A novel mechanism regulating DNA-damage repair in *Mycobacterium tuberculosis*

Lisa Frances Dawson

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

September 2005

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

The intracellular pathogen Mycobacterium tuberculosis (M. tuberculosis) resides in macrophages and is the causative agent of human tuberculosis. Infected macrophages produce reactive oxygen and nitrogen intermediates, known to damage DNA; therefore DNA damage repair is thought to be important in survival of M. tuberculosis in the host. The expression of many bacterial DNA repair genes is often regulated by the SOS response, in which RecA is an integral part; however, in M. tuberculosis the majority of genes in the DNA-damage regulon are regulated independently of the RecA/LexA system. In this study two potential mechanisms for this alternative mode of gene regulation were investigated.

The first hypothesis addressed was that regulation of expression following DNA-damage is controlled by an alternative sigma factor. Sigma factors are protein subunits of RNA polymerase, which confer specificity of binding to particular promoters. The function/expression of alternative sigma factors is usually regulated by various mechanisms. The sigma factor SigG is the most highly induced of all 13 sigma factors of M. tuberculosis in response to DNA-damage in both wild-type and ΔrecA strains. A knockout of sigG in M. tuberculosis was constructed, and found to be more susceptible to mitomycin C stress than wild-type H37Rv and attenuated in mice. ruvC was shown to possesses 2 transcriptional start sites, and although neither were regulated by SigG, the P1 promoter appeared to be dual regulated by LexA and the RecA/LexA independent mechanism. Microarray analysis revealed that SigG was not significantly involved in regulation of the RecA/LexA independent DNA-damage regulon, but that SigG directly or indirectly regulated expression of 127 genes in the absence of DNA-damage.
The other possible mode of RecA/LexA independent regulation was via a repressor/activator protein. Gel shifts assays using *M. tuberculosis* cell free extracts were used to attempt to identify a repressor or activator protein that bound to the operator of the *recA* P1 promoter, known to be induced independently of RecA, but failed to detect specific binding. Published microarray data revealed that Rv2017, a predicted regulatory protein, was upregulated in both wild-type and Δ*recA* strains of *M. tuberculosis* in response to DNA-damage. Therefore, a gene inactivation knockout of Rv2017 was constructed and analysed in *M. tuberculosis*. The ΔRv2017 strain was hyposensitive to mitomycin C stress and preliminary mouse *in-vivo* infection data suggested that the ΔRv2017 strain may be hypervirulent. Microarray data revealed that Rv2017 plays a direct/indirect role in regulation of a large regulon, including some genes in the DNA-damage repair regulon.
Acknowledgements

I would like to take this opportunity to thank my supervisor Dr. Elaine Davis for giving me the opportunity to study for my PhD, and would like to thank her, along with my second supervisor Dr. Roger Buxton for their critical analysis and guidance. I would also like to extend my thanks to all the members of the mycobacterial research group, past and present who have helped, guided and discussed both practical and technical problems.

My thanks are extended to Dr. Krishna Gopaul for the pGardee2 construct, from which I constructed the $\text{sigG}$ mutant strain of $\textit{M. tuberculosis}$. I would also like to thank Dr. Ricardo Tascon and the animal technicians for performing in-vivo mouse infections and Dr. Jason Hinds at the BµGS microarray facility in St. Georges Hospital Medical School for the $\textit{M. tuberculosis}$ microarrays, and for technical assistance.

Finally I would like to thank my family and friends for their unmoving support throughout my PhD. I would particularly like to thank Jonny for all his Mac technical support, without which, I would have been lost, and I would like to also thank Paul for making me smile. I would also like to thank Rox, Jude and Kat for listening relentlessly to my problems.
2.7.6 Digestion of DNA with Restriction Endonucleases ................................................... 68
2.7.7 Dephosphorylation of linearised plasmids................................................................. 69
2.7.8 Ethanol Precipitation of Nucleic Acids ..................................................................... 69
2.7.9 Phenol-Chloroform extraction of DNA .................................................................... 69
2.7.10 Ligation of DNA ....................................................................................................... 70
2.7.11 Transformation of chemically competent E. coli ..................................................... 70
  2.7.11.1 DH5α Subcloning efficiency/Library efficiency (Invitrogen) ................................ 70
  2.7.11.2 One-shot chemically competent cells ................................................................. 70
2.7.12 Automated DNA Sequencing: ABI Prism BigDye Terminator Cycle Sequencing
  Ready Reaction protocol: ............................................................................................. 71
2.7.13 Site directed Mutagenesis (SDM): QuickChange Site directed mutagenesis Kit
  (Stratagene) .................................................................................................................. 71

2.8 Protein analysis ............................................................................................................. 71
  2.8.1 Synthesis of antibodies ........................................................................................... 71
  2.8.2 SDS page gel electrophoresis .................................................................................. 72
  2.8.3 Dot blot .................................................................................................................... 72
  2.8.4 Western Blot ............................................................................................................ 73

2.9 Transcriptional analysis .............................................................................................. 73
  2.9.1 cDNA synthesis ..................................................................................................... 73
  2.9.2 RT-PCR .................................................................................................................. 74
  2.9.3 Taqman real-time PCR ........................................................................................ 74
  2.9.4 End-labelling of oligonucleotides ......................................................................... 75
  2.9.5 Primer extension reactions ..................................................................................... 76
  2.9.6 Manual Sequencing: T7 Sequenase V2.0 (Amersham LS) .................................... 76
2.9.7 RNase protection ...............................................................................................................77
  2.9.7.1 In-vitro transcription of the radio-labelled probe ........................................................77
  2.9.7.2 RNase protection assay................................................................................................77
2.9.8 Denaturing polyacrylamide gel electrophoresis ..........................................................78

2.10 Detection of DNA-Protein binding .................................................................................79
  2.10.1 Extraction of Cell Free Extract (CFE) from M. tuberculosis ...............................79
  2.10.2 Preparation of probes for gel retardation assay ......................................................81
  2.10.3 Gel Retardation........................................................................................................81

2.11 Constructing a knockout in M. tuberculosis .................................................................81
  2.11.1 Cloning into pBackbone .......................................................................................81
  2.11.2 Competent cells and electroporation of M. tuberculosis ....................................82
  2.11.3 Selection of a knockout .........................................................................................82
    2.11.3.1 Probe Labelling for Southern analysis ..............................................................83
    2.11.3.2 Southern Blot ..................................................................................................83
  2.11.4 M. tuberculosis liquid cultures and Ziehl-Nielson stain ..................................84

2.12 Phenotypic analysis of M. tuberculosis knockouts ...................................................... 85
  2.12.1 Determination of in vitro growth ..........................................................................85
  2.12.2 Viability assays in vitro .........................................................................................85
  2.12.3 Viability assays in vivo .........................................................................................86

2.13 Expression analysis by microarray .............................................................................. 87
  2.13.1 Poly-L-lysine coating of microarray slides ............................................................87
  2.13.2 Post processing ....................................................................................................88
  2.13.3 RNA labelling, hybridization and washing ............................................................88
  2.13.4 DNA versus RNA microarray ...............................................................................89
3 Analysis of a sigma factor mutant in *M. tuberculosis*..... 92

3.1 Introduction................................................................. 92

3.2 Construction of a *sigG* knockout............................................. 93

3.2.1 PCR used to detect potential knockouts................................. 94

3.2.2 Southern blot used to detect potential knockouts.................. 94

3.2.3 Chromosomal location of *sigG*...................................... 99

3.3 Predicted protein domains of SigG, Rv0181c and Rv0180c........... 102

3.4 Complementation of the *sigG* knockout.................................. 104

3.4.1 Checking expression of the SigG, Rv0181c and Rv0180c............. 106

3.4.2 Design and production of antibodies specific to *sigG*, Rv0181c and Rv0180c 106

3.4.3 Analysis of SigG, Rv0181c and Rv0180c, using antibodies........... 109

3.5 *In-vitro* analysis of *ΔsigG* strain and complements.................. 112

3.5.1 Growth curves *in-vitro*.................................................. 112

3.5.1.1 Viability of *ΔsigG*, complements and wild-type to DNA damaging agents 112

3.6 *In-vivo* phenotype of *ΔsigG* strain compared to wild-type......... 121

3.7 Discussion........................................................................... 124

3.7.1 Inspection of the SigG operon.......................................... 127

3.7.2 Viability of the *ΔsigG* strain......................................... 130

4 Regulation of *sigG* and identification of its regulon............. 133

4.1 Identification of the *sigG* regulon by microarray analysis........... 133
4.1.1 Analysis and input of the microarray slides ................................................................. 134
4.1.2 Statistical analysis of microarray data ........................................................................... 135
  4.1.2.1 SigG and the genes involved in the DNA-damage response ........................................ 135
  4.1.2.2 Detecting differential gene expression in the ΔsigG strain ........................................ 139

4.2 Identification of sigG transcriptional start site(s) ....................................................... 149
  4.2.1 Primer extension of sigG .......................................................................................... 149
  4.2.2 RNase protection of sigG .......................................................................................... 156
  4.2.3 Precise identification of the sigG transcriptional start site(s) ...................................... 158
  4.2.4 Identification and analysis of promoter motifs ............................................................ 159

4.3 Discussion ...................................................................................................................... 169
  4.3.1 Analysis of the sigG regulon ..................................................................................... 169
  4.3.2 Dissection of the upstream region of sigG ................................................................. 176

5 Detailed analysis of ruvC ................................................................................................. 183
  5.1 Identification of ruvC transcriptional start site(s) by primer extension ......................... 184
    5.1.1 Potential promoter motifs upstream region of ruvC .................................................. 187
  5.2 Quantitative analysis of the transcriptional start sites of ruvC compared to recA by RNase protection ................................................................. 190
    5.2.1 Design and optimisation of the RNase protection assay ............................................ 192
    5.2.2 RNase protection of ruvC ....................................................................................... 199
    5.2.3 RNase protection of recA ....................................................................................... 204
  5.3 Validation of quantitation .............................................................................................. 208
  5.4 Discussion ...................................................................................................................... 213
6 Regulation of gene expression from the P1 promoter of recA

6.1 Gel retardation assays

6.2 Bandshifts using large P1 and P2 probes

6.3 Bandshifts using small P1 and P2 probes

6.4 Competitive bandshifts

6.5 Discussion

7 Construction of a knockout in a predicted regulatory protein

7.1 Introduction

7.2 Analysis of homology of the predicted regulatory proteins

7.3 Construction of Gene inactivation knockouts

7.3.1 Chromosomal location of Rv2017 and Rv2884

7.4 Designing and testing of knockout constructs

7.5 in-vitro and in-vivo phenotyping of the Rv2017 knockout

7.5.1 Growth curve

7.5.2 Susceptibility of the Rv2017 knockout and wild-type to DNA-damage

7.5.3 in-vivo phenotyping of the Rv2017 knockout compared to wild-type H37Rv

7.6 Discussion
7.6.1 Functional classification of the five predicted regulatory proteins ..................... 261
7.6.2 Analysis of the Rv2017 gene inactivation knockout ............................................ 263

8 Identification of the Rv2017 regulon by microarray analysis........................................ 267

8.1 Introduction........................................................................................................... 267
8.2 Analysis and input of the microarray slides......................................................... 268
  8.2.1 Statistical analysis of microarray data.............................................................. 271
    8.2.1.1 Rv2017 and the genes involved in the DNA-damage response ................ 271
    8.2.1.2 Analysis of the Rv2017 regulon ............................................................... 283
8.3 Discussion............................................................................................................. 292

9 Discussion.............................................................................................................. 301

9.1 The role of SigG................................................................................................. 302
  9.1.1 Identification of the SigG regulon................................................................. 306
9.2 The role of an activator or repressor in DNA-damage........................................ 311
9.3 Where to begin.................................................................................................... 316

10 References............................................................................................................ 318

Appendix I: Media and solutions .......................................................... 335
Appendix II: Primers and probes ............................................................. 337
Appendix III: Microarray fold change Rv2017................................. 340
Appendix IV - Keystone abstract ........................................ 343
List of figures

Figure 1.1: A schematic representation of the cell wall of *M. tuberculosis* ..............................32

Figure 2.1: Cell free extract protocol *M. tuberculosis* ........................................................................80

Figure 3.1: A schematic representation of the design of the PCR reaction to screen for potential knockout colonies .............................................................................................................................95

Figure 3.2: (a) A schematic representation of the size of products produced in a Southern blot for both a Δ*sigG* strain and the H37Rv wild-type (b) A Southern blot .................................................96

Figure 3.3: (a) A schematic representation of the size of product produced in a Southern blot from the deleted region of *sigG* (b) A Southern Blot .............................................................98

Figure 3.4: A schematic representation of the location of *sigG* .....................................................100

Figure 3.5: (a) A schematic representation of the relative position of *sigG* (b) An agarose gel showing co-transcription studies of *sigG* .....................................................................................101

Figure 3.6: Microarray expression for *sigG* operon in Δ*sigG* strain compared to H37Rv wild-type........................................................................................................................................107

Figure 3.7: Western blots to test specificity of SigG, Rv0181c and Rv0180c antibodies.... 110

Figure 3.8: The *in-vitro* growth curves of the *sigG* knockout, complements and wild-type H37Rv ...................................................................................................................................................113

Figure 3.9: Histogram depicting the viability of *sigG* knockout treated with different DNA damaging agents ...........................................................................................................................................115

Figure 3.10: A graph comparing the viability of the *sigG* knockout to H37Rv in response to paraquat stress ....................................................................................................................................117

Figure 3.11: A graph comparing the viability of the *sigG* knockout, *sigG* full operon complement and H37Rv wild-type in response to mitomycin C stress .........................119
Figure 3.12: Colony counts in lung and spleen of BALB/c mice infected with ΔsigG strain and H37Rv wild-type .................................................................................................................................................. 123

Figure 4.1: (a) Microarray data showing decrease in induction ratio of genes in the sigG strain following DNA-damage (b) Microarray data showing increase in induction ratio of genes in the sigG strain following DNA-damage ........................................................................................................................................ 138

Figure 4.2: A Venn diagram of genes significantly different in H37Rv and ΔsigG under induced and uninduced conditions .................................................................................................................................................. 144

Figure 4.3: (a) Genes significantly different under both uninduced and induced conditions for sigG compared to H37Rv wild-type (b) Induction ratio for these genes ........................................................................................................................................ 145

Figure 4.4: Genes significantly different in the ΔsigG compared to H37Rv under (a) uninduced and (b) induced conditions ........................................................................................................................................ 147

Figure 4.5: Venn diagram showing the overlap between genes found to be significantly different in ΔsigG compared to H37Rv ........................................................................................................................................ 148

Figure 4.6: Schematic representation of the primer designed to perform the primer extension and manual sequencing ........................................................................................................................................ 153

Figure 4.7: Relative positions of primers used in primer extension and the potential translational start sites ........................................................................................................................................ 155

Figure 4.8: RNase protection assay to identify the sigG promoters in H37Rv, ΔrecA and ΔsigG ........................................................................................................................................ 157

Figure 4.9: Primer extension of sigG ........................................................................................................................................ 160

Figure 4.10: Identification of the transcriptional start sites of sigG ........................................................................................................................................ 161

Figure 4.11: MEME motif search using the P3 promoter region of sigG ........................................................................................................................................ 164

Figure 4.12: A comparison between genes identified with conserved upstream regions, which are significantly different in ΔsigG strain and H37Rv ........................................................................................................................................ 167
Figure 5.1: (a) A schematic representation of the position of ruvC and potential transcriptional start sites (b) An autoradiograph showing the transcriptional start sites of ruvC .......... 186

Figure 5.2: (a) Primer extension of ruvC (b) transcriptional start sites of ruvC (c) transcriptional start sites of recA .............................................................................................................................. 188

Figure 5.3: Bioanalyser data from RNA extractions .............................................................................. 191

Figure 5.4: Schematic of an RNase protection assay ................................................................. 194

Figure 5.5: In-vitro transcription of the ruvC probe with varying concentrations of unlabelled UTP ..................................................................................................................................................... 195

Figure 5.6: In-vitro transcription of the ruvC and recA probes at altered temperature.......... 196

Figure 5.7: Optimisation of the RNase protection ................................................................. 198

Figure 5.8: Synthesis and sizing of the ruvC, recA and sigG probes under optimised conditions .................................................................................................................................................................... 200

Figure 5.9: RNase protection assay to identify ruvC promoters in H37Rv, ΔrecA and ΔsigG ............................................................................................................................................................................. 202

Figure 5.10: (a) The expression level of P1 and P2 promoters of ruvC, under induced conditions in H37Rv, ΔrecA and ΔsigG (b) A graph showing the differences in expression level of the P1 and P2 promoters of ruvC ................................................................................................................................. 203

Figure 5.11: RNase protection assay to identify recA promoters in H37Rv, ΔrecA and ΔsigG ............................................................................................................................................................................. 205

Figure 5.12: RNase protection assay to identify recA promoters in H37Rv and ΔsigG ...... 207

Figure 5.13: (a) Expression of P1 and P2 promoters of recA under induced conditions in H37Rv, recA and sigG (b) A graph showing the differences in expression level of the P1 and P2 promoters of recA as determined using the phosphorimager ................. 209

Figure 5.14: (a) A comparison between the expression level of recA, ruvC and sigG in the wild-type H37Rv and ΔsigG mutant of M. tuberculosis (b) A comparison between induction
ratio of recA, ruvC and sigG in the wild-type H37Rv and ΔsigG mutant (c) A comparison between the expression level of recA, ruvC and sigG in the wild-type H37Rv and ΔrecA mutant of M. tuberculosis ........................................210

Figure 5.15: (a) Taqman expression levels of ruvC in H37Rv wild-type, ΔrecA and ΔsigG strains (b) Taqman expression level of ruvC expressed as a proportion of sigA .......................212

Figure 6.1: (a) A schematic representation of P1 and P2 promoter region of recA (b) a gel retardation using 60mer oligonucleotide probes .................................................................230

Figure 6.2: (a) A schematic representation of P1 and P2 promoter region of recA (b) a gel retardation using 30mer oligonucleotide probes .................................................................232

Figure 6.3: (a) A schematic representation of P1 and P2 promoter region of recA (b) a gel retardation comparing binding buffers using 30mer oligonucleotide probes (c) a competitive gel retardation using the 30mer oligonucleotide probes .........................233

Figure 7.1: Conserved domains present ion the 5 predicted regulatory proteins .................240

Figure 7.2: An alignment with IrrE from D. radiodurans and Rv2017 from M. tuberculosis using (a) EMBOSS and (b) BLAST .........................................................................................243

Figure 7.3: Chromosomal location of (a) Rv2017 and (b) Rv2884 ......................................245

Figure 7.4: (a) Schematic representation of the relative position of the predicted transcriptional regulatory protein Rv2017 (b) An agarose gel showing co-transcription of the Rv2017 operon ......................................................................................................................246

Figure 7.5: A schematic representation of formation of single and double crossovers for the Rv2017 knockout ..................................................................................................249

Figure 7.6: (a) Southern blot to detect potential knockouts of Rv2017 (b) A schematic representation of the size of fragments produced in the Southern blot with ClaI digest ..................................................................................................................250
Figure 7.7: (a) A schematic representation of the size of the product produced by a Southern blot for Rv2017 (b) A Southern blot using Clal digest to detect potential Rv2017 knockout ........................................... 252

Figure 7.8: (a) A schematic representation of the size of the product produced by a Southern blot for Rv2017 following Xhol/Hpal double digest (b) A Southern blot using Xhol/Hpal double digest to detect potential Rv2017 knockout ........................................... 253

Figure 7.9: The in-vitro growth curve of ΔRv2017 strain compared to H37Rv wild-type...255

Figure 7.10: A graph comparing the viability of the ΔRv2017 strain with H37Rv wild-type to mitomycin C stress........................................................................................................................... 256

Figure 7.11: In-vivo mouse data for ΔRv2017 strain and H37Rv comparing (a) bacterial levels in spleen and lung and (b) organ weights for the spleen and lung................................. 260

Figure 8.1: PCR to detect DNA contamination in ΔRv2017 strain RNA extractions .......... 269

Figure 8.2: Schematic representation of the spread of microarray data for H37Rv, ΔsigG and ΔRv2017 strains .......................................................................................................................... 270

Figure 8.3: Schematic representation of the analysis of genes in the DNA-damage regulon .......................................................................................................................... 273

Figure 8.4: 30 genes with significantly different induction ratios, also significantly different under induced conditions.......................................................................................... 274

Figure 8.5: 11 genes with significantly different induction ratios, also significantly different under uninduced conditions................................................................................ 276

Figure 8.6: 5 genes with significantly different induction ratios, which are not significantly different under uninduced or induced conditions........................................... 277

Figure 8.7: 13 genes with significantly different induction ratios, also significantly different under both uninduced and induced conditions.................................................. 279
Figure 8.8: 26 genes without significantly different induction ratios, but which were significantly different under induced conditions ................................................................. 281

Figure 8.9: 4 genes without significantly different induction ratios, but which were significantly different under uninduced conditions ........................................................................... 282

Figure 8.10: 20 genes without significantly different induction ratios, but which are not significantly different under uninduced or induced conditions ........................................... 284

Figure 8.11: 9 genes without significantly different induction ratios, yet were significantly different under both uninduced and induced conditions .................................................. 285

Figure 8.12: Venn diagram showing the genes significantly different in ARv2017 strain and H37Rv .......................................................................................................................................... 287

Figure 8.13: Transcriptional regulators with altered expression in ARv2017 strain compared to H37Rv under both uninduced and induced conditions ...................................................... 289
List of tables

Table 2.1: Media supplements ................................................................. 63
Table 2.2: Plasmid descriptions ............................................................. 63
Table 2.3: Optimization primer concentrations .................................... 75
Table 2.4: Mitomycin C and paraquat concentrations .......................... 86
Table 3.1: Comparison between the full operon and sigG only complements .................................................. 104
Table 3.2: Percentage viability of H37Rv in comparison to sigG knockout in response to paraquat stress ................................................................. 117
Table 3.3: Statistical analysis of the response to paraquat stress of the sigG knockout compared to H37Rv wild-type ................................................................. 118
Table 3.4: Percentage viability of H37Rv in comparison to sigG knockout and sigG full operon complement in response to mitomycin C stress ................................................................. 119
Table 3.5: Statistical analysis of mitomycin C stress (a) sigG compared to H37Rv wild-type, (b) sigG complement compared to H37Rv wild-type and (c) sigG knockout compared to sigG complement ................................................................. 120
Table 3.6: Sigma factors regulated by transmembrane anti-sigma factors ................................................................. 125
Table 4.1: Gene list of genes showing decreased expression in uninduced conditions in ΔsigG compared to H37Rv wild-type ................................................................. 140
Table 4.2: Gene list of genes showing increased expression in uninduced conditions in ΔsigG compared to H37Rv wild-type ................................................................. 141
Table 4.3: Gene list of genes showing decreased expression in induced conditions in ΔsigG compared to H37Rv wild-type ................................................................. 142
Table 4.4: Gene list of genes showing increased expression in induced conditions in ΔsigG compared to H37Rv wild-type ................................................................. 143
Table 4.5: Microarray data of genes with a fold change of greater than or equal to 1.7 for H37Rv compared to ΔsigG ............................................................................................................. 150
Table 4.6: Microarray data of genes with a fold change of less than 0.59 for H37Rv compared to ΔsigG ............................................................................................................................................... 151
Table 4.7: Alignment of the possible consensus sequences for sigG and potentially SigG regulated genes ................................................................................................................................... 165
Table 4.8: Interesting genes with the sigG/recA consensus in close proximity to predicted translational start sites ...................................................................................................................... 168
Table 5.1: Probe size, orientation and test templates designed for the three genes of interest ........................................................................................................................................................................ 192
Table 7.1: Relative induction ratio in response to mitomycin C induction of 5 predicted regulatory genes in both wild-type and ΔrecA strains of M. tuberculosis ............................................. 238
Table 7.2: Percentage viability of ΔRv2017 strain compared to H37Rv in response to mitomycin C stress .................................................................................................................................................. 256
Table 7.3: Statistical analysis of the response to mitomycin C stress of the Rv2017 knockout compared to H37Rv .................................................................................................................................................. 258
Table 7.4: Viable CFU counts of H37Rv wild-type and Rv2017 knockout ........................................................................................................................................................................ 259
Table 8.1: Functional classification of 595 genes significantly different under uninduced and induced conditions in ΔRv2017 strain and H37Rv .................................................................................................................................................. 288
Table 8.2: Genes with the highest fold change in ΔRv2017 strain compared to H37Rv ........................................................................................................................................................................ 291
Table 8.3: Expression level of two sigma factors with decreased expression in ΔRv2017 strain compared to H37Rv .................................................................................................................................................. 295
Table 8.4: Classification of 59 genes with significantly different induction ratios according to expression patterns seen in the ΔrecA strain ........................................................................................................................................ 297
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>Acquired immunodeficiency syndrome</td>
</tr>
<tr>
<td>APC</td>
<td>Antigen presenting cell</td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulphate</td>
</tr>
<tr>
<td>AraLAM</td>
<td>Arabinose capped lipoarabinomannan</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triose phosphate</td>
</tr>
<tr>
<td>BCG</td>
<td>Bacillus Calmette-Guerin</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>CAP</td>
<td>Catabolite activator protein</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>CFE</td>
<td>Cell free extract</td>
</tr>
<tr>
<td>CFU</td>
<td>Colony forming units</td>
</tr>
<tr>
<td>CR</td>
<td>Complement receptor</td>
</tr>
<tr>
<td>dATP</td>
<td>Deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>ddATP</td>
<td>Dideoxyadenosine triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>Deoxycytosine triphosphate</td>
</tr>
<tr>
<td>ddCTP</td>
<td>Dideoxyctosine triphosphate</td>
</tr>
<tr>
<td>ddGTP</td>
<td>Dideoxyguanine triphosphate</td>
</tr>
<tr>
<td>ddTTP</td>
<td>Dideoxytyrosine triphosphate</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethylpyrocarbonate</td>
</tr>
<tr>
<td>dGTP</td>
<td>Deoxyguanine triphosphate</td>
</tr>
<tr>
<td>dH₂O</td>
<td>Deionized water</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s modified Eagles media</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulphoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>DOTS</td>
<td>Directly observed short course therapy</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double stranded deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>dTTP</td>
<td>Deoxytyrosine triphosphate</td>
</tr>
<tr>
<td>EDTA</td>
<td>Diaminoethanetetraacetic acid</td>
</tr>
<tr>
<td>FCS</td>
<td>Foetal calf serum</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>H-NS</td>
<td>Histone like nucleoid structuring protein</td>
</tr>
<tr>
<td>Hyg&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Hygromycin resistant</td>
</tr>
<tr>
<td>LAM</td>
<td>lipoarabinomannan</td>
</tr>
<tr>
<td>INF-γ</td>
<td>Gamma interferon</td>
</tr>
<tr>
<td>IL</td>
<td>interleukin</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Kanamycin resistant</td>
</tr>
<tr>
<td>kb</td>
<td>Kilo base (1Kb = 1000bp)</td>
</tr>
<tr>
<td>KCl</td>
<td>Potassium chloride</td>
</tr>
<tr>
<td>KDa</td>
<td>Kilo dalton</td>
</tr>
<tr>
<td>LB-Broth</td>
<td>Luria-Bertani medium</td>
</tr>
<tr>
<td>log</td>
<td>Logarithmic</td>
</tr>
<tr>
<td>ManLAM</td>
<td>Mannose capped lipoarabinomannan</td>
</tr>
<tr>
<td>MBq</td>
<td>MegaBequerels</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>MHC1</td>
<td>Major histocompatibility complex class 1</td>
</tr>
<tr>
<td>MMLV</td>
<td>Moloney murine leukemia virus derived thermostable reverse transcriptase</td>
</tr>
<tr>
<td>MOPS</td>
<td>4-Morpholinepropanesulfonic acid</td>
</tr>
<tr>
<td>MR</td>
<td>Mannose receptor</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NaOAC</td>
<td>Sodium acetate</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>Paraquat</td>
<td>Methy viologen (1,1-dimethyl 1-4'-dipryriminim dichloride methyl viologen)</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PiLAM</td>
<td>Phospho-myoinositol capped lipoarabinomannan</td>
</tr>
<tr>
<td>ppGpp</td>
<td>alarmone guanosine 3', 5'-bisphosphate</td>
</tr>
<tr>
<td>RE</td>
<td>Restriction endonuclease</td>
</tr>
<tr>
<td>RecA-ND</td>
<td>RecA non-dependent promoter</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>RNAP-σ</td>
<td>RNA polymerase-sigma factor complex</td>
</tr>
<tr>
<td>RNase A</td>
<td>Ribonuclease A</td>
</tr>
<tr>
<td>RNase T1</td>
<td>Ribonuclease T1</td>
</tr>
<tr>
<td>RNI</td>
<td>Reactive nitrogen intermediates</td>
</tr>
<tr>
<td>ROI</td>
<td>Reactive oxygen intermediates</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcriptase PCR</td>
</tr>
<tr>
<td>σ</td>
<td>Sigma factor</td>
</tr>
<tr>
<td>SSC</td>
<td>Standard short course chemotherapy</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>Abbreviations</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SSC</td>
<td>Standard saline citrate</td>
</tr>
<tr>
<td>SSPE</td>
<td>Standard saline phosphate EDTA</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate EDTA</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-borate EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>Tris EDTA</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll like receptor</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris (hydroxymethyl) aminomethane (2-amino-2-hydroxymethylpropane-1,3-diol)</td>
</tr>
<tr>
<td>Tris-Cl</td>
<td>Tris chloride</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>Tween 80</td>
<td>Polyoxyethylene sorbitan monooleate</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>X-Gal</td>
<td>5-Bromo-4-chloro-3-indolyl-β-D-galactoside</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Tuberculosis

1.1.1 Global incidence of *M. tuberculosis*

Today, tuberculosis is one of the top three killers in developing countries, along with malaria and HIV (Saeyers, 2002), and is responsible for over 2 million deaths a year (WHO, 2002). After the discovery of a treatment for *M. tuberculosis* in the early 1950’s, health programs were set up in Europe and the US to combat and eradicate *M. tuberculosis* (O’Brien and Nunn, 2001). The initial success of these programmes led to the steady decline of *M. tuberculosis* in these areas, and the subsequent transfer of resources away from the *M. tuberculosis* programme, to other health programmes. However in the mid 1980’s the cases of *M. tuberculosis* were on the increase, then in 1993, the WHO proclaimed *M. tuberculosis* a global emergency (WHO, 1993).

1.1.2 Efficacy of *M. tuberculosis* vaccine bacillus Calmette-Guerin (BCG).

Since the 1920’s over 3 billion vaccinations with BCG have been widely administered to combat *M. tuberculosis*. Certain aspects of the BCG vaccine have fuelled an ongoing debate about the suitability and safety of BCG as a vaccine; this live attenuated vaccine was derived by 230 passages over 13 years of a virulent strain of *Mycobacterium bovis* (M. bovis) (Agger and Andersen, 2002). The efficacy varies between 0 and 80%, with widespread ineffectiveness in preventing pulmonary tuberculosis (O’Brien and Nunn, 2001). The nature of the vaccine renders the PPD skin test for *M. tuberculosis* ineffective, resulting in vaccinated individuals being indistinguishable from infected individuals (Agger and Andersen, 2002). It has been
hypothesised that the ineffectiveness of the BCG vaccine is linked to exposure of individuals to environmental mycobacteria, which elicits an immune response, whereby antibodies to environmental mycobacteria result in clearance of BCG prior to establishment of protective immunity (Agger and Andersen, 2002). This could explain the wide variation in efficacy, particularly as the lowest level of protection was observed in developing countries. However, BCG exhibits good efficacy at preventing *Mycobacterium leprae* (*M. leprae*) infections along with miliary and meningeal tuberculosis in children (Agger and Andersen, 2002).

### 1.1.3 Monitoring strategy to limit the spread of *M. tuberculosis*

Tuberculosis is particularly rife among the homeless and refugees worldwide, with over 50% of the worldwide refugee population suffering from TB (WHO, 2002), due to poor living conditions and malnutrition. In developing countries, inadequate control practices for tuberculosis alongside the widespread ineffectiveness of the BCG vaccine (O'Brien and Nunn, 2001), has resulted in a global surge in the number of cases, reaching almost pandemic levels. In an effort to tackle the growing problem, the world health organisation (WHO) set up a strategy called directly observed therapy, short course (DOTS); a strategy for detection, monitoring and cure of tuberculosis. It encompasses 5 elements: surveillance/monitoring systems, microscopy services, drug supplies, directly observed effective treatment regimes, and political commitment (WHO, 2002). Without extensive multiple drug treatment, the fatality rate of those with acute tuberculosis is as high as 50% (Saeyers, 2002).
1.1.4 Co-infections with HIV

*M. tuberculosis* infections are most prevalent in south-east Asia, sub-Saharan Africa and Eastern Europe (WHO, 2002). The rise of *M. tuberculosis* cases has been fuelled by the HIV epidemic sweeping through Africa and Asia; 70% of the total 36.1 million people worldwide suffering from HIV and AIDS, localised to Sub-Saharan Africa (WHO, 2003). Co-infected patients are more likely to develop acute pulmonary tuberculosis (Dye *et al*., 2002a), due to a depletion of CD4+ T-cells and subsequent decreased levels of interferon-γ (IFN-γ) in HIV patients. CD4+ T-cells are known to secrete IFN-γ, which has been shown in mouse models to be key in combating acute pulmonary tuberculosis (Agger and Andersen, 2002). This absence of CD4+ T cells results in *M. tuberculosis* being the leading cause of death of HIV positive patients (WHO, 2002).

1.1.5 Emergence of multiple-drug resistant *M. tuberculosis* (MDR-TB)

In recent years, there has been a rapid increase in the prevalence of MDR-TB, which is characterised by resistance to at least two of the most important and widely used first line drugs for *M. tuberculosis* treatment, namely rifampicin and isoniazid (Bone, 2001), which are given as part of combination drug therapy (Dye *et al*., 2002a). Standard short course chemotherapy (SCC) consists of a cocktail of 4 first line drugs, including rifampicin, isoniazid, pyrazinamide and ethambutol given for 2 months, followed by just rifampicin and isoniazid for a further 4 months. These drugs are more effective, cheaper and less toxic than second line drugs, such as capreomycin, kanamycin, ethionamide, cycloserine, and para-aminosalicylate (Lordi, 2000), which are used to treat MDR-TB, for as long as 2 years (Dye *et al*., 2002b). The treatment of
MDR-TB costs up to 50 times more than the standard short course chemotherapy regime used to treat drug susceptible cases (Dye \textit{et al.}, 2002a). However, the rapid spread of MDR-TB is a cause for concern, as in the latter stages of acute pulmonary infections with MDR-TB there is no real effective treatment, therefore, even with good compliance during the 2 year chemotherapy course, the prognosis is poor. There is a particular problem with the spread of MDR-TB in Eastern Europe (Bone, 2001), namely in prisons in Russia, due to poor living conditions and lack of compliance with drug regimes, compounded by the lack of constant availability of anti-tuberculosis drugs (Bone, 2001).

The decline in the number of new cases of pulmonary tuberculosis in developed countries led to the dismantling of the health programmes in the 1970’s, which in turn lead to the withdrawal of major pharmaceutical companies from novel drug discovery for \textit{M. tuberculosis}, as it was no longer deemed a threat in the developed world, and therefore thought to be no longer profitable (Saeyers, 2002). This resulted in a gap in the research, causing a short fall in new drug therapy to treat the rapidly increasing worldwide burden of MDR-TB.

1.2 \textit{M. tuberculosis} classification

\textit{M. tuberculosis} is a member of the actinomycetes family of Gram-positive, rod shaped bacteria; it is an obligate aerobe, which can be visualised by a Ziehl Nielsen stain, as it is acid-fast (Saeyers, 2002). \textit{M. tuberculosis} is also highly GC-rich and forms the \textit{M. tuberculosis} complex along with; \textit{M. bovis}, \textit{Mycobacterium microti}, \textit{M. bovis} (BCG), \textit{Mycobacterium africanum}, and \textit{Mycobacterium canetti} (Wayne, 1982). Mycobacteria are generally defined as slow growing bacteria in comparison to other species, such as \textit{Bacillus}, or \textit{Escherichia coli} (\textit{E. coli}): the generation time varies from 30 minutes in \textit{E. coli} to 17 hours in \textit{M. tuberculosis}. However,
there are inter-species differences in generation time for mycobacterial species. Fast-growing mycobacteria are defined by the ability to produce colonies on a plate in one week, whereas slow growing mycobacteria take considerably longer (up to 4 weeks) (Goodfellows, 1986). The saprophyte *Mycobacterium smegmatis* is a fast growing mycobacteria, which is often used as a model of *M. tuberculosis*. The slow-growing mycobacteria include the pathogenic mycobacteria; *M. tuberculosis, M. leprae and M. bovis*. However, *M. leprae* is non-culturable *in-vitro* and can only be cultured outside its natural host (humans) using mouse foot pads, or the nine banded armadillo (Ravisse *et al.*, 1984).

A defining feature of *M. tuberculosis* is the lipid rich envelope, which makes it relatively impenetrable (Jarlier and Nikaido, 1994) and is thought to contribute to the ability of *M. tuberculosis* to survive in phagocytes (Kaufmann, 2001).

