glucose-6-Phosphate	C2 D-Galactonic Acid-γ-Lactone	C3 D,L-Malic Acid	C4 D-Ribose	C5 Tween 20	C6 L-Rhamnose	C7 D-Fructose	C8 Acetic Acid	C9 α-D-Glucose	C10 Maltose	C11 D-Melibiose	C12 Thymidine
D-1 L-Asparagine	D2 D-Aspartic Acid	D3 D-Glucosaminic Acid	D4 1,2-Propanediol	D5 Tween 40	D6 α-Keto-Glutaric Acid	D7 α-Keto-Butyric Acid	D8 α-Methyl-D-Galactoside	D9 α-D-Lactose	D10 Lactulose	D11 Sucrose	D12 Uridine
E1 L-Glutamine	E2 M-Tartaric Acid	E3 D-Glucose-1-Phosphate	E4 D-Fructose-6-Phosphate	E5 Tween 80	E6 α-Hydroxy Glutaric Acid-γ-Lactone	E7 α-Hydroxy Butyric Acid	E8 β-Methyl-D-Glucoside	E9 Adonitol	E10 Maltotriose	E11 2-Deoxy Adenosine	E12 Adenosine
F1 Glycyl-L-Aspartic Acid	F2 Citric Acid	F3 M-Inositol	F4 D-Threonine	F5 Fumaric Acid	F6 Bromo Succinic Acid	F7 Propionic Acid	F8 Mucic Acid	F9 Glycolic Acid	F10 Glyoxylic Acid	F11 D-Cellobiose	F12 Inosine
G1 Glycyl-L-Glutamic Acid	G2 Tricarballylic Acid	G3 L-Serine	G4 L-Threonine	G5 L-Alanine	G6 L-Alanyl-Glycine	G7 Acetoacetic Acid	G8 N-Acetyl-β-D-Mannosamine	G9 Mono Methyl Succinate	G10 Methyl Pyruvate	G11 D-Malic Acid	G12 L-Malic Acid
H1 Glycyl-L-Proline	H2 p-Hydroxy Phenyl Acetic Acid	H3 m-Hydroxy Phenyl Acetic Acid	H4 Tyramine	H5 D-Psicose	H6 L-Lyxose	H7 Glucuronamide	H8 Pyruvic Acid	H9 L-Galactonic Acid-γ-Lactone	H10 D-Galacturonic Acid	H11 Phenylethyl-amine	H12 2-Aminoethanol