1.2.1 Structure of the cell wall

The lipid rich cell wall of *M. tuberculosis* is comprised of at least 4 layers, which overlay the cell membrane (see figure 1.1). The membrane consists of a lipid bi-layer (plasma membrane), encased in peptidoglycan. The peptidoglycan is covalently linked via phosphodiester bonds to arabinogalactan. Mycolic acids are attached to the distal portion of the arabinogalactan (Fenton and Vermeulen, 1996), and glycolipid surface molecules are associated with the mycolic acid. The mannose capped lipoarabinomannan (ManLAM), is tethered via a phosphatidylinositol anchor to the plasma membrane, and extends to the glycolipid surface molecules (Karakousis *et al.*, 2004). There are three classes of mycobacterial lipoarabinomannan (LAM), ManLAM, present in virulent strains, which contains extensive mannose capping, at the arabinan termini, whereas the rapid growing mycobacteria such as *M. smegmatis* possess phospho-myoo-inositol
Figure 1.1: A schematic representation of the cell wall of *M. tuberculosis*. Adapted from (Karakousis *et al.*, 2004). The components are plasma membrane (A), peptidoglycan (B), arabinogalactan (C), Mannose-capped lipoarabinan (D), plasma membrane and cell envelope associated proteins (E), mycolic acids (F) and glycolipid surface molecules associated with mycolic acid (G).
capped LAM (PILAM). The third type of LAM, arabinose capped LAM (AraLAM), was described in *M. chelonae*, a rapidly growing mycobacteria, and lacks mannosylation at the arabinan termini. Constituent parts of the mycobacterial cell wall have potential immunogenic properties. Interestingly, AraLAM, is a more potent inducer of pro-inflammatory cytokines from macrophages, and unlike *M. tuberculosis*, induced the production of TNF-α (Karakousis et al., 2004). ManLAM from *M. bovis* and *M. tuberculosis* do not elicit the toll like receptors (TLR) dependent activation of macrophages (Means et al., 1999). ManLAM can also induce the production of TGF-β from macrophages, which inhibits macrophage and T-cell activation, potentially leading to a T-helper type 2 immune response, and decreased immunity to *M. tuberculosis* (Karakousis et al., 2004). The sulpholipids in the cell wall of *M. tuberculosis* have been associated with increased resistance to killing in the macrophage, thought to be elicited via inhibition of phago-lysosome fusion (Goren et al., 1976). Mycolic acids bind to the acid stain, and render it inaccessible to the action of the destain in the Ziehl-Nielson stain. Lipids comprise over 60% of the mycobacterial cell wall; interestingly, mycolic acid constitutes 50% of the dry weight of the mycobacterial cell envelope. Mycolic acids are strong hydrophobic molecules, which affect permeability of the mycobacteria (Chatterjee, 1997).

### 1.3 Pathogen-host interactions

Tuberculosis or consumption as it was known was initially thought to be caused by a virus, until 1892 when Robert Koch isolated the bacteria from infected lung tissue. The causative agent of tuberculosis, *M. tuberculosis* is a slow growing intracellular human pathogen (Mariani et al., 2000) which is transmitted via aerosol (Saeyers, 2002); tiny water droplets produced by coughing or sneezing enter the lungs via the upper respiratory tract, carrying between 1 and 3 bacilli into the lung alveoli (Graham and Clark-Curtiss, 1999). The bacteria are then
phagocytosed by unactivated macrophages, where they reside in the phagosomal compartment (Goren et al., 1976). After the alveolar macrophages have engulfed the bacteria, specific T cells are stimulated in the draining lymph nodes, which induces containment of the bacteria in granulomas in the lung. These granulomas are comprised of *M. tuberculosis* infected macrophages, which sometimes fuse to create giant cells, along with various T cell populations (Kaufmann, 2001). Interestingly, *M. tuberculosis* prevents phago-lysosome production by inhibiting the fusion of the lysosome with the infected phagosome (Goren et al., 1976). Unactivated macrophages fail to harm *M. tuberculosis*, however activated macrophages (in presence of IFN-γ) can control the growth of the bacteria, but rarely achieve sterile eradication of the pathogen (Kaufmann, 2001). Therefore the bacteria can remain in a dormant state, in the phagosomal compartment of the macrophage, resulting in latent *M. tuberculosis*, or can replicate and form active/acute *M. tuberculosis* (Saeyers, 2002). It has been observed particularly in immuno-compromised patients that most cases of tuberculosis emerge from reactivation of a latent infection (Flynn and Chan, 2001), rather than progression to direct primary tuberculosis (Kaufmann, 2001). It is relevant to note that *M. tuberculosis* is an obligate aerobe but when exposed to gradual oxygen deprivation, like inside a forming granuloma, they are able to survive for extended periods (Sohaskey and Wayne, 2003).

Macrophages are the primary effector cells of the innate immune system, activated following IFN-γ exposure, which subsequently produce reactive oxygen and reactive nitrogen intermediates as part of their anti-microbial repertoire (Kaufmann, 2001). The cytokines TNF-α and lymphotoxin-α3 are important in formation and maintenance of the granuloma, as patients treated with anti-TNF-α for Rheumatoid Arthritis, suffer from reactivation of latent tuberculosis (Kaufmann, 2001). The anti-microbial effects of the macrophage are present via two distinct pathways, both activated by IFN-γ, NADPH phagocyte oxidase (phox) and
inducible nitric oxide synthase (iNOS). Phox catalyses the reduction of molecular oxygen to superoxide ($O_2^-$), which can be further reduced to hydrogen peroxide ($H_2O_2$), thus forming the reactive oxygen intermediates. The reactive nitrogen intermediates are formed using iNOS, whereby nitric oxide (NO), formed directly is oxidised to derivatives such as $NO_2^-$ and $N_2O_3$. The reactive oxygen and nitrogen intermediates can combine to produce compounds such as peroxynitrite (ONOO$^-$), which have potent antimicrobial properties (Lindgren et al., 2005). The intracellular pathogen *Francisella tularensis* is killed in macrophages by ONOO$^-$, however interestingly, other intracellular pathogens such as *M. tuberculosis* and *S. enterica* serovar *typhimurium* have the ability to detoxify ONOO$^-$ via peroxyredoxins (Lindgren et al., 2005).

### 1.3.1 Receptor mediated phagocytosis

Phagocytosis of *M. tuberculosis* is mediated by receptors. There is evidence to suggest that receptor mediated phagocytosis takes place via complement receptor (CR), and mannose receptor (MR) (Fenton and Vermeulen, 1996; Means et al., 1999).

The CR1, CR3 and CR4 receptors have been associated with receptor-mediated phagocytosis of *M. tuberculosis* (Fenton and Vermeulen, 1996). Lipoarabinomannan is a major surface lipoglycan of *M. tuberculosis* (Chatterjee et al., 1992), which is recognised by the MRs (Schlesinger et al., 1994), that are important for phagocytosis of intracellular pathogens and it has been suggested that they contribute to antigen presentation and granuloma formation (Kang and Schlesinger, 1998). The role of Toll-like receptor (TLR) mediated activation of macrophages is not completely understood, yet TLR activation induces production of interleukin-12 (IL-12), a pro-inflammatory cytokine, whose activation of T-cells is thought to be important, therefore the loss of IL-12 induction may be a potential mechanism for avoidance of the immune response by *M. tuberculosis* (Karakousis et al., 2004). TLR's are expressed
mainly on dendritic cells and macrophages, and prime the adaptive immune response, by controlling activation of antigen presenting cells (APC) (Barton and Medzhitov, 2002; Medzhitov, 2001). Different TLR's recognise different antigens, TLR2 and 4 generally recognise bacterial lipopolysaccharide (LPS), whereas TLR9 recognise bacterial CpG DNA: however, TLR2 and TLR4 have been shown to recognise the mycobacterial glycoprotein LAM (Means et al., 1999). Interestingly, TLR4 does not require the CD14 ligand when binding LAM from *M. tuberculosis* (Means et al., 1999).

### 1.3.2 Active versus latent tuberculosis infection

The development of active tuberculosis infection, or the reactivation of a latent infection only happens in a small proportion of individuals (Song et al., 2003), particularly those that are immuno-compromised. It has therefore been suggested that it is important for *M. tuberculosis* to sense the environment of the host, and possess the ability to adapt to the physiological changes of the host, which would result in carefully regulated gene expression in response to the changing host signals during the course of the infection (Shi et al., 2003). It has been suggested that the low ratio of active tuberculosis compared to latent infections is important for the successful survival of the pathogen (Kaufmann, 2001).

### 1.3.3 Immunogenicity of *M. tuberculosis*

Recent publication of the entire genome sequence (Cole et al., 1998), revealed two distinct gene family clusters, PE-PEGRS. These clusters encode glycine-rich proteins, and comprise about 10% of the genome. They share a conserved amino terminus, and are thought to be immunologically important because either these regions could cause antigenic variation,
resulting in evasion from the host immune response or they could impair antigen processing (Kaufmann, 2001). This is particularly relevant as they have homology to some Pro-Glu family proteins in Epstein-Barr virus which interfere with major histocomptability complex class 1 (MHC1) presentation for T cells (Kaufmann, 2001). Deletion of certain PE-PGRS genes in M. marinum impaired growth in granulomas and macrophages (Ramakrishnan et al., 2000).

1.3.4 Immunology of tuberculosis infection

Control of the M. tuberculosis infection is brought about by CD4+ and CD8+ T cells, along with natural killer cells, macrophages, which produce INF-γ, and dendritic cells, which activate other cell-types via IL-12 and IL-18 (Kawakami et al., 2004). This activates the antimicrobial pathways in macrophages, producing a hostile environment for the bacteria, particularly as the phagosomal compartments are low in both nutrients and oxygen, and the activated macrophages produce reactive oxygen (ROI) and nitrogen intermediates (RNI), known to damage lipids, proteins, DNA (Cabisco et al., 2000; Park and Imlay, 2003; Storz and Imlay, 1999) and RNA (Farr and Kogoma, 1991).

1.4 Exposure to DNA-damage

DNA is dynamic, and is continually exposed to a variety of types of DNA-damage, which have lead to the development of mechanisms designed to tolerate and repair the damage to DNA in all living organisms (Friedberg, 2003). An example of the kind of DNA-damage incurred is deamination, which results in miscoding lesions, hence leading to mutations during replication.
1.4.1 Types, effects and repair of DNA-damage

In the case of *M. tuberculosis*, the bacilli are thought to be exposed to a number of DNA damaging agents during infection and persistence in the macrophage, including exposure to ROI and RNI (Nathan and Shiloh, 2000). There are a number of cellular processes, which result in the formation of ROI/RNI, including peroxisomal metabolism, mitochondrial respiration, nitric oxide synthesis, and phagocytic leukocyte metabolism (Friedberg, 1995), suggesting oxidative stress is both an environmental phenomenon and an unavoidable by-product of an aerobic existence (Farr and Kogoma, 1991; Storz and Imlay, 1999). However, it has been suggested that during host-pathogen interactions, it is primarily within phagocytes that the bacteria are exposed to oxidants such as peroxide radicals (\( \bullet O_2 \)), hydrogen peroxide (\( H_2O_2 \)), hydroxyl radicals (\( \bullet OH \)) and singlet oxygen (Friedberg, 1995; Miller and Britigan, 1997). ROI and RNI cause damage to DNA, proteins, lipids and RNA, so pathogens have three options: to interfere with host production of ROI/RNI, to catabolize them or to repair damage caused by these agents (Nathan and Shiloh, 2000). Therefore the repair of damaged DNA is thought to be important for survival and replication of the tubercle bacillus (Davis et al., 2002a).

There are three main DNA repair mechanisms that have been extensively studied in *E. coli*, by which damaged bases are repaired, base excision repair (BER), nucleotide excision repair (NER) and homologous recombination (Friedberg, 2003). More recently non-homologous end joining (NHEJ) was also identified in some bacteria with a role in double-stranded break repair.

BER is a two stage process involved in repair of oxidised and alkylated DNA (Sancar, 1994). Initially the damaged bases are recognised by a DNA glycosylase e.g. *fpg* or *mutY*, and excised
then an Apurinic/apyrimidinic (AP) endonuclease \textit{xthA} or \textit{end} catalyses the formation of single stranded breaks, repaired via DNA polymerase I and ligase (Friedberg, 2003).

NER involves \textit{uvrABC} and \textit{D}, which function synergistically to remove pyrimidine dimers, whereby the \textit{uvrA2B} complex detects lesions in the DNA by sensing disruption of the double helix (Smith \textit{et al.}, 2002), the strands are then separated, and \textit{uvrB} remains bound, while \textit{uvrA} dissociates. \textit{uvrC} then binds and functions in unison with \textit{uvrB} and the \textit{uvrD} to excise a small 13bp region of ssDNA, including the damaged base, which is then repaired by DNA polymerase I and ligase (Friedberg, 2003; Smith \textit{et al.}, 2002).

Homologous recombination is involved in both generation of novel genetic combinations, brought about by processes such as horizontal gene transfer, and the repair of DNA-damage (Colston M. J., 2000; Lorenz and Wackernagel, 1994). Allelic exchange has been utilised to produce gene inactivation knockouts in \textit{M. tuberculosis} among other species. RecA plays a central role in repair, along with RecBCD complex, to form the major pathway of repairing double stranded breaks. RecBCD can act as an ATP-dependent dsDNA exonuclease and a helicase, to produce a single stranded DNA tail, which allows the binding of RecA (Friedberg, 2003). However, RecBCD can also function as a Chi (\chi) specific endonuclease, which recognises specific chi sites (Chedin and Kowalczykowski, 2002). The second or minor pathway requires RecA along with the RecFOR complex to repair lesions at replication forks, and single stranded breaks (Friedberg, 2003). There are a number of other recombination repair genes, \textit{recQ}, \textit{recJ}, \textit{recG} and \textit{ruvABC}. RecQ is a helicase and RecJ an endonuclease involved in pre-synaptic stages. RecG and RuvAB are helicas, which function along with the nuclease RuvC, in resolution of cross-overs post synapsis (Friedberg, 2003).
1.4.2 Response to DNA-damage in prokaryotes

The response of exposure to DNA damaging agents has been most extensively studied in *E. coli*. In this organism, there are two pathways, which modulate gene expression in response to oxidative damaging agents; OxyR and SoxRS. In addition there is the SOS response, a further regulatory system, which responds to DNA-damage (Raivio and Silhavy, 2001).

1.4.2.1 OxyR and SoxRS pathways

Redox regulation is the modulation of protein activity by oxidation and reduction, brought about by redox signalling molecules such as ROI and RNI, e.g. superoxide, hydrogen peroxide and nitric oxide (Zheng and Storz, 2000). OxyR and SoxR are redox sensing transcription factors, extensively studied in *E. coli*, which detect elevated levels in ROI and act by regulating expression of bacterial antioxidant genes, to combat the effects of the ROI (Zheng and Storz, 2000). The OxyR regulon in *E. coli* is thought to control at least eight genes (Farr and Kogoma, 1991). The transcriptional activation of these genes is thought to be mediated directly via activated OxyR (activated by oxidation) binding to the $\alpha$ subunit of the RNA polymerase (Pomposiello and Demple, 2001; Zheng and Storz, 2000), which binds to the promoter regions of the OxyR regulon (Pomposiello and Demple, 2001). SoxR responds to superoxide and nitric oxide in a two step activation cascade (Zheng and Storz, 2000), where SoxR regulates transcription of *soxS*, then SoxS in turn activates transcription of genes involved in the antioxidant response (Zheng and Storz, 2000) by binding to the promoter regions and recruiting RNA polymerase (Pomposiello and Demple, 2001). It has been suggested that both OxyR and SoxRS are autoregulated (Georgiou, 2002; Zheng and Storz, 2000). However, interestingly the sigma factor, $\sigma^E$, is implicated in the regulation of gene expression of a number of genes involved in the stress responses, and has been shown to regulate expression of several
antioxidant genes, including some of those regulated by OxyR and SoxRS (Storz and Imlay, 1999).

OxyR homologues have been characterised in a number of bacteria (Storz and Imlay, 1999), yet a functional homologue has not been found in *M. tuberculosis* (Deretic et al., 1997). It has been shown that all the members of the *M. tuberculosis* complex contain multiple deletions and termination codons in the *oxyR* pseudogene. There is also no homologue in *M. tuberculosis* of *soxR*, suggesting *M. tuberculosis* may have an alternative mechanism of regulation in response to DNA damaging agents such as ROI/RNI.

*M. tuberculosis*, like other bacterial pathogens have evolved protective mechanisms against reactive oxygen and nitrogen intermediates (Zahrt and Deretic, 2002), to which they are exposed in their intracellular lifestyle. It has been suggested that the response to ROI is of particular importance, which involves catalase-peroxidase (*KatG*) and catalytic subunit of alkyl hydroperoxide reductase (*ahpC*). KatG is involved in peroxide stress response and in many bacteria, is generally positively regulated by OxyR (Zahrt et al., 2001), however *M. tuberculosis* lacks a functional *oxyR* (Deretic et al., 1995; Deretic et al., 1997). It was discovered that regulation of *katG* in *M. tuberculosis* takes place via FurR, a negative regulator, which forms part of the KatG locus (Zahrt et al., 2001). KatG and AphC are implicated in the sensitivity of *M. tuberculosis* to isoniazid, particularly as *M. leprae* lacks a functional *katG*, but retains functional *aphC* and *oxyR* genes, and is insensitive to isoniazid (Eiglmeier et al., 1997). Oxidative stress response genes often exhibit coupled expression with genes involved in iron metabolism, such as *sodA* and *sodB*; metalloproteins, which encode superoxide dismutase (Niederhoffer et al., 1990). It has been suggested that Fur negatively regulates *sodA* via Fe$^{2+}$
dependent repression, whereas Fur positively regulates sodB in an indirectly manner (Niederhoffer et al., 1990).

### 1.4.2.2 DNA damage repair in organisms other than *E. coli*

Many basic genetic mechanisms for replication, repair and recombination are conserved between bacterial species. However, there appears to be many differences between species in the mechanistic properties, biological functions and molecular components involved in these processes (Kline et al., 2003). One mechanism that appears to be relatively conserved is the SOS response.

### 1.4.2.3 SOS response

The term SOS response was first coined by Mirolsav Radman in the 1970's (Radman, 1975), and this response has been extensively studied in Gram-negative bacteria, with *E. coli* being the main focus (Courcelle et al., 2001). Current literature indicates that at least 40 genes have been defined as co-ordinately expressed in the *E. coli* LexA regulon (Courcelle et al., 2001; Fernandez De Henestrosa et al., 2000).

The *recA* gene is part of the *E. coli* SOS regulon, which has been shown to be under the control of LexA (Little, 1982), a transcriptional repressor, which binds to a specific region termed the SOS box located in the promoter of SOS response genes (Bertrand-Burggraf et al., 1987; Fernandez De Henestrosa et al., 2000). Some of the genes in the SOS regulon are still expressed to significant levels in the uninduced state, where LexA is bound as a repressor molecule to the SOS box. This is particularly notable with *recA*, which appears to be basally
expressed under uninduced conditions (Friedberg, 1995). This basal level of RecA is possibly required for detection and response to DNA damage, as the SOS response is a dual component system: RecA is an activator (Fernandez De Henestrosa et al., 2000), which recognises single-stranded DNA, often created by blockage of replication forks following DNA-damage (Courcelle et al., 2001). RecA binds to the single-stranded regions and becomes activated, which causes the autocatalytic cleavage of LexA (Sassanfar and Roberts, 1990), in the presence of a nucleoside triphosphate (Friedberg, 1995). The cleaved LexA is no longer able to bind effectively to the SOS box to suppress expression of genes in the SOS regulon (Bertrand-Burggraf et al., 1987), resulting in de-repression of these genes and therefore increased expression. Some proteins have been identified that are DNA-damage inducible, but are independent of the LexA/RecA system (Courcelle et al., 2001).

1.4.2.4 SOS response in \textit{M. tuberculosis}

Although a classical SOS system has been identified in \textit{M. tuberculosis}, where a number of genes have been shown to be DNA-damage inducible independently of LexA and RecA, including \textit{recA} itself. \textit{recA} has two promoters; P2 possesses a defined palindromic LexA binding site, which is regulated by the LexA/RecA system, but the other promoter, P1, does not possess a LexA binding site, and remains DNA-damage inducible in both a \textit{recA} mutant as well as in wild-type H37Rv (Laboratory strain of \textit{M. tuberculosis}) (Davis et al., 2002b).

\textit{M. tuberculosis} \textit{recA} contains an intein (Davis et al., 1991), which is removed from the precursor protein while the mature protein is re-ligated concomitantly. The intein is not thought to play a role in regulation of \textit{recA} (Frischkorn \textit{et al.}, 1998; Papavinasasundaram \textit{et al.}, 1998).
Interestingly, a knockout of recA in *M. bovis* BCG shows little attenuation in mouse model (Sander *et al.*, 2003).

1.4.2.5 Absence of the SOS response in other pathogenic microorganisms

RecA is thought to be essential for transformation, where integration of foreign DNA is required (Koomey and Falkow, 1987) with RecA and RecX being involved in both recombinational repair pathways, via RecBCD and RecF (Koomey and Falkow, 1987; Stohl and Seifert, 2001). However, the part that recA plays in recombinational repair may explain the necessity of the gene in transformation and therefore genetic diversity (Kline *et al.*, 2003). This is thought to be of importance to the human pathogen *Neisseria* (*Neisseria meningitides* and *Neisseria gonorrhoea*) as gene transfer occurs frequently, via transformation of DNA. The transformed DNA was initially thought to be made available by cell lysis, but recent studies show that *Neisseria* is able to secrete DNA via a type IV secretion system, encoded on a genetic island. The ability of *Neisseria* to take up and integrate DNA is especially interesting as it is missing homologues to a substantial number of key SOS response genes; alkylation repair, recombinational repair, replication repair, as well as a few genes involved in; methyl-dependent mismatch repair and base excision repair (Kline *et al.*, 2003).

*Neisseria* resides solely in the human body; the nasopharynx, blood stream, or the genital tract. Therefore, it has been suggested that *Neisseria* is not subjected to extreme environmental conditions, UV radiation or desiccation (Kline *et al.*, 2003). On that basis, it has been predicted that *Neisseria* has developed a DNA-damage repair system specialised to the type of DNA-damage that would occur in its niche (Kline *et al.*, 2003). Notably, genome comparisons between *N. meningitides*, *N. gonorrhoea* and *E. coli* have shown the absence of both a lexA
homologue, and LexA binding sites/SOS boxes, upstream of the promoters of known *E. coli* LexA regulated SOS response genes. It has therefore been suggested that *Neisseria* does not possess the classical SOS response (Black et al., 1998), but may have an alternative damage inducible repair system other than the classical SOS response (Kline et al., 2003).

### 1.5 Regulation of bacterial gene expression

Bacterial gene regulation takes place at a number of different levels, but it has been suggested that particularly in eubacteria, regulation of transcription initiation is the most significant control point (Hughes and Mathee, 1998), owing to the rapid turnover of transcripts. However regulation also occurs at other levels, including transcriptional termination (Henkin, 1996), as well as post-translational processing (Paget *et al.*, 1999).

#### 1.5.1 Transcription

The transcription process comprises of three stages, initiation, elongation and termination (Borukhov and Nudler, 2003). During transcription the RNAP covers approximately a 35bp region of DNA, 12-15bp of this region are unwound, thus forming the transcription bubble. Inside the melted region the template forms a constant heteroduplex with the 3' region of the transcript, covering approximately 8-9bp (Toulme *et al.*, 1999). Site-specific repressor proteins and chromosomal proteins can block transcription elongation, and RNA polymerase has the ability to oscillate back or forward at each template position, stabilising the complex before re-initiating the elongation process. Nevertheless, the DNA template sequence can cause lateral oscillation of the ternary complex, which can result in the slowing down of elongation of RNA, or can result in a pause, or even cessation of transcription (Toulme *et al.*, 1999).
1.5.2 Transcriptional activators/repressors

A common mechanism of gene regulation involves proteins, which act as transcriptional activators or repressors to initiate or suppress gene expression respectively. A repressor binds to the operator, which often overlaps the promoter, and in these instances this inhibits the interaction of the promoter with RNA polymerase, preventing it forming an open complex (Gralla, 1996; Lodish H, 1995). Activator proteins come in two distinct forms: those that bind upstream of, but close to promoters and interact with either the carboxy-terminal domain of the $\alpha$-subunit of RNA polymerase, or with the $\sigma$ factor (Gomez M, 2000) to activate or enhance transcription (Lodish H, 1995) and those that act as distant enhancers (Gralla, 1996). Although distant enhancers are rare, in *E. coli*, they are associated with a holoenzyme containing an alternative sigma factor known as $\sigma^{34}$ (Gralla, 1996).

Master regulatory proteins work as part of the cell cycle machinery to organise and co-ordinate multiple proteins, required to function co-ordinately in order to execute a particular function (McAdams and Shapiro, 2003). Examples of master regulatory proteins are the general stress response proteins in *E. coli* (Hengge-Aronis, 1999). The presence of these master regulatory proteins, directing modular or co-ordinate functions, may simplify the control signals in communication pathways, and may also facilitate the short and long term abilities of an organism to adapt to changes in environmental conditions (McAdams and Shapiro, 2003).

SoxR and OxyR are examples of transcriptional activators, which enhance transcription (Volkert and Landini, 2001), whereas LexA is a repressor protein, which binds to the operator to repress transcription (Bertrand-Burggraf et al., 1987). Many genes and operons are subject to dual regulation, by repressors and activators, as seen with *uspA* in *E. coli* (Kvint et al., 2000).
1.5.3 Architectural proteins

H-NS is a histone like nucleoid structuring protein and pleiotropic/global regulator. It works by adapting cellular responses to external signals, such as osmotic stress, oxidative stress, cold shock, motility and acid tolerance (Hansen et al., 2005) and binds DNA with a preference for bent DNA of AT rich sequence (Dole et al., 2004). The binding of H-NS to DNA forms an oligomeric structure, which prevents binding of the RNA polymerase, or traps it at the promoter (Dole et al., 2004). The way in which the activity of the H-NS is regulated is unknown, but hns is autoregulated, whereby H-NS acts to repress transcription from its own promoter. However, the DNA binding protein FIS activates transcription of hns (Dole et al., 2004). The bgl operon in E. coli is an example of regulation partially controlled by H-NS and is also responsible for regulating the sigma factor RhoS, which is expressed during early stationary phase. This sigma factor is responsible for regulating gene expression in early stationary phase in response to nutrient starvation/ limitation and decreased pH (Hansen et al., 2005).

1.5.4 Multi-factorial concordant regulation

The expression of certain genes can include more than one regulatory mechanism, as shown in the bgl promoter in E. coli. Expression from the bgl promoter is controlled by an activator protein, known as CAP (catabolite gene activator), whose binding site is located upstream of the promoter. However, under normal physiological conditions, the gene is repressed by the presence of silencing elements flanking the promoter (Schnetz and Wang, 1996). DNA topology, in conjunction with cellular factors such as H-NS is important for the expression from the bgl promoter. Reporter assays were performed with a wild-type bgl promoter, a bgl promoter with enhanced CAP binding, and a bgl promoter containing a deletion in the upstream
silencing element. These assays showed that crude cell extract in the absence of CAP represses transcription, where as negative supercoiling overcomes this repression and enables transcription in the presence of the CAP-cAMP complex (Schnetz and Wang, 1996). It has been suggested that H-NS, DNA topology and other cellular factors are involved in the formation of a nucleoprotein structure in the region of the gene silencer, which renders the promoter inactive, until it is activated by the presence of CAP (Schnetz and Wang, 1996), therefore the regulation of the $bgl$ protein is multi-factorial.

Temporal control of gene expression can also be achieved by alarmone guanosine 3', 5'-bisphosphate (ppGpp), which can negatively affect transcription of $rrn$ operon in $E. coli$, but can positively affect transcription of a large number of $\sigma^70$ regulated genes. ppGpp affects the induction of these genes in response to environmental stress (Nystrom, 2004b). When bacteria enter into nutrient limitation, the levels of ppGpp increase, and this enhances transcription of genes regulated in response to stress by alternative sigma factors. The hypothesis is that there is a transcriptional trade off between stress defence and growth and proliferation.

1.5.5 Negative feedback regulation

Different types of physiological and environmental stresses cause different types of damage to a cell, and therefore require different mechanisms to repair the damage. Molecular chaperones (DnaK, GroE, DnaJ and GrpE) and ATP dependent proteases (ClpAP, ClpXP and Lon) are responsible for disposing of potentially toxic protein aggregates, which are formed by damaged or misfolded proteins. These malformed proteins can be formed under normal conditions, but aggregate in response to stress, such as heat shock or pathophysiological stresses (Bucca et al., 2003). The tight regulation of the response to this type of stress is vital, and unravelling the
The regulation of heat shock gene expression in *Streptomyces* is of particular interest, as like *M. tuberculosis*, *Streptomyces coelicolor* (*S. coelicolor*) is a highly GC rich member of the actinomycetes family of gram positive bacteria, and unlike in *E. coli*, the expression of heat shock genes in *Streptomyces* is controlled by negative regulation (Bucca et al., 2003).

The *dnaK* operon in *S. coelicolor* is negatively autoregulated by the binding of a repressor molecule to a specific inverted repeat sequence in the operator (Bucca et al., 2003). It has been suggested that DnaK binds HspR (co-expressed with DnaK), to form a complex, which interacts with the specific inverted repeat sequences, to repress transcription of the *dnaK* regulon (which includes the *dnaK* operon, *clpB* and *lon*). Under normal growth conditions, DnaK depletion experiments showed high level transcription of the *dnaK* operon (Bucca et al., 2003). This method of autoregulation, termed the HspR repressor/operator system is found in other actinomycetes species, including *M. tuberculosis* (Stewart et al., 2001).

### 1.5.6 Bacterial promoters

The promoter of a particular gene is generally defined by two regions, the −10 and −35 sites, which are 5' non-coding regions upstream of the transcriptional start site and which have sequence-specific affinity to RNA polymerase holoenzyme (Hughes and Mathee, 1998). Some promoters also include an element UP, which lies upstream of the −35 site (Gralla, 1996); these
are generally AT-rich elements and interact with the RNA polymerase (Hughes and Mathee, 1998), to enhance transcription.

1.5.6.1 Control of bacterial promoters

There are a number of different mechanisms that have been described for modulating bacterial promoters; DNA topology, activator or repressor molecules (outlined in 1.5.1 and 1.5.1.1), enhancers, UP elements. These regulators control both the temporal and spatial expression of a given promoter.

Temporal control of bacterial promoters is vital for regulation of expression of a given gene at a particular time. Temporal control is important for the maintenance and viability of an organism, particularly with regard to cell cycle, as improper timing of cell-cycle genes has proved to have fatal consequences (McAdams and Shapiro, 2003).

In Caulobacter, ctrA, encodes a protein CtrA, which is responsible for binding and silencing the chromosomal origin of replication in Caulobacter. Like M. tuberculosis recA, the gene has two promoters, P1 and P2; however, unlike recA, the mechanisms of regulation of expression from both promoters have been elucidated. The P1 promoter can only be activated when it is hemimethylated (the intrinsic state of newly replicated DNA). The P2 promoter is autoactivated, as a result of elevated CtrA levels produced from the P1 promoter. This autoactivation takes place when a threshold level of phosphorylated CtrA is reached (McAdams and Shapiro, 2003).
1.5.7 RNA polymerase and formation of an active holoenzyme

The core RNA polymerase is comprised of 5 subunits, $\beta\beta'\alpha_2\omega$ (for a review see (Borukhov and Severinov, 2002). Core RNA polymerase is unable to initiate transcription from promoters (Borukhov and Severinov, 2002), but when it combines with a sigma ($\sigma$) factor, an active holoenzyme is formed, able to initiate transcription of the sigma factor’s regulon. The sigma factor forms the promoter recognition site, with region 2 of the sigma factor interacting with the $-10$ region of the promoter, and region 4 of the sigma factor interacting with the $-35$ site of the promoter (Gralla, 1996), thus giving the RNA polymerase its specificity (for a review see (Severinov, 2000). Shortly after the initiation of transcription, within about the first 10 nucleotides, the $\sigma$ subunit is discharged from the RNA polymerase (RNAP), and transcription continues with the core RNA polymerase (Haldenwang, 1995; Travers and Burgess R.R, 1969).

1.5.7.1 Core RNA polymerase subunits

The core RNA polymerase has a molecular mass of approximately 400kDa (Darst, 2001). The $\alpha$, $\beta$ and $\beta'$ subunits are the larger subunits, whereas $\omega$ is the smallest subunit (Darst, 2001). The $\alpha$ subunits are comprised of two functional domains, a C-terminal domain, and an N-terminal domain (Finn et al., 2000). The C-terminal region of the $\alpha$ subunits, present as a dimer, forms the site of interaction between UP elements and RNAP and with upstream transcriptional activators and the RNAP (Busby and Ebright, 1994). The $\alpha$ subunits form a platform, which binds to the $\beta$ and $\beta'$ subunits, which are involved in transcriptional activation (Zhang and Darst, 1998). The $\beta$ and $\beta'$ subunits form the catalytic centre of the RNAP (Finn et al., 2000) they form a trough for 12bp of duplex DNA, downstream of the ternary elongation complex.
(Darst, 2001). The role of the $\omega$ subunit is unclear, although it is thought to have an effect on core formation, and interacts solely with the $\beta'$ subunit (Naryshkin et al., 2000).

There are differences in *E. coli* and *B. subtilis* between the structure and function of the different subunits, which comprise to form the RNAP. In *B. subtilis*, there is a delta (\(\delta\)) subunit, which reduces non-specific initiation of the RNAP when complexed with the housekeeping sigma factor $\sigma^\text{A}$. However in *E. coli* this reduction of non-specific initiation is thought to be carried out by $\sigma^{70}$, along with the usual role of a sigma factor as promoter recognition site (Haldenwang, 1995). There are also differences in antibiotic targeting for the different RNAP’s but there do not seem to be any major differences between the core enzymes (Haldenwang, 1995).

### 1.5.7.2 Sigma factor domains

Sigma factors are comprised of 4 regions. Region 1 of $\sigma^{70}$ prevents the sigma factor binding directly to the promoter in the absence of the RNAP-$\sigma$ complex (Lonetto et al., 1994). Region 1 can be subdivided into two sections region 1.1, which solely affects DNA binding by $\sigma^{70}$ and is found only in primary sigma factors (Baldwin and Dombroski, 2001), and region 1.2, which affects promoter binding, open complex formation, along with initiation and the transition to elongation (Baldwin and Dombroski, 2001) and unlike region 1.1, is also present in alternative sigma factors. Region 2 and 4 are the most highly conserved regions (Lonetto et al., 1994); region 2 is further subdivided into 2.1, 2.2, 2.3 and 2.4, although recent publications also annotate region 2.5 (renamed from region 3) (Checroun et al., 2004). Region 2.4 and 4.2 recognise the $-10$ and $-35$ promoter elements respectively (Daniels et al., 1990; Siegele et al., 1989), whereas region 2.5 (3) recognises the extended $-10$ motif (Checroun et al., 2004).
Regions 2.1, 2.2 and 3.2 are thought to be involved in core binding (Sharp et al., 1999). ECF sigma factors appear to have lost most of region 3 (3.1 and 3.2), in comparison to primary sigma factors (Lonetto et al., 1994). As mentioned, region 4 is required for recognition and binding to the –35 sites, particularly 4.2 region.

1.5.8 The role of sigma factors

There are two main types of sigma factor: primary sigma factors, also known as housekeeping sigma factors, and alternative sigma factors.

When complexed with core RNA polymerase, primary sigma factors such as $\sigma^{70}$ in *E. coli*, or $\sigma^A$ in *M. tuberculosis* are responsible for transcription of housekeeping genes (Hu and Coates, 1999; Missiakas and Raina, 1998), such as essential biosynthetic pathways including amino acid biosynthesis (Hughes and Mathee, 1998).

Alternative sigma factors, recognise different –10 and –35 promoter sequences compared to primary sigma factors (Hughes and Mathee, 1998). Extra cytoplasmic family (ECF) sigma factors belong to a sub-family of the sigma 70 ($\sigma^{70}$) class of sigma factors, which are highly conserved through many Gram-positive and Gram-negative species (Helmann, 2002), including *Mycobacteria*, *B. subtilis*, *Sulpholobus acidocaldarius*, *S. coelicolor* and *Pseudomonas aeruginosa* (Helmann, 2001; Missiakas and Raina, 1998).

Sequencing of the *M. tuberculosis* genome (Cole et al., 1998), revealed that *M. tuberculosis* has 13 putative sigma factors, each belonging to the $\sigma^{70}$ class of sigma factors. The principle sigma factor is $\sigma^A$ (Fernandes et al., 1999), whereas $\sigma^B$ and $\sigma^F$ belong to the sporulation/stress response
family of sigma factors and the remaining 10 sigma factors are ECF family (Manganelli et al., 2001).

Sigma factors compete to bind with core RNAP and initiate transcription of their regulon suggesting the levels of competing sigma factors provide an additional layer to the regulation of transcription at different stages during growth (Nystrom, 2004a).

1.5.9 Regulation of alternative Sigma factors

Extra cytoplasmic family (ECF) $\sigma$ factors are thought to be key regulatory molecules in the bacterial adaptive response to the host (Graham and Clark-Curtiss, 1999). They provide a response mechanism to various environmental stresses (Song et al., 2003), by co-ordinating transcription of their specific regulons (Missiakas and Raina, 1998), which are otherwise physically unlinked genes (Song et al., 2003). Sigma factors are themselves regulated in order to control temporal gene expression of the sigma factors regulon in response to particular stimuli. There are a number of different modes of regulation for specific sigma factors, usually involving different types of both transcriptional and post-translational regulation:

1.5.9.1 Transcriptional and post-translational regulation of sigma factors by proteolysis

Some sigma factors are both transcriptionally and post-translationally regulated, i.e. they are transcribed as a pro-protein, which contains an amino acid extension at the amino terminus. This extension enables the pro-protein to bind to the cytoplasmic membrane where it remains inactive. The pro-protein is then converted into an active sigma factor by developmentally
triggered proteolysis, which removes the membrane bound pro sequence from the sigma factor pro-protein (Ju and Haldenwang, 2003). This type of regulation has been observed in *B. subtilis*, where the sigma factor SigE is transcribed as a pro-protein. SigE is a sporulation specific sigma factor, which responds to nutrient deprivation known to initiate spore formation (Haldenwang, 1995), as part of a signal transduction cascade. SigE is co-transcribed with a protease SpoIIGA, which is also membrane bound, and remains inactive until septum formation, when SpoIIR is synthesised in the forespore, which in turn activates the protease SpoIIGA. SpoIIGA then causes proteolytic cleavage of the pro-protein, to release the active sigma factor SigE (Ju and Haldenwang, 2003). The transcription of the SigE operon is controlled by an activator protein, Spo0A, which binds to the promoter of *spoIIG* to initiate transcription (Haldenwang, 1995). It has been shown in *M. tuberculosis*, that SigH is required for transcription of the SigE operon, however, RNAP-σ^A^ complex not RNAP-σ^II^ complex was responsible for transcription from spoIIG promoter, suggesting SigH has an indirect effect on the expression of SigE (Raman et al., 2001).

The post-translational modification of SigE into its active form, requires products from *ftsZ*, *divIC*, *spoIIGA*, *spoIIAA*, *spoIAC*, *spoIIE* and *spoIIA* (SigF). However the roles of many of the proteins have not been defined, with the data indicating SigF and SpoIIE affect the processing of pro-SigE indirectly (Haldenwang, 1995). This example highlights the complexity and multi-level regulation involved in control of bacterial gene expression.

This process is not limited to *B. subtilis*, as interestingly, an ECF sigma factor BldN in *S. coelicolor* is regulated at both the transcriptional and post-transcriptional level. On the one hand, its transcription is repressed by BldD, whereas BldN is synthesised as a pro-protein, with an 86 residue N-terminal extension, which is cleaved by proteolysis (Bibb and Buttner, 2003).
1.5.9.2 Two component regulators of sigma factors

The sigma factor $\sigma^{54}$ is involved in a two-component signal transduction cascade to initiate flagella production, in *Campylobacter jejuni* (Hendrixson and DiRita, 2003). The flagella transcription cascade in *Helicobacter pylori*, *Vibrio cholerae* and *Pseudomonas aeruginosa* requires two sigma factors, $\sigma^{54}$ and $\sigma^{28}$. FlgR is part of a two component regulatory system, which acts as a transcriptional activator for the $\sigma^{54}$ regulon, after phosphorylation by FlgS. FlgS is a sensor kinase, which may be able to detect the correct conditions to initiate transcription of the $\sigma^{54}$ regulon (Hendrixson and DiRita, 2003). It has been suggested that FlgS detects proper formation of the flagella secretory apparatus in the cytoplasm, before initiating $\sigma^{54}$ dependent expression of the flagella genes. The other sigma factor $\sigma^{28}$ is regulated via the anti-sigma factor FlgM, which binds to $\sigma^{28}$, preventing transcription of $\sigma^{28}$ regulon (Hendrixson and DiRita, 2003).

In *S. coelicolor*, $\sigma^{E}$ is regulated by a two-component system and $\sigma^{E}$ forms an operon with CseA, CseB and CseC, where CseB is a response regulator and CseC is a sensor kinase (Hong *et al.*, 2002) and CseA is a negative regulator of sigE expression (Hutchings *et al.*, 2004).

1.5.9.3 Post-translational regulation by anti-sigma factors and anti-anti sigma factors

Sigma factors can be post-translationally regulated by environmental stimuli interacting directly with an anti-sigma factor or via an anti-anti-sigma factor, to regulate a particular sigma factors activity (Hughes and Mathee, 1998). The ECF family of sigma factors are usually negatively regulated by their cognate anti-sigma factors (Raivio and Silhavy, 2001), some of which are membrane bound, which often exhibit transcriptionally coupled expression, with their
sigma factors. Studies in *B. subtilis* have shown that anti-sigma factors are negatively regulated by a co-expressed protein termed an anti-anti-sigma factor (Hughes and Mathee, 1998). Also, many ECF sigma factors are positively auto-regulated; they drive transcription from their own promoters (Helmann, 1999). This is observed in *M. tuberculosis*, where transcription of the ECF sigma factor, SigH is initiated from an auto-regulated promoter, as a result of oxidative and heat stress (Song *et al.*, 2003).

Anti-sigma factors are regulated by a number of different stimuli, depending on the type of sigma factor they regulate. Non-ECF anti-sigma factors such as AsiA in *E. coli*, FlgM in *S. typhimurium*, SpoIIAB in *B. subtilis*, are cytoplasmic proteins. Some ECF family of anti-sigma factors are inner membrane proteins, with at least one transmembrane domain (Hughes and Mathee, 1998), which bind sigma factors as part of a signalling cascade (Hughes and Mathee, 1998). Therefore anti-sigma factors function to negatively regulate their cognate sigma factors, by binding reversibly to the sigma factor (Duncan and Losick, 1993; Song *et al.*, 2003). Studies using the *M. tuberculosis* sigma factor SigH and its cognate anti-sigma factor RshA, have elucidated the importance of conserved cysteine residues in the binding of anti-sigma factors to their cognate sigma factors (Song *et al.*, 2003).

SigF in *B. subtilis* is regulated by the anti-sigma factor SpoIIAB (Haldenwang, 1995; Ju and Haldenwang, 2003) and the anti anti-sigma factor SpoIIAA. These have homology to the SigB anti-sigma factor RsbW and the anti anti-sigma factor RsbV respectively (Haldenwang, 1995).

One of the best characterised ECF σ factor mediated stress responses is that of SigE in *E. coli* (Hughes and Mathee, 1998; Raivio and Silhavy, 2001) and forms an example of the interaction between a sigma factor (SigE) and its cognate anti-sigma factor RseA. RseA is situated in the
cytoplasmic membrane, with predicted periplasmic and cytoplasmic domains. Under uninduced conditions, the cytoplasmic domain of RseA binds to SigE, preventing it from binding to core RNA polymerase. Under environmental stimulation, a protease cascade is initiated which results in the proteolysis RseA. DegS and RseP respond to stress signals generated in the envelope, and alongside ClpX, degrade RseA by proteolysis (Alba et al., 2002, Grigorova et al., 2004), which results in the release of SigE and initiation of sigE dependent transcription. RseB is a sensor, which binds to the periplasmic domain of RseA, and DseG, which detects cellular changes in the form of overexpressed outer membrane porins induced by environmental stresses.

1.5.9.4 The role of anti anti-sigma factors

The mode of action of anti anti-sigma factors varies, and has been shown to range from enzymatic modification of the anti-sigma factor, such as phosphorylation (Helmann, 1999), to export of the anti-sigma factor out of the cell (Hughes and Mathee, 1998), to interactions with extracytoplasmic or small proteins (Helmann, 1999). The anti anti-sigma factor regulating SigF in *M. tuberculosis*, is regulated by redox (Beaucher et al., 2002).

The importance of phosphorylation as a method of regulating sigma factors is outlined in the case of the *B. subtilis* sigma factor, SigB. SigB is co-transcribed with and controlled by its cognate anti-sigma factor (RsbW) and anti anti-sigma factor (RsbV). The interaction between RsbW and RsbV is regulated by phosphorylation. In this case, RsbW is only able to bind unmodified RsbV, which suggests that either RsbW is phosphorylating RsbV to prevent the formation of a RsbV-RsbW complex, or the phosphorylation of RsbV is a consequence of the ability of RsbW to change its specificity for either SigB or RsbV (Haldenwang, 1995).
However under certain stresses, an additional co-factor is required, RsbU. It is not clear however which regulatory protein receives the signal to initiate binding of RsbW to RsbV (Haldenwang 1995).

In *E. coli*, the heat shock response elicits expression of over 40 genes, including *dnaK*, *dnaJ*, *grpE*, *clpB*, *lon* and *clp* operon, under the control of $\sigma^H$, an alternative sigma factor (Bucca *et al.*, 2003), so called due to its molecular weight.

### 1.5.10 The importance and mode of detecting environmental signals

SigH, an alternative sigma factor in *M. tuberculosis* is involved in heat and oxidative stress responses (Fernandes *et al.*, 1999; Raman *et al.*, 2001). Its cognate anti-sigma factor, RshA, is part of the SigH operon and inhibits SigH-dependent transcription. The action of the anti-sigma factor RshA is redox dependent, such that it only negatively regulates SigH under reducing conditions. The interaction of SigH and RshA is also disrupted by elevated temperature suggesting the involvement of SigH in response to heat shock, suggesting RshA reacts to both oxidative stress and heat shock (Song *et al.*, 2003). SigH is a particularly interesting sigma factor, as it induces expression of two mycobacterial stress response ECF sigma factors, *sigE* and *sigB* (Hu and Coates, 1999; Manganelli *et al.*, 2001; Raman *et al.*, 2001; Wu *et al.*, 2004). The *rshA* gene is located downstream of *sigH*, and a motif is situated between the coding regions of *sigH* and *rshA*, which is also present downstream of the *M. tuberculosis* sigma factors, *sigE* and *sigL* (Song *et al.*, 2003).

Redox sensing is important for bacteria other than *M. tuberculosis*, which are subject to oxidative or disulphide stresses. This is observed in *S. coelicolor*, whereby SigR-RsrA is
equivalent to the SigH-RshA system in *M. tuberculosis*. SigR, a redox dependent sigma factor, which is regulated via its redox dependent cognate anti-sigma factor RsrA encoded by rsrA (Kang *et al*., 1999). It has been observed that certain cysteine residues are required for the function of the anti-sigma factor RsrA (Kang *et al*., 1999; Paget *et al*., 2001). Interestingly the *S. coelicolor* rsrA shares homology to rshA, an anti-sigma factor from *M. tuberculosis*, also thought to be involved in redox sensing (Kang *et al*., 1999; Paget *et al*., 1998; Paget *et al*., 2001). Interestingly RT-PCR has shown that sigH is co-transcribed with rshA (Song *et al*., 2003) and both SigR from *S. coelicolor* and SigH from *M. tuberculosis* are subject to autoregulation from one transcriptional start site (Paget *et al*., 1998; Raman *et al*., 2001).

1.6 The use of microarrays for global gene expression profiling

It appears that regulation of gene expression is a multi-factorial process; therefore it is conceivable that microarray technology and other high through-put experimental methods may provide a starting point to unravel regulation of gene expression, or a particular regulon. However, it is important to realise that gene expression analysis does not provide information regarding the effects of post-transcriptional mechanisms, which may affect the half life of mRNA, translation initiation, progression, or other modifications, at the structural or chemical level, which may result in differential protein expression (McAdams and Shapiro, 2003).

Microarray technology provides the means to assess genes expression of an entire genome in response to a particular type of stress, or under certain growth conditions. The use of DNA versus RNA arrays also provides the potential to perform inter- and intra- strain comparisons of global gene expression under the desired conditions.
Potentially, it would be a metabolic disadvantage to express DNA-damage repair genes when they are not required. However, there needs to be a tight controlling mechanism, which enables the rapid response of the required DNA-damage repair genes. Therefore it is possible that one or more master regulatory proteins govern the transcriptional regulation of DNA-damage repair genes in *M. tuberculosis*.

### 1.7 Aims

The overall aim of this project is to investigate the alternative mechanisms of regulation of DNA-damage repair genes in *M. tuberculosis*. Two different hypotheses are investigated:

a) That regulation is controlled by an alternative sigma factor. The expression of the sigma factor SigG has been demonstrated to be induced following DNA-damage. Therefore, any role of SigG in expression of DNA-damage inducible genes is assessed by construction and analysis of a *sigG* mutant strain of *M. tuberculosis*, including a detailed analysis of expression of *ruvC* in comparison to *recA*.

b) That induction is determined by an activator or repressor protein. The P1 promoter of *recA* is known to be DNA-damage inducible independently of RecA/LexA, so any proteins interacting with this region are sought using bandshift assays. In addition, the role of the damage induced regulatory protein Rv2017 is investigated following construction of a gene inactivation mutant of *M. tuberculosis*. 
2 Materials and Methods:

2.1 E. coli bacterial strains:

*E. coli* bacterial strains used in cloning and expression are listed below:

DH5α subcloning/library efficiency (Invitrogen): F'\phi80lacZΔM15 Δ(lacZYA-argF) U169 deoR recA1 endA1 hsdR17(rK-, mk+) phoA supE44 λ-thi-1 gyrA96 relA1

XL1 Blue Supercompetent cells (Stratagene): recA1 endA1 gyrA96 thi-Z hsdR1 7supE44 relA1 lac[F'proABlacZAΔM15 Tn10(TetR)]

XL10 Gold Ultra competent cells (Stratagene): TetR Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacIqZΔM15 Tn10 (TetR)] Amy CamR

One-shot chemically competent cells (Invitrogen): F- mcrA Δ(mrr-hsdRMS-mcrBC) \phi80lacZΔM15 ΔlacX74 recA1 ara139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG

2.2 M. tuberculosis bacterial strains:

H37Rv: Laboratory strain

ΔrecA recA knockout strain of *M. tuberculosis*

ΔsigG sigG knockout strain of *M. tuberculosis*

ΔsigG full operon comp 8T: Complement of ΔSigG containing entire operon

ΔsigG partial comp Δ1: Complement of ΔSigG containing sigG only

ΔRv2017: Rv2017 knockout strain of *M. tuberculosis*

2.3 Bacterial Media:

*M. tuberculosis* was grown in Modified Dubos medium (Difco, see appendix I) containing 10% albumin (v/v) for liquid culture, and *E. coli* was grown in L-Broth supplemented with the
appropriate antibiotics (see table 2.1). For solid culture, *M. tuberculosis* and *E. coli* were grown on 7H11 or L-Agar respectively containing the appropriate supplements (see table 2.1):

**Table 2.1: Media supplements**

<table>
<thead>
<tr>
<th>Media supplements</th>
<th>Concentration in <em>E. coli</em> µg/ml</th>
<th>Concentration in <em>M. tuberculosis</em> µg/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kanamycin</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>Hygromycin</td>
<td>250</td>
<td>50</td>
</tr>
<tr>
<td>Ampicillin</td>
<td>100</td>
<td>n/a</td>
</tr>
<tr>
<td>X-gal</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

### 2.4 Plasmids

**Table 2.2: Plasmid descriptions**

<table>
<thead>
<tr>
<th>Vector</th>
<th>Description</th>
<th>Source of reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vectors for <em>M. tuberculosis</em> knockout and complementation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC-Hyg</td>
<td>Hyg&lt;sup&gt;R&lt;/sup&gt; cassette</td>
<td>(Mahenthiralingam et al., 1998)</td>
</tr>
<tr>
<td>pGoal17</td>
<td><em>aph</em> (Km&lt;sup&gt;R&lt;/sup&gt;), <em>sacB/lacZ</em> cassette</td>
<td>(Parish and Stoker, 2000)</td>
</tr>
<tr>
<td>pBackbone</td>
<td>Skeleton plasmid + Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Gopaul, 2002)</td>
</tr>
<tr>
<td>pGarthdee2</td>
<td>pBackbone (Km&lt;sup&gt;R&lt;/sup&gt;) + 5’ and 3’ <em>sigG</em></td>
<td>(Parish and Stoker, 2000)</td>
</tr>
<tr>
<td>pLD1</td>
<td>pGarthdee 2 (Km&lt;sup&gt;R&lt;/sup&gt;) + Hyg&lt;sup&gt;R&lt;/sup&gt; + <em>sacB/lacZ</em></td>
<td>(Gopaul, 2002)</td>
</tr>
<tr>
<td>pKP186</td>
<td>pMV306 (Km&lt;sup&gt;R&lt;/sup&gt;) (Stover et al., 1991), no integrase</td>
<td>Papavinassasundaram</td>
</tr>
<tr>
<td>pBSINT</td>
<td>contains integrase, Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Springer et al., 2001)</td>
</tr>
<tr>
<td>pLD1comp8T</td>
<td>pKp186 (Km&lt;sup&gt;R&lt;/sup&gt;) including full <em>sigG</em> operon</td>
<td>This study</td>
</tr>
<tr>
<td>pLD1compD1</td>
<td>pKp186 (Km&lt;sup&gt;R&lt;/sup&gt;) including only <em>sigG</em></td>
<td>This study</td>
</tr>
<tr>
<td>pLD2</td>
<td>pBackbone (Km&lt;sup&gt;R&lt;/sup&gt;) + 5’ and 3’ <em>Rv2017</em></td>
<td>This study</td>
</tr>
<tr>
<td>pLD3</td>
<td>pLD2 (Km&lt;sup&gt;R&lt;/sup&gt;) + Hyg&lt;sup&gt;R&lt;/sup&gt; + <em>sacB/lacZ</em></td>
<td>This study</td>
</tr>
<tr>
<td><strong>Vectors for Primer extension and RNase protection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pBluescript SK-</td>
<td><em>Amp&lt;sup&gt;R&lt;/sup&gt;</em></td>
<td>Stratagene</td>
</tr>
<tr>
<td>pLDseq1</td>
<td>pBluescript <em>Amp&lt;sup&gt;R&lt;/sup&gt;</em> + 639bp <em>sigG</em> inc 512bp upstream</td>
<td>This study</td>
</tr>
<tr>
<td>pLDseq2</td>
<td>pBluescript <em>Amp&lt;sup&gt;R&lt;/sup&gt;</em> + 322bp <em>ruvC</em>, inc 242bp upstream</td>
<td>This study</td>
</tr>
<tr>
<td>pCR4-Blunt</td>
<td><em>Amp&lt;sup&gt;R&lt;/sup&gt;, Km&lt;sup&gt;R&lt;/sup&gt;</em></td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pLDRNAse1</td>
<td>pCR4-Blunt *Amp&lt;sup&gt;R&lt;/sup&gt;, Km&lt;sup&gt;R&lt;/sup&gt; + 339bp <em>sigG</em>, inc 151bp upstream</td>
<td>This study</td>
</tr>
<tr>
<td>pLDRNAse2</td>
<td>pCR4-Blunt *Amp&lt;sup&gt;R&lt;/sup&gt;, Km&lt;sup&gt;R&lt;/sup&gt; + 322bp <em>ruvC</em>, inc 242bp upstream</td>
<td>This study</td>
</tr>
<tr>
<td>pLDRNAse3</td>
<td>pCR4-Blunt *Amp&lt;sup&gt;R&lt;/sup&gt;, Km&lt;sup&gt;R&lt;/sup&gt; + 358bp <em>recA</em>, inc 271bp upstream</td>
<td>This study</td>
</tr>
</tbody>
</table>
2.5 DNA preparation

2.5.1 Purification of Plasmid DNA: S.N.A.P. Miniprep Kit (Invitrogen)
This Miniprep kit is designed to purify up to 20μg of high copy number plasmid DNA from cultures of E. coli. Silica-gel membranes are used to adsorb plasmid DNA from bacterial lysates. The protocol incorporates a modified alkaline lysis method of the bacteria, the lysate is then neutralized and adjusted to high salt concentrations, which enables the selective adsorption of plasmid DNA to the Silica-gel membrane in the Miniprep columns. RNA, bacterial DNA, proteins and metabolites flow through the silica gel membrane and are therefore discarded. A wash step ensures that endonucleases are removed from the preparation. DNA extraction was performed in accordance with the manufacturer's guidelines.

2.5.2 Purification of plasmid DNA: QIAprep Spin Miniprep Kit Protocol (Qiagen)
This high speed Miniprep kit is designed to purify up to 20μg of high copy number plasmid DNA from cultures of E. coli. The manufacturer's QIAprep Spin Miniprep Kit Protocol was followed for the preparation of Plasmid DNA.

2.5.3 Large scale plasmid extraction: HiSpeed Plasmid purification Kit (QIAgen)
This Midiprep kit is designed to purify up to 200μg of high/low copy number plasmid DNA from cultures of E. coli. DNA extraction was performed in accordance with the manufacturer's guidelines.
2.6 Nucleic acid preparations from *M. tuberculosis*

2.6.1 DNA extractions

A single *M. tuberculosis* colony was streaked onto either quarter, half or whole plate (7H11) and incubated at 37°C for 3 to 4 weeks. The quarter or half plate (7H11 plates) of *M. tuberculosis* colonies were harvested using a sterile loop, and incubated at 80-90°C for 1 hour in TE. Then a final concentration of 2mg/ml of lysozyme and lipase, and 5μl RNase (DNase free RNase, Boehringer) was added and incubated at 37°C for 2 hrs. Samples were then frozen using dry ice and ethanol, then incubated at 75°C for 10 mins. Samples were cooled to room temperature, then 500μg/ml proteinase K was added, along with 0.5% SDS and 2μl RNase, and were incubated at 50°C for 1hr. DNA was purified with two extractions using an equal volume of phenol:chloroform:isoamyl alcohol, followed by one extraction with an equal volume of chloroform. DNA was then precipitated by the addition of 1/50th volume of 5M NaCl, and 2 volumes of 100% ethanol.

2.6.2 RNA extraction: FastRNA Pro Blue Kit (BIO 101 systems)

RNA was extracted from 100-200ml cultures of an appropriate strain of *M. tuberculosis* in accordance with the manufacturer’s guidelines. Cultures were harvested at room temperature for 15 minutes and 10,000rpm. Pellets were resuspended in 1ml RNApro solution (provided in the kit) per 25mls of culture, and transferred to Lysing Matrix B (1ml per tube of Martix B). Samples were then ribolysed at speed 6 for 40 seconds and then centrifuged at 13,000rpm for 10 minutes, after which the supernatent was transferred to 300μl chloroform and vortexed. Samples were centrifuges at 13,000rpm, 4°C for 15 minutes, then the top phase was transferred to 500μl 100% EtOH and ethanol precipitated overnight at -20°C, and after washing in 70%
Materials and Methods

EtOH, samples were resuspended in 50µl DEPC dH₂O. DNA was then removed from the sample by addition of 20 units RNase-free DNase (Roche), 20 units RNase inhibitor (Invitrogen), 5.7mM MgSO₄ and 0.1M NaOAc (final concentration), and the samples were incubated at 37°C for 1 hour. After 1hr, additional DNase and RNase inhibitor were added as outlined above and incubated for an additional hour at 37°C. Samples were purified using RNeasy Minikit (QIAgen), and RNA was analysed by agarose gel electrophoresis and quantitated using a spectrophotometer.

2.7 Preparation of DNA for Cloning

2.7.1 Polymerase Chain Reaction (PCR):
Reactions were carried out in a final volume of 50µl, containing 2 units of *Pfu* turbo DNA Polymerase, 10-250ng template DNA/a single bacterial or yeast colony (Colony PCR), 0.2mM each dATP, dCTP, dGTP and dTTP, 20 pmol each primer, 1X *Pfu* reaction buffer (Stratagene). Samples were run in a Hybaid Omn-E or Omnigene Thermal Cycler, with an initial step of 95°C for 5 min, followed by 30-35 cycles at 94°C for 30 sec, annealing temp (Tm –5) for 30 sec, 68°C for 1 min per kb. The annealing temperatures were differed depending on the predicted Tm for the specific primers. All programmes were run with the Hybaid Thermal Cycler lid heated to 110°C to prevent evaporation of the samples. PCR products were purified and either restriction digested for cloning 2.7.6. Alternatively they were cloned directly into pCR4-Blunt in accordance with the manufacturers protocol (Invitrogen).
2.7.2 Touch-down PCR:

PCR reactions were set up as outlined in 2.7.1. Samples were run in a Hybaid Omn-E or Omnigene Thermal Cycler with an initial step of 95°C for 5 min, followed by 10 cycles of 94°C for 30 sec, annealing temperature (Tm + 10°C) for 30 sec, 68°C for 1 min per kb with step wise decrease of the annealing temperature (1°C per cycle). The final 25 cycles were carried out at the specific Tm.

2.7.3 Ligation and cloning of PCR products

PCR products were either gel extracted as outlined in section 2.7.5, or were purified using QIAquick PCR purification Kit (Qiagen), to remove unincorporated nucleotides and enzymes from the reaction. Purified PCR products were incubated at 37°C for 1 hour with 1X T4 polynucleotide kinase buffer (70mM Tris-HCl pH 7.6, 10mM MgCl₂, 5mM dithiothreitol with 1mM ATP) final concentration and 10 units T4 polynucleotide kinase (New England Biolabs), the polynucleotide kinase was then heat inactivated at 65°C for 10 min. Kinase reaction enables the transfer and exchange of a phosphate group form the ATP to the 5' hydroxyl terminus of the PCR product, which enables downstream ligation of the insert into the vector of choice, outlined in section.

2.7.4 Agarose Gel Electrophoresis

A solution of agarose powder, 0.8-4% (w/v) was prepared by heating to boiling point with electrophoresis buffer, 1x TBE (40mM Tris-borate pH8.3, 1M EDTA) or 1X TAE (40mM Tris-acetate pH8.3, 20mM sodium acetate, 1mM EDTA), and 0.3µg/ml ethidium bromide was
Materials and Methods

added. Samples were prepared by adding 1/5th loading buffer (30% (v/v) Glycerol, 0.25% (w/v) Bromophenol blue, 0.25% (w/v) Xylene cyanol FF) and were loaded along with either 1kb-DNA-ladder, λEcoHind, λHind or 100bp ladder marker (Life-Technologies) used to size DNA fragments. A Consort E833 electrophoresis power supply was used to run the agarose gel immersed in 1X TBE or 1X TAE, at 45 to 90 Volts. The DNA in the gel was visualized on 2UV transilluminator at 302nm and photographed with a UPV BioDoc It System.

2.7.5 Extraction of DNA from Agarose gel

Extraction of DNA fragments was achieved using one of two methods: Microcon columns (Micon), were used with the specific NCO (nucleotide cut off) depending on the size of the required fragment. These columns were also used for concentrating samples and removal of unincorporated nucleotides/primers following PCR. Alternatively QIAquick Gel extraction kit (Qiagen) was used to extract 70bp to 10kb fragments of DNA from certified molecular biology agarose (Bio-Rad) in either TAE or TBE buffer. Each method was performed in accordance with the relevant manufacturers’ protocol.

2.7.6 Digestion of DNA with Restriction Endonucleases

One unit of restriction endonuclease (RE) is required to cleave 1μg λDNA in 1hour at 37°C. A final concentration of 10 units of RE (Boehringer Mannheim, or New England Biolabs) per μg DNA was incubated with the manufacturer’s recommended buffer at the appropriate temperature (usually 37°C) for 2 hours-overnight.
2.7.7 Dephosphorylation of linearised plasmids

Plasmids were linearised using restriction digestion (as outlined in section 2.6.5), then dephosphorylated by the addition of 6 units of calf alkaline phosphatase (Roche) per 500ng linearized plasmid DNA, samples were incubated at either 37°C for 1 hour, or 50°C for 45 mins. This prevents re-ligation of the plasmid, by removal of the terminal phosphate group. Samples were then extracted with phenol chloroform and ethanol precipitated (as outlined in sections 2.7.8 and 2.7.9).

2.7.8 Ethanol Precipitation of Nucleic Acids

DNA was precipitated and purified from solution by incubation at -20°C for 15 min-o/n, following the addition of 2.5 volumes absolute ethanol (EtOH) with 0.1 volumes 3M NaOAc pH 6.0. RNA was precipitated and purified at -80°C over night following the addition of 2 volumes absolute EtOH with 0.1 volumes 3M NaOAc pH 4.8. Samples were then centrifuged at 13,000 rpm for 15 min at 4°C. The pellet was washed with 70% EtOH (made with DEPC treated water (Sigma) for the RNA), and centrifuged at 13,000 rpm for 5 min, at 4°C. The pellet was air dried and resuspended in TE (10mM Tris, 1mM EDTA), dH₂O (for DNA) or DEPC treated water (for RNA).

2.7.9 Phenol-Chloroform extraction of DNA

Phenol chloroform extraction was used to remove protein contaminants from nucleic acid preparations. A solution of 50% phenol, 48% chloroform and 2% isoamylalcohol was used to purify nucleic acids prior to ethanol precipitation. 1 volume of phenol-chloroform was added to the DNA preparation, and centrifuged at 13,000 rpm for 5 min. The top layer containing
aqueous DNA was removed and added to an equal volume of chloroform (24:1, Chloroform:Isoamylalcohol), and centrifuged at 13,000 rpm for 5 min. The top layer was then removed and the DNA was precipitated with ethanol (see section 2.7.8). Alternatively, Phase lock tubes (Eppendorf) were used to separate the phenol-chloroform purification of the sample in accordance with the manufacturer’s protocol, after which, a chloroform extraction and EtOH precipitations were performed as previously described.

2.7.10 Ligation of DNA

Ligations were carried out with either blunt-ended or sticky-ended linear DNA produced as a result of restriction digest with the required restriction enzyme(s) (Outlined in section 2.7.5). The inserts were cloned into dephosphorylated linearized vectors in a ratio of either 1:1, 1:2 or 1:4, (vector: insert respectively) depending on the construct. Ligation reactions were carried out using Rapid DNA ligation Kit (Boehringer) in accordance with the manufacturers protocol and were transformed immediately into chemically competent cells as outlined in section 2.7.11.

2.7.11 Transformation of chemically competent *E. coli*

2.7.11.1 DH5α Subcloning efficiency/Library efficiency (Invitrogen)

Transformations were carried out with 10-100ng of plasmid DNA (derived for section 2.5.1) in accordance with the manufacturers’ protocol.

2.7.11.2 One-shot chemically competent cells

Transformations were carried out with 10-100ng of plasmid DNA, containing a PCR product cloned directly into pCR4-Blunt. The transformations were carried out in accordance with the manufacturer’s protocol.
2.7.12 Automated DNA Sequencing: ABI Prism BigDye Terminator

Cycle Sequencing Ready Reaction protocol:

After cloning, constructs were checked for accuracy by restriction digest and sequencing. Automated sequencing reactions were carried out in accordance with the manufacturer's protocol. The unincorporated fluorescently labelled ddNTP's were removed using EtOH precipitation, or Dye X kits (QIAgen) in accordance with the manufacturers protocol. The resultant samples were then analysed using an ABI Prism 377 DNA sequencer, in a denaturing polyacrylamide sequencing gel, and analyzed with the recommended software, Sequence Analysis 3.4.

2.7.13 Site directed Mutagenesis (SDM): QuickChange Site directed mutagenesis Kit (Stratagene)

SDM is a rapid in vitro method of producing site-specific mutations in a gene of interest, using Pfu DNA polymerase. Primers were designed for the mutagenesis reaction and the experiment was performed using the sigG operon in pKP186 as a template, in accordance with the manufacturer's protocol. The plasmid was then DpnI digested and transformed into either XL1 Blue supercompetent cells or XL10 Gold Cells in accordance with the manufacturers protocol.

2.8 Protein analysis

2.8.1 Synthesis of antibodies

Protein sequences were downloaded form the TubercuList website for SigG (RV0182c), Rv0181c and Rv0180c, and potential antigenic peptides were designed and synthesised by
trained technicians at Sigma-Genosys. The three best candidates were then used to raise antibodies in two rabbits per peptide (Sigma-Genosys).

2.8.2 SDS page gel electrophoresis

15% non-denaturing polyacrylamide SDS-PAGE gels were used for protein visualization as outlined in (Sambrook et al., 1989), or Nu-Page Bis-tris gels (Invitrogen) were used in accordance with the manufacturers protocol, using the XCell sure lock Mini-cell gel system (Invitrogen). In brief, samples were heated to 90°C for 2 mins in NuPAGE buffer (Invitrogen) and reducing agent was added (50mM dithiothreitol), then samples were electrophoresed using Nu-Page Bis-tris gels (Invitrogen), with NuPAGE running buffer (Invitrogen). Gels were run at 200volts for 45 mins-1 hour and visualised using Coomassie (Sambrook et al., 1989).

2.8.3 Dot blot

Dot blots were performed using 0.1, 0.3 and 0.6µg of peptide, alongside 10µg M. tuberculosis CFE, E. coli CFE and SigG inclusion body preparation. Sample volumes were made up to 2µl using phosphate buffered saline (PSB), then 2µl 2X SDS sample buffer (0.15M Tris (pH 6.8), 1.2% SDS, 30% v/v glycerol, 15% β-mercaptoethanol, 1.8µg/ml bromophenol blue) was added. The samples were spotted onto nitrocellulose membrane, and allowed to dry for 30 mins before a western blocking antibody binding and wash steps were performed at outlined in (Sambrook et al., 1989). Dot blots were visualised using Supersignal West Pico chemiluminescence in accordance with the manufacturers protocol (Pierce).
2.8.4 Western Blot

Western blots were performed using the XCell sure lock Mini-cell gel system (Invitrogen), with the XCell II module for transfer of the Western blot (Invitrogen). Briefly, samples were prepared in SDS-loading buffer, heated to 90°C for 5 mins and electrophoresed at 200 volts for 40 mins to 1 hour, as outlined in section 2.8.2. The PVDF membrane (Invitrogen) was soaked in methanol, followed by reducing NuPAGE Transfer Buffer (with 10% methanol). The transfer was performed in the XCell II module at 30V for 1 hour. The PVDF membrane was then washed in dH₂O, before being blocked overnight in 5% milk, 0.1% Tween-20 in PBS, with gentle agitation. The primary antibody was added at either 1/500 (anti-Rv0181c and Rv0180c) or 1/250 (anti-SigG), and incubated for 1 hour with gentle agitation. Four wash steps were performed with 0.1% tween in PBS, followed by the addition of the secondary antibody into blocking solution (described above) all with gentle agitation. Horse Radish Peroxidase (HRP) goat-anti-rabbit antibodies (BD Biosciences) were used as the secondary antibody, at a 1/5000 dilution. The PVDF membrane was then washed a total of 5 times in 0.1% tween-20 in PBS and visualised using Supersignal West Pico chemiluminescence in accordance with the manufacturers protocol (Pierce).

2.9 Transcriptional analysis

2.9.1 cDNA synthesis

For cDNA synthesis random primers (7.5μg) were annealed to 1μg RNA in a reaction volume of 10μl. Samples were incubated at 65°C for 10 min, and then cooled to room temperature. To the reaction, 1X reaction buffer (Roche), 1μg Acetylated BSA, 0.2mM dNTP’s (with DEPC treated dH₂O), 125mM DTT and 1μl RNase inhibitor was added. To the RT+ reaction (with
RNA), 1.5µl Superscript was added. To the negative control RT- reaction (with RNA), 1.5µl of DEPC dH₂O was added. Samples were then incubated at 37°C for 1 hour and the Superscript was heat inactivated at 70°C for 15 mins.

### 2.9.2 RT-PCR

RT-PCR was used to determine if genes were co-transcribed or to check for expression. RT-PCR reactions were performed using the cDNA (see section 2.8.1). 2µl of RT + or - reactions, were added to 0.2mM dNTPs, 300nM forward primer, 300nM reverse primer, 1X *Pfu* turbo buffer (Stratagene), 2.5µl DMSO, 2.5units *pfu* turbo. Samples were run in a Hybaid OmE or Omnigene Thermal Cycler, with an initial step of 94°C for 3 min, followed by 35 cycles at 94°C for 45 sec, 58°C for 45 sec, 72°C for 2 min, and finally 72°C for 7 mins. The programme was run with the Hybaid Thermal Cycler lid heated to 110°C to prevent evaporation of the samples.

### 2.9.3 Taqman real-time PCR

Taqman is a real-time quantitative PCR method, which enables quantitation of both DNA and cDNA to determine a quantitative analysis of gene expression. The method involves a DNA probe labelled with a fluorophore (FAM) and a quencher (TAMRA), as previously outlined in (Brooks *et al.*, 2001). The labelled probe binds to the amplification product. During each PCR cycle, extension from the primer, causes the 5’ end of the probe to be displaced and degraded, which separates the fluorophore and the quencher, resulting in fluorescence of the fluorophore, which is in turn detected by the TaqMan machine (ABI7700). Primers and probe were designed using Primer Express as outlined in the manufacturer’s protocol, and experiments were performed using the Taqman Universal PCR Master Mix (PE Applied Biosystems). 1µg RNA
was reverse transcribed into cDNA as outlined in section 2.9.1. This was then diluted 1/10 or 1/20 and used in the Taqman reactions with the probe and relevant primer pairs. Both the \textit{sigA} and test probes were used on the same samples of RNA in order to quantify the test PCR product. For each of the test probes an optimization was carried out using a standard 5\(\mu\)M probe along with differing concentrations of each primer (see table 2.3). All reactions were carried out in 96 well plates alongside DNA standards of known concentration ranging from 1ng to 8pg, for each of the probes (\textit{sigA} and test probe).