PM2A MicroPlate™ Carbon Sources

A1 Negative Control	A2 Chondroitin Sulfate C	A3 α-Cyclodextrin	A4 β-Cyclodextrin	A5 γ-Cyclodextrin	A6 Dextrin	A7 Gelatin	A8 Glycogen	A9 Inulin	A10 Laminarin	A11 Mannan	A12 Pectin
B1 N-Acetyl-D-Galactosamine	B2 N-Acetyl-Neuraminic Acid	B3 β-D-Allose	B4 Amygdalin	B5 D-Arabinose	B6 D-Arabitol	B7 L-Arabitol	B8 Arbutin	B9 2-Deoxy-D-Ribose	B10 i-Erythritol	B11 D-Fucose	B12 3-O-β-D-Galacto-pyranosyl-D-Arabinose
C1 Gentiobiose	C2 L-Glucose	C3 Lactitol	C4 D-Melezitose	C5 Maltitol	C6 α-Methyl-D-Glucoside	C7 β-Methyl-D-Galactoside	C8 3-Methyl Glucose	C9 β-Methyl-D-Glucuronic Acid	C10 α-Methyl-D-Mannoside	C11 β-Methyl-D-Xyloside	C12 Palatinose
D1 D-Raffinose	D2 Salicin	D3 Sedoheptulosan	D4 L-Sorbose	D5 Stachyose	D6 D-Tagatose	D7 Turanose	D8 Xylitol	D9 N-Acetyl-D-Glucosaminitol	D10 γ-Amino Butyric Acid	D11 α-Amino Valeric Acid	D12 Butyric Acid
E1 Capric Acid	E2 Caproic Acid	E3 Citraconic Acid	E4 Citramalic Acid	E5 D-Glucosamine	E6 2-Hydroxy Benzoic Acid	E7 4-Hydroxy Benzoic Acid	E8 β-Hydroxy Butyric Acid	E9 γ-Hydroxy Butyric Acid	E10 α-Keto Valeric Acid	E11 Itaconic Acid	E12 5-Keto-D-Gluconic Acid
F1 D-Lactic Acid Methyl Ester	F2 Malonic Acid	F3 Melibionic Acid	F4 Oxalic Acid	F5 Oxalomalic Acid	F6 Quinic Acid	F7 D-Ribono-1,4-Lactone	F8 Sebacic Acid	F9 Sorbic Acid	F10 Succinamic Acid	F11 D-Tartaric Acid	F12 L-Tartaric Acid
G1 Acetamide	G2 L-Alaninamide	G3 N-Acetyl-L-Glutamic Acid	G4 L-Arginine	G5 Glycine	G6 L-Histidine	G7 L-Homoserine	G8 Hydroxy-L-Proline	G9 L-Isoleucine	G10 L-Leucine	G11 L-Lysine	G12 L-Methionine
H1 L-Ornithine	H2 L-Phenylalanine	H3 L-Pyroglutamic Acid	H4 L-Valine	H5 D,L-Carnitine	H6 Sec-Butylamine	H7 D,L-Octopamine	H8 Putrescine	H9 Dihydroxy Acetone	H10 2,3-Butanediol	H11 2,3-Butanone	H12 3-Hydroxy 2-Butanone

PM4A MicroPlate™ Phosphorus and Sulfur Sources

A1 Negative Control	A2 Phosphate	A3 Pyrophosphate	A4 Trimeta-phosphate	A5 Tripoly-phosphate	A6 Triethyl Phosphate	A7 Hypophosphite	A8 Adenosine- 2'-monophosphate	A9 Adenosine- 3'-monophosphate	A10 Adenosine- 5'-monophosphate	A11 Adenosine- 2',3'-cyclic monophosphate	A12 Adenosine- 3',5'-cyclic monophosphate
B1 Thiophosphate	B2 Dithiophosphat e	B3 D,L-α-Glycerol Phosphate	B4 β-Glycerol Phosphate	B5 Carbamyl Phosphate	B6 D-2-Phospho-Glyceric Acid	B7 D-3-Phospho-Glyceric Acid	B8 Guanosine- 2'-monophosphate	B9 Guanosine- 3'-monophosphate	B10 Guanosine- 5'-monophosphate	B11 Guanosine- 2',3'-cyclic monophosphate	B12 Guanosine- 3',5'-cyclic monophosphate
C1 Phosphoenol Pyruvate	C2 Phospho-Glycolic Acid	C3 D-Glucose-1-Phosphate	C4 D-Glucose-6-Phosphate	C5 2-Deoxy-D-Glucose 6-Phosphate	C6 D-Glucosamine-6-Phosphate	C7 6-Phospho-Gluconic Acid	C8 Cytidine- 2'-monophosphate	C9 Cytidine- 3'-monophosphate	C10 Cytidine- 5'-monophosphate	C11 Cytidine- 2',3'-cyclic monophosphate	C12 Cytidine- 3',5'-cyclic monophosphate
D1 D-Mannose-1-Phosphate	D2 D-Mannose-6-Phosphate	D3 Cysteamine-S-Phosphate	D4 Phospho-L-Arginine	D5 O-Phospho-D-Serine	D6 O-Phospho-L-Serine	D7 O-Phospho-L-Threonine	D8 Uridine- 2'-monophosphate	D9 Uridine- 3'-monophosphate	D10 Uridine- 5'-monophosphate	D11 Uridine- 2',3'-cyclic monophosphate	D12 Uridine- 3',5'-cyclic monophosphate
E1 O-Phospho-D-Tyrosine	E2 O-Phospho-L-Tyrosine	E3 Phosphocreatin e	E4 Phosphoryl Choline	E5 O-Phosphoryl-Ethanolamine	E6 Phosphono Acetic Acid	E7 2-Aminoethyl Phosphonic Acid	E8 Methylene Diphosphonic Acid	E9 Thymidine- 3'-monophosphate	E10 Thymidine- 5'-monophosphate	E11 Inositol Hexaphosphate	E12 Thymidine 3',5'- cyclic monophosphate
F1 Negative Control	F2 Sulfate	F3 Thiosulfate	F4 Tetrathionate	F5 Thiophosphate	F6 Dithiophosphat e	F7 L-Cysteine	F8 D-Cysteine	F9 L-Cysteinyl-Glycine	F10 L-Cysteic Acid	F11 Cysteamine	F12 L-Cysteine Sulfinic Acid
G1 N-Acetyl-L-Cysteine	G2 S-Methyl-L-Cysteine	G3 Cystathionine	G4 Lanthionine	G5 Glutathione	G6 D,L-Ethionine	G7 L-Methionine	G8 D-Methionine	G9 Glycyl-L-Methionine	G10 N-Acetyl-D,L-Methionine	G11 L- Methionine Sulfoxide	G12 L-Methionine Sulfone
H1 L-Djenkolic Acid	H2 Thiourea	H3 1-Thio-β-D-Glucose	H4 D,L-Lipoamide	H5 Taurocholic Acid	H6 Taurine	H7 Hypotaurine	H8 p-Amino Benzene Sulfonic Acid	H9 Butane Sulfonic Acid	H10 2-Hydroxyethane Sulfonic Acid	H11 Methane Sulfonic Acid	H12 Tetramethylene Sulfone