### Table 2.3: Optimization primer concentrations

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Concentration of forward and reverse primer (nM)</th>
<th>Reaction</th>
<th>Concentration of forward and reverse primer (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50F/50R</td>
<td>F</td>
<td>300F/900R</td>
</tr>
<tr>
<td>B</td>
<td>50F/300R</td>
<td>G</td>
<td>900F/50R</td>
</tr>
<tr>
<td>C</td>
<td>50F/900R</td>
<td>H</td>
<td>900F/300R</td>
</tr>
<tr>
<td>D</td>
<td>300F/50R</td>
<td>I</td>
<td>900F/900R</td>
</tr>
<tr>
<td>E</td>
<td>300F/300R</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

#### 2.9.4 End-labelling of oligonucleotides

End-labelling of oligonucleotides was carried out in a final volume of 10\(\mu\)l, using 10pmol specific primer/oligonucleotide, in a 1X T4 polynucleotide kinase (PNK) buffer (Promega), with 10 units of T4 PNK (Promega), and 30\(\mu\)Ci [\(\gamma\)-32P]ATP. Reactions were incubated at 37°C for 10 mins, and heated at 90°C for 2 mins to heat inactivate the T4 PNK. 9\(\mu\)l were removed, and added to 41\(\mu\)l nuclease free water, then applied to a G25 Sephadex spin column (Life Tech), to remove any unincorporated [\(\gamma\)-32P]ATP. The volume was adjusted to give a final concentration of the end-labelled primer of 180fmol/\(\mu\)l.
2.9.5 Primer extension reactions

Primers used in primer extension (PE) reactions were end-labelled (as outlined in section 2.9.4). 40-100µg total RNA was EtOH precipitated o/n at -80°C, and resuspended in 5µl DEPC treated dH₂O. In brief, RNA was precipitated overnight as outlined in section 2.7.8, and was resuspended in 5µl DEPC dH₂O. Annealing reactions were performed in a final volume of 11µl, with 5µl RNA, 1X avian myeloblastosis virus (AMV) reverse transcriptase buffer and 180fmol of end-labelled primer and were incubated at 58°C for 30 mins (or Tm-5°C). Samples were left for 1hour to cool to room temperature, and the extension reactions was performed with 1X AMV buffer (Promega), 2.8mM Sodium Pyrophosphate and 1 unit of AMV reverse transcriptase to make a total volume of 20µl, then incubated at 42°C for 30-40mins. Samples were EtOH precipitated over night, and resuspended in Formamide loading dye (Promega). Primer extension reactions were carried out in accordance with the manufacturer’s protocol (Promega). PE reactions were run on an 8% denaturing polyacrylamide gel, alongside manual sequencing reactions (outlined in section 2.9.6), and visualized using autoradiography.

2.9.6 Manual Sequencing: T7 Sequenase V2.0 (Amersham LS)

5µg of template DNA was denatured with 0.1 volumes (2M NaOH, 2mM EDTA) at 37°C for 30 min. The reaction was neutralized by adding 0.1 volumes 3M NaOAc (pH 4.8), and EtOH precipitated. Samples were resuspended in 7µl dH₂O. Sequencing reactions were carried out in accordance with the manufacturer’s protocol, and run on an 8% denaturing polyacrylamide gel, alongside primer extension reactions (see section 2.9.5).
2.9.7 RNase protection

2.9.7.1 In-vitro transcription of the radio-labelled probe

In-vitro transcription reactions were performed using the MAXIscript in-vitro transcription kit (Ambion), with either T3 or T7 polymerase in accordance with the manufacturers protocol. The templates were linearized using either NotI (for ruvC and recA), or SpeI (for sigG) restriction digests. These were then purified using a PCR clean up column (Qiagen) in accordance with the manufacturers protocol. 2.5μg of template was used in a 20μl in-vitro transcriptional reaction, with 1X transcription buffer, 0.5mM ATP, CTP and GTP, and either 0.25, 0.5, 1 or 5mM UTP, 10mCi/ml [α-32P] UTP, and either 30 units T7 polymerase or 60 units T3 polymerase. Reactions were incubated for 1 hour at 37°C (or 15°C), 2μl yeast RNA was added as a carrier, then DNase treated with 2 units DNase I, incubated at 37°C for 15-30mins, to stop the reaction, 1μl 0.5M EDTA was added. Reactions were purified using NucAway columns (Ambion) in accordance with the manufacturers protocol. In the initial test probe synthesis reactions, 0.25mM, 0.5mM, 1mM and 5mM un-labelled UTP was added to the in-vitro transcription reaction. For subsequent probe synthesis reactions 1mM non-labelled UTP was used. The test templates were synthesised using the same protocol except the radio-labelled [α-32P] UTP was omitted. 1μl probe was visualised using autoradiography (denaturing polyacrylamide gel), and then quantified using a scintillation counter and used directly in the RNase protection assays at 5x10^4 CPM. The century marker (Ambion) was in-vitro transcribed in accordance with the manufacturer’s guidelines.

2.9.7.2 RNase protection assay

RNase protection assays were performed to identify transcriptional start site, and to approximate the level of transcription from each promoter. 20-50μg RNA were used in the
RNase protection assays with $5 \times 10^4$ CPM of purified probe, negative control (no RNA) and positive control (test templates) were performed alongside. The RNA/controls were co-precipitated with the probe, as outlined in the manufacturers guide lines, then resuspended in 10μl hybridization buffer (Ambion), and incubated at 95°C for 4 mins, followed by an overnight incubation at 42°C in a hybaid oven to ensure even distribution of heat. Yeast RNA was added to all samples as a carrier and digestions were initially tested with a range of RNase A/T1 mix, 1:50, 1:100 and 1:500, alongside RNase T1 alone at 1:50, 1:100 and 1:500. For subsequent reactions, 1:100 ratio of the RNase A/T1 mix was used in all reactions except the negative control (probe only). Digestions were performed at 37°C for 30 mins, then reactions were stopped using the RNase inactivation/precipitation buffer (Ambion). A further 2μl Yeast carrier RNA was added prior to precipitation at -20°C for 15 mins. After precipitation, samples were eluted in 4μl gel loading buffer II (Ambion), the heated for 3 mins at 95°C and loaded onto the 5% denaturing polyacrylamide gel, alongside marker (produced by *in-vitro* transcription), then visualised by autoradiography or using a phosphorimager.

### 2.9.8 Denaturing polyacrylamide gel electrophoresis

Denaturing polyacrylamide gels 8% (w/v) were used to visualize manual sequencing and primer extension reaction, and 5% for RNase protection assays. The 8% polyacrylamide gel was made using Long Ranger in accordance with the manufacturer's protocol (FMC – Long Ranger Gel Solution). The gel was cast between 2 glass plates, one of which was silicon coated, which were separated by plastic spacers, with both edges and the bottom sealed with gel sealing tape (Life Technologies). A comb was inserted into the top of the gel to enable loading of the samples. The gel was then run in 1X TBE in a vertical tank (Life Technologies, Gibco BR1, SA model) at 5 Volts per cm$^2$, with a Consort Flowgen E734 power supply. The gel was then
transferred to a 3MM Whatman filter paper and dried using a Savent Slab Gel Dryer SDG4050. Gels were then visualised by autoradiography.

2.10 Detection of DNA-Protein binding

2.10.1 Extraction of Cell Free Extract (CFE) from \textit{M. tuberculosis}

CFE is extracted from \textit{M. tuberculosis} to be used in gel retardation assays to look for proteins binding to specific regions of DNA. CFE were obtained as outlined in figure 2.1, in brief, cultures were grown the modified Dubos (Difco) containing 10% albumin, to an OD\textsubscript{600} of 0.4. these were then pooled and split. One was induced by the addition of 0.2\textmu g/ml mitomycin C, and both were incubated for a further 24 hours at 37°C. Samples were then harvest at 10,000rpm for 30 mins and 4°C. Pellets were resuspended in Z-buffer (appendix I) containing pefabloc protease inhibitor (Roche) and pelleted as described. The wash step was repeated three times. Samples were added to 2ml tubes containing 0.5-1mg glass beads and ribolysed at speed 6.5 for 30 sec, these were then placed on ice for 2 mins, then the ribolysing step was repeated. The samples were then applied to Ultrafree-MC 0.22\mu M spin columns (Millipore) in accordance with the manufacturer's protocol to remove cell debris. Quantitation was achieved using BCA protein assay kit (Pierce) in accordance with the manufacturer's guidelines.
**Figure 2.1: Cell free extract protocol in *M.tuberculosis***
2.10.2 Preparation of probes for gel retardation assay

100pmol of single-stranded oligonucleotides were mixed in a 1:1 ratio, with TEN buffer (10mM Tris-HCl, 1mM EDTA, 0.1M NaCl pH8.0), and incubated in a beaker of water at 95°C for 10 mins, then allowed to cool to room temperature in the water. 25µM annealed oligonucleotides were incubated with 1X DNA polymerase buffer (Promega), 5mM [dTTP, dATP, dGTP], 1.1Mbq [α-32P] dCTP, 2 units of Klenow (Promega) for 15 mins at 37°C. The reaction was stopped by the addition of 250mM EDTA, and the reaction was diluted 1:2 with dH2O. Unincorporated label was removed using G25 spin columns (Pharmacia) in accordance with the manufacturer’s guidelines.

2.10.3 Gel Retardation

Binding reactions were carried out at room temperature with 20-30µg CFE, 5X binding buffer (20mM HEPES, 120mM KCl, 0.8mM DTT, 20% glycerol, 0.4µg/µl BSA), 1µg l-lysine, 1.5µg poly d(I-C), 4µM [α-32P]dCTP labelled annealed oligos. The binding reactions were analysed using a Hoefer Se400 Vertical slab electrophoresis unit, with an 8% non-denaturing polyacrylamide gel, containing 1% glycerol.

2.11 Constructing a knockout in *M. tuberculosis*

2.11.1 Cloning into pBackbone

PCR primers were designed to amplify approximately 2kb regions containing the 5' and 3' flanking regions of the gene of interest. The primers were designed such that a region from the internal coding region of the gene was deleted. These regions of interest were then sequentially cloned into pBackbone (Gopaul, 2002). A 1.6kb hygromycin resistance cassette from pUC-Hy
(Mahenthiralingam et al., 1998) was then inserted between the 5' and 3' regions of the gene. Then a 6.3kb sacBlacZ cassette from pGolal7 (Parish and Stoker, 2000) was ligated into the PacI site in pBackbone, to enable screening and counter-selection in M. tuberculosis. sacB encodes levansucrase, which confers sucrose sensitivity and lacZ encodes β-galactosidase, which enables blue-white screening with 7H11 plates containing 1.6mg X-Gal (a lactose analogous substrate 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside which, when cleaved by β-galactosidase produces a blue product). The construct was then electroporated into competent M. tuberculosis to select and screen for a mutant strain (see sections 2.11.2 and 2.11.3).

2.11.2 Competent cells and electroporation of M. tuberculosis

Cultures were grown to an OD of 1.0 (as outlined in section 2.12.1) and then 0.1 volumes 2M glycine were added and cultures were incubated in the rolling incubator for 24 hours at 37°C. Cells were harvested by centrifugation at 10,000rpm for 30 mins, at room temperature. The pellet was washed 3 times in 1 volume 10% glycerol (filter sterile), then resuspended in 10ml 10% glycerol. 400μl competent cells were electroporated, with 2-5μg plasmid DNA at 25μF, 1000Ω (Bio-Rad gene pulser). Cells were then incubated overnight in static culture at 37°C, then harvested by centrifugation, resuspended in 100μl supernatant, then plated onto 7H11 plus relevant antibiotic. These were incubated at 37°C for 3-4wks. Knockout constructs were selected on 7H11 with 50μg/ml hygromycin and X-Gal.

2.11.3 Selection of a knockout

Initial electroporations were plated onto 7H11 + Hyg +X-Gal for 3/4 weeks. Blue colonies were potential single crossovers, and white colonies were potential double crossovers. The
Materials and Methods

White colonies were replica plated onto 7H11 + Kan and 7H11 + Hyg. A double crossover would be Hyg<sup>R</sup>, but Kms<sup>S</sup>, indicating the loss of the plasmid. Blue colonies were plated onto 7H11 + Hyg for 3/4 weeks, to allow the second cross-over to take place. Colonies were then serially diluted and plated onto 7H11 + Hyg + X-Gal + 2% sucrose, to select for loss of the plasmid. Only colonies that have lost the plasmid should be able to grow on sucrose, and remain white. White colonies were then replica plated on 7H11 + Km and 7H11 + Hyg + X-Gal. Kms<sup>S</sup>, Hyg<sup>R</sup> colonies which remained white were then grown on 7H11 + Hyg for DNA extraction and analysis by PCR and/or Southern blot.

2.11.3.1 Probe Labelling for Southern analysis

50 ng of probe DNA was labelled with [α-32P] dCTP using the Ready-to-go DNA labelling beads (Amersham-Pharmacia) following the manufacturer’s protocol. Unincorporated [α-32P] dCTP was then removed using Sephadex G-50 Nick Columns (Pharmacia) in accordance with the manufacturer’s protocol.

2.11.3.2 Southern Blot

A Southern blot, developed by (Southern, 1975), is a method of transferring restriction digested DNA onto a membrane, then visualizing it by hybridization with a radio-labelled probe. Restriction digested genomic DNA was subjected to electrophoresis in a 1% w/v TBE agarose gel, with EthBr, as outlined in section 2.6.3 at 20 volts overnight. The gel was visualized using an ultraviolet transilluminator at 302 nm and photographed with a ruler marker. The gel was then treated with 0.25M HCl for 15 min, then placed in denaturation solution (1.5M NaCl, 0.5M NaOH) for 30 min, then transferred to neutralisation solution (0.5M Tris-HCl pH 7.2, 1.5M NaCl)
Materials and Methods

NaCl, 1mM EDTA) for 30 min. The blot was then constructed with a wick (Whatman 3MM filter paper) placed in a glass dish containing 20X SSC (3M NaCl, 0.6M trisodium citrate pH7), the gel was placed on top of a filter paper the same size as the gel, on top of the wick. The membrane Hybond N+ (Amersham) was placed on top of the gel, then another three filter papers of the correct size were placed on top of the Hybond N+ membrane. All the air bubbles were removed, then Saran Wrap was placed around the sides of the membrane filter sandwich to prevent drying out and 10mm of paper towels were placed on the filter papers. These were then covered with a glass plate and weighted down with two 500ml bottles. The transfer was then left to take place overnight. The wells were marked, and the membrane was then washed in 2XSSC, and crosslinked using optimal crosslink programme, on a SpectrolinkerXL-1500UV crosslinker (Spectronics Corporation). The membrane was then wrapped in gauze and placed in a hybridization bottle with pre-hybridization solution B (5X SSC, 5X Denhardt’s solution, 2% (w/v) SDS) and 0.2mg denatured Salmon Sperm DNA and incubated at 65°C overnight. The α-32P labelled denatured probe was then added, and incubated at 65°C overnight. The membrane was then washed twice in wash buffer 1b (2X SSC, 1.0% SDS) at room temperature for 30min, then in wash buffer 2b(1X SSC, 0.1% SDS) at 65°C for 30 min, then in wash buffer 3b (0.1X SSC, 0.1% SDS) at 65°C for 30 min. The membrane was then autoradiographed.

2.11.4 *M. tuberculosis* liquid cultures and Ziehl-Nielson stain

Cultures of *M. tuberculosis* were grown in modified Dubos medium (Difco) containing glycerol and 10% Albumin, at 37°C in rolling bottles at 2rpm. A sample of the culture was applied to a glass slide, which was dried for 10-15 mins, before being fixed for a few seconds in a Bunsen burner flame. Slides were then placed in a formalin jar for at least 15mins for fumigation, then
a Ziehl-Nielson stain was performed (as outlined in Sambrook et al., 1987). Slides were then visualised using light microscopy to look for contaminants.

2.12 Phenotypic analysis of *M. tuberculosis* knockouts

2.12.1 Determination of *in vitro* growth

Static 5ml cultures of modified Dubos medium (Difco), were inoculated using a loop from a 7H11 plate. Theses were incubated for two weeks at 37°C. 1ml of the static culture was used to inoculate a 100ml rolling culture, to an OD of 0.005. These were then incubated at 37°C, in a rolling incubator at 2rpm. 1ml aliquots were removed every 24 hours and the optical density was measured. When the optical density reached 0.4, the aliquot was diluted, either 1/5, 1/10 or 1/15 in modified Dubos medium (Difco), before the optical density was taken.

2.12.2 Viability assays *in vitro*

Two 200ml cultures of H37Rv were grown in modified Dubos medium (Difco), in rolling bottles (1000ml bottles, irradiated to 30Kgy - Nalgene, Techmate) at 37°C, in a rolling incubator at 2rpm, to an OD of 0.3-0.4, then combined. 100ml aliquot was removed for the sodium nitrite stress, and centrifuged at 10,000g for 30 mins. The pellet was resuspended in an equal volume of modified Dubos medium pH 5.4. From this aliquots of 40mls were transferred into fresh rolling bottles, one was left untreated and a final concentration of 3mM NaNO₂ was added to the other. The remaining 300ml culture was split into 40ml aliquots in rolling bottles, one was left untreated, 0.2μg/ml ofloxacin (final concentration) was added to one aliquot, 0.02μg/ml mitomycin C (final concentration) to another aliquot, 10μM Cumene hydroperoxide.
(final concentration) to another, and 25mM paraquat (final concentration) to the final aliquot. These were incubated for 24 hrs at 37°C in a rolling incubator at 2rpm. After 24 hours, serial dilutions were set up using 50µl of each culture, added to 450µl of pre-autoclaved DMEM (Dulbecco's modified Eagles media) containing 50% FCS (heat inactivated foetal calf serum) and 2.5-3.5 mm glass beads. 10µl of the 10^2 and 10^3 dilutions were plated in duplicate onto 1/4 plates of 7H11. Plates were incubated at 37°C for 13-15 days, and colonies were counted.

For some of the stress agents, further experiments were performed using a range of paraquat and mitomycin C (see table 2.4).

**Table 2.4: Mitomycin C and paraquat concentrations**

<table>
<thead>
<tr>
<th>Mitomycin C stress</th>
<th>Paraquat stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>untreated</td>
<td>Untreated</td>
</tr>
<tr>
<td>0.02 µg/ml</td>
<td>10mM</td>
</tr>
<tr>
<td>0.05 µg/ml</td>
<td>20mM</td>
</tr>
<tr>
<td>0.1 µg/ml</td>
<td>30mM</td>
</tr>
<tr>
<td>0.2 µg/ml</td>
<td>40mM</td>
</tr>
<tr>
<td></td>
<td>50mM</td>
</tr>
</tbody>
</table>

**2.12.3 Viability assays in vivo**

Single colonies of H37Rv, ΔsigG and ΔsigG full operon complement (8T) were grown in duplicate in 15ml of supplemented 7H9 media (7H9 + glycerol + tween + ADC), standing culture for 14 days, then were sub-cultured into another 10ml of supplemented 7H9 media, and grown to an optical density (OD) of 0.02. These were then passed onto the animal handling
Materials and Methods

A trained animal technician injected 200μl of culture intravenously into 8 week old BALB/C female mice. The lungs and spleen were then harvested at day 1, 30, 60, 90 and 120 to determine the colony forming units (CFU), by serial dilution. A total of four mice were harvested per-strain, per time point.

Alternatively, single colonies of H37Rv, ΔsigG and cΔsigG, cΔsigGΔ1 and ΔRv2017 were grown in duplicate in 10ml of modified Dubos medium (Difco), and were incubated in static culture for 14 days at 37°C. 1ml of static culture was then used to inoculate 100mls of modified Dubos medium (Difco), which were incubated at 37°C in a rolling incubator at 2rpm, until an OD of 0.3 was reached. These were then pelleted in a centrifuge, and re-suspended in (phosphate buffered saline) PBS, the OD was re-measures, and were diluted to an OD of 0.02 in PBS. These were then passed onto the animal handling unit, where 200μl of culture was injected intravenously into 3 month old BALB/C female mice. The initial inoculum was plated to determine the dose, then lungs and spleen were then harvested at day 1, 30, 60, 90 and 120 to determine the CFU, by serial dilution. A total of four mice were harvested per-strain, per time point.

2.13 Expression analysis by microarray

2.13.1 Poly-L-lysine coating of microarray slides

Prior to printing of microarray slides, they are coated in poly-L-lysine. The slides (Sigma) were placed in a rack, and were cleaned by being stirred for 2 hours in a specialised container, with an alkaline solution (70g NaOH dissolved in 300ml filtered dH₂O, with 400ml absolute EtOH). Slides were vigorously rinsed 6 times in filtered dH₂O, for 1 minute per rinse. The slides were
then transferred to poly-L-lysine solution (70ml poly-L-lysine, 70ml filtered 10X PBS, 560ml filtered dH<sub>2</sub>O) and stirred for 1 hour. Slides were rinsed as outlined above, then dried overnight at 37°C. This protocol has been previously described at: http://cmgm.stanford.edu/pbrown/protocols/l_slides.html. For early experiments I performed the slide coating and processing, for later experiments, pre-coated Corning GAP slides were ordered by the BμG@S bacterial microarray group, before printing.

2.13.2 Post processing

After poly-L-lysine coating, the slides were printed at St. Georges Hospital by the BμG@S bacterial microarray group. The arrays were post-processed by being rehydrated over stream (boiling dH<sub>2</sub>O) for 5 sec, followed by snap drying on a hot plate at 100°C for a few seconds. Slides were then cross-linked to fix the DNA at 1mJ. The array slides were then blocked in succinate anhydride/sodium borate solution (5g succinate anhydride (Sigma) was dissolved in 315ml of N-methyl-pyrrolidinone (Sigma), then 35ml of 0.2M Sodium borate pH 8.0 was added) for 15 minutes. The slides were then washed vigorously in a 95°C water bath (filtered dH<sub>2</sub>O) for 2 mins, followed by 1 min in 95% EtOH. Slides were then air dried and stored in an airtight container.

2.13.3 RNA labelling, hybridization and washing

4-5μg RNA was used as template for 1st strand cDNA synthesis using Superscript II (Invitrogen), with fluorescently labelled cy3-dCTP and cy5-dCTP (Amersham Pharmacia Biotech). Random priming was carried with 6μg random primers (Invitrogen), and was incubated at 95°C for 5 mins, then snap cooled on ice. This was then added to a labelling
reaction, carried out with 1X first strand buffer (Invitrogen), 250mM DTT, 11.5mM [dATP, dGTP, dTTP], 4.6mM dCTP, 1.7μl cy3 or cy5 dCTP and 2.5μl Superscript II (Invitrogen). Samples were incubated for 10 mins at 25°C, then at 42°C for 90 mins. Prehybridized slides (in buffer 3.5X SSC, 0.1% SDS, 10mg/ml BSA, for 20 mins at 60°C) were washed in dH2O for 1 min, then propan-2-ol for 1 min, and were dried. cy3 and cy5 labelling reactions were pooled, and purified using MinElute spin columns (QIAgen), and eluted in 13.5μl DEPC dH2O. Samples were added to a final concentration of 4X SSC and 0.3% SDS, denatured at 95°C for 2mins, then cooled briefly (1min), before being applied to the microarray slide. A cover slip was place over the sample, and slides were placed in a waterproof hybridization chamber at 65°C over night. Slides were then washed in 1X SSC with 0.05% SDS for 2 mins, at 65°C, then twice in 0.06X SSC for 2 mins. Slides were then scanned using a GenePix Axon 4000A scanner (Axon Instruments), with dual wavelengths set to 600V. Image data was quantified using GenePix Pro 3.0 software, where any bad/unusable spots were removed. The data was then analysed using GeneSpring 4.1.2 (Silicon Genetics).

2.13.4 DNA versus RNA microarray

The protocol is similar to the RNA versus RNA arrays except the RNA samples are competitively hybridised with DNA to the microarray slides. 1μg DNA is labelled with cy3 after annealing of the random primer (as outlined in section 2.13.3), using 8 units Klenow (1X reaction buffer 50mM Tris-HCl (pH7.2 at 25°C), 10mM MgSO4, 0.1mM DTT) and incubated at 37°C for 90 mins in the dark, whereas the RNA is labelled as outline in section 2.13.3 except with only cy5, and the use of superscript III alters the incubation to 5 mins at 25°C, followed by 60 mins at 50°C. Samples are then pooled and purified as outlined in section 2.13.3.
2.13.5 Microarray data analysis

The microarray slides were scanned using a GenePix Axon 4000A scanner (Axon Instruments) and the collated image data was then processed using GenePix Pro 3.0 software, where the control spots, along with any absent or occluded spots were flagged as absent. The results obtained from GenePix were then transferred to the GeneSpring 4.1.2 (Silicon Genetics) analysis software. The data obtained from all the microarray slides was included in a single experiment in Gene Spring, and the input was annotated as follows:

The experimental normalisations used were:

• Data transformations: set measurements less than 0.01 to 0.01
• Per spot: Divide by control channel
  - Change the cut-off value from 10 to 0.01, to include data where genes are expressed at a low level.
• Per chip: Normalise to the 50th percentile
• Use only measured flags: Anything but absent.

The experimental interpretations used were:

• Log of ratio
• Present
  - This is useful if the data is only present in 5 out of the 6 replicates, it only uses the 5 present data values, thus decreasing the background noise.
• The cross gene error model was not active
The experimental parameters used were:

- Array (slide identification number)
- Strain (H37Rv, ΔsigG or ΔRv2017)
- Induction (mitomycin C + or -)
- Biological replicate (1, 2 or 3)
- Technical replicate (A or B)

Altering the interpretation using change display parameters enabled the data to be displayed differently. Initially the data was grouped by strain, in which the array was displayed as continuous, the strain and induction were displayed as non-continuous and the biological and technical replicates were not displayed, this was saved as the group by strain. Then the data was displayed with the strain as non-continuous, induction (mitomycin C +/-) as continuous, and the biological (1, 2 and 3) and technical replicates (A and B) were not displayed, this was saved as the log sample replicates. This enabled one to observe the difference between the strains, and the induction. Statistical analysis was in Gene Spring, which enables a range of parametric and non-parametric tests to be performed. A parametric test student T-test, assuming equal variance for the H37Rv and ΔsigG strains was performed, however for the analysis comparing the ΔRv2017 and H37Rv strains the Welsh approximation to a Student T-test was performed, which does not assume equal variance, due to the difference in spread of the data observed. The significance value used was either P<0.05 or P<0.01. The data were further scrutinised using the Benjamini and Hockberg False Discovery Rate correction (FDR), p<0.01.
3 Analysis of a sigma factor mutant in *M. tuberculosis*

3.1 Introduction

Regulation of bacterial gene expression takes place at a number of different levels. Sigma factors regulate gene expression at the transcriptional level: they are protein subunits that combine with core RNA polymerase to initiate transcription of their particular regulon, by forming specific interactions with specific promoters (Borukhov and Severinov, 2002). Sigma factors can be divided into 2 major groups, the primary sigma factors (house keeping) or the alternative sigma factors (Lonetto *et al.*, 1994). *M. tuberculosis* has 13 $\sigma^{70}$ class sigma factors (Cole *et al.*, 1998), of which 10 are ECF family (Manganelli *et al.*, 2001). ECF sigma factors are thought to be important in regulation of adaptive responses to environmental stresses by co-ordinating transcription of a stress response regulon (Missiakas and Raina, 1998; Song *et al.*, 2003). An example is *sigH* in *M. tuberculosis*, which is positively autoregulated in response to oxidative and heat stress (Song *et al.*, 2003) and induces the expression of a set of genes, including known heat shock proteins. Of the 13 sigma factors in *M. tuberculosis*, *sigG* was the most highly upregulated sigma factor in response to DNA-damage (Rand *et al.*, 2003), suggesting that it may have a role in regulation of the response to DNA-damage. Therefore, in this study, a strain of *M. tuberculosis* was constructed, in which *sigG* was inactivated. The phenotype of the knockout was then analysed with respect to *in-vitro* growth and susceptibility to DNA damaging agents, along with *in-vivo* analysis of virulence assessed using a mouse model of infection.
3.2 Construction of a \textit{sigG} knockout

A gene inactivation knockout of \textit{sigG} was constructed in the laboratory strain of \textit{M. tuberculosis}, H37Rv. In brief, the knockout was constructed using a non-integrating plasmid containing a deletion in the \textit{sigG} coding region, replaced by an antibiotic resistance cassette: this was electroporated into H37Rv, where homologous recombination resulted in the replacement of the functional chromosomal copy of \textit{sigG} with the inactivated copy present on the plasmid, as outlined in Parish \textit{et al.}, (1999).

A 4368bp fragment containing the coding region of \textit{sigG} was ligated into the \textit{Km} vector pBackbone, which was derived from pBluescript by K. Gopaul (Gopaul, 2003). Inverse PCR was then used to remove a 691bp region from within the 1110bp predicted coding region of \textit{sigG}, whilst also incorporating AvrII sites. A previous member of the laboratory produced this construct (Gopaul, 2003). The AvrII site was utilised to enabled cloning of a hygromycin resistance cassette into the deleted region of \textit{sigG}. The construct consisted of a 5' 1706bp fragment containing part of the 5' region of \textit{sigG} separated from the 1970bp fragment containing part of the 3' region of \textit{sigG}, by a hygromycin resistance cassette. A \textit{sacB/lacZ} cassette was cloned into the \textit{PacI} site within the vector part of this construct, to enable blue-white screening from the \textit{lacZ} gene (which encodes \(\beta\)-galactosidase) and sucrose counter-selection from the sucrose sensitivity gene \textit{sacB} (encodes levansucrase (Jager \textit{et al.}, 1992) in \textit{M. tuberculosis}). This construct (pLD1) was then electroporated into \textit{M. tuberculosis}, H37Rv Mill Hill strain.
The selection and counter selection process was carried out as outlined in section 2.11, to detect potential knockouts, exhibiting Km\textsuperscript{s} and Hyg\textsuperscript{r} phenotypes. DNA extractions were then carried out on these potential knockouts, and screened by PCR and Southern blotting.

### 3.2.1 PCR used to detect potential knockouts

PCR reactions were designed to amplify separately the 5' flanking region of the mutation and the 3' flanking region. One primer for each pair was located outside the cloned region of DNA, and the other was located inside the hygromycin resistance cassette. The design of the PCR reaction enabled screening of potential double crossovers, whilst enabling one to distinguish between single crossovers, double cross-overs and random integrations (figure 3.1). Genomic DNA was extracted from potential knockouts and used in both PCR reactions. The PCR reactions gave inconclusive results (data not included). The drawback to this approach of detecting potential knockouts is the lack of a positive control, from which to optimise the PCR reactions, as the PCR reactions potentially produce large products, which may be difficult to amplify.

### 3.2.2 Southern blot used to detect potential knockouts

A Southern blot was designed such that double cross-over events could be distinguished from wild-type (control), single 3' or 5' cross-over events and random integrations. The design required the identification of a restriction enzyme that cuts twice, once in the genomic region, upstream of the knockout cassette and again in the 3' region of the $\text{sigG}$ knockout construct (see figure 3.2a). A DNA probe (probe a) was designed to overlap the 5' coding region of $\text{sigG}$ and the 5' upstream flanking region contained within the construct (see figure 3.2a). This design
Figure 3.1: A schematic representation of the design of the PCR reaction to screen for potential knockout colonies. The black arrows indicate the PCR primers for the 5' region of the construct and the red arrows indicate the PCR primers for the 3' region of the construct. The dotted arrows indicate the regions that lack homology to the primers and therefore indicate where binding will not take place. The primer locations for the wild-type, knockout, single crossover and random integrations are shown.
Figure 3.2a: A schematic representation of the size of product produced in a Southern blot for both a ΔsigG strain and the H37Rv wild-type. The probe is positioned in the 5' flanking region of sigG, and the restriction sites (XmnI) are marked with arrows. The size of the expected wild-type and mutant bands are indicated in base pairs (bp).

Figure 3.2b: A southern blot to detect a knockout of sigG. DNA was extracted from potential sigG knockout colonies and were digested with XmnI. A radio-labelled probe was then used to detect potential double cross-overs. ΔsigGW1 (track 1) ΔsigGW2 (track 2), ΔsigGW3 (track 3) and ΔsigGW4 (track 4). M. tuberculosis genomic DNA (track C) was used as a positive control.
enables the differentiation of 3' and 5' cross-over events, double cross-overs, random integration and wild-type. Mutant and wild-type are differentiated by size of a single band whereas single crossovers and random integrations are differentiated by the production of 2 bands. A random integration would produce 2 bands of different size from the 3' and 5' single crossovers, due to the positioning of the restriction sites.

*XmnI* restriction digests were performed on DNA isolated from potential knockout colonies, alongside an *M. tuberculosis* H37Rv genomic DNA control. These digests were then run on an agarose gel and a Southern blot was performed. Probe a, when hybridized to the restriction fragment is expected to visualise a 5678bp fragment with a knockout, and a 4739bp fragment with the wild type (see figure 3.2a and 3.2b). DNA isolated from colonies 1, 2 and 4 showed a band of 5678bp, therefore indicating a knockout (see figure 3.2b), whereas, colony 3 exhibited the same size band as the genomic control, of 4739bp, indicating a wild-type genotype (see figure 3.2b).

A second Southern blot was designed to confirm the identification of *sigG* knockout isolates, by determining if the 691bp deletion of the internal coding region of *sigG* had taken place. *Scal* sites were identified that cut once in the genomic region upstream of 5' *sigG*, and once in the genomic region downstream of 3' *sigG* (see figure 3.3a). Probe b was designed within the 691bp deletion from the *sigG* coding region and therefore only binds to wild-type DNA in a Southern blot (see figure 3.3b). The DNA isolated from colonies 1 and 2 showed no binding of the probe, confirming they were knockouts, whereas colony 3 showed binding to the probe, producing a 3235bp fragment the same as the wild-type genomic control, indicating that colony 3 was wild-type (see figure 3.3b). The knockout *sigGl (ΔsigG)* was used in all downstream experiments.
Figure 3.3a: A schematic representation of the size of product produced in a Southern blot using a probe designed in the deleted region of $\text{sigG}$. The probe will only bind to the wild-type, and not the $\text{sigG}$ mutant. The restriction sites (Scal) are indicated along with the size of the wild-type band.

Figure 3.3b: A southern blot to detect a knockout of $\text{sigG}$, using a probe designed in the deleted region of $\text{sigG}$. DNA was extracted from potential $\text{sigG}$ knockout colonies DNA was digested with Scal and run on an agarose gel. This was transferred to a nylon membrane, probed with radio-labelled probe b. $\Delta\text{sigG}W1$ (track 1) $\Delta\text{sigG}W2$ (track 2), $\Delta\text{sigG}W3$ (track 3) and M. tuberculosis genomic DNA (track C), was used as a positive control.
3.2.3 Chromosomal location of \textit{sigG}

Current literature suggests some sigma factors are co-transcribed as polycistronic RNA with their cognate anti-sigma factors and even anti-anti sigma factors (Raivio and Silhavy, 2001). If there are genes co-transcribed with \textit{sigG}, the deletion and insertion of a hygromycin resistance cassette may have polar effects on any co-transcribed downstream genes. Therefore, it was important to determine if \textit{sigG} is co-transcribed with other genes. Visual analysis of the locus revealed that 4 genes were transcribed in the same orientation, with \textit{sigG} being positioned furthest upstream (see figure 3.4). The gene directly upstream of \textit{sigG} was divergently transcribed, and therefore, would not form part of the same cistron (see figure 3.4). The size of the intergenic regions, between genes in the same genomic orientation, can be used to predict if genes are co-transcribed as a polycistronic mRNA (Strong \textit{et al.}, 2003). Therefore, genes that are separated by a few base pairs tend to be co-transcribed (Price \textit{et al.}, 2005). A distance model has been produced for \textit{E. coli} and \textit{B. subtilis}, however, intergenic difference may vary across species for conserved operons (Price \textit{et al.}, 2005). The size of the intergenic regions between \textit{sigG} and the 3 downstream genes, Rv0181c, Rv0180c and IprO are indicated in figure 3.5a. The intergenic region between Rv0181c and \textit{sigG} is 17bp and the predicted coding regions of Rv0180c and Rv0181c overlap by 27bp, suggesting they may be co-transcribed. However, IprO is situated 80bp downstream of Rv0180c, making it more difficult to predict whether they are co-transcribed. One must also take into account that the open reading frames (ORF's) are predictions, and the actual coding regions may vary slightly from these predictions.

To determine whether these genes formed a polycistronic mRNA, primers were designed in neighbouring genes to amplify the intergenic regions from RNA (see figure 3.5a). Internal control primers (see figure 3.5a, primer set D) were designed in the internal coding region of
Figure 3.4: A schematic representation of the location of sigG. The figure was adapted from the data available on the TubercuList website.
Figure 3.5a: A schematic representation of the relative position of the sigG. The direction of transcription of these genes is indicated with the red arrow. The intergenic distances are indicated, along with the primers used in the RT-PCR reaction, A (LD sigGAf/r), B (LD sigGBf/r), C (LD sigG Cf/r) and D (LD sigGf/r). D was used as a positive control for the RT-PCR.