Appendix IV – Biolog Phenotype Microarray Data

Substrates with a significant fold change between the *M. tuberculosis* wild-type H37Rv and ΔsigGoperon deletion strains – PMs 1, 2 & 4

		WT			ΔSigGoperon			Fold Change				
PM	Compound	1	2	3	1	2	3	1	2	3	Average Fold Change	p-value
4	B9 Guanosine3'monophosphate	0.048	0.190		0.019	0.028		0.403	0.148		0.275	0.408
4	E8 Methylene Diphosphonic Acid	0.171	0.199		0.059	0.054		0.345	0.272		0.308	0.061
4	B12 Guanosine3',5'-cyclic monophosphate	0.086	0.071		0.048	0.014		0.558	0.199		0.378	0.178
1	D5 Tween 40	0.509	0.530	0.371	0.197	0.182	0.174	0.386	0.343	0.470	0.400	0.028
4	B7 D-3-Phospho-Glyceric Acid	0.071	0.168		0.064	0.011		0.904	0.064		0.484	0.311
1	E5 Tween 80	0.450	0.470	0.517	0.194	0.283	0.279	0.430	0.602	0.539	0.524	0.005
4	B3 D,L-α -Glycerol Phosphate	0.111	0.102		0.094	0.022		0.849	0.220		0.535	0.405
4	C7 6-Phospho-Gluconic Acid	0.100	0.150		0.077	0.050		0.774	0.332		0.553	0.203
4	B5 Carbamyl Phosphate	0.063	0.138		0.071	0.002		1.132	0.018		0.575	0.339
4	C6 DGlucosamine-6Phosphate	0.083	0.135		0.077	0.032		0.927	0.239		0.583	0.257
2	G2 L-Alaninamide	0.106	0.194		0.068	0.103		0.642	0.531		0.587	0.361
4	A5 Tripoly-phosphate	0.133	0.113		0.112	0.038		0.840	0.339		0.590	0.407
4	E11 Inositol Hexaphosphate	0.151	0.129		0.128	0.282		0.843	2.182		1.512	0.556
1	G12 L-Malic Acid	0.093	0.104	0.095	0.105	0.173	0.170	1.134	1.656	1.783	1.525	0.141
2	B4 Amygdalin	0.049	0.092		0.064	0.162		1.315	1.759		1.537	0.539
2	C11 β-Methyl-D-Xyloside	0.032	0.043		0.060	0.053		1.867	1.240		1.553	0.120
4	D10 Uridine5'monophosphate	0.053	0.097		0.060	0.190		1.146	1.971		1.558	0.576