Figure 3.5b: An agarose gel showing co-transcription studies of sigG. PCR reactions were performed using cDNA (track 1), positive DNA control (track 2) and a negative control, RT-reaction (tracks -). The primers used in each reaction are marked under the relevant tracks. The sizes were as expected A = 263bp, B = 653bp, C = 649bp and D = 571bp.
sigG to check the integrity of the cDNA template. The PCR reactions were performed on cDNA, with genomic DNA as a positive control and RT- (omitting superscript) reactions as a negative control. RT- reactions, were performed on RNA, omitting the Superscript (see figure 3.5b). Oligo set A was designed to determine if sigG (Rv0182c) and Rv0181c were co-transcribed, oligo set B to determine if Rv0180c and Rv0181c were co-transcribed, and oligo set C to determine if Rv0180c and lprO were co-transcribed. Figure 3.5b shows that sigG, Rv0181c and Rv0180c are most likely co-transcribed as part of an operon, but lprO forms a separate cistron.

3.3 Predicted protein domains of SigG, Rv0181c and Rv0180c

The SMART database, (available on-line at www.sambit.embl-heidelberg.de/smart) was used to predict protein domains of SigG, Rv0181c and Rv0180c that might allude to their functions.

SigG contains a PFAM sigma 70 region 2 domain, located at 61-129aa, which corresponds to the most highly conserved of the sigma 70 regions, as region 2 forms the site of interaction with the core-RNAP as well as the −10 recognition helix (Lonetto et al., 1994). SigG from M. tuberculosis processes an ECF subfamily signature (TubercuList) and shows high levels of homology to SigG (MB0188C) from M. bovis (8e-96), and M. paratuberculosis (MAP3621c) e-value (8e-73) and to 2SCK8.21C, a putative ECF family sigma factor in Streptomyces coelicolor (3e-65). Alignment of the sequences using the programme Needle of the EMBOSS-Align suite, showed that SigG from M. tuberculosis and M. bovis shared 100% identity, whereas SigG from M. paratuberculosis showed 73.9% identity (Blosum62 matrix, with 10.0 open gap penalty, 0.5 gap extension penalty).
Chapter 3  Analysis of a sigma factor mutant in *M. tuberculosis*

Rv0181c contains the PFAM Pirin domain from position 6-122aa, and shows sequence homology to MB0187C in *M. bovis* (hypothetical protein), (6e-76), which is located in the corresponding location downstream of *sigG*. Rv0181c also shows homology to two putative cytoplasmic proteins, one from *Salmonella typhimurium*, STM3544 (YHHW) (5e-72) and the other from *E. coli*, B3439 (YhhW) (6e-72).

Rv0180c is predicted to contain an N-terminal signal peptide (1-43aa) and five C-terminal transmembrane domains, interspersed with undefined regions, spanning the region 218-414aa. A homology search revealed that Rv0180c shows homology to MB0186C (e-152), a predicted membrane protein from *M. bovis* and ML2600 of *M. leprae*, also a probable conserved membrane protein (e-145). Rv0180c also shows some homology with phage infection protein, BC3083 from *Bacillus cereus* (2e-38), and also to the ABC transporter, FTT0729 of *Francisella tularensis* (subsp. tularensis) (9e-25), a pathogenic aerobic gram negative bacterium, which causes Tularemia.

The domain predictions of Rv0181c and Rv0180c along with the fact that they are co-transcribed with *sigG* in *M. tuberculosis* suggests the possibility that the two downstream genes are involved in regulation of the sigma factor. A number of sigma factors have been found to be co-transcribed with their cognate anti-sigma factors (Raivio and Silhavy, 2001), which negatively regulate them, preventing binding of the sigma factor to the core-RNAP and therefore preventing transcription of the sigma factor's regulon. Anti-sigma factors are in turn regulated by an environmental signal, or by binding of their cognate anti-anti-sigma factors, both of which result in a conformational change of the anti-sigma factor (Humphreys *et al.*, 1999), thus inhibiting binding of the sigma factor and the anti-sigma factor, which enables the sigma factor to bind to core-RNAP and initiate transcription. The predictions regarding the
structure of Rv0180c indicate the presence of a transmembrane domain suggesting that Rv0180c is a transmembrane protein, which may act to detect environmental signals, which could, in turn, result in a conformational change in the structure of the anti-sigma factor (Rv0181c), thus releasing the sigma factor (SigG) to enable transcription of the sigma factor’s regulon.

3.4 Complementation of the sigG knockout

As previously stated inactivation of sigG via deletion and insertion may have polar effects on downstream genes in the operon. Any effects observed in the ΔsigG strain may be attributed to a culmination of the global effect of sigG knockout and the polar effects this may have on the co-transcribed genes.

To determine whether the genotypic and phenotypic effects exhibited by the knockout are solely due to the effects of constructing an inactivation mutant of sigG rather than any polar effects, a full operon complement was produced along side a complement that contained only the coding region of sigG (see table 3.1).

<table>
<thead>
<tr>
<th>(sigG)</th>
<th>Rv0182c</th>
<th>Rv0181c</th>
<th>Rv0180c</th>
<th>Deletion size</th>
<th>Construct size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full operon (8T)</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>0</td>
<td>6870</td>
</tr>
<tr>
<td>Partial δ1</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>2125</td>
<td>4754</td>
</tr>
</tbody>
</table>

Table 3.1: Comparison between the full operon and sigG only complements.

Full operon complementing plasmid was constructed in pKP186. Site directed mutagenesis was performed on the full operon complementing plasmid to produce a deletion containing only sigG.
Chapter 3 – Analysis of a sigma factor mutant in *M. tuberculosis*

The constructs for complementation were produced in a modified version of the integrating plasmid pMV306 (Stover *et al.*, 1991), called pKP186, in which the integrase gene (*int*) had been deleted in (Papavinasasundaram, personal communication). pMV306 contains genes for both attachment site (*att*) and integrase (*int*), derived from mycobacteriophage L5. These enable integration of the plasmid into the *attB* site of the mycobacterial chromosome. The absence of the excisionase (*xis*) gene means the plasmid should be maintained without the need for selection (Stover *et al.*, 1991). However, problems with maintenance of this plasmid have been observed by others (Springer *et al.*, 2001), which lead to the construction of vectors, in which, the integrase gene has been removed and incorporated into a separate plasmid pBSint (Springer *et al.*, 2001), resulting in increased stability of the vector. pKP186 is one such plasmid in which the integrase has been deleted to increase stability and maintenance.

The full operon complement was constructed, using PCR primers designed to amplify the entire *sigG* operon including enough upstream region to contain the putative *sigG* promoter and downstream of the *sigG* operon, including part of *lprO* gene. This was then cloned into pKP186. The *sigG* only complement was constructed by deletion: primers were designed as outlined in section (2.7.13) to enable site directed mutagenesis to remove a region including the coding regions of Rv0181c and Rv0180c, using the full operon complement as a template (see table 3.1). These constructs were then individually co-electroporated into the Δ*sigG* strain of *M. tuberculosis*, along with pBSint, which contains the integrase gene, thus enabling the complementing plasmid to integrate at *attB* site in the bacterial chromosome. The electroporations were plated onto 7H11 plus kanamycin to select for the presence of the pKP186 construct (see table 3.1 and 2.11.3). The selection was not maintained after integration, thus resulting in loss of non-integrated plasmids, after which, colonies were grown on 7H11 + Km to select for the complementing plasmid.
### 3.4.1 Checking expression of the SigG, Rv0181c and Rv0180c

To determine if the effects of the deletion and insertion had partial, full or no polar effects on downstream gene expression, data from the DNA vs. RNA microarrays (described in chapter 5) were analysed. These suggest that the levels of sigG, Rv0181c and Rv0180c are decreased under both uninduced and induced (0.02μg/ml mitomycin C) conditions, in the ΔsigG strain compared to H37Rv wild-type, but not eliminated (see figure 3.6). One would have expected the level of sigG expression to have been eliminated in the ΔsigG strain, however, sigG still appears to be expressed under both uninduced and induced conditions, albeit to a much lower level (significantly different p<0.01 Students t-test), and sigG does not appears to be induced. Suggesting that expression of sigG is not totally SigG dependent. Examination of the location of the microarray probe indicates that the probe overlaps with the retained coding region of sigG by 48bp, sufficient to facilitate binding of the probe.

The level of expression of Rv0181c is decreased significantly under uninduced and induced conditions (p<0.01 - two-tailed Student t-test), and Rv0180c is significantly decreased under induced conditions (p<0.01), (see figure 3.6). It appears that only Rv0181c is still induced, but to a much lower level in ΔsigG strain compared to H37Rv.

### 3.4.2 Design and production of antibodies specific to sigG, Rv0181c and Rv0180c

An alternative method of detecting expression of SigG and the downstream co-transcribed genes Rv0181c and Rv0180c was identified, whereby peptides were used to raise antibodies against SigG, Rv0181c and Rv0180c.
Figure 3.6

A comparison of the gene expression of the genes in the $\operatorname{sig}G$ operon in H37Rv wild-type and $\Delta \operatorname{sig}G$ strain of $M. \text{tuberculosis}$

Figure 3.6: Microarray expression data for $\operatorname{sig}G$ operon in $\Delta \operatorname{sig}G$ strain compared to H37Rv wild-type. DNA versus RNA microarrays were performed with triplicate biological cultures, repeated in duplicate. The uninduced and induced (0.02μg/ml mitomycin C) conditions were compared for both the $\Delta \operatorname{sig}G$ strain and H37Rv. The uninduced conditions were significantly decreased in $\Delta \operatorname{sig}G$ strain compared to H37Rv, at p<0.01 (Student t-test) for $\operatorname{sig}G$ and Rv0181c, and p<0.05 for Rv0180c. The induced values were significantly decreased for $\operatorname{sig}G$, Rv0181c and Rv0180c at p<0.01 (Students t-test).
Protein sequences for SigG, Rv0181c and Rv0180c were sent to Sigma Genosys for the design of potentially immunogenic peptides. There was only one possible peptide detected for Rv0180c at position 88-102, within the region between the signal peptide and the transmembrane domain. Three alternative peptides were identified for both SigG and Rv0181c, at amino acid positions 54-68, 139-153 and 240-254 for SigG, whereas the peptides for Rv0181c were identified at 1-15, 52-66 and 206-220. These were then assessed for their relative positions in the genes to determine which would potentially form the most reactive antibody. For Rv0181c peptide 1 (1-15aa) was thought to be inappropriate, as it has been shown that anti-sigma factors interact with their cognate sigma factors via their N-terminal region (Yoshimura et al., 2004), peptide 2 (52-66aa) was also thought to be inappropriate, as it overlapped with the PFAM domain, which may affect the specificity of the antibody, as PFAM domains have a reasonable degree of conservation. Therefore peptide 3 (206-220) was chosen as it was located at the carboxy terminus of the protein and did not overlap with any specific domains. For SigG peptide 1 (54-68) was not chosen for antibody synthesis as it overlaps with the sigma 70 PFAM domain, which is conserved in the sigma 70 family of sigma factors. Peptide 3 (240-254) was chosen arbitrarily over peptide 2 (139-153) for antibody production. A carboxy terminal cysteine residue was added to all of the peptides to enable conjugation to the carrier protein KLH (keyhole limpet hemocyanin, derived from marine mollusk). Freunds complete adjuvant was omitted from the antibody production protocol as it contains mycobacterial components, which may have compromised antigen specificity, as a result, Freunds incomplete adjuvant was utilised. The synthesised peptides were used to raise antibodies in rabbits. Two rabbits were used per peptide, and the final bleed was used in affinity purification of the antibodies.
3.4.3 **Analysis of SigG, Rv0181c and Rv0180c, using antibodies.**

Initially, the reactivity and an indication of the specificity of the antibodies raised was assessed using dot blots that included the synthesised peptide specific for each antibody, alongside *M. tuberculosis* cell free extract (CFE) and *E. coli* CFE. The terminal bleeds and purified fractions were compared with the pre-bleeds (null serum) to select fractions and specific dilutions to use in subsequent analysis (data not shown). Western blots were performed to simultaneously examine the expression in the ΔsigG strain compared with that of the H37Rv wild-type, to determine if the reactivity of the antibodies raised against SigG, Rv0181c and Rv0180c were specific, using CFE from both uninduced and induced conditions for both ΔsigG strain and H37Rv. The western blots were performed with 30μg of H37Rv, ΔsigG strain and *E. coli* CFE. For Rv0181c and Rv0180c the purified fractions (F1) were used at a 1/500 dilution for the primary antibody, while for SigG the purified fraction was used at 1/250 dilution (assessed using dot blot, data not shown). The secondary antibody was used at a 1/5000 dilution (goat-anti-rabbit).

The antibody raised against the SigG peptide was not able to detect SigG in any of the cell free extracts, even from H37Rv CFE. The anti-SigG antibody did not cross react with *E. coli* CFE.

In the case of Rv0181c, a single band was obtained, for both the uninduced and induced CFE from H37Rv and ΔsigG strains. Although the size of Rv0181c is 26.3kDa, the band appears to run at approximately 48kDa (see figure 3.7, arrow 1), which suggests that Rv0181c maybe running as a dimer, however, even though the anti-Rv0181c antibody did not cross react with *E. coli* CFE, the binding may be non-specific, as a denaturing loading buffer was used.
Figure 3.7: Western blot to test specificity of SigG, Rv0181c and Rv0180c antibodies. The Western blots were performed with 30µg H37Rv uninduced CFE (1), 30µg H37Rv induced CFE (2), 30µg ΔsigG uninduced CFE (3), 30µg ΔsigG induced CFE (4), 30µg E. coli uninduced CFE (5). The samples were run on four 10% NuPage Bis-Tris gel (Invitrogen), alongside a low molecular weight marker (M) (BioRad). The transfers were performed with PVDF membrane (Invitrogen). The primary antibodies were used at a dilution of 1/250 SigG, 1/500 Rv0181c and 1/500 Rv0180c. The secondary antibodies were used at a dilution of 1/5000 HRP goat anti-rabbit (SigG, Rv0181c, Rv0180c), Western blots were visualised using chemiluminescence (Pierce). The SigG Western blot was stripped and re-probed with Rv0180c antibody using the protocol as outlined above.
For Rv0180c, three bands were present, for both the uninduced and induced CFE from H37Rv and ΔsigG strains. The size of Rv0180c is 47.6kDa, however, bands appear to run at 47kDa (see figure 3.7 arrow 2), approximately 92kDa (see figure 3.7, arrow 3), and the third band is of unknown size, as it was too large to size on this gel system, especially as the size markers are not in the correct range (see figure 3.7, arrow 4). The band indicated with arrow 2 may be a monomer of Rv0180c, arrow 3 may be a dimer, and arrow 4 may indicate a trimer or tetramer. The anti-Rv0180c antibody did not cross react with E. coli, however, without a deletion mutant of Rv0180c as a negative control, one can not determine whether the binding is specific.

To determine whether the binding were specific and therefore to assess whether the sigG deletion caused polar effects on the downstream genes, a denaturing SDS-page gel would have to be performed, ideally using either a deletion mutant of Rv0181c as a negative control, or M. smegmatis instead of E. coli as M. smegmatis does not possess orthologues of the sigG operon.

To control for the integrity of the cell extracts used and the efficiency of protein transfer in the blot probed with the anti-SigG antibody, this blot was stripped and re-probed with anti-Rv0180c antibody (see figure 3.7). The Rv0180c antibody bound to the H37Rv and ΔsigG strain CFEs, yielding the same three-band pattern as observed previously. This indicates that the CFEs used for this blot were good quality, and that protein transfer from the gel to the membrane had occurred; thus the lack of binding seen with the anti-SigG antibody must be due to the nature of the antibody rather than a problem with the CFE on the blot.
3.5  *In-vitro* analysis of Δ*sigG* strain and complements

3.5.1  Growth curves *in-vitro*

Growth curves for the Δ*sigG* strain, wild-type H37Rv and the two *sigG* complements were produced under normal growth conditions, in a rolling incubator, as outlined in section (Methods – growth curve). The Δ*sigG* strain and two complement strains (8T and Δ1), grew at the same rate as the wild-type H37Rv (see figure 3.8), suggesting that the knockout and complements have the same ability to grow and divide as the wild-type under normal *in-vitro* growth conditions.

3.5.1.1  Viability of Δ*sigG*, complements and wild-type to DNA damaging agents

Alternative sigma factors have been implicated to respond to a number of different environmental signals, as outlined in section 1.5.9.1. Knockouts of some alternative sigma factors have been shown to be more susceptible to certain types of environmental stimuli than wild-type.

To determine if SigG could be involved in the response to DNA-damage in *M. tuberculosis*, the survival of the Δ*sigG* strain following exposure to a number of different DNA damaging agents was compared with that of the wild-type H37Rv. Ofloxacin and mitomycin C are chemical DNA damaging agents; ofloxacin is a quinolone antibiotic, which inhibits DNA gyrase and is toxic to mycobacteria (Movahedzadeh *et al.*, 1997), whereas mitomycin C alkylates guanines to form inter-strand cross-links in DNA (Iyer and Szybalski, 1964). Paraquat and cumene
Figure 3.8: The *in-vitro* growth curves of the sigG knockout, sigG full operon complement, partial complement and H37Rv wild type. For each strain, single colonies were inoculated into 5ml of Dubos plus albumin in triplicate and incubated in static culture for 10 days at 37°C. Static cultures were then used to inoculate a 100ml rolling culture in Dubos plus albumin, to an OD of 0.005 (approx 1ml). Triplicate cultures were returned to the rolling incubator for a total of 14 days. Optical density readings were taken at the same time daily, and dilutions were made in Dubos + albumin after an OD of 0.4 to enable accurate reading by the spectrophotometer.
hydroperoxide are oxidative agents, which resemble environmental stresses produced in the macrophage (see section 1.4.1). Paraquat generates superoxide, by catalytically diverting electrons from NAD(P)H to oxygen, whereas cumene hydroperoxide generates peroxy stress by oxidising bases and deoxyribose (Demple and Harrison, 1994). Acidified sodium nitrite mimics the exposure of the bacteria to nitric oxide in an activated macrophage (Nathan and Shiloh, 2000), viability assays were performed as outlined in section 2.12.2. Cultures were grown to exponential phase (0.3 to 0.4 optical density ($OD_{600}$)) and were split into aliquots of 40mls, one was left untreated while the others were stressed with the damaging agent. In the case of nitrosative stress the cultures were harvested at the required OD (0.3-0.4) and resuspended in acidified media (Dubos + albumin), pH 5.4. Experiments were performed on two distinct biological replicates and duplicate colony counts were recorded for each. The viability is presented as percentage viable counts, with treated viable counts expressed as a proportion of mean untreated viable counts (see table 3.2 and figure 3.9). The data for $\Delta \text{sig}G$ strain was compared with that for wild-type H37Rv. Paraquat and mitomycin C showed the most dramatic decrease in viability in the knockout compared to the wild-type, with ofloxacin also causing a decrease in viability of the knockout compared to the wild-type (see figure 3.9). However, there was a significant decrease in viability in the $\Delta \text{sig}G$ strain for paraquat ($p<0.01$) and mitomycin C ($p<0.06$) using a Student t-test, but there was no significant difference in viability in $\Delta \text{sig}G$ strain compared to H37Rv, when exposed to ofloxacin ($p=0.20$). There was also no significant difference at the 1% level ($p<0.01$) (Student t-test) between wild-type and knockout in response to cumene ($p=0.26$) and sodium nitrite stress ($p=0.74$).

To analyse the enhanced susceptibility of the $\Delta \text{sig}G$ strain to paraquat and mitomycin C in more detail, further experiments were performed using a titration to determine the effects of different concentrations of the DNA damaging agents on $\Delta \text{sig}G$ strain, wild-type H37Rv and full operon
Figure 3.9: Histogram depicting the viability of sigG knockout treated with different DNA damaging agents as a percentage of the untreated control.

Cultures of ΔsigG and H37Rv were grown to an OD 0.3-0.4. 40ml aliquots were removed. One was untreated, and the others were treated with 0.2μg/ml ofloxacin, 0.02μg/ml mitomycin C, 10mM cumene hydroperoxide, 25mM paraquat, or 3mM NaN02 (for NaN02 stress, both untreated and test samples were resuspended in Dubos media pH5.4). Cultures were incubated @ 37°C for 24 hours. 50μl of treated culture was serial diluted in DMEM/FCS (50%). Serial dilutions were plated on 7H11 plates and incubated at 37°C for 13-15 days to obtain colony counts. Colony counts are expressed as a percentage of untreated, with mean and standard error for the replicates.
complement (mitomycin C only). Both strains showed a decrease in viability with increasing concentration of paraquat (see table 3.3 and figure 3.10), however, no significant difference was observed between the ΔsigG strain and wild-type H37Rv (see table 3.4). The titration with mitomycin C confirmed that the ΔsigG strain is considerably more susceptible than wild-type H37Rv (see table 3.5 and figure 3.11). The ΔsigG strain exhibits an average viability of only 14% at 0.02μg/ml mitomycin C, whereas the wild-type H37Rv exhibits a 41% viability, this susceptibility increases with increasing concentrations of mitomycin C. Statistical analysis was used to determine whether there was a significant difference between the mean viable count of ΔsigG and H37Rv at different concentrations of mitomycin C. An F-test was performed on the percentage viability of ΔsigG compared to H37Rv, which showed the variances of the two sets of data were not equal, therefore the data was transformed using log10. The F-test was repeated using the log data, and showed the variances were equal (see table 3.6a, 3.6b and 3.6c), which indicated the transformed data approximates to a normal distribution, and therefore a parametric test such as the t-test was a valid test to perform on the data (see table 3.6a, 3.6b and 3.6c). T-tests were performed with the following hypotheses:

H₀: P value > 0.05. Accept H₀: There is no significant difference between the mean viable counts of ΔsigG and H37Rv in response to mitomycin C stress at a given concentration.

H₁: P value < 0.05. Accept H₁: There is a significant difference between the mean viable counts of ΔsigG and H37Rv in response to mitomycin C stress at a given concentration.

Two-tailed unpaired t-tests were performed between ΔsigG strain and H37Rv untreated transformed data. There was no significant difference between the means at the 95%
Table 3.2

<table>
<thead>
<tr>
<th>Paraquat (mM)</th>
<th>H37Rv</th>
<th>sigG knockout</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard error</td>
</tr>
<tr>
<td>Unreated</td>
<td>100</td>
<td>± 14.754</td>
</tr>
<tr>
<td>10</td>
<td>72.600</td>
<td>± 11.241</td>
</tr>
<tr>
<td>20</td>
<td>52.693</td>
<td>± 8.197</td>
</tr>
<tr>
<td>30</td>
<td>29.040</td>
<td>± 5.152</td>
</tr>
<tr>
<td>40</td>
<td>30.913</td>
<td>± 2.810</td>
</tr>
<tr>
<td>50</td>
<td>13.665</td>
<td>± 2.758</td>
</tr>
</tbody>
</table>

Figure 3.10

Table 3.2: Percentage viability of H37Rv in comparison to sigG knockout in response to paraquat stress. Duplicate cultures were grown to exponential phase (0.3 to 0.4 OD), 40ml aliquots were incubated for 24 hours untreated or with the relevant concentration of Paraquat. After 24 hours, the cultures were serial diluted and plated on 7H11 plates in duplicate and incubated at 37°C for 16-18 days. Colony counts were taken of at least 2 dilutions. The viable CFU was then expressed as a percentage of the untreated for each sample, and the mean and standard error was calculated.

Figure 3.10: A graph comparing the viability of the sigG knockout to H37Rv wild-type in response to paraquat stress. In Prim 4, the percentage viability data as outlined above was transformed using log₁₀ and plotted on a linear scale. A linear regression sigmoidal dose response curve was fitted to the data.
Table 3.3: Statistical analysis of the response to Paraquat stress of the sigG knockout compared to H37Rv wild-type. The percentage viability for \( \Delta\text{sigG} \) and H37Rv was transformed using \( \log_{10} \). An F-test was performed to determine if both samples had equal variance (H_0) (a requirement for parametric tests), when the p-value >0.05, H_0 is accepted, the populations have equal variance and a t-test is appropriate. Then an unpaired, two-tailed t-test was performed between each sample for both \( \Delta\text{sigG} \) knockout and H37Rv wild-type, under each stress condition individually, to determine if the means for each sample were the same (H_0) or the means of \( \Delta\text{sigG} \) and H37Rv were different (H_1). Where the p-values >0.05, H_1 is rejected and H_0 is accepted. There is no significant difference between the means, therefore there is no significant difference between the effects of paraquat on \( \Delta\text{sigG} \) knockout or H37Rv.

<table>
<thead>
<tr>
<th>Table Analyzed</th>
<th>Untreated</th>
<th>10mM Paraquat</th>
<th>20mM Paraquat</th>
<th>30mM Paraquat</th>
<th>40mM Paraquat</th>
<th>50mM Paraquat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transformed ( \log_{10} )</td>
<td>Transformed ( \log_{10} )</td>
<td>Transformed ( \log_{10} )</td>
<td>Transformed ( \log_{10} )</td>
<td>Transformed ( \log_{10} )</td>
<td>Transformed ( \log_{10} )</td>
</tr>
<tr>
<td>H37Rv wild-type</td>
<td>H37Rv U</td>
<td>H37Rv 10</td>
<td>H37Rv 20</td>
<td>H37Rv 30</td>
<td>H37Rv 40</td>
<td>H37Rv 50</td>
</tr>
<tr>
<td>vs</td>
<td>vs</td>
<td>vs</td>
<td>vs</td>
<td>vs</td>
<td>vs</td>
<td>vs</td>
</tr>
<tr>
<td>sigG knockout</td>
<td>sigG U</td>
<td>sigG 10</td>
<td>sigG 20</td>
<td>sigG 30</td>
<td>sigG 40</td>
<td>sigG 50</td>
</tr>
<tr>
<td>Unpaired t test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td>0.9479</td>
<td>0.2934</td>
<td>0.5437</td>
<td>0.0609</td>
<td>0.1463</td>
<td>0.4268</td>
</tr>
<tr>
<td>P value summary</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Significant difference at 95% (P &lt; 0.05)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>One- or two-tailed P value?</td>
<td>Two-tailed</td>
<td>Two-tailed</td>
<td>Two-tailed</td>
<td>Two-tailed</td>
<td>Two-tailed</td>
<td>Two-tailed</td>
</tr>
<tr>
<td>F test to compare variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F, Df, Df</td>
<td>490.4, 1, 1</td>
<td>9.046, 1, 1</td>
<td>1.046, 1, 1</td>
<td>1.378, 1, 3</td>
<td>4.151, 3, 1</td>
<td>1.487, 3, 3</td>
</tr>
<tr>
<td>P value</td>
<td>0.0575</td>
<td>0.4087</td>
<td>0.9858</td>
<td>0.6504</td>
<td>0.6856</td>
<td>0.7521</td>
</tr>
<tr>
<td>P value summary</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Are variances significantly different?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Table 3.4

<table>
<thead>
<tr>
<th>Mitomycin C (µg/ml)</th>
<th>H37Rv Mean ± Standard error</th>
<th>sigG knockout Mean ± Standard error</th>
<th>sigG complement 8T Mean ± Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untreated</td>
<td>100.00 ± 12.74</td>
<td>100.00 ± 7.78</td>
<td>100.00 ± 7.18</td>
</tr>
<tr>
<td>0.02</td>
<td>41.16 ± 6.89</td>
<td>14.87 ± 1.03</td>
<td>29.48 ± 2.87</td>
</tr>
<tr>
<td>0.05</td>
<td>7.70 ± 0.98</td>
<td>2.34 ± 0.28</td>
<td>4.40 ± 0.6</td>
</tr>
<tr>
<td>0.1</td>
<td>1.74 ± 0.46</td>
<td>0.13 ± 0.02</td>
<td>0.75 ± 0.17</td>
</tr>
<tr>
<td>0.2</td>
<td>0.23 ± 0.03</td>
<td>0.02 ± 0.006</td>
<td>0.03 ± 0.004</td>
</tr>
</tbody>
</table>

Figure 3.11

Table 3.4: Percentage viability of H37Rv in comparison to sigG knockout and sigG full operon complement in response to Mitomycin C stress. Duplicate cultures were grown to exponential phase (0.3 to 0.4 OD), 40ml aliquots were incubated for 24 hours untreated or with the relevant concentration of Mitomycin C. After 24 hours the cultures were serially diluted and plated on 7H11 plates in duplicate and incubated at 37°C for 16-18 days. Colony counts were taken of at least 2 dilutions. The viable CFU was then expressed as a percentage of the untreated for each sample, and the mean and standard error was calculated.

Figure 3.11: A graph comparing the viability of the sigG knockout, the sigG whole operon complement and H37Rv wild-type to Mitomycin C stress. In Prism 4, the percentage viability data as outlined above was transformed using log₁₀ and plotted on a linear scale. A non-linear regression sigmoidal dose response curve was fitted to the data.
### Table 3.5a: Statistical analysis of the response to mitomycin C stress of the \( \Delta \text{sigG} \) strain compared to H37Rv wild-type.

<table>
<thead>
<tr>
<th>Table Analyzed</th>
<th>Untreated</th>
<th>0.02ug/ml Mlt C</th>
<th>0.05ug/ml Mlt C</th>
<th>0.1ug/ml Mlt C</th>
<th>0.2ug/ml Mlt C</th>
</tr>
</thead>
<tbody>
<tr>
<td>H37Rv wild-type</td>
<td>H37Rv u</td>
<td>H37Rv 0.02</td>
<td>H37Rv 0.05</td>
<td>H37Rv 0.1</td>
<td>H37Rv 0.2</td>
</tr>
<tr>
<td>sigG knockout</td>
<td>H37Rv u</td>
<td>H37Rv 0.02</td>
<td>H37Rv 0.05</td>
<td>H37Rv 0.1</td>
<td>H37Rv 0.2</td>
</tr>
</tbody>
</table>

**Unpaired t test**
- P value: 0.8427
- P value summary: ns
- Are means significantly different? (P < 0.05): No
- One or two-tailed P value?: Two-tailed
- df = 0.2060 df= 14
- F value: 8.894, 5, 2
- P value: 0.2083
- P value summary: ns
- Are variances significantly different? No

### Table 3.5b: Statistical analysis of the response to mitomycin C stress of the \( \Delta \text{sigG} \) whole operon complement compared to H37Rv wild-type.

<table>
<thead>
<tr>
<th>Table Analyzed</th>
<th>Untreated</th>
<th>0.02ug/ml Mlt C</th>
<th>0.05ug/ml Mlt C</th>
<th>0.1ug/ml Mlt C</th>
<th>0.2ug/ml Mlt C</th>
</tr>
</thead>
<tbody>
<tr>
<td>H37Rv wild-type</td>
<td>H37Rv u</td>
<td>H37Rv 0.02</td>
<td>H37Rv 0.05</td>
<td>H37Rv 0.1</td>
<td>H37Rv 0.2</td>
</tr>
<tr>
<td>sigG full operon complement</td>
<td>Comp BT u</td>
<td>Comp BT 0.02</td>
<td>Comp BT 0.05</td>
<td>Comp BT 0.1</td>
<td>Comp BT 0.2</td>
</tr>
</tbody>
</table>

**Unpaired t test**
- P value: 0.8142
- P value summary: ns
- Are means significantly different? (P < 0.05): No
- One or two-tailed P value?: Two-tailed
- df = 0.2407 df= 11
- F value: 4.371. 5. 6
- P value: 0.1008
- P value summary: ns
- Are variances significantly different? No

### Table 3.5c: Statistical analysis of the response to mitomycin C stress of the \( \Delta \text{sigG} \) strain compared to \( \Delta \text{sigG} \) whole operon complement.

<table>
<thead>
<tr>
<th>Table Analyzed</th>
<th>Untreated</th>
<th>0.02ug/ml Mlt C</th>
<th>0.05ug/ml Mlt C</th>
<th>0.1ug/ml Mlt C</th>
<th>0.2ug/ml Mlt C</th>
</tr>
</thead>
<tbody>
<tr>
<td>sigG knockout</td>
<td>H37Rv u</td>
<td>H37Rv 0.02</td>
<td>H37Rv 0.05</td>
<td>H37Rv 0.1</td>
<td>H37Rv 0.2</td>
</tr>
<tr>
<td>sigG full operon complement</td>
<td>Comp BT u</td>
<td>Comp BT 0.02</td>
<td>Comp BT 0.05</td>
<td>Comp BT 0.1</td>
<td>Comp BT 0.2</td>
</tr>
</tbody>
</table>

**Unpaired t test**
- P value: 0.94
- P value summary: ns
- Are means significantly different? (P < 0.05): No
- One or two-tailed P value?: Two-tailed
- df = 0.1789 df= 7
- F value: 2.035. 5. 2
- P value: 0.7312
- P value summary: ns
- Are variances significantly different? No

### Table 3.5a: Statistical analysis of the response to mitomycin C stress of the \( \Delta \text{sigG} \) strain compared to H37Rv wild-type.

The percentage viability for \( \Delta \text{sigG} \) and H37Rv was transformed using log10. An F-test was performed to determine if both samples had equal variance (H0) (a requirement for parametric tests), when the p-value >0.05, H0 is accepted, the populations have equal variance and a t-test is appropriate. Then unpaired, two-tailed t-tests were performed between each sample for both \( \Delta \text{sigG} \) knockout and H37Rv wild-type, under each stress condition individually, to determine if the means for each sample were the same (H0) or the means of \( \Delta \text{sigG} \) and H37Rv were different (H1). Where the p-values <0.05, H0 is rejected and H1 is accepted. There is a significant difference between the means.

### Table 3.5b: Statistical analysis of the response to mitomycin C stress of the \( \Delta \text{sigG} \) whole operon complement compared to H37Rv wild-type.

Data was treated as outlined in table 3.2a, except the comparison was made between the \( \Delta \text{sigG} \) whole operon complement and H37Rv.

### Table 3.5c: A statistical analysis of the response to mitomycin C of the \( \Delta \text{sigG} \) strain compared to \( \Delta \text{sigG} \) whole operon complement.

Data was treated as outlined in table 3.2a, except the comparison was made between the \( \Delta \text{sigG} \) and the \( \Delta \text{sigG} \) whole operon complement.
confidence level (see Table 3.6a). This shows that the data were comparable, as the untreated samples are not significantly different, as you would expect, as the untreated is set to 100%, and the viability is calculated as a proportion of the untreated. However, there is a significant difference between viability of ΔsigG and H37Rv for all concentrations of mitomycin C at the 99% confidence interval (p<0.01) (see table 3.6a).

Statistical analysis was also performed to determine whether the sigG whole operon complement (8T) was significantly different from H37Rv (see table 3.6b) and ΔsigG (see table 3.6c). The data was transformed (log10) to give an approximation to a normal distribution, confirmed by an F-test (see tables 3.6b and 3.6c), and t-tests showed there was no significant difference between the means of H37Rv and the whole operon complement for the untreated and 0.02μg/ml mitomycin C stress, however, there were significant differences at 0.05, 0.1 and 0.2 μg/ml mitomycin C (see table 3.6b). When the t-test was performed between the ΔsigG strain and ΔsigG full operon complement, H0 was accepted for untreated and 0.2μg/ml mitomycin C (no significant difference), but H1 was accepted at 0.02, 0.05 and 0.1μg/ml mitomycin C (significant difference) (see table 3.6c). This suggests the complement does not completely restore sensitivity of the ΔsigG strain back to wild-type H37Rv.

### 3.6 In-vivo phenotype of ΔsigG strain compared to wild-type

A mouse model of infection was used to determine the in-vivo phenotype of the ΔsigG strain, in comparison to wild-type H37Rv. Six to eight week old BALB/C female mice were injected intravenously in the tail vain with 200μl of each strain of bacteria, by a trained animal technician (see section 2.12.3). Initial infection colony counts (CFU) were determine from the
inoculum, whereas bacterial load was determined from harvested lungs and spleen at regular intervals throughout the time course, beginning at day 1.

The method of injection accounts for the difference in CFU observed initially between the lungs and the spleen, with the higher CFU being present in the spleen. In a human infection, the route of infection determines that the CFU is highest in the lungs. The only way to mimic this would be to use an aerosol route of infection, which is currently not available at our animal facility.

Figure 3.12a and 3.12b show a difference between CFU in both the lung and the spleen for ΔsigG strain compared to the H37Rv wild-type. The ΔsigG strain shows greater than a log decrease in CFU at both day 95 and 150.

Statistic analysis using a two-tailed Student T-test shows a significant difference between ΔsigG strain and H37Rv: In the spleen, there is a significant difference in CFU in ΔsigG strain compared to H37Rv at day 60, 95, 125 and 150 at p<0.1, however there is no significant difference at day 2, day 30. In the lung there is a significant difference after inoculation at day 2, p<0.05 and also at days 60, 95 and 125 (p<0.05), however, there is no significant difference at p<0.05 for day 30 or day 150. This analysis suggests that the ΔsigG strain is attenuated in the mouse model of infection in comparison to H37Rv.
Figure 3.11: Colony counts in lung of BALB/C mice infected with △sigG strain and H37Rv wild-type. Six to eight week old female BALB/C mice were infected intravenously via the tail vein with △sigG or wild-type strains of M. tuberculosis. Initial colony counts were taken from the inoculum to determine the dose administered. Colony counts were then obtained from the lung (a) and spleen (b) at day 2, 30, 60, 95, 125 and finally day 150.
3.7 Discussion

SigG is of particular interest, as it was the most highly induced sigma factor in the \( \Delta recA \) strain of \( M. \) tuberculosis in response to DNA-damage (Rand et al., 2003), suggesting it may play a role in the regulation of DNA-damage repair. The level of induction in the \( \Delta recA \) strain indicates that the induction of SigG is not dependent on RecA, and thus may form part of an alternative mechanism of regulation of genes involved in DNA-damage repair, particularly as the majority of genes upregulated in response to DNA-damage are regulated independently of the RecA/LexA system (Rand et al., 2003), extensively studied in both \( E. \) coli and \( B. \) subtilis.

In order to determine whether SigG played a role in regulation of DNA-damage repair, the obvious mode of analysis was to produce a gene inactivation mutant of SigG, and assess the response of the mutant to a variety of DNA damaging agent to determine whether the \( \Delta sigG \) strain was more susceptible to DNA-damage therefore, indicating a possible role in regulation of DNA-damage repair.

As previously mentioned, sigma factors regulate gene expression by forming the specific interaction between a promoter and RNA polymerase, which enables initiation of transcription. However, it is important that sigma factors themselves are regulated, in order to provide specific temporal and spatial control of gene expression. The majority of sigma factors are negatively regulated by their cognate anti-sigma factors, which, via a number of alternative mechanisms inhibit binding of the sigma factor to the RNAP, and therefore inhibit transcription of the sigma factor’s regulon. The number of different methods the various anti-sigma factors employ to inhibit sigma factor-RNAP binding is wide ranging, as sigma factors can be regulated at transcriptional, translational and post-translational levels (Helmann, 1999). The pathways
controlling anti-sigma factors can be broadly grouped into three categories, export from the cell, as seen with the flagella biosynthesis anti-sigma factor FlgM of *Salmonella typhimurium*, partner switching modules as outlined with SigF of *B. subtilis* (i.e. regulation by an anti-anti sigma factor) and interactions with small molecules or protein ligands (Helmann, 1999). A limited number of sigma factor regulators have been identified, whereby the anti-sigma factor is a transmembrane protein with an intracellular inhibitory domain, which binds and inhibits the cognate sigma factor (Yoshimura *et al.*, 2004), see table 3.7 for a few examples.

<table>
<thead>
<tr>
<th>Bacterial strain</th>
<th>Sigma factor</th>
<th>Transmembrane anti-sigma factor</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td>SigE</td>
<td>RseA</td>
</tr>
<tr>
<td></td>
<td>FecI</td>
<td>FecR</td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td>SigX</td>
<td>RsIX</td>
</tr>
<tr>
<td><em>P. aeruginosa</em></td>
<td>AlgU</td>
<td>MucA</td>
</tr>
<tr>
<td><em>R. myxococcus</em></td>
<td>CarQ</td>
<td>CarR</td>
</tr>
</tbody>
</table>

Table 3.6: Sigma factors regulated by transmembrane anti-sigma factors

Interestingly 5 of the 7 ECF family sigma factors in *B. subtilis* are co-transcribed with downstream genes (Yoshimura *et al.*, 2004). SigM in *B. subtilis* is co-transcribed with two downstream genes, known to negatively regulate SigM, *yhdL* and *yhdK*. It appears that YhdL possesses 1 possible transmembrane region, whereas, YhdK the last gene in the operon possesses multiple transmembrane regions. Yeast-2-hybrid studies revealed that SigM interacts with the N-terminal region of YhdL, but not with YhdK. It has also been shown that YhdL and YhdK interact with a highly specific interaction despite the transmembrane regions. Interestingly in deletion experiments removal of regions of the trans-membrane domains nullified the interaction, which suggests the interactions of these proteins YhdL and YhdK, may take place in the cytoplasmic membrane (Yoshimura *et al.*, 2004).
Interestingly, two component regulatory systems are also able to regulate sigma factor activity, as is the case for SigE of *Streptomyces coelicolor*, which is not regulated by a membrane bound anti-sigma factor, but by a two component regulatory system (Paget *et al*., 1999). CseB and CseC are the response regulator and histidine kinase respectively that regulate SigE (Hong *et al*., 2002). It is also worth noting that *sigE* is co-transcribed with *cseA*, *cseB* and *cseC* (Hutchings *et al*., 2004). SigE has been shown to not autoregulate, but is regulated by the sensor kinase CseC and the response regulator CseB, whereby CscC responds to cell wall damage, resulting in autophosphorylation, which, in turn, leads to phosphorylation of the response regulator CseB, which activates transcription from the SigE promoter (Hutchings *et al*., 2004). It has been suggested, that CseA is a negative regulator of SigE, hypothesised to interact with the sensor kinase CseC, either directly or indirectly negatively regulating SigE (Hutchings *et al*., 2004).

ECF sigma factors do not solely respond to environmental signals, in the case of SigR from *S. coelicolor*, the regulon is stimulated by the response of the cytoplasmic anti-sigma factor to redox (Paget *et al*., 1998). RsrA is a member of the ZAS family of zinc binding anti-sigma factors (Li *et al*., 2003). Under reducing conditions, RsrA binds to the N-terminal region of SigR, encompassing the region 2, thus preventing association with RNAP (Li *et al*., 2002), then under oxidising conditions (oxidation by molecular oxygen, hydrogen peroxide or diamide), disulfide bonds are formed between the cysteine residues that form the Zn binding motif, this results in release of the Zn, causing a conformational change in RsrA, which results in release of SigR, thus enabling transcription of the SigR regulon (Bae *et al*., 2004; Li *et al*., 2003).

The examples of sigma factor regulation have shown that the regulation of activity is diverse. However, *sigG* is co-transcribed with two downstream genes, so, it is possible that these genes
are involved in regulation of the sigma factor, a hypothesis supported by analysis of the domain predictions for the proteins which form the SigG operon.

### 3.7.1 Inspection of the SigG operon

The close proximity of the two downstream genes (Rv0181c and Rv0180c) to sigG, initially indicated that they may be transcribed as a part of a polycistron, especially due to the small intergenic distance between sigG and Rv0181c, and the overlap of Rv0181c and Rv0180c. RT-PCR results confirmed that these genes were part of a polycistron. This is of particular interest, as it has been suggested that sigma factors are co-transcribed as polycistronic RNA with their cognate anti-sigma factors and anti-anti sigma factors (Raivio et al., 2001). It is therefore possible that Rv0181c and Rv0180c are required for regulation of the sigma factor SigG.

The domain analysis showed that SigG possesses homology to a sigma 70 region 2 domain, which forms the specific contact between the sigma factor-RNAP complex and the promoter – 10 region (Lonetto et al., 1994). Interestingly there are homologues to all three proteins of the M. tuberculosis SigG operon in M. bovis, yet only to SigG in M. paratuberculosis. Rv0181c shows homology to cytoplasmic proteins. Rv0180c contains multiple transmembrane domains, which suggest the protein winds in and out of the membrane, possibly to detect environmental signals. It has been shown that some anti-sigma factors are held at the cytoplasmic membrane by accessory proteins, therefore one could speculate that Rv0181c is an anti-sigma factor which may remain tethered to the cytoplasmic membrane via interactions with the anti-anti sigma factor Rv0180c, then by the mechanism of partner switching, the environmental signal could induce a conformational change in the anti-anti sigma factor, which would then have a downstream effect on the conformation of the anti-sigma factor, thus resulting in the release of...
the sigma factor, to initiate transcription of the sigma factor's regulon. This is however, pure speculation, but experiments were designed with the view to analysing the interactions between SigG and the two downstream co-transcribed proteins.

The design of the peptide to produce the antibodies took into account the potential binding sites reported for sigma factors and their putative anti-sigma factors, with anti-sigma factors interacting with sigma factors via their N-terminal region (Yoshimura et al., 2004). The Western did not provide conclusive evidence to determine whether the Rv0181c and Rv0180c antibodies were specific for the respective proteins; however, the SigG antibody did not react with its corresponding peptide. The problems with the SigG antibody may have been due to the poor immunogenic properties of the peptide used in the antibody synthesis. However, there are a number of considerations for the poor reactivity of the antibody. The problems with the sigma factor antibody specificity may lie partially with the location of the antigen. The sigma factors have been reported to bind the anti-sigma factors with the C-terminal region, therefore when the sigma factor is bound to the anti sigma factor, the antibody recognition site may be occluded. Alternatively there may be insufficient SigG present to be detected by a Western blot, or the antibody does not bind under the conditions used. One possibility would be to repeat the Western using more CFE, and use the terminal bleed which appeared to react more strongly by dot blot, rather than the purified fraction.

The antibodies raised to Rv0181c and Rv0180c could also be used along with an improved antibody to SigG in co-immunoprecipitation experiments to determine whether either of the two downstream genes interacts with SigG. This would be a potentially exciting method to determine the interaction of the genes in the SigG operon.
The Westerns did not show any difference between the bands present under uninduced and induced conditions for CFE from either the H37Rv or ΔsigG strain of *M. tuberculosis*. This suggests that the two downstream genes Rv0181c and Rv0180c may undergo both transcription and translation in the ΔsigG strain, however to determine this properly, the anti-bodies would need to be tested for specificity using deletion mutants of the relevant proteins. It appeared by Western, that the levels of protein are approximately equivalent in both the H37Rv and ΔsigG strains for both Rv0180c and Rv0181c, suggesting any potential polar effects do not play a role in the expression of these proteins, despite RT-PCR indicating that the genes are co-transcribed. Thus, to be able to analyse any role of the downstream genes in controlling SigG activity, it would be necessary to construct a new strain of *M. tuberculosis*, in which all three genes are deleted. This would then allow the assessment of expression of a SigG-dependent promoter following the re-introduction of the individual and pairs of genes. The antibodies were also intended for use in co-immunoprecipitations to determine if under certain environmental stimuli such as DNA-damage, the proteins formed an interaction, with the initial hypothesis that sigG may interact with Rv0181c (under uninduced conditions), and that Rv0181c and Rv0180c may interact (under induced conditions). The co-immunoprecipitations were not performed due to the inability of the synthesised SigG antibody to recognise its target.

The construction of the ΔsigG strain was such that a non-functional protein would be produced, due to the deletion of part of the coding region and the insertion of the antibiotic resistance cassette. The protein analysis revealed that contrary to microarray analysis, the levels of Rv0181c and Rv0180c were similar in the ΔsigG strain to H37Rv, therefore suggesting there were no polar effects of the knockout on the downstream protein expression. There may however, be differences at the transcriptional level as observed by the microarray analysis, which, may not be observed in the translation. The lack of induction detectable by Western
analysis may reflect the limited discrimination of the method in quantitation. The Western may have been saturated, therefore to determine whether the apparent lack of induction was a valid observation, a titration of both the antigen (CFE) and the antibody would need to be performed.

3.7.2 Viability of the $\Delta$sgG strain

There does not appear to be any difference between the in-vitro growth rates of the $\Delta$sgG strain or $\Delta$sgG strain complements in comparison to the wild-type H37Rv. However, the $\Delta$sgG strain is more susceptible than H37Rv to the DNA damaging agent, mitomycin C (significantly different $p<0.01$). Interestingly, mitomycin C causes inter-strand cross-links of complementary DNA by alkylation of guanine residues, which makes GC rich organisms more susceptible (Iyer and Szybalski, 1964).

Titration experiments revealed that $\Delta$sgG strain was not significantly more susceptible to paraquat than H37Rv, contradicting the preliminary results. For the preliminary experiments, using the range of DNA damaging agents, the viability for the wild-type and $\Delta$sgG strains were performed separately using different freshly made stocks of the spectrum of chemical damaging agents, whereas the wild-type and $\Delta$sgG strain for the titration experiments were performed in parallel using the same freshly made stocks and dilutions of the reagents. This could therefore account for the differences observed in viability across the two experiments. The titration experiments are presumably more reliable due to the use of the same stocks and dilutions for both H37Rv and $\Delta$sgG strain, leading to the conclusion that the $\Delta$sgG strain is no more susceptible to paraquat stress than H37Rv.
The mitomycin C titration showed that the ΔsigG strain is significantly more susceptible to mitomycin C stress than wild-type. However, the full operon complement of ΔsigG did not completely restore susceptibility back to wild-type levels, although the complemented strain was significantly less susceptible to mitomycin C than the ΔsigG strain. The variation cannot be accounted for simply by the variability in the separate stocks of mitomycin C used for the biological replicates, as the initial experiment for H37Rv and ΔsigG strain used one freshly made stock of mitomycin C, which was repeated including the full operon complement, with a different freshly made stock, yet the biological replicates for the H37Rv and the ΔsigG strain were almost identical in their viability (data not shown). Strangely there was more variability in the biological replicates of the full operon complement, with one of the replicates mirroring the H37Rv wild-type at all concentrations except 0.2μg/ml. This incomplete restoration to wild-type viability may be due to the location of the integrated plasmid, whereby the topology of the DNA has a negative effect on the gene expression. This type of altered expression is known as position effect variegation in gene expression is readily observed in the more complicated system of eukaryotes, whereby transgenes integrated into or close to heterochromatin (condensed/closed chromatin) show decreased levels of transcription (Festenstein et al., 1996).

It may be possible that the complementing construct has integrated (at the att site), close to supercoiled DNA, which would hinder the access of transcription factors, and may influence gene expression such that it does not fully return to wild-type. The other possibility is that the complementing construct does not function to fully complement gene expression, or viability.

Decreased viability of the ΔsigG strain was also observed in the mouse in-vivo model of infection, where the CFU in both the lung and the spleen were significantly decreased in the ΔsigG strain compared to H37Rv, suggesting that ΔsigG strain was less virulent. This
experiment was repeated using the full and partial operon complements alongside the ΔsigG strain and H37Rv. However, due to technical problems with obtaining the CFU in both the lung and spleen after early time points (day 34), no H37Rv colony counts were available, after this time point and high variation was observed in very limited counts that were available for the ΔsigG strain and the various complements. The samples were re-streaked from frozen stocks, but technical problems in the animal facilities have prevented CFU reading.

In brief, a gene inactivation knockout of sigG was successfully constructed in M. tuberculosis; in-vitro analysis showed that there was no difference in in-vitro growth, although ΔsigG strain was more susceptible to mitomycin C than the wild-type, and preliminary mouse in-vivo growth indicates that the ΔsigG strain is attenuated in the mouse model. Transcriptional analysis revealed that sigG is co-transcribed with two downstream genes, which potentially may be involved in regulation of SigG, particularly as other sigma factors are co-transcribed with their regulatory partners. Rv0180c appears to be a transmembrane protein, whereas Rv0181c appears to be a cytoplasmic protein. One could hypothesise that Rv0180c may be involved in detection of environmental signals, which could be transmitted via Rv0181c, to activate SigG; this may be in the form of anti-anti-sigma factor (Rv0180c) and anti-sigma factor (Rv0181c). The antibodies raised to Rv0181c and Rv0180c gave inconclusive data regarding their specificity, due to the E. coli CFE not being an ideal negative control. However, the antibodies potentially showed that the levels of Rv0181c and Rv0180c in the ΔsigG strain appear to be similar to that in H37Rv, with only slightly decreased levels, therefore suggesting the ΔsigG mutation may have limited polar effects on the downstream genes.
4 Regulation of \textit{sigG} and identification of its regulon

The expression level of the sigma factor, SigG (encoded by \textit{sigG}) has been shown to be induced in response to DNA-damage, in both the H37Rv and \textit{\Delta recA} strains of \textit{M. tuberculosis} (Rand \textit{et al.}, 2003), suggesting that it may be involved in the response to DNA-damage. The induction of the gene in the \textit{\Delta recA} strain of \textit{M. tuberculosis} indicates that any role in DNA-damage repair is thought to be independent of the RecA/LexA system.

It is possible that \textit{sigG} plays a role in the response to DNA-damage, in which case, DNA-damage repair genes may form part of this sigma factor's regulon. To determine whether this was the case, microarray experiments were performed to compare the \textit{\Delta sigG} strain of \textit{M. tuberculosis} with the H37Rv wild-type under both uninduced and induced conditions.

As \textit{sigG} is DNA-damage inducible at the transcriptional level, it was important to determine if \textit{sigG} possesses one or more damage inducible transcriptional start sites. The expression from these transcriptional start site(s) could then be assayed in the H37Rv wild-type, \textit{\Delta recA}, and \textit{\Delta sigG} strains of \textit{M. tuberculosis}, to determine if RecA plays any role in transcription, and to see if \textit{sigG} is autoregulated, which is of particular interest, as several sigma factors drive transcription from their own promoters, including \textit{sigD} from \textit{M. tuberculosis} (Raman \textit{et al.}, 2004).

4.1 Identification of the \textit{sigG} regulon by microarray analysis

Microarray experiments were designed such that inter- and intra-strain comparisons could be made between the levels of expression under uninduced and induced (0.02\textmu g/ml mitomycin C) conditions. This meant that RNA versus DNA arrays were performed, whereby each RNA
sample was labelled with cy5, and competitively hybridized with cy3-labelled DNA, obtained from Colorado State University. The design of the experiment took into account the difficulties observed with differential labelling with one of the dyes, as DNA was used as a normaliser across all samples, thereby enabling direct comparisons between transcriptional levels in ΔsigG strain compared to wild-type H37Rv, under both uninduced and induced conditions individually. Traditional microarray data gives an induction ratio, which may be skewed when the expression level of a gene is particularly low. Therefore, the ability to dissect both the uninduced and induced conditions, procures the ability to calculate the induction ratio from the uninduced and induced data. Thus enabling a more thorough picture of the expression levels in response to different environmental conditions to be constructed.

The microarray slides used were PCR spotted whole genome M. tuberculosis arrays, produced by the BµGS microarray unit at St. George’s Hospital Medical School. Triplicate cultures for RNA samples were harvested from exponentially grown (OD₆₀₀ 0.15) H37Rv and ΔsigG strains of M. tuberculosis, and were either uninduced or induced with 0.02 µg/ml mitomycin C. Both the uninduced and induced cultures were incubated for a further 24 hours after the addition of the chemical DNA damaging agent mitomycin C. Any DNA contamination was removed by DNase treatment. The RNA samples were then analysed using a Bioanalyzer (Agilent Technologies) to determine the quantity and quality of the RNA preparation.

4.1.1 Analysis and input of the microarray slides

Each triplicate biological sample was used in duplicate for both the uninduced and induced (mitomycin C 0.02 µg/ml) cultures of H37Rv wild-type and the ΔsigG strain, resulting in a
minimum of 6 slides per strain, per treatment. (Slides were repeated where there was a problem with either scratching of the slide or background fluorescence).

The slides were scanned and analysed as outlined in the methods (see section 2.13.5). The spread of the data was similar in both the H37Rv strain and the ΔsigG strain, therefore the statistical analysis performed on the data, was a parametric test; the Student’s T-test, which assumes equal variance for the H37Rv and ΔsigG strains.

4.1.2 Statistical analysis of microarray data

The microarray data was analysed in two ways, to answer two different questions. The first was to identify whether SigG played a role in the regulation of genes induced by DNA-damage and the second was to look at the difference in expression between genes in the ΔsigG and H37Rv strains under each condition separately, producing a comparison between the expression levels of genes in the ΔsigG strain compared to H37Rv under uninduced and induced conditions.

4.1.2.1 SigG and the genes involved in the DNA-damage response

To address the possibility that SigG was involved in gene regulation following DNA-damage, a list of genes induced two fold or more in the H37Rv wild-type in response to DNA-damage, was collated using the ‘filtering on fold change’ option in Genespring, with the minimum cut off value of 2-fold induction. A parametric Student’s T-test (equal variance) was then performed on the 2-fold gene list, using the log-of-ratio values, to determine if the expression of these genes was significantly different between uninduced and induced samples of H37Rv (p<0.01). A parametric test with equal variance (t-test) was chosen for H37Rv uninduced versus
induced comparisons, as transformation of data (log of ratios) by either \( \log_{10} \) or \( \log_e \) transforms the data to approximate a normal distribution (see 2.12.2), and is widely recognised as an acceptable approximation of a normal distribution. The resultant gene list contained 115 genes, whose induction ratio was greater than or equal to 2 and the uninduced values were significantly different from the induced values (\( p<0.01 \)) for H37Rv.

Combining the H37Rv and \( \Delta\text{sig}G \) data into one experiment in Genespring was very useful when it came to producing gene lists. The gene list of 115 genes was exported into Excel, and contained both the H37Rv and the \( \Delta\text{sig}G \) values, which enabled a direct comparison to be made in Excel to determine whether there were any significant differences in the induction ratios of H37Rv and the \( \Delta\text{sig}G \) strain. The data downloaded was in the format of the normalised data for each biological (1, 2 and 3) and technical replicate (A and B), for each condition (+/- mitomycin C) in both strains (H37Rv and \( \Delta\text{sig}G \)). Calculations were performed to determine the mean, standard deviation and standard error for both the uninduced and induced values for each strain individually. Any standard errors greater than 3 standard deviations from the mean were highlighted, and the individual values were assessed. In some cases, spurious data spots had not been effectively removed in Gene Pix, which could be seen as huge numbers (>1000), usually in one out of the 6 replicates. When the data for 5 out of the 6 biological and technical replicates were within one standard deviation of the mean, and the last replicate was greater than 3 standard deviations away from the mean, these data points were removed as errors. However, outliers that were less than 3 standard deviations from the mean were included in the data.

After the removal of outliers, the induction ratios were calculated for each of the biological and technical replicates individually. The mean, standard deviation and standard error of the
induction ratios were calculated for each of the 115 genes in the list. A two-tailed Student T-test was performed in Excel to determine if there were any significant differences in induction ratio between the H37Rv wild-type and ΔsigG strains, using a P<0.05 as a cut off. Of the 115 genes tested, 17 showed a significantly different induction ratio in the ΔsigG strain compared to H37Rv (P<0.05). Of those 17, 4 had a significantly lower induction ratio in the ΔsigG strain compared to H37Rv (see figure 4.1a), whereas 13 had a significantly higher induction ratio in the ΔsigG strain compared to H37Rv (see figure 4.1b).

The four genes with the lower induction ratio fmt, lexA, Rv1956 and sigG, were examined more closely to reveal that all three of these genes other than sigG were expressed at a similar level following DNA-damage in the two strains, H37Rv and ΔsigG (see figure 4.1a); however, these genes were expressed at a higher level in the ΔsigG strain under uninduced conditions, resulting in a reduced induction ratio. The elevated level of expression observed under normal growth conditions might be attributed to the sigma factor competition, whereby the sigma factor responsible for the expression of these genes may be better able to compete for RNAP in the absence of SigG. In the case of SigG itself, expression was detected, as the probe overlaps with the retained 5' region of SigG by approximately 48bp.

Of the 13 genes with significantly higher induction ratios, only 6 genes appear to be expressed at a higher level following DNA-damage when looking at the dissected uninduced and induced data; the other 7 genes appear to be expressed to a similar extent after DNA-damage in both the ΔsigG strain and H37Rv. Those genes which appear to be upregulated in ΔsigG strain to a greater extent than in H37Rv under induced conditions are: dnaB - a probable replicative DNA helicase, Rv2884 - a predicted transcriptional regulator and Rv0059, Rv2734, Rv3075c and Rv3467 which are all conserved hypothetical proteins. Further analysis revealed that only
Figure 4.1a: Microarray data showing decrease in induction ratio of genes in \(\Delta\text{sigG}\) strain, following in DNA damage. Four genes were identified that were 2-fold or more upregulated in H37Rv in response to DNA damage (mitomycin C stress), where the induction ratio in \(\Delta\text{sigG}\) was decreased in comparison to H37Rv with \(p<0.05\) (student t-test). The top graph shows the individual values under induced and uninduced conditions for each strain, whereas the bottom graph shows the induction ratio.

Figure 4.1a: Microarray data showing an increase in induction ratio of genes in \(\Delta\text{sigG}\) strain, following in DNA damage. Thirteen genes were identified that were 2-fold or more upregulated in H37Rv in response to DNA damage (mitomycin C stress), where the induction ratio in \(\Delta\text{sigG}\) was increased in comparison to H37Rv with \(p<0.05\) (student t-test). The top graph shows the individual values under induced and uninduced conditions for each strain, whereas the bottom graph shows the induction ratio.
Rv2884 under uninduced and Rv3467 under induced conditions were significantly different using a Student t-test p<0.01 in the ΔsigG strain compared to H37Rv.

4.1.2.2 Detecting differential gene expression in the ΔsigG strain

It does not appear that SigG plays a significant role in the response to DNA-damage. Therefore, the second question was addressed, to determine, which genes were differentially expressed in the ΔsigG strain compared to the H37Rv strain. To answer this question, a comparison was made between the expression levels of genes under uninduced conditions in each strain, then a separate comparison of expression level was made between the ΔsigG and H37Rv strains under induced conditions.

Under uninduced conditions, 81 genes were identified as significantly different between the two strains at the P<0.01 level, using a parametric Student’s T-test (assume equal variance). Of these, 52 were expressed less in ΔsigG compared to H37Rv (see table 4.1), whereas 29 were expressed to a higher level in ΔsigG compared to H37Rv (see table 4.2). A comparison between the expression level of induced H37Rv and induced ΔsigG, revealed that 50 genes were significantly different at the p<0.01; of those, 15 were induced to a lesser extent in the ΔsigG compared to H37Rv (see table 4.3), and 35 were expressed to a higher degree in ΔsigG compared to H37Rv (see table 4.4). A Venn diagram was produced to determine if any of the genes were significantly different under both uninduced and induced conditions (see figure 4.2): The expression of only 7 genes were significantly different under both conditions in the ΔsigG compared to H37Rv strain of M. tuberculosis (see figure 4.2). Of these 7 genes, 5 showed significantly decreased expression in ΔsigG, whereas 2 genes showed a significant increase in expression for ΔsigG compared to the H37Rv wild-type (see figure 4.3a and 4.3b).
Table 4.1: Gene list of genes showing decreased expression in uninduced conditions in ΔsigG compared to H37Rv wild-type. The gene list was produced in Gene spring, using a parametric t-test, with cut off p value <0.01, to determine which genes were significantly different in ΔsigG compared to H37Rv wild-type under uninduced conditions. The normalised expression levels were then exported into excel. The data was divided into genes significantly lower in the ΔsigG and genes significantly higher in ΔsigG (table 4.2) compared to H37Rv wild-type. The highlighted genes are significantly different in ΔsigG for both uninduced and induced (0.02μg/ml mitomycin C) conditions.

<table>
<thead>
<tr>
<th>Common name</th>
<th>Gene ID</th>
<th>H37Rv uninduced</th>
<th>H37Rv induced</th>
<th>ΔsigG uninduced</th>
<th>ΔsigG induced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
</tr>
</tbody>
</table>

- **fadD5**
  - Rv0166: 1.441 ± 0.158
  - Rv0168: 2.938 ± 0.341

- **mce1**
  - Rv0169: 3.552 ± 0.464

- **Rv0232**
  - Rv0232: 0.704 ± 0.066

- **rmbB**
  - Rv2295: 0.876 ± 0.099

- **Rv0312**
  - Rv0312: 1.717 ± 0.183

- **murB**
  - Rv0482: 0.841 ± 0.082

- **Rv0655**
  - Rv0655: 8.305 ± 1.261

- **PPE**
  - Rv1204c: 0.304 ± 0.029

- **Rv1230c**
  - Rv1230c: 2.037 ± 0.141

- **Rv1232c**
  - Rv1232c: 1.827 ± 0.176

- **Rv1348**
  - Rv1348: 0.811 ± 0.120

- **IprF**
  - Rv1368: 1.655 ± 0.128

- **Rv1566c**
  - Rv1566c: 2.387 ± 0.179

- **Rv1776c**
  - Rv1776c: 1.279 ± 0.141

- **Rv1802**
  - Rv1802: 0.330 ± 0.023

- **Rv1883c**
  - Rv1883c: 1.040 ± 0.123

- **Rv1891**
  - Rv1891: 2.198 ± 0.164

- **Rv1986**
  - Rv1986: 0.913 ± 0.044

- **Rv1987**
  - Rv1987: 2.287 ± 0.112

- **Rv2180c**
  - Rv2180c: 0.284 ± 0.030

- **fabD**
  - Rv2243: 4.389 ± 0.534

- **Rv2252**
  - Rv2252: 0.723 ± 0.051

- **Rv2262c**
  - Rv2262c: 0.314 ± 0.022

- **Rv2293c**
  - Rv2293c: 0.592 ± 0.036

- **gvy5**
  - Rv2357: 0.449 ± 0.038

- **lipQ**
  - Rv2485c: 0.627 ± 0.044

- **Rv2563**
  - Rv2563: 0.563 ± 0.013

- **Rv2599**
  - Rv2599: 0.714 ± 0.087

- **Rv2616**
  - Rv2616: 0.617 ± 0.061

- **Rv2633c**
  - Rv2633c: 1.601 ± 0.166

- **Rv2690c**
  - Rv2690c: 0.550 ± 0.074

- **aid**
  - Rv2780: 2.855 ± 0.178

- **efpA**
  - Rv2846c: 1.918 ± 0.194

- **Rv2884**
  - Rv2884: 1.030 ± 0.081

- **ppsE**
  - Rv2935: 1.182 ± 0.118

- **Rv3050c**
  - Rv3050c: 1.732 ± 0.215

- **nuoB**
  - Rv3146: 2.343 ± 0.650

- **nuoL**
  - Rv3156: 2.356 ± 0.215

- **Rv3402c**
  - Rv3402c: 4.299 ± 0.372

- **Rv3616c**
  - Rv3616c: 3.978 ± 0.067

- **Rv3633**
  - Rv3633: 5.762 ± 0.310

- **Rv3719**
  - Rv3719: 2.672 ± 0.169

- **Rv3764c**
  - Rv3764c: 2.513 ± 0.454
Table 4.2

<table>
<thead>
<tr>
<th>Common name</th>
<th>Gene ID</th>
<th>H37Rv uninduced</th>
<th>H37Rv induced</th>
<th>AsigG uninduced</th>
<th>AsigG induced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
</tr>
<tr>
<td>fadE6</td>
<td>Rv0271c</td>
<td>1.455 ± 0.320</td>
<td>1.876 ± 1.004</td>
<td>2.706 ± 0.343</td>
<td>2.080 ± 0.510</td>
</tr>
<tr>
<td>Rv0540</td>
<td>Rv0540</td>
<td>0.362 ± 0.039</td>
<td>0.747 ± 0.209</td>
<td>0.888 ± 0.164</td>
<td>0.531 ± 0.073</td>
</tr>
<tr>
<td>Rv0997</td>
<td>Rv0997</td>
<td>0.982 ± 0.061</td>
<td>1.917 ± 0.229</td>
<td>1.544 ± 0.206</td>
<td>1.733 ± 0.092</td>
</tr>
<tr>
<td>Rv1057</td>
<td>Rv1057</td>
<td>0.509 ± 0.152</td>
<td>0.595 ± 0.088</td>
<td>1.208 ± 0.120</td>
<td>0.959 ± 0.052</td>
</tr>
<tr>
<td>sigE</td>
<td>Rv1221</td>
<td>2.723 ± 0.438</td>
<td>3.138 ± 0.412</td>
<td>5.521 ± 0.358</td>
<td>3.231 ± 0.199</td>
</tr>
<tr>
<td>Rv1261c</td>
<td>Rv1261c</td>
<td>1.438 ± 0.145</td>
<td>2.333 ± 0.798</td>
<td>3.078 ± 0.594</td>
<td>1.474 ± 0.097</td>
</tr>
<tr>
<td>Rv1265</td>
<td>Rv1265</td>
<td>1.706 ± 0.216</td>
<td>2.932 ± 0.805</td>
<td>2.695 ± 0.294</td>
<td>1.564 ± 0.155</td>
</tr>
<tr>
<td>Rv1288</td>
<td>Rv1288</td>
<td>0.520 ± 0.098</td>
<td>0.823 ± 0.191</td>
<td>1.316 ± 0.208</td>
<td>1.118 ± 0.082</td>
</tr>
<tr>
<td>Rv1457c</td>
<td>Rv1457c</td>
<td>0.370 ± 0.137</td>
<td>0.392 ± 0.036</td>
<td>1.170 ± 0.116</td>
<td>0.951 ± 0.123</td>
</tr>
<tr>
<td>PPE</td>
<td>Rv1809</td>
<td>0.674 ± 0.051</td>
<td>0.706 ± 0.070</td>
<td>0.947 ± 0.080</td>
<td>0.911 ± 0.077</td>
</tr>
<tr>
<td>Rv1813c</td>
<td>Rv1813c</td>
<td>0.474 ± 0.137</td>
<td>0.392 ± 0.036</td>
<td>1.170 ± 0.116</td>
<td>0.951 ± 0.123</td>
</tr>
<tr>
<td>ureD</td>
<td>Rv1853</td>
<td>0.533 ± 0.067</td>
<td>0.637 ± 0.043</td>
<td>0.876 ± 0.085</td>
<td>0.871 ± 0.083</td>
</tr>
<tr>
<td>aao</td>
<td>Rv1905c</td>
<td>0.556 ± 0.049</td>
<td>1.145 ± 0.202</td>
<td>1.028 ± 0.154</td>
<td>0.879 ± 0.095</td>
</tr>
<tr>
<td>rpmB2</td>
<td>Rv2058c</td>
<td>3.246 ± 0.465</td>
<td>2.805 ± 0.399</td>
<td>6.418 ± 0.393</td>
<td>4.675 ± 0.659</td>
</tr>
<tr>
<td>rpe</td>
<td>Rv2450c</td>
<td>0.904 ± 0.136</td>
<td>1.115 ± 0.062</td>
<td>1.755 ± 0.175</td>
<td>1.844 ± 0.119</td>
</tr>
<tr>
<td>Rv2604c</td>
<td>Rv2604c</td>
<td>1.466 ± 0.166</td>
<td>3.141 ± 0.796</td>
<td>3.616 ± 0.712</td>
<td>1.458 ± 0.186</td>
</tr>
<tr>
<td>Rv2743c</td>
<td>Rv2743c</td>
<td>0.441 ± 0.032</td>
<td>0.622 ± 0.165</td>
<td>0.668 ± 0.055</td>
<td>0.634 ± 0.099</td>
</tr>
<tr>
<td>fadD26</td>
<td>Rv2930</td>
<td>2.185 ± 0.267</td>
<td>4.592 ± 1.114</td>
<td>4.129 ± 0.399</td>
<td>6.616 ± 0.980</td>
</tr>
<tr>
<td>ppsD</td>
<td>Rv2934</td>
<td>1.249 ± 0.150</td>
<td>1.520 ± 0.143</td>
<td>2.278 ± 0.224</td>
<td>2.762 ± 0.347</td>
</tr>
<tr>
<td>Rv3123</td>
<td>Rv3123</td>
<td>0.485 ± 0.053</td>
<td>0.865 ± 0.206</td>
<td>1.074 ± 0.198</td>
<td>0.709 ± 0.085</td>
</tr>
<tr>
<td>Rv3133c</td>
<td>Rv3133c</td>
<td>0.658 ± 0.049</td>
<td>0.978 ± 0.192</td>
<td>1.108 ± 0.095</td>
<td>1.110 ± 0.011</td>
</tr>
<tr>
<td>Rv3482c</td>
<td>Rv3482c</td>
<td>0.792 ± 0.052</td>
<td>0.851 ± 0.052</td>
<td>1.239 ± 0.107</td>
<td>1.006 ± 0.089</td>
</tr>
<tr>
<td>cpsA</td>
<td>Rv3484</td>
<td>0.925 ± 0.052</td>
<td>0.968 ± 0.098</td>
<td>2.557 ± 0.273</td>
<td>1.864 ± 0.340</td>
</tr>
<tr>
<td>Rv3485c</td>
<td>Rv3485c</td>
<td>0.408 ± 0.028</td>
<td>0.528 ± 0.080</td>
<td>0.835 ± 0.155</td>
<td>0.783 ± 0.108</td>
</tr>
<tr>
<td>otsA</td>
<td>Rv3490</td>
<td>0.502 ± 0.083</td>
<td>0.706 ± 0.065</td>
<td>0.912 ± 0.165</td>
<td>0.769 ± 0.119</td>
</tr>
<tr>
<td>Rv3836</td>
<td>Rv3836</td>
<td>0.771 ± 0.057</td>
<td>1.903 ± 0.711</td>
<td>2.389 ± 0.355</td>
<td>1.802 ± 0.188</td>
</tr>
<tr>
<td>bfrB</td>
<td>Rv3841</td>
<td>1.030 ± 0.172</td>
<td>2.603 ± 1.213</td>
<td>2.070 ± 0.330</td>
<td>2.571 ± 0.539</td>
</tr>
</tbody>
</table>

Table 4.2: Gene list of genes showing increased expression under uninduced conditions in ΔsigG compared to H37Rv wild-type. The gene list was produced using Gene spring software, with a parametric t-test, using a cut off p value of <0.01, to determine which genes were significantly different in ΔsigG compared to H37Rv wild-type under uninduced condition. The normalised expression levels were then exported into Excel where the mean standard deviation and standard errors were calculated. The genes with increased mean expression levels in the ΔsigG strain compared to H37Rv are listed in the table, and the genes marked in bold are significantly different in ΔsigG compared to H37Rv under both uninduced and induced (0.02µg/ml mitomycin C) conditions.
Table 4.3: Gene list of genes showing decreased expression under induced conditions in ΔsigG compared to H37Rv wild-type. The gene list was produced using Gene spring software, with a parametric t-test, using a cut off p value of <0.01, to determine which genes were significantly different in ΔsigG compared to H37Rv wild-type under induced (0.02µg/ml mitomycin C) conditions. The normalised expression values were then exported into Excel where the mean, standard deviation and standard errors were calculated. The genes with decreased expression under induced conditions in ΔsigG strain compared to H37Rv are listed above and the genes marked in bold are significantly different in ΔsigG compared to H37Rv under both uninduced and induced (0.02µg/ml mitomycin C) conditions.
Table 4.4: Gene list of genes showing increased expression under induced conditions in $\Delta$sigG compared to H37Rv wild-type. The gene list was produced using Gene spring software, with a parametric t-test, using a cut off p value of <0.01, to determine which genes were significantly different in $\Delta$sigG compared to H37Rv wild-type under induced (0.02 pg/ml mitomycin C) condition. The normalised expression values were then exported into Excel where the mean, standard deviation and standard error were calculated. The genes with increased expression under induced conditions in the $\Delta$sigG strain compared to H37Rv are listed above, and the genes marked in bold are significantly different in $\Delta$sigG compared to H37Rv under both uninduced and induced (0.02 pg/ml mitomycin C) conditions.
Figure 4.2

All genes in *M. tuberculosis* genome

Genes significantly different in H37Rv induced versus ΔsigG induced p<0.01

3800

Genes significantly different in H37Rv uninduced versus ΔsigG uninduced P<0.01

43

74

0

0

Figure 4.2: A Venn diagram of genes significantly different in H37Rv and ΔsigG under induced and uninduced conditions. Microarrays were performed with RNA extracted from H37Rv and ΔsigG strains under uninduced and induced (0.02 μg/ml mitomycin C) conditions. The Venn diagram was constructed in Gene Spring microarray analysis programme, using genes lists derived by ANOVA statistical analysis. The Venn diagram illustrates that 7 genes are significantly different under both restrictions, whereas 43 genes are only significantly different between H37Rv and ΔsigG under induced conditions, and 74 genes are significantly different between H37Rv and ΔsigG under uninduced conditions.
Figure 4.3a: Genes significantly different under both uninduced and induced conditions for ΔsigG compared to H37Rv wild-type. Data obtained from DNA vs RNA microarrays. Venn diagram was used to show which genes were significantly different with p<0.01 using a student t-test between ΔsigG and H37Rv. The normalised values of uninduced and induced expression were plotted for each strain. The first 5 genes show decreased expression in ΔsigG strain (nuoB, Rv0655, Rv0312, ppsE and papA3), whereas the last two genes (ppsD and Rv2450c) show increased expression in ΔsigG. The box indicates genes where the expression level is less than 1.