Substrates with a significant fold change between the *M. tuberculosis* wild-type H37Rv and ΔsigGoperon deletion strains – PMs 1, 2 & 4 continued

4	F7 L-Cysteine	0.100	0.110		0.107	0.226		1.067	2.054		1.561	0.489	
4	B11 Guanosine2',3'-cyclic monophosphate	0.045	0.046		0.020	0.122		0.458	2.675		1.566	0.698	
4	H4 D,L-Lipoamide	0.088	0.045		0.099	0.091		1.124	2.018		1.571	0.404	
4	C10 Cytidine5'monophosphate	0.042	0.083		0.035	0.192		0.824	2.320		1.572	0.633	
1	B12 L-Glutamic Acid	0.082	0.078	0.049	0.109	0.081	0.119	1.328	1.035	2.414	1.592	0.100	
1	D12 Uridine	0.090	0.106	0.037	0.071	0.101	0.115	0.786	0.953	3.108	1.616	0.517	
4	F3 Thiosulfate	0.024	0.022		0.041	0.035		1.684	1.554		1.619	0.107	
1	H11 Phenylethylamine	0.055	0.114	0.102	0.117	0.157	0.154	2.112	1.384	1.514	1.670	0.082	
2	E6 2-Hydroxy Benzoic Acid	0.081	0.032		0.105	0.066		1.292	2.050		1.671	0.460	
4	F11 Cysteamine	0.105	0.088		0.081	0.230		0.768	2.617		1.692	0.574	
4	E12 Thymidine 3',5'-cyclic monophosphate	0.088	0.098		0.097	0.228		1.110	2.330		1.720	0.478	
1	H9 L-Galactonic Acid-γ -Lactone	0.121	0.061	0.088	0.089	0.117	0.219	0.734	1.927	2.500	1.720	0.323	
2	F11 D-Tartaric Acid	0.120	0.027		0.114	0.068		0.955	2.539		1.747	0.774	
2	B6 D-Arabitol	0.023	0.016		0.018	0.043		0.800	2.711		1.755	0.527	
2	F4 Oxalic Acid	0.088	0.114		0.103	0.275		1.177	2.418		1.798	0.488	
4	G10 N-Acetyl-D,L-Methionine	0.091	0.100		0.129	0.224		1.416	2.237		1.826	0.337	
2	E10 α -Keto Valeric Acid	0.108	0.029		0.124	0.077		1.145	2.633		1.889	0.573	
2	F7 D-Ribono-1,4Lactone	0.072	0.027		0.091	0.070		1.270	2.618		1.944	0.377	
4	B1 Thiophosphate	0.062	0.035		0.106	0.083		1.710	2.387		2.049	0.125	
2	G12 L-Methionine	0.132	0.059		0.165	0.170		1.251	2.885		2.068	0.299	
4	G9 Glycyl-L-Methionine	0.097	0.084		0.105	0.260		1.077	3.102		2.090	0.446	
2	C4 D-Melezitose	0.039	0.087		0.122	0.099		3.137	1.135		2.136	0.265	

Substrates with a significant fold change between the *M. tuberculosis* wild-type H37Rv and ΔsigGoperon deletion strains – PMs 1, 2 & 4 continued