Figure 4.3b: Induction ratio of genes significantly different under both uninduced and induced conditions for ΔsigG compared to H37Rv wild-type. Data obtained from DNA vs RNA microarrays. Venn diagram was used to show which genes were significantly different with p<0.01 using a student t-test between ΔsigG and H37Rv, and individual induction ratios were calculated and plotted. The asterisk marks the induction ratios that are misleading due to very low levels of expression observed with individual normalised values (figure 4.3a).
When the strains were compared using a Student T-test, p<0.05, alongside a Benjamini and Hochberg false discovery rate (FDR) correction, only three genes were identified that were significantly different between the induced samples (sigG, papA3 and ppsE (see figure 4.4a)), and five genes were identified that were significantly different under uninduced conditions (papA3, Rv2633c, ppsE, cpsA and Rv2293c (see figure 4.4b). The significance of these genes are discussed later.

The data obtained from analysis of the uninduced and induced conditions were used in conjunction with the data obtained from the DNA-damage regulon to determine if there was any overlap. To that end, a Venn diagram was produced using the genes that were significantly different in the ΔsigG strain compared to H37Rv under either uninduced or induced conditions, alongside the genes induced following DNA-damage (see figure 4.5). Only 4 genes that were significantly different under either uninduced or induced conditions for sigG compared to H37Rv were in fact in the DNA-damage regulon (115 genes, upregulated 2-fold or more in H37Rv in response to DNA-damage). One gene, Rv2884, is a predicted transcriptional regulatory protein, the others were sigG and Rv0181c, a predicted hypothetical protein, co-transcribed with sigG (see chapter 3) and Rv3467 a conserved hypothetical protein. However, none of the genes were significantly different under both induced and uninduced conditions and present in the gene list of genes involved in DNA-damage, thus adding more proof that SigG did not play a role in regulation of the DNA-damage regulon.

Identification of a possible role for SigG required focusing on the 131 genes, that were significantly different under either the uninduced or induced conditions, and assigning them into functional groups. This was achieved using fold change (Geiman et al., 2004; Raman et al., 2004) whereby ΔsigG gene expression was expressed as a proportion of H37Rv gene
Figure 4.4a: Genes significantly different in the ΔsigG compared to H37Rv under uninduced conditions at P<0.05 with Benjamini false discovery rate correction. Genes were identified using Genes spring where there was a significant difference at p<0.05 between the normalised expression levels under uninduced conditions for ΔsigG compared to H37Rv, then a Benjamini and Hockberg false discovery rate correction was applied to the data.

Figure 4.4b: Genes significantly different in the ΔsigG compared to H37Rv under induced conditions at P<0.05 with Benjamini false discovery rate correction. Genes were identified using Genes spring where there was a significant difference at p<0.05 between the normalised expression levels under induced conditions (0.02μg/ml mitomycin C) for ΔsigG compared to H37Rv, then a Benjamini and Hockberg false discovery rate correction was applied to the data.
Figure 4.5

2 fold induced in response to DNA damage in H37Rv wild-type p<0.01

Genes significantly different in ΔsigG compared to H37Rv under induced conditions, p<0.01

Gene lists were combined in Gene spring software to produce a venn diagram, showing genes common to each list used. The data added to the venn diagram were the gene list showing 115 genes 2 fold or more upregulated in H37Rv (p<0.01), the gene list with 50 genes significantly different in ΔsigG versus H37Rv under induced conditions (0.02μg/ml mitomycin C) and the gene list of 81 genes significantly different in ΔsigG compared to H37Rv under uninduced conditions. No genes were present in all three gene lists, however, Rv2884 was 2 fold or more upregulated in H37Rv and significantly different under uninduced conditions in ΔsigG compared to H37Rv. Three genes, sigG, Rv3467c and Rv0181c were 2 fold or more upregulated in H37Rv and significantly different under induced conditions in ΔsigG compared to H37Rv.

Figure 4.5: Venn diagram showing the overlap between genes found to be significantly different in ΔsigG compared to H37Rv under various conditions.

Gene lists were combined in Gene spring software to produce a venn diagram, showing genes common to each list used. The data added to the venn diagram were the gene list showing 115 genes 2 fold or more upregulated in H37Rv (p<0.01), the gene list with 50 genes significantly different in ΔsigG versus H37Rv under induced conditions (0.02μg/ml mitomycin C) and the gene list of 81 genes significantly different in ΔsigG compared to H37Rv under uninduced conditions. No genes were present in all three gene lists, however, Rv2884 was 2 fold or more upregulated in H37Rv and significantly different under uninduced conditions in ΔsigG compared to H37Rv. Three genes, sigG, Rv3467c and Rv0181c were 2 fold or more upregulated in H37Rv and significantly different under induced conditions in ΔsigG compared to H37Rv.
expression. The cut-off values were set at >1.7 fold and <0.59 (1/1.7) fold. This revealed that 36 genes were expressed significantly less in the ΔsigG strain, with p<0.01, cut-off >1.7 (see table 4.5), and 16 genes, were significantly higher in ΔsigG strain with a fold change of <0.59 (see table 4.6). This data was then collated into functional groups, using the data available on the TuberculList website (see table 4.5 and 4.6).

4.2 Identification of sigG transcriptional start site(s)

Microarray analysis of ΔrecA strain compared to H37Rv strain of *M. tuberculosis* had shown that SigG was the most highly induced sigma factor in response to DNA-damage (Rand *et al.*, 2003), with only a slight decrease in induction in the ΔrecA strain compared to wild-type, suggesting that the regulation of sigG was not hindered by the absence of RecA. However, even though the microarray data presented in section 4.1 showed that SigG was not involved in regulation of genes involved in DNA-damage response, sigG was induced in H37Rv wild-type yet sigG did not appear to be induced in response to DNA-damage in the ΔsigG strain, which may have been a result of the position of the microarray probe. Therefore, experiments were performed to determine the number and location of transcriptional start site(s), and to identify which were DNA-damage inducible, as well as to determine whether sigG is subject to autoregulation.

4.2.1 Primer extension of sigG

Primer extension reactions utilise radioactively end-labelled primers combined in a reaction with reverse transcriptase to produce cDNA from an RNA template, in order to determine the quantity and position of the 5' end of the RNA molecule. Primer extension reactions were carried out using 50µg of RNA, extracted from uninduced and induced (mitomycin C, final
Table 4.5

<table>
<thead>
<tr>
<th>Functions:</th>
<th>Lipid metabolism</th>
<th>Intermediary metabolism and respiration</th>
<th>Information pathways</th>
<th>Regulatory proteins</th>
<th>Conserved and hypothetical proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Mean Fold Change</td>
<td>SEM Fold Change</td>
<td>Functions:</td>
<td>Mean Fold Change</td>
<td>SEM Fold Change</td>
</tr>
<tr>
<td>ppsE</td>
<td>19.21</td>
<td>± 5.71</td>
<td>PHENOLPHOSOCEROL SYNTHESIS TYPE-I POLYKETIDE SYNTHASE PPSE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pgA3</td>
<td>11.69</td>
<td>± 1.73</td>
<td>PROBABLE CONSERVED POLYKETIDE SYNTHASE ASSOCIATED PROTEIN PAPA3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pks4</td>
<td>3.73</td>
<td>± 1.20</td>
<td>PROBABLE POLYKETIDE BETAL-KETOACYL SYNTHASE PKS4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fabD5</td>
<td>3.19</td>
<td>± 0.50</td>
<td>PROBABLE FATTY-ACID-CoA LIGASE FADDs (FATTY-ACID-CoA SYNTHETASE) (FATTY-ACID-CoA SYNTHASE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kasB</td>
<td>2.04</td>
<td>± 0.27</td>
<td>3-OXOKETACYL (ACYL-CARRIER PROTEIN) SYNTHASE 2 KASB (BETA-KETOACYL-ACP SYNTHASE) (KAS 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mceA</td>
<td>2.90</td>
<td>± 0.36</td>
<td>MCE-FAMILY PROTEIN MCEA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>murB</td>
<td>2.33</td>
<td>± 0.45</td>
<td>MURINE N-ACETYLPYRUVOLSAMINE REDUCTASE MRB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1348</td>
<td>2.07</td>
<td>± 0.25</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pprF</td>
<td>1.93</td>
<td>± 0.46</td>
<td>PROBABLE LIPOPROTEIN PPRF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv2690c</td>
<td>1.93</td>
<td>± 0.30</td>
<td>PROBABLE CONSERVED INTEGRAL MEMBRANE ALANINE AND VALINE AND LEUCINE RICH PROTEIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ekpA</td>
<td>1.74</td>
<td>± 0.12</td>
<td>POSSIBLE INTEGRAL MEMBRANE EFFLUX PROTEIN EFPA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lipQ</td>
<td>2.68</td>
<td>± 0.28</td>
<td>PROBABLE CARBOXYLESTERASE LIQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nuoB</td>
<td>2.30</td>
<td>± 0.55</td>
<td>PROBABLE NADH DEHYDROGENASE I (CHAIN B) NUOB (NADH-UBIQUINONE OXIDOREDUCTASE CHAIN B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv2251</td>
<td>1.81</td>
<td>± 0.26</td>
<td>POSSIBLE FLAVOPROTEIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nuoL</td>
<td>1.79</td>
<td>± 0.34</td>
<td>PROBABLE NADH DEHYDROGENASE I (CHAIN L) NUOL (NADH-UBIQUINONE OXIDOREDUCTASE CHAIN L)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv3083</td>
<td>1.75</td>
<td>± 0.21</td>
<td>PROBABLE NONOXYLANE (HYDROXYLASE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aid</td>
<td>1.74</td>
<td>± 0.21</td>
<td>SECRETED L-ALanine DEHYDROGENASE ALD (40 KDA ANTIGEN) (TB43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>glyS</td>
<td>2.28</td>
<td>± 0.41</td>
<td>PROBABLE GLYCOSAMINE REDUCTASE GLYS (GLYCOSAMINE-tRNA LIGASE) (GLYRS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ndtB</td>
<td>2.01</td>
<td>± 0.17</td>
<td>PROBABLE NADPH-RELATED ATP-BINDING PROTEIN (BETA CHAIN) NDB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv250c</td>
<td>2.29</td>
<td>± 0.38</td>
<td>PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN (PROBABLY ASCN-FAMILY)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0232</td>
<td>2.06</td>
<td>± 0.30</td>
<td>PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN (PROBABLY TETR/ACR-FAMILY)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv3764c</td>
<td>2.03</td>
<td>± 0.38</td>
<td>POSSIBLE TWO COMPONENT SENSOR KINASE TCRY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0655</td>
<td>3.79</td>
<td>± 0.40</td>
<td>POSSIBLE NADPH-RELATED TRANSPORT ATP-BINDING PROTEIN ABC TRANSPORTER MKL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0312</td>
<td>3.49</td>
<td>± 0.52</td>
<td>CONSERVED HYPOTHETICAL PROLINE AND THREONINE RICH PROTEIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv2633c</td>
<td>3.04</td>
<td>± 0.49</td>
<td>HYPOTHETICAL PROTEIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv166c</td>
<td>2.88</td>
<td>± 0.52</td>
<td>CONSERVED HYPOTHETICAL ALANINE AND GLYCINE RICH PROTEIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv004c</td>
<td>2.28</td>
<td>± 0.31</td>
<td>SECRETED PROLINE RICH PROTEIN MTC28 (PROLINE RICH 28 KDA ANTIGEN)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0166</td>
<td>2.28</td>
<td>± 0.45</td>
<td>CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBEB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv2616</td>
<td>1.75</td>
<td>± 0.21</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1883c</td>
<td>1.74</td>
<td>± 0.25</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1891</td>
<td>1.73</td>
<td>± 0.17</td>
<td>PROBABLE MEMBRANE PROTEIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0887c</td>
<td>1.73</td>
<td>± 0.14</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.5:** Microarray data of genes with a fold change of greater than or equal to 1.7 for H37Rv compared to sigG strains of *M. tuberculosis*. The normalised data for H37Rv was expressed as a proportion of sigG, for uninduced cultures, giving a fold change, all data with a fold change of ≥ 1.7 were identified. This indicated that the expression level in H37Rv was ≥ 1.7 times the expression level in sigG strain.
Table 4.6: Microarray data of genes with a fold change of less than 0.59 for H37Rv compared to ΔsigG stain of M. tuberculosis. The normalised data for H37Rv was expressed as a proportion of ΔsigG, for uninduced samples, resulting in a fold change. The data above is therefore expressed 1.7 times higher in ΔsigG compared to H37Rv strains of M. tuberculosis.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mean</th>
<th>SEM</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lipid metabolism</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fadD26</td>
<td>0.58</td>
<td>± 0.03</td>
<td>FATTY-ACID-CoA LIGASE FADD26 (FATTY-ACID-COA SYNTHETASE)</td>
</tr>
<tr>
<td>ppsD</td>
<td>0.55</td>
<td>± 0.05</td>
<td>PHENOLPHIIOCEROL SYNTHESIS TYPE-I POLYKETIDE SYNTHASE PPSD</td>
</tr>
<tr>
<td><strong>Information pathways</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rpmB2</td>
<td>0.52</td>
<td>± 0.10</td>
<td>PROBABLE 50S RIBOSOMAL PROTEIN L28 RPMB2</td>
</tr>
<tr>
<td>sigE</td>
<td>0.49</td>
<td>± 0.06</td>
<td>ALTERNATIVE RNA POLYMERASE SIGMA FACTOR SIGE</td>
</tr>
<tr>
<td><strong>Cell wall and cell processes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rpfE</td>
<td>0.54</td>
<td>± 0.10</td>
<td>PROBABLE RESUSCITATION-PROMOTING FACTOR RPFE</td>
</tr>
<tr>
<td><strong>Intermediary metabolism</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv3485c</td>
<td>0.58</td>
<td>± 0.09</td>
<td>PROBABLE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE</td>
</tr>
<tr>
<td>bfrB</td>
<td>0.53</td>
<td>± 0.10</td>
<td>POSSIBLE BACTERIOFERRITIN BFRB</td>
</tr>
<tr>
<td>Rv2604c</td>
<td>0.50</td>
<td>± 0.07</td>
<td>PROBABLE GLUTAMINE AMIDOTRANSFERASE SNOP</td>
</tr>
<tr>
<td><strong>Conserved and hypothetical proteins</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cpsA</td>
<td>0.40</td>
<td>± 0.08</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
</tr>
<tr>
<td>Rv0540</td>
<td>0.45</td>
<td>± 0.06</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
</tr>
<tr>
<td>Rv1057</td>
<td>0.39</td>
<td>± 0.07</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
</tr>
<tr>
<td>Rv1261c</td>
<td>0.54</td>
<td>± 0.10</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
</tr>
<tr>
<td>Rv1288</td>
<td>0.44</td>
<td>± 0.11</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
</tr>
<tr>
<td>Rv1813c</td>
<td>0.45</td>
<td>± 0.15</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
</tr>
<tr>
<td>Rv3123</td>
<td>0.57</td>
<td>± 0.14</td>
<td>HYPOTHETICAL PROTEIN</td>
</tr>
<tr>
<td>Rv3836</td>
<td>0.35</td>
<td>± 0.05</td>
<td>CONSERVED HYPOTHETICAL PROTEIN</td>
</tr>
</tbody>
</table>
concentration 0.02µg/ml) cultures of wild-type H37Rv M. tuberculosis, alongside a no RNA control. Prior to the primer extension, the RNA samples were quantified using a spectrophotometer, and the primer extension was performed using a primer sigGmidH, designed from within the predicted coding region of sigG (see figure 4.6a and b), based on the predictions available on the Tuberculist website. The primer extension reactions were run alongside manual sequencing reactions, where the template used for the sequencing reactions was a clone, constructed in pBluescript, containing the upstream region and predicted coding region of sigG (see table 2.2). The same primer was used in the primer extension and sequencing reactions so one could determine the exact transcriptional start site. A commercially available control (1.2kb kanamycin positive control RNA, Promega) was used which produced a band of 87bp. However, although the manual sequencing reactions with the sigG primer produced a clear manual sequencing trace, the primer extension reactions failed to produce a product (data not shown).

A test was performed whereby the radiolabelled primer was either ethanol precipitated after the labelling reaction, cleaned up using a G25 column, or left as they were. These were then run on a polyacrylamide gel, either undiluted or diluted 1/10 (data not shown). The G25 spin columns appeared to remove all the unincorporated label (usually present as a smear), but also appeared to remove the primer, despite the fact that the primer was larger than the cut off of the G25 spin columns. The unpurified primer, diluted 1/10 gave the least smearing, so, the remaining primer extension reactions were carried out using 1/10 diluted primer, as outlined in figure 4.6b.

The primer extension reactions were repeated using different quantities of RNA, 25, 45 and 85µg of RNA, the quantity of the RNA was increased, due to the low level of expression of sigG (Manganelli et al., 1999). However, no transcriptional start sites were identified.
Figure 4.6a: Schematic representation of the primer designed to perform the primer extension assay and the manual sequencing reactions. Arrow A, corresponds to the primer extension primer sigGM2, located 47bp downstream of the translational start site, and arrow B corresponded to primer extension primer sigGmidH, located 89bp downstream of the translational start site. The region of interest was cloned into pBluescript and the sequenced clone was used in the manual sequencing reactions.

Figure 4.6b: Schematic representation of the primer extension reactions. The primer extension primers A and B as outlined above, were end-labelled with γ32P-ATP. Primer A (or B) was then annealed to the total RNA from H37Rv wild-type *M. tuberculosis*. Extension reactions were carried out using AMV reverse transcriptase to produce cDNA, which was then ethanol precipitated, run on a denaturing polyacrylamide gel and visualised by autoradiography.
Therefore an alternative primer sigGM2 use used (see figure 4.6a), in case the restriction site on the 3' end of the original primer was interfering with annealing of the primer to the template. In addition, the temperature of the extension reaction was altered, whereby duplicate samples were incubated for 30 minutes at 60°C alongside reactions incubated for 45 minutes at 54°C. However, once again no transcriptional start sites(s) were identified (data not shown). The positive control reactions produced a band, thus indicating that the reactions were successful.

There were a number of possible explanations for the lack of production of a primer extension product. Either both primers used were unable to bind effectively to the RNA template at the reaction temperatures tested, or the transcriptional start site(s) was further upstream or downstream than predicted. The sequencing reaction produced very clear sequence traces with both primers, which indicated that the problem may have been the position of the transcriptional start site(s). Therefore, the upstream and coding region of sigG was assessed for the presence of additional potential translational start sites downstream of the predicted translational start site.

Closer analysis of the upstream and coding regions of sigG revealed the possibility of a number of different translational start sites, downstream of the predicted translational start site (see figure 4.7), introducing the possibility that the predicted translational start site was incorrect and therefore, the primers used in the primer extension reactions were actually placed upstream of the translational start site. The primers may have been located upstream of the transcriptional start site(s), which would have resulted in no primer extension product, or if the primers were located in too close a proximity to the transcriptional start site(s), the resulting primer extension products would have been so small that they would not have been resolved from the unincorporated labelled primers. Due to these possibilities, an alternative method was sought, whereby the position of the predicted translational start site was not such a vital issue. RNase
Figure 4.7: Relative positions of primers used in Primer extension and the potential translational start sites. The primers indicated in yellow were used to clone the sigG region into pBluescript for manual sequencing. The green primer (sigGM2) and the yellow primer (sigGMidH) were used in both the sequencing and primer extension reactions. The predicted translational start site is marked in red with an arrow indicating the direction of translation. The alternative translational start sites are outlined with a pink box, and a potential ribosome binding site gtgaga is underlined in red.
protection was chosen as it enables production of a probe of approximately 500bp used to
detect the transcriptional start site(s), which gives a larger region of the gene to use as a probe,
thus relinquishing the need to know the exact position of the translation start site.

4.2.2 RNase protection of \( \text{sigG} \)

The possibility of a discrepancy between the predicted translational start site and the actual
translational start site was taken into account in the design of the RNase protection assay for
\( \text{sigG} \). A region of 386bp was cloned into pCR4-Blunt, containing 311bp downstream of the
predicted translational start site and 75bp upstream of the translational start site. This region
was cloned between the T3 and T7 promoters of pCR4-Blunt, to be used in an \textit{in-vitro}
transcription reaction to produce a complementary radiolabelled RNA probe (see chapter 5,
figure 5.4). An outline of the RNase protection assay can be found in chapter 5 (section 5.2.1).
A number of optimisation steps were performed prior to the RNase protection assay, to
determine the optimum conditions for the reaction (outlined in section 5.2.1, figures 5.5, 5.6, 5.7
and 5.8).

The RNase protection assay was performed using 40\( \mu \)g of total RNA extracted at mid-
exponential phase (OD 0.35), from H37Rv wild-type, \( \Delta \text{recA} \) and \( \Delta \text{sigG} \) strains, under both
uninduced and induced (mitomycin C 0.02\( \mu \)g/ml) conditions. The assay in the \( \Delta \text{sigG} \) strain
indicated that \( \text{sigG} \) possesses 2 possible transcriptional start sites, indicated by arrows 1 and 2
(figure 4.8), both of which appeared to be inducible, whereas in the H37Rv and \( \Delta \text{recA} \) strains,
one of these was detected (see figure 4.8). Interestingly, the level of expression appeared to
differ between strains, yet closer examination of the autoradiograph showed a band co-
migrating with the probe control for the induced sample (track 2, figure 4.8), suggesting in this
Figure 4.8: RNase protection assay to identify the $\text{sigG}$ promoters in H37Rv, $\Delta\text{recA}$ and $\Delta\text{sigG}$ strains of $\text{M. tuberculosis}$. Assays were performed with 40μg of RNA, harvested at an optical density of 0.3, under uninduced and induced (0.02μg/ml mitomycin C) conditions for H37Rv (tracks 1 and 2), $\Delta\text{recA}$ (tracks 3 and 4) and $\Delta\text{sigG}$ (tracks 5 and 6) strains of $\text{M. tuberculosis}$. The strains are indicated below the tracks. The undigested probed was used as a positive control (track 7) and yeast RNA was used as a negative control (track 8). Samples were run alongside in-vitro transcribed markers (track M1 and M2), sizes are indicated. Two products are marked with arrows 1 and 2.
case that the RNase digestion may have been incomplete. Therefore, even though it appeared that the level of expression under induced conditions was greater in the ΔrecA strain, and greater still in the ΔsigG strain, in comparison to H37Rv, this result is not conclusive. It is also possible that this difference may be due to differential loading. However, the same RNA samples were also used in the ruvC and recA RNase protection assays (see section 5, figures 5.9 and 5.11), where weaker signals were observed for the ΔsigG strain than H37Rv, suggesting that if differential loading due to inaccurate quantitation of RNA were the case, one would have expected the level to be lower in the ΔsigG strain (see chapter 5, figures 5.9, 5.11 and 5.12). This indicates that equal quantities of RNA were most likely used in all the RNase protection assays. Identification of two potential transcriptional start sites of approximately 80bp and 120bp (see figure 4.8) indicate that the prediction for the translational start site is probably wrong. This approximate sizing obtained from the RNase protection assay, along with the clone used to produced the probe for the RNase protection assay, facilitated the performance of a new primer extension reaction, in order to identify the precise positions of the transcriptional start sites.

4.2.3 Precise identification of the sigG transcriptional start site(s)

The primer extension reactions were repeated, using the primers designed to produce the construct of sigG in pCR4-Blunt for the RNase protection assay. The clone constructed to produce the complementary RNA probe for the RNase protection assays was also used in the sequencing reactions, which were run alongside the primer extension reactions to determine the precise transcriptional start site(s) of sigG. The primer extension reactions were performed using 80μg of total RNA extracted at mid-exponential phase (OD 0.35), from H37Rv wild-type and ΔsigG strains of M. tuberculosis, under both uninduced and induced (mitomycin C
0.02µg/ml) conditions. In the H37Rv wild-type, three bands were visible, indicated with arrows 1, 2, and 3 (figure 4.9). In the ΔsigG strain, only two bands were present, indicated with arrows 1 and 2 (figure 4.9). In both H37Rv wild-type and ΔsigG strains, one of the promoters P1, was clearly DNA-damage inducible (arrow 1, figure 4.9), whilst the P2 promoter may also be induced, but to a lesser extent (arrow 2, figure 4.9). The level of transcription and induction (where applicable) remains the same for both the P1 and P2 promoters, in both H37Rv wild-type and ΔsigG strains (arrows 1 and 2 respectively, figure 4.9). Interestingly there appeared to be no expression from the P3 promoter in the ΔsigG strain, suggesting that this promoter may be SigG dependent. It is also noteworthy that the level of expression from the P3 promoter is particularly low, especially when taking into account the amount of total RNA (80µg) used in the primer extension reaction.

4.2.4 Identification and analysis of promoter motifs

Analysis of the upstream region of sigG using the manual sequencing trace and the primer extension products, made it was possible to locate and identify the three transcriptional start sites (see figure 4.10). Closer analysis of the region enabled identification of potential σ70 consensus motifs at both the −10 and −35 sites for the P1 promoter (see figure 4.10), although there are two different possibilities for the location of the −35 site for the P1 promoter. One of these is highlighted in the green box, and shows homology to the consensus for M. tuberculosis σ70 -10 TA(G/T)(A/G)aT and -35 sites TtGaCa (Gomez M, 2000), the other potential −35 site identified is underlined in green (Gamulin et al., 2004) (see figure 4.10). The motif defined by Gamlin et al., (2004) tTGTCRgtg-N₆-TAnnnT (R indicates a G/A), was suggested to define a novel RecA/LexA independent promoter. The potential −10 site for the P2 promoter is indicated in blue (figure 4.10), with homology to the σ70 consensus, however, no consensus...
Figure 4.9: Primer extension of *sigG*. Assays were performed on 80μg of RNA under uninduced (track 1 and 3) and induced (0.02μg/ml mitomycin C) (tracks 2 and 4) conditions for H37Rv (tracks 1 and 2) and Δ*sigG* (tracks 3 and 4) strains of *M. tuberculosis*, alongside a negative control (track 5). Cultures were induced at an OD of 0.3, both uninduced and induced cultures were incubated for a further 24 hours. The primer extension and sequencing reactions were carried out with the *sigG*RNaseR2 primer. Reactions were run on a denaturing polyacrylamide gel and visualised by autoradiography.
Figure 4.10: Identification of the transcriptional start sites of \textit{sigG}. The transcriptional start sites are highlighted green, blue and yellow, with a +1 indication above. The promoter consensus sequences are indicated with a box, colour corresponding to the transcriptional start site. Therefore the potential -10 and -35 consensi for P1 are indicated in green and show homology to the \textit{M. tuberculosis} $\sigma^{70}$ consensus TA(G/T)(A/G)aT (-10) and TtGaCa (-35). The green underline and $\sigma^{70}$ -35 label indicate an alternative prediction of the -35 promoter regions (Gamulin \textit{et al.}, 2004). Regions of homology are underlined in black. The $\sigma^{70}$ -10 denoted by the blue box representing P2, shows limited homology to the \textit{M. tuberculosis} $\sigma^{70}$ consensus TA(G/T)(A/G)aT (-10). The possible extended -10 is underlined in light blue. The dark blue line that extends under the $\sigma^{70}$ -10 shows homology to the region between the -10 of the P2 promoter of \textit{recA} and the transcriptional start site. The possible -10 and -35 regions of the P3 promoter are underlined in yellow. The potential ribosome binding site is underlined in red (GTGAG). The originally predicted translational start site is marked in red, and the alternative-potential translational start sites are outlined in pink.
motif was identified at the −35 site for the P2 promoter. There is an alternative possibility, that the position of the −10 sites is closer to the transcriptional start site, and therefore forms an extended −10 region (see figure 4.10, underlined in light blue).

The P3 promoter was not transcribed in the ΔsigG strain, which suggests it may be autoregulated. Therefore it was important to look at the −10 and −35 consensus regions for the previously identified alternative sigma factor recognition sites. Promoter motifs have been identified or suggested for a number of the *M. tuberculosis* alternative sigma factors: SigF (GTTC-N_{17}-GGGTAT), SigC ((G/C)(G/C)(G/C)AAT-N_{16-20}-CGT(G/C)(G/C)(G/C)), SigE (GG(A/G)(A/C)C-N_{18}-(G/C)GTTG), SigH ((G/C)GGAAC-N_{17-22}-(G/C)GTT(G/C)) (Manganelli *et al.*, 2004), and SigD (GTAACGCT) −35 sites. The −35 site of sigD was highly GC rich, but the −10 site of SigD was AT rich (Raman *et al.*, 2004), and shows less homology to the sigA consensus. A recent publication gives an altered consensus for the SigD promoter motif AGAAAG-N_{16-20}-CGTTAA (Calamita *et al.*, 2005). It appears that the consensus sequences identified for the alternative sigma factors differ somewhat from the general σ^{70} consensus of *M. tuberculosis* and appear to have a higher GC content that the σ^{70} consensus (sigA). Therefore, if sigG is autoregulated from the P3, the promoter consensus may differ somewhat from the general σ^{70} consensus of *M. tuberculosis*, which may indicate why it was not possible with only one promoter to define a −10 or −35 promoter consensus sequence. However GC rich regions were identified at both the −10 and −35 regions, underlined in yellow (see figure 4.10). It is clear that the SigG P3 promoter does not exhibit homology to any of the previously defined sigma factor consensus motifs. Without the upstream regions of other genes in the SigG regulon for comparison it is not possible to predict the promoter consensus for the SigG specific −10 and −35 regions. The microarray data presented in section 4.1 from the ΔsigG strain indicated there are a number of genes that appear to be regulated by SigG, such as
papA3 and ppsE. Primer extension studies could be performed on genes identified by the microarray analysis as part of the SigG regulon, with a view to identify their transcriptional start sites, and use this information to predict a potential consensus recognition site for SigG.

To determine if there was any homology between the P3 promoter region and the upstream region of genes thought to be in the SigG regulon, a 42bp region, upstream of and including the transcriptional start site (−41 to +1) of the P3 promoter of sigG was used in a MEME software (Bailey and Gribskov, 1998) motif search to identify any regions of homology. Sequences incorporating 150bp upstream and 100bp downstream of the predicted translational start sites for nuoB, Rv0655, Rv0312, ppsE and papA3 were assembled in MEME, along with the 42bp region of the P3 promoter of sigG. A motif search was performed to identify a maximum number of 3 motifs, with the motif width limited to 4-11bp. For nuoB and Rv0655, both motif 1 and 2 were identified in relatively close proximity, in the correct orientation, whereas for Rv0312, both motifs were identified, but in the incorrect orientation (see figure 4.11). It is possible however, that one of these is correct. For ppsE, the motif 2 was identified upstream too far of the motif 1, however problems have been encountered with the predictions of the translational start sites as mentioned. For papA3, none of the motif searches revealed a possible homologous motif to the P3 promoter region of sigG. However, there appear to be some similarities, as indicated with the red boxes.

Interestingly further analysis using part of the predicted motif from MEME revealed that a potential SigG consensus of CGACC(W)(t/c)C-N_{13-22}TGTCCG, where w is a/t/g, was present within 100bp of the predicted translational start sites, see below:
Figure 4.11: MEME motif search using the P3 promoter region of sigG. Motif 1 (yellow) represents a possible -10 region, and motif 2 (pink) represents a possible -35 site. The predicted translational start sites are highlighted with a box, and other potential -10/-35 sites are highlighted with a red box. Homology between part of the -35 motif and the MEME consensus revealed an alternative for the -35 site consensus CGACC, boxed in orange, and the corresponding -10 sites are boxed in light blue.
Chapter 4 – Regulation of $\sigma$G and identification of its regulon

Table 4.7: Alignment of the possible consensus sequences for $\sigma$G and potentially SigG regulated genes.

There appears to be a greater degree of similarity at the –35 sites rather than the –10 site, which has been observed before with other ECF family sigma factor (Raman et al., 2004). Interestingly, $nuoB$, $Rv0655$ and $Rv0312$ show a high degree of homology at both the –10 and –35 sites. The SigG consensus for $papA3$ was 71bp downstream of the predicted translational start site and 37bp downstream of the predicted translational start site of $nuoB$ (see figure 4.11). Conformation of the consensus would require either RNase protection or primer extension to identify the transcriptional start sites of the aforementioned genes, which could then be used in conjunction with the microarray findings to search for other genes regulated by SigG using the defined consensus.

165
An additional interesting feature was identified within the promoter region for \( \text{sigG} \), upstream of the second transcriptional start site (figure 4.10, underlined in dark blue). The motif GTGGTAGTCATT is also found with minor variation (GTGGTAGTCATT) upstream of the \(-10\) site of the P2 promoter of \( \text{recA} \) (see chapter 5, figure 5.2c). The consensus sequence GTGGTAGTCATT was used in the DNA pattern search tool, available on the Tuberculist website, to determine whether this motif occurred upstream of the predicted translational start sites of other genes. The search was limited to allow 1bp mismatch in the first part of the consensus, and 1bp mismatch in the second part of the consensus, and a filter was applied to restrict the search to sequences between 150bp before and 100bp after the predicted translational start sites. These parameters were chosen rather than simply restricting the search to the upstream regions of the predicted translational start sites due to potential problems with mis-identification of the translational start sites as observed with \( \text{sigG} \). Only the two genes \( \text{sigG} \) and \( \text{recA} \), used to define the consensus were identified with no mismatches, while 24 genes were identified with 1 mismatch, and 207 genes were identified with 2 mismatches (see appendix III).

None of the 24 genes with the 1bp mismatch were significantly different in the \( \Delta \text{sigG} \) microarray data compared to H37Rv. However, 4 of the genes significantly different in uninduced \( \Delta \text{sigG} \) compared to H37Rv were present in the list of genes with 2bp mismatch to the consensus: Rv1232c, Rv2616, Rv2743c and \( \text{ald} \) (see figure 4.12 and table 4.7). The expression level of Rv1232c, \( \text{ald} \) and Rv2616 is decreased in \( \Delta \text{sigG} \) strain under uninduced conditions, whereas the expression level of Rv2743c is increased under uninduced conditions with \( p<0.01 \) (Student t-test, assume equal variance – see section 4.3.2). Curiously, Rv1232c appears to be induced in \( \Delta \text{sigG} \) strain, but not in H37Rv; however, this difference is not significant at \( p<0.01 \) between the strains.
Figure 4.12: A comparison between genes identified with conserved upstream regions which are significantly different in ΔsigG strain and H37Rv. These regions of homology GTGGTn2TCATT were defined by Gamulin et al., (2004) as the RecA-NDP or RecA non dependent promoter, These regions were identified upstream of 4 genes which were significantly different in the ΔsigG strain under uninduced conditions. The fold change was calculated as H37Rv as a proportion of ΔsigG strain.
Table 4.8

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>Pattern sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>sigG</td>
<td>(+) 42</td>
<td>cagtcggGTGGTagTCATTacccgga</td>
</tr>
<tr>
<td>recA</td>
<td>(-) 102</td>
<td>cggctactGTGGTgaTCATTcggagca</td>
</tr>
<tr>
<td>mismatch 1bp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>end</td>
<td>(+) 3</td>
<td>agacttggGTGTgcTCATTggttcgc</td>
</tr>
<tr>
<td>mismatch 2bp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ald</td>
<td>(-) 9</td>
<td>tatcgagaGGGGTagTCATGCgcgtcg</td>
</tr>
<tr>
<td>dnaE1</td>
<td>(+) 17</td>
<td>tcatctcGGGGTGccTCTTTtgcgcac</td>
</tr>
<tr>
<td>Rv1765c</td>
<td>(+) 27</td>
<td>ctggcgcagaGGGGTagTCAGTcctgcac</td>
</tr>
<tr>
<td>Rv1232c</td>
<td>(+) 38</td>
<td>gccctgcccGTGGCGcGcATTgtgattg</td>
</tr>
<tr>
<td>Rv2015c</td>
<td>(+) 27</td>
<td>ctggcgcagaGGGGTagTCAGTcctgcac</td>
</tr>
<tr>
<td>Rv2616c</td>
<td>(+) 0</td>
<td>ctggcagGGGGAccTCAAGtcgcttg</td>
</tr>
<tr>
<td>Rv2743c</td>
<td>(+) 29</td>
<td>cgccggcGGGGTgTCGGTtgctgcag</td>
</tr>
<tr>
<td>Rv2792c</td>
<td>(-) 34</td>
<td>cgatctttgcTGGGTctTCATGcgcgtcg</td>
</tr>
<tr>
<td>uvrB</td>
<td>(+) 21</td>
<td>gtcacttgcAGGTgTCAGTcgcgcac</td>
</tr>
<tr>
<td>uvrD2</td>
<td>(+) 134</td>
<td>tgagcgcagaGGGGTagTCATGaCacgc</td>
</tr>
</tbody>
</table>

Table 4.8: Interesting genes with the sigG/recA consensus in close proximity to predicted translational start sites. Genes identified with the GTGGTN2 TCATT consensus within close proximity to predicted translational start sites. These genes were either significantly different in the ΔsigG strain compared to H37Rv in microarray analysis, or were identified as possessing a RecA-NDp as outlined in Gamulin et al., (2004), or are involved in DNA damage repair.
The list of 47 genes identified with the RecA-NDp consensus (Gamulin et al., 2004), were analysed for the presence of the potential conserved region, GTGGT-n2-TCATT, with 1 or 2 mismatches. A total of 6 genes were identified that possessed both the RecA-NDp and the conserved region: *dnaE1*, *uvrB*, *uvrD2*, *Rv1765c*, *Rv2015c* and *Rv2792c*, with the conserved region located within a maximum of 34bp proximity of the predicted translational start sites, except for *uvrD2*, located 134bp upstream of the translational start site (see table 4.7). One other gene, *end*, involved in DNA repair, also possessed the conserved region (see table 4.7), in this case, with only a single mismatch. Thus, it is possible that this motif plays a role in the regulation of expression of a subset of genes following DNA-damage.

### 4.3 Discussion

#### 4.3.1 Analysis of the *sigG* regulon

The initial hypothesis was that SigG was involved in regulation of genes involved in DNA-damage repair. This hypothesis was based on viability data (see chapter 3), which indicated that the Δ*sigG* strain was significantly more susceptible to the chemical DNA damaging agent mitomycin C, than the H37Rv wild-type strain. Previous microarray analysis has also revealed that *sigG* was the most highly induced sigma factor following DNA-damage (Rand *et al.*, 2003). Therefore, in order to determine whether SigG was responsible for regulating genes involved in DNA-damage repair, the induction ratios in the Δ*sigG* strain were compared with those in the wild-type strain. Of the 115 genes that were induced 2-fold or more in the wild-type, only 4 genes exhibited reduced induction ratios in the Δ*sigG* strain. However, closer analysis of the data revealed that discrimination by induction ratio was possibly not the best method of examining the data. The use of RNA vs DNA hybridizations in the microarray experiments
meant it was possible to determine whether there were any differences in expression under uninduced and induced conditions individually for the ΔsigG and H37Rv strains. This revealed that the apparent decrease in induction ratio was affected by an increase in the expression under uninduced conditions in the ΔsigG strain. This is highlighted with Rv1956, a predicted transcriptional regulatory protein. The induction ratio is 3.09 ± 0.37 in H37Rv, and only 1.59 ± 0.28 in ΔsigG, these are significantly different with a p=0.0093. Although, when one looks at the normalised data for the uninduced and the induced conditions individually, there appears to be little difference between the induced expression level in ΔsigG strain compared to H37Rv; rather the expression level of Rv1956 appears to be higher in the ΔsigG than in H37Rv under uninduced conditions, however, this difference was not significantly different at P<0.01. This disparity between the uninduced and induced comparisons compared to the induction ratio becomes particularly apparent when looking at genes whose expression is almost completely ablated, of which, papA3 is a prime example. Using induction ratio, it appears that papA3 expression is induced to a greater extent in the ΔsigG strain than in H37Rv; however, when looking at the uninduced and induced values separately, it became evident that the expression level of papA3 under both uninduced and induced conditions is greatly reduced in ΔsigG compared to H37Rv. This difference in induction ratio could be the result of dividing by a small value; any ratio whereby a very small denominator is used results in a large ratio, thus skewing the data. This observation is particularly important when looking for genes whose expression in dramatically reduced in a mutant. Examining the induction ratio gives the impression that the induction levels of lexA, fmt, and Rv1956 are all decreased in ΔsigG compared to H37Rv; nevertheless, when dissecting the expression level under uninduced and induced conditions separately, it becomes clear that although expression varies slightly under uninduced and induced conditions between the strains, this difference is not significant.
Therefore, in spite of the induction ratio suggesting that *fmt*, Rv1956 and *lexA* are regulated by SigG, the individual evidence suggests this is unlikely to be the case.

This highlights the potential difficulties in analysing microarray data in determining which comparisons are meaningful. From the set of 17 genes with a significantly different induction ratio in the Δ*sigG* strain compared to H37Rv, only Rv2884 was also significantly different in the uninduced versus uninduced comparisons, and Rv0181c, *sigG* and Rv3467 were significantly different in the induced versus induced comparisons. Interestingly, *sigG* is co-transcribed with Rv0181c and Rv0180c (see chapter 3); however, only *sigG* and Rv0181c appear to be induced greater than 2-fold in response to DNA-damage in H37Rv wild-type, while Rv0180c is induced to a lesser extent (1.7-fold) in H37Rv. Even though *sigG*, Rv0181c and Rv0180c are all significantly different under induced conditions at P<0.01, none of the 17 genes with significantly different induction ratios were significantly different under both uninduced and induced conditions P<0.01. Therefore, it would appear that SigG does not play a significant role in the control of gene expression following DNA-damage.

The differential expression observed in Rv0181c, one of the downstream co-transcribed genes of the *sigG* operon could show decreased expression in Δ*sigG* compared to H37Rv, due to the polar effects of the SigG deletion and insertion (hyg resistance cassette). The hygromycin resistance cassette may affect the stability of the mRNA, thus making it more prone to degradation.

The problems observed with examination of the data using the induction ratio lead to a different focus of analysis, whereby the uninduced samples were directly compared between strains, then the induced samples were also directly compared between strains, to determine whether the
expression of any genes were significantly different in the ΔsigG strain compared to H37Rv. This revealed that 7 genes were significantly different under both uninduced and induced conditions in the ΔsigG strain compared to H37Rv. These included genes involved in lipid metabolism (papA3, ppsE, ppsD), respiration (nuoB), transport across the membrane (Rv0655), a probable resuscitation promoting factor (rpfE/Rv2450c) and a hypothetical protein (Rv0312). Interestingly both papA3 and Rv0312 were predicted to be essential by transposon mutagenesis (Sassetti et al., 2003), suggesting that the low levels of expression observed in ΔsigG compared to H37Rv are sufficient for the knockout to remain viable. This residual level of expression may be explained by compensation by an alternative sigma factor, which may be able to initiate sub-optimal levels of transcription from papA3 and Rv0312, in the absence of SigG. The residual level of expression of Rv0655 in the ΔsigG strain is higher than that of the other 4 genes whose expression is decreased in the ΔsigG strain compared to H37Rv. This may be because these genes are expressed from multiple promoters, one or more of which is recognised by a different sigma factor, conversely, it may be that more than one sigma factor can recognise a single promoter. An alternative possibility is that SigG may regulate a repressor of Rv0655.

There appear to be conflicting results regarding the expression of genes in the pps operon. The expression level of ppsE is all but abolished in the ΔsigG strain, while the upstream, cotranscribed gene ppsD appears to be upregulated in ΔsigG compared to H37Rv. In previous dye-swap RNA vs RNA M. tuberculosis arrays, problems have been observed with incorporation of the fluorescent labels cy3/cy5 with ppsE, resulting in an apparent decrease in expression, when in reality this decrease was due to differential labelling with cy3/cy5 (Roger Buxton- personal communication). However, the design of the DNA vs RNA arrays discounts any preferential labelling with either cy3/cy5, by using DNA as the normalising control, always labelled with cy3. Therefore it is unlikely that the almost complete abolition of expression of
ppsE in ΔsigG occurred as a result of preferential labelling with cy3/cy5. The validity of the data values for ppsE are even more convincing taking into account that ppsE was also significantly different in the ΔsigG strain compared to H37Rv, under the increased stringency of the Benjamini and Hockberg false discovery rate correction. Nevertheless it is surprising, that the level of ppsD should be increased and the level of ppsE decreased, given that they have been demonstrated experimentally to be co-transcribed (Camacho et al., 2001). However, these are huge genes; ppsE is 4.467kb and ppsD is 5.484kb. It is possible therefore, that either the microarray probes are not specific for the genes, due to their size, or the RT-PCR may be flawed. Closer analysis of the predicted co-transcribed region revealed that not only ppsA, ppsB, ppsC, ppsD and ppsE were predicted to form an operon, based on RT-PCR data between adjacent pairs of genes (Camacho et al., 2001), but that fadD26 upstream of the pps operon was shown to also be co-transcribed with ppsA, as were ppsE and drrA, therefore suggesting that the predicted polycistronic region includes; fadD26, ppsA-E, drrA-C and papA5. The RT-PCR results visually do not seem conclusive, yet they are reported to have been sequenced. However, it seems unlikely that an operon as large as 32.4kb could be transcribed from a single promoter. It may, therefore, be possible that ppsE is transcribed independently from ppsD. It could be that ppsE is transcribed from an alternative promoter located in the coding region of ppsD, which is regulated directly or indirectly by SigG.

Potentially, the most interesting set of genes are those that are expressed to a lesser extent in ΔsigG strain, as these may be directly or indirectly regulated by SigG. Examination of the data obtained in the absence of DNA-damage, revealed that 52 genes showed reduced expression in the ΔsigG strain. These include 6 genes involved in lipid metabolism and 8 genes in cell wall and cellular processes. The reduced expression of these genes may account for the increased susceptibility of M. tuberculosis ΔsigG strain to mitomycin C. Lipids comprise a large
proportion of the cell wall of *M. tuberculosis*, which provides a protective layer to help defend the bacterium against the hostile external environment and the components of this barrier may be altered by as a consequence of the changes in gene expression observed. In addition, the cell wall and cell processes genes downregulated in Δ*sigG* strain include two genes involved in membrane transport: Rv1348 is thought to be an ABC transporter most likely involved in drug transport and *epfA* is an integral membrane protein, involved in efflux. Therefore, a decrease in the expression of membrane transporters may increase the levels of damaging agents in the cell, or may prevent agents such as mitomycin C being removed from the cell.

Among the genes with decreased expression in Δ*sigG* strain are four transcriptional regulators, Rv3050c, Rv0232, which are annotated to be AsnC family and TetR family respectively, Rv3764c a sensor of a two component regulatory system and Rv2884 which is predicted to contain a response regulator receiver domain. TetR family of transcriptional repressors respond to tetracycline, whereby in the absence of tetracycline, the TetR repressor binds to the operator and suppresses transcription (Yan *et al.*, 2001). The AsnC family, often referred to as the Lrp/AsnC family of transcriptional regulators respond to leucine/asparagine synthase C (Shrivastava *et al.*, 2004). It is possible, that *sigG* could regulate the genes described directly or indirectly via any one of the four transcriptional regulators identified in the *sigG* regulon.

Another interesting gene shown to decrease in expression level in Δ*sigG* strain, is *mcelA*, which is involved in host cell invasion and is thought to be involved in entry and survival in macrophages. This may suggest why preliminary mouse *in-vivo* data has indicated a decrease in virulence of Δ*sigG* strain compared to H37Rv (see chapter 3).
Although the main focus has so far been on genes whose expression is deceased in the ΔsigG strain, one gene in particular was of interest that showed increased expression in ΔsigG strain, the stress response sigma factor sigE. The increased expression of sigE, might be expected to result in increased expression of genes in the SigE regulon, although there was no overlap in the set of genes showing increased expression in uninduced ΔsigG strain here, and those reported to be reduced in expression in exponential growth in ΔsigE strain (Manganelli et al., 2001). Alternatively, one could speculate that SigE may be able to partially compensate for the lack of SigG. Interestingly two of the genes whose expression is increased in ΔsigG strain to greater than 1.7 fold are involved in lipid metabolism, fadD26 and ppsD with ppsA increased to a lesser extent in ΔsigG strain. These are of particular interest as they have been suggested to be co-transcribed as part of the pps gene cluster (Camacho et al., 2001). Another interesting genes is the resuscitation promotion factor rpfE whose expression was increased under both uninduced and induced conditions in the ΔsigG strain, suggesting SigG may directly or indirectly regulate, rpfE, possibly via a repressor, which potentially may be emergence from dormancy.

To confirm the possibility that the genes identified by microarray analysis did belong to the SigG regulon, the expression of these genes could be validated using Taqman real-time PCR in both the ΔsigG strain and H37Rv wild-type. The upstream regions of these genes could then be used to identify a consensus motif for the recognition site of SigG, initially by alignment, then using primer extension assay to determine the precise transcriptional start site(s). Since the level of sigG expression decreased, but was not abolished in the ΔsigG strain, it is possible that sigG may be subject to autoregulation. However, due to the detection of expression of sigG in the ΔsigG strain it is likely that sigG is transcribed from multiple promoters, one of which may be subject to autoregulation.
4.3.2 Dissection of the upstream region of sigG

Identification of the transcriptional start sites of sigG proved difficult initially due to the incorrect prediction of the translational start site, in the TubercuList database. Both primer extension and RNase protection have revealed that the major transcriptional start sites are in fact downstream of the originally predicted translational start site, indicating that the actual translational start site is further downstream than initially thought.

Primer extension is a useful assay for the precise identification of the transcriptional start site(s) of a gene to the nucleotide, due to the ability to run a manual sequence alongside the primer extension reactions using the same primer, thus enabling a direct comparison between the band(s) produced in the primer extension reactions and the manual sequence.

The initial primer extension reactions performed on sigG were unsuccessful, although the positive control yielded a band, indicating the technique was working. A number of possibilities were checked out to determine the cause of the primer extension reaction failing for sigG. The primer used in the sigG primer extension worked in the sequencing reactions, so it was not that the primer was not annealing to the template (RNA). It has been suggested by Manganelli (1999) that the level of sigG expression is very low during exponential phase, therefore, it was thought that more template RNA was required for the primer extension, but, even after increasing the template to 85µg, no product was observed.

A closer look at the sigG translational start site revealed there were a number of other possible translational start sites further into the coding region of sigG. Therefore, another technique was employed, whereby a 500bp RNA probe was used in an RNase protection assay. This assay is
more quantitative than the primer extension assay, nevertheless, it does not allow such accurate location of the transcriptional start site. Markers used alongside the RNase protection assay enable the approximate sizing, but the size range of the makers is such that only an approximation can be made, thus necessitating a subsequent primer extension assay to define the transcriptional start sites to the nucleotide. This meant that both assays were used sequentially to firstly identify if the transcriptional start site(s) were in fact downstream of the predicted translational start site, then secondly to identify precisely the transcriptional start sites based on this information.

The primer extension reaction showed the presence of two inducible promoters in both the ΔsigG strain and H37Rv strain, as observed in the RNase protection assay. In addition, there appeared to be a constitutively expressed promoter in the H37Rv strain, that was potentially autoregulated, as indicated by the absence of a corresponding band in the ΔsigG strain in the primer extension assay. Intriguingly, the finding from the primer extension and RNase protection do not tie up with the microarray data, in which it appeared that sigG was not induced in response to DNA-damage in the ΔsigG strain, yet the primer extension and RNase protection assays clearly show that sigG possesses two DNA-damage inducible promoters, regulated independently of SigG, due to their induction in ΔsigG strain.

As expression of sigG from the P1 and P2 promoters increased following DNA-damage, if the P3 promoter is autoregulated, one would have expected its level of expression to also increase. However, the RNase protection and primer extension showed that sigG possesses three promoters, two of which are DNA-damage inducible, and a third which does not appear to be DNA-damage inducible, but may potentially be autoregulated. Therefore, the data indicates that SigG is not required for the induction of expression from the two inducible promoters,
which begs the question, why was the level of sigG transcription decreased in the knockout. The answer may lie in the position of the sigG probe for the microarray analysis. The probe appears to overlap with the deletion of the sigG coding region, but also overlaps by approximately 48bp with the remaining region of sigG, therefore the binding of the probe to the sigG message may be transient, or unstable. However, it is perhaps more likely that the decreased induction of sigG in the AsigG strain, is due to the decreased stability of the transcript brought about by the insertion of the hygromycin resistance gene, which may have made the mRNA more prone to digestion with RNase.

Analysis of the upstream region of sigG revealed that there were a number of potential promoter recognition sites upstream of the three transcriptional start sites identified by the primer extension. Interestingly, the P1 promoter was the most highly induced, and showed the greatest homology to a σ70 consensus for the −10 and −35 regions. However, there appear to be two potential sites of the −35 region, that show a similar degree of homology to the σ70 consensus. One in particular was of interest, as it was identified by Gamulin et al., (2004) and is similar to a promoter motif upstream of DNA repair genes, such as recA, ruvC and uvrA/B, along with some mobile genetic elements, such as Rv1148c. This motif, termed the RecA-NDp (RecA non-dependent promoter) has been suggested to be involved in an alternative mechanism of regulation, in a RecA/LexA independent fashion, thought to be by sigma factor specificity (Gamulin et al., 2004).

There are two main hypotheses, which have been put forward to explain how the RecA-NDp is regulated. The first hypothesis is that an undefined sigma factor is responsible for regulating transcription from this motif, and the other is that this motif is recognised by a major sigma factor, through interactions of the −35 site with a positive regulatory protein (activator),
whereby the activator protein modifies the preference of the sigma factor for the -35 region (Gamulin et al., 2004).

The RecA-NDp motif was identified upstream of two ECF family sigma factors, sigG and sigH. Many ECF sigma factors have been shown to be autoregulated, suggesting that this motif may be regulated by either SigG or SigH. RNase protection experiments performed in chapter 5 demonstrate that SigG is not responsible for the expression from the recA P1 promoter. The primer extension data for sigG shows that the P1 promoter, which contains the aforementioned motif, is induced in response to DNA-damage in the ΔsigG strain, thus indicating that SigG is not involved in regulation of gene expression from the P1 promoter. Microarray data for the ΔsigG strain revealed that SigG does not regulate expression from any of the genes with this identified motif, thus suggesting that SigG does not recognise promoters with the previously described RecA-NDp motif. Therefore, current literature was surveyed for the possibility that SigH regulates expression of DNA-damage inducible genes possessing this upstream motif. Analysis of a sigH knockout strain of M. tuberculosis revealed that it was more susceptible to heat shock and oxidative stress (diamide) than H37Rv. Furthermore, transcriptional profiling using microarray technology revealed that SigH directly or indirectly regulated 39 genes, some of which were involved in heat shock and others were predicted to be involved in thiol metabolism. Interestingly two other sigma factors showed decreased induction (5mM diamide) in the ΔsigH strain, SigE and SigB (Manganelli et al., 2002). It was later hypothesised that SigH was a central regulator in the response to oxidative stress and heat shock in M. tuberculosis, and that SigH was responsible for regulation of these two sigma factors, sigE and sigB (Song et al., 2003). Although SigH is involved in the response to oxidative stress, it appears that RecA-NDp genes, some of which fall under the DNA-damage repair system, are not regulated directly by SigH, or indirectly by SigE or SigB based on currently available
microarray data describing the regulons of these sigma factors. Microarray analysis of ΔsigE strain identified genes regulated by SigE under untreated conditions during exponential phase, and after SDS treatment. This showed that sigB was not upregulated during exponential growth in the ΔsigE strain, thus indicating sigB is part of the SigE regulon (Manganelli et al., 2001). It has been shown that sigE is regulated by SigH and sigB is dependent on both SigE and SigH for both basal and inducible expression. The analysis of these mutants lead to the identification of possible −35 and −10 recognition sites for these sigma factors, none of which match the motif consensus defined by Gamulin et al., (2004). Interestingly the oxidative stress response genes sodA and sodC were induced to a lesser extent in the ΔsigH strain, but no SigH promoter consensus was observed upstream of sodA, sodC or ahpC (Raman et al., 2001), suggesting that sigH may indirectly affect expression of these genes, or may be able to recognise more than one promoter motif. Although the RecA-NDp motif was identified upstream of sigG and sigH it does not appear that these motifs are recognised by either SigG or SigH.

Another noteworthy observation was made during the analysis of the upstream regions of sigG (figure 5.4), and recA (see chapter 5, figure 5.2c): There is a conserved motif GTGGT-N$_2$-TCATT (SigRec motif), which overlaps with the P2 promoter of sigG and is located downstream of the −10 site of the distal P2 promoter of recA. This conserved region may act as a binding site for a repressor or activator protein, which may be involved in regulation of transcription from these promoters. The close proximity of the RecA-NDp motif and SigRec motif, to the P2 promoter of recA and the P1 promoter of sigG, suggests they may be involved in regulation. However, it was shown by Gopaul et al., (2003) that the region containing the SigRec motif was not essential for induction of the RecA P1 promoter. Therefore, SigRec motif could represent an activator binding site, the presence of which, merely enhances transcription rather than being a pre-requisite for transcription. As mentioned in the results, 24 genes were
identified with 1bp mismatch to the consensus outlined above, and 207 genes were identified with 2 mismatches. None of the 24 genes with the 1bp mismatch were uncovered in the ΔsigG microarray data as significantly different in the ΔsigG strain compared to H37Rv. Nevertheless, a total of 5 genes were identified that possessed both the RecA-NDp and the conserved region, comprised of the potential DNA repair genes; *uvrB* helicase, subunit B, *uvrD2* ATP dependent DNA helicase II, Rv2792c a possible resolvase, *dnaE* a DNA polymerase III, and Rv1765c and Rv2015c, conserved hypothetical proteins. In addition, one DNA repair gene without a RecA-NDp was identified as possessing the conserved region, *end*, an endonuclease involved in base excision repair. It does not appear that the SigRec motif identified upstream of *recA* and *sigG* is required for regulation of DNA-damage repair genes that are induced independently of the RecA/LexA system, nor does it appear that the RecA-NDp is solely responsible for regulation of these genes, as the motif is not present in all cases. Therefore, it appears that there is a complex network of interconnected regulation of genes responding to DNA-damage, including those predicted to be dual regulated.

Although sigma factors often autoregulate, the primer extension and RNase protection assays performed on the ΔsigG strain have revealed they may possess more than one promoter and that only one of the promoters may in fact be autoregulated, as is the case of *sigG*. In the case of *sigG*, this could be tested using lacZ-promoter fusion assays, whereby the three promoters of *sigG* could be cloned upstream of a lacZ reporter gene, and various mutations could be introduced to define the regions of importance in the potential -10 and -35 promoter sites. These could then be assayed in both the H37Rv wild-type (as a control) and ΔsigG strain, to address the hypothesis that the P3 promoter is regulated by SigG and therefore would not be expressed in the ΔsigG strain. If SigG regulated expression from the P3 promoter by recognising and binding to the motif outlined above, then there would be either abolition of
expression or decreased expression (due to compensation) in the $\Delta$sig$G$ strain compared to H37Rv.

The SigG dependent promoter, identified upstream of sig$G$, P3, reveals potential -10 and -35 motifs that do not resemble any of the previously described sigma factor motifs. Motif searches using MEME database (Bailey and Gribskov, 1998), with the 5 genes though to be regulated by SigG, under both induced and uninduced conditions did not reveal any concrete consensus for the consensus motif of SigG, but when the smaller -35 region CGACC (from the MEME search) was used to search the sequences, potential -10 and -35 sites were identified upstream of the predicted translational start sites for all of the genes except nuo$B$ and papA3, where the consensus was identified downstream of the predicted translational start site. This revealed the potential consensus CGACC(R)(t/c)C-N$_{13-22}$-TGTCCG (R represents g/c/a). There was less homology observed at the -10 site than the -35 site for all genes, except nuo$B$ which was identical to the P3 promoter of sig$G$ at the -10 sites. Therefore, the best course of action would be to perform primer extension assays on the 5 genes predicted to be downregulated in $\Delta$sig$G$ strain under both induced and uninduced conditions, to identify the transcriptional start site(s), thus enabling an alignment to be produced to define a consensus for the recognition motif of SigG. This could then be tested using the wild-type sequence and mutated sequences upstream of a lac$Z$ reporter gene to determine which of the residues are vital from the recognition of the sigma factor SigG.
5 Detailed analysis of *ruvC*

The RecA/LexA system coordinating expression of the DNA-damage regulon has been extensively studied in *E. coli*, whereby the majority of genes induced following DNA-damage appear to be regulated by the RecA/LexA system (Walker 1984). The same appears to hold true for a number of different species (Eisen and Hanawalt, 1999). However, the DNA-damage regulon is not controlled by the RecA/LexA system in all bacterial species. A number of bacteria whose annotated genome sequence is complete lack a *lexA* homologue, including; *Aquifex aeolicus*, *Borrelia burgdorferi*, *Campylobacter jejuni*, *Chlamidia pneumoniae*, *Helicobacter pylori*, *Mycoplasma pneumoniae* and *Porphyromonas gingivalis* (Fernandez de Henestrosa *et al.*, 2002), and even those that possess *lexA/recA* homologues, such as the plant pathogen *Xylella fastidiosa*, show independent induction of some genes involved in the DNA-damage regulon (Campoy *et al.*, 2002).

Analysis of the DNA-damage repair regulon in H37Rv wild-type compared to \(\Delta recA\) strain of *M. tuberculosis* showed that although a proportion of DNA-damage repair genes were regulated by the classical RecA/LexA system, some DNA-damage inducible genes were regulated independently of the RecA/LexA system. In H37Rv wild-type, 112 genes were induced greater than or equal to 3 fold in the wild-type in response to DNA-damage (Rand *et al.*, 2003). Of those 112, 21 genes were not expressed in the \(\Delta recA\) strain, suggesting they are solely regulated by RecA/LexA system. However, 28 genes were expressed less in the \(\Delta recA\) strain, suggesting these genes were partially regulated by RecA/LexA system, and partially by an alternative mechanism. Interestingly 50 genes were expressed to the same extent in the \(\Delta recA\) strain, suggesting these genes are solely regulated by an alternative mechanism (Rand *et al.*, 2003). Some of the genes shown to be partially induced in the \(\Delta recA\) strain have been shown to possess
5.1 Identification of \textit{ruvC} transcriptional start site(s) by primer extension

Preliminary RNA versus RNA microarray data from the \textit{\Delta sigG} strain of \textit{M. tuberculosis} compared to the wild-type H37Rv, had suggested a significant decrease (at the 2\% level) in induction of \textit{ruvC} in the \textit{\Delta sigG} strain compared to the wild-type in response to DNA-damage (mitomycin C 0.2\mu g/ml) (data not shown). Therefore, the expression level of \textit{ruvC} appeared to be decreased in \textit{\Delta recA} and \textit{\Delta sigG} strains, which suggests that \textit{ruvC} may be dual regulated by RecA/LexA and SigG. One could hypothesise, that this dual regulation takes place via two promoters, one regulated by LexA/RecA and the other by SigG. It has been demonstrated that \textit{ruvC} has a DNA-damage inducible promoter, which contains a palindromic LexA binding site (Brooks \textit{et al.}, 2001). Consequently, to test the theory that two promoters drive \textit{ruvC}
expression and to investigate their dependence on RecA and SigG, RNA was extracted from induced (mitomycin C, final concentration 0.02μg/ml) and uninduced cultures of wild-type H37Rv, ΔrecA, and ΔsigG strains of M. tuberculosis. The RNA samples were quantified using a spectrophotometer, and primer extension reactions were performed using a primer designed from within the coding region of ruvC (see figure 5.1a). The primer extension reactions were run alongside manual sequencing reactions, where a clone containing the upstream region and coding region of ruvC in pBluescript (see table 2.2) was used as a template. The same primer was used in the primer extension and sequencing reactions so one could determine the exact transcriptional start site. However, the sequencing reaction failed and sizing could not be performed. In the uninduced isolates of H37Rv wild-type, ΔrecA, and ΔsigG strains, two bands were visible (see figure 5.1b), suggesting that ruvC possesses two transcriptional start sites. The P1 promoter (indicated with arrow A, figure 5.1b) possesses a LexA binding site (Brooks et al., 2001), and was clearly DNA-damage inducible in the H37Rv wild-type. However, this induction was not apparent with the P2 promoter (indicated with arrow B, figure 5.1b). Nevertheless, the P2 promoter may be DNA-damage inducible, as the P2 promoter of recA did not appear to be induced by mitomycin C when samples were used in primer extension reactions (Movahedzadeh et al., 1997), although induction of the P2 promoter was observed using a transcriptional fusion to lacZ (Davis et al., 2002b). The apparent lack of induction of the P2 promoter of ruvC in the primer extension reactions may have been a result of poor levels of quantitation available by primer extension. Surprisingly, in the ΔrecA strain, the P1 promoter (labelled A, figure 5.1b), regulated by LexA, remained partially inducible; suggesting the repression of ruvC by LexA may be partially overcome by an alternative mode of regulation. In the ΔsigG strain, the promoter P1 was induced, but to a lesser extent than in the H37Rv wild-type and ΔrecA strain. In contrast, the P2 promoter was not induced in any of the three strains
Figure 5.1a: A schematic representation of the position of \textit{ruvC} and potential transcriptional start sites. The arrow below \textit{ruvC} indicates the primer designed in the coding region of \textit{ruvC} for the primer extension reaction to detect the transcriptional start sites, indicated by A and B. The P1 promoter contains a putative LexA binding site, defined in red and labelled the SOS box.

Figure 5.1b: An autoradiograph showing the transcriptional start sites of \textit{ruvC} in \textit{H37Rv} wild type, \textit{\textDelta recA} and \textit{\textDelta sigG} strains of \textit{M. tuberculosis}. Two cultures of each strain were grown to mid exponential phase (0.3-0.4) in a rolling incubator at 37°C. These were pooled and split, one remained untreated, while the other was induced with 0.02\,\mu g/ml mitomycin C, all cultures were incubated for a further 24 hours in a rolling incubator at 37°C. RNA was then extracted from the cultures, and primer extension reactions were carried out with 100\,\mu g of RNA from uninduced (track U), induced (track I), and a negative control with no RNA (track -). The strains of \textit{M. tuberculosis} used in the primer extension are indicated below the autoradiograph.
tested and the level of transcription from the P2 promoter appeared to be less under induced rather than uninduced conditions (see figure 5.1b).

The original primer extension (figure 5.1) had indicated that *ruvC* possesses two transcriptional start sites; however, technical problems encountered with the manual sequencing run along side the primer extension meant the exact transcriptional start sites were not identified. The sequencing reactions were repeated using an increased quantity of template. Half of the sequencing reactions were run on a polyacrylamide gel and visualised by autoradiography (data not included). The remaining halves were stored at −20°C overnight before being run alongside the primer extension reactions. The same primer used in the sequencing reactions was used for the primer extension reactions. The primer extension was repeated using uninduced and induced (0.02μg/ml mitomycin C) H37Rv wild-type RNA, from a duplicate culture, alongside a sequencing reaction (see figure 5.2a), enabling precise positioning of the transcriptional start sites (see figure 5.2b). This revealed that the LexA binding site overlaps the P1 transcriptional start site, suggesting that LexA binding would interfere with transcription initiation (see figure 5.2b).

### 5.1.1 Potential promoter motifs upstream region of *ruvC*

Primer extension analysis showed that *ruvC* is expressed from two promoters (see figure 5.2a). Analysis of the upstream sequence revealed that the P1 promoter exhibits some similarity to sigma 70 (σ70) −10 and −35 promoter elements (see figure 5.2b), with the *E. coli* consensus TATAAT and TTGACA respectively (Gross *et al.*, 1992). Interestingly there is a high degree of similarity between the P1 promoter regions of *recA* and *ruvC*: the −10 region of the P1 promoter of *ruvC* (pink box, figure 5.2b) CTAGcGT exhibits similarity to the −10 region
Figure 5.2a: Primer extension of *ruvC*. Primer extension reactions were carried out using 80 μg of RNA from H37Rv uninduced and induced (0.02 μg/ml mitomycin C) cultures. These were run alongside manual sequencing reactions.

Figure 5.2b: Transcriptional start sites of *ruvC*. The first transcriptional start site (+1) indicated in yellow, and the second is indicated in red (+1). The blue box and line indicate the position of the LexA binding site, with the consensus TCGAAC, GTTCGA. The green highlighted base in the LexA binding site indicates the deviation from the LexA binding site consensus. The potential ribosome binding site GAGGG is underlined in red. The light blue box indicates the coding region of *ruvC*. The potential -35 sites are indicated with light grey boxes, and the underlined regions indicate regions of homology to the *M. tuberculosis* *lexA* -35 promoter at -10 (TACATT) and -35 (TTG GTC). The regions of homology to the RecA-NP promoter Gamulin *et al.*, (2004) are underlined in purple. Homology to the *recA* P1 promoter is boxed in pink.

Figure 5.2c: Transcriptional start sites of *recA*. Amended from Movahzedadeh 1996. The first transcriptional start site (+1) is indicated in red and the second is indicated in green (+1). The blue box and line indicates the position of the LexA binding site, with the consensus TCGAAC, GTTCGA. The green highlighted base in the LexA binding site indicates the deviation from the LexA binding site consensus. The potential ribosome binding site GAGAG is underlined in red. The light blue box indicates the coding region of *recA*. The potential -10 and -35 sites are indicated with light grey boxes, with the consensus -10 (TATAAT) and -35 (TTGACA) for *E. coli*. Regions of homology to the RecA-NP promoter Gamulin *et al.*, (2004) are underlined in purple. The potential -10 and -35 sites are indicated in dark grey, with the *E. coli* consensus -10 (CCCCATNTA) and -35 (TCTCNCCCTTGAA). Homology with the *ruvC* P1 promoter is boxed in pink. The SigRec motif is underlined in yellow.
upstream of the P1 promoter of \textit{recA} CTAGtGT, with only 1bp mismatch, which is separated by an 11bp spacer from a 4bp region of homology TGTC (pink box figure 5.2b) at the -35 site of \textit{recA} P1. The $\sigma^{70}$ consensus of \textit{M. tuberculosis} was initially outlined as TAYgAT (-10), where Y indicates a pyrimidine (Bashyam et al., 1996), however, that was slightly modified to TA(G/T)(A/G)aT (-10) and TtGaCa (-35) (Gomez M, 2000). The -10 and -35 regions of the P2 promoter of \textit{ruvC}, show a high degree of homology to the \textit{M. tuberculosis} consensus for a $\sigma^{70}$ promoter (dark grey box, figure 5.2b), this homology is also observed at the -10 region of the P1 promoter of \textit{ruvC} (light grey box, figure 5.2b). However the -35 site of the P1 promoter shows a lesser degree of homology to the \textit{M. tuberculosis} consensus (see figure 5.2b).

The \textit{recA} P1 promoter is of particular interest as there appears to be a shorter spacer between the -10 and -35 sites (see figure 5.2c) than observed in \textit{E. coli} (Movahedzadeh et al., 1997), which has been noted for a few promoters regulated by $\sigma^{70}$, although the spacer between the proposed -10 and -35 sites for the P2 promoter of \textit{ruvC} may follow the \textit{E. coli} consensus of a 16-19bp spacer region (Gross et al., 1992). There are two possibilities for the P1 promoter of the \textit{ruvC}, the spacer between the proposed -10 promoter element and the $\sigma^{70}$ like -35 element is 23bp (see figure 5.2b light grey boxes), whereas the other option, underlined in pink (figure 5.2b) pertains to the RecA non-dependent promoter (RecA-NP) consensus defined