4	F6 Dithiophosphate	0.032	0.040		0.030	0.133		0.936	3.351		2.144	0.538	
1	C12 Thymidine	0.020	0.110	0.066	0.081	0.099	0.109	4.002	0.901	1.642	2.182	0.356	
4	G12 L-Methionine Sulfone	0.043	0.033		0.019	0.133		0.447	4.036		2.241	0.624	
2	F10 Succinamic Acid	0.103	0.023		0.094	0.089		0.913	3.921		2.417	0.603	
4	C11 Cytidine2’,3’cyclic monophosphate	0.083	0.045		0.045	0.202		0.536	4.458		2.497	0.588	
4	G8 D-Methionine	0.114	0.051		0.095	0.214		0.831	4.181		2.506	0.427	
4	G7 L-Methionine	0.076	0.046		0.097	0.176		1.282	3.796		2.539	0.279	
4	F5 Thiophosphate	0.047	0.043		0.045	0.192		0.947	4.499		2.723	0.501	
2	E7 4-Hydroxy Benzoic Acid	0.112	0.008		0.126	0.037		1.132	4.629		2.880	0.780	
2	F12 L-Tartaric Acid	0.068	0.055		0.077	0.268		1.120	4.896		3.008	0.453	
2	D11 δ-Amino Valeric Acid	0.022	0.015		0.068	0.047		3.144	3.174		3.159	0.143	
4	A10 Adenosine5'monophosphate	0.015	0.006		0.013	0.037		0.889	5.746		3.317	0.423	
4	G11 L-Methionine Sulfoxide	0.069	0.038		0.039	0.264		0.558	6.921		3.740	0.544	
2	B8 Arbutin	0.033	0.007		0.038	0.051		1.140	7.617		4.379	0.282	
1	A3 N-Acetyl-D-Glucosamine	0.061	0.093	0.091	0.145	1.115	0.055	2.377	12.040	0.610	5.009	0.403	

Values represent absorbance at 570nm normalised to a negative control well and Fold Change was calculated by dividing the normalised absorbance for the ΔsigGoperon deletion strain by that of wild-type *M. tuberculosis* H37Rv (WT). Columns 1-3 represent replicate experiments. p-values were calculated using a Student’s t-test to compare the absorbance values of the WT strain and ΔsigGoperon deletion strain.

PMs 1 & 2 comparing the *M. tuberculosis sigG* only complement and ΔsigGoperon deletion strains

ΔsigGoperon/SigG only complement fold change		
PM	**Compound**	**Fold Change**
1	D5 Tween 40	0.52
2	H6 Sec-Butylamine	0.58
2	B1 N-Acetyl-D-Galactosamine	0.61
1	A11 D-Mannose	0.68
2	B11 D-Fucose	0.69
2	H12 3-Hydroxy 2Butanone	0.71
1	E4 D-Fructose-6Phosphate	0.75
1	C6 L-Rhamnose	0.78
2	H10 2,3-Butanediol	0.79
2	D1 D-Raffinose	0.80
2	G1 Acetamide	0.82
1	D12 Uridine	0.84
1	C1 D-Glucose-6Phosphate	0.84
2	D12 Butyric Acid	0.86
1	C5 Tween 20	0.86
2	H1 L-Ornithine	0.86
2	C1 Gentiobiose	0.87
1	G10 Methyl Pyruvate	0.90
1	B12 L-Glutamic Acid	0.94
1	E5 Tween 80	0.97
1	A12 Dulcitol	0.99
2	G11 L-Lysine	1.00
1	E1 L-Glutamine	1.03
1	D6 α -Keto-Glutaric Acid	1.03
1	H12 2-Aminoethanol	1.08
1	F1 Glycyl-L-Aspartic Acid	1.13
2	C4 D-Melezitose	1.16
1	C12 Thymidine	1.17
1	G5 L-Alanine	1.17
2	B12 3-0-β -D-Galactopyranosyl-D-Arabinose	1.20
2	D9 N-Acetyl-D-Glucosaminitol	1.21
1	G8 N-Acetyl-β -D-Mannosamine	1.30
1	H10 D-Galacturonic Acid	1.33
2	G9 L-Isoleucine	1.36
1	F12 Inosine	1.38
1	G1 Glycyl-L-Glutamic Acid	1.42
1	H1 Glycyl-L-Proline	1.45
2	H7 D.L-Octopamine	1.46
1	C8 Acetic Acid	1.48
1	G3 L-Serine	1.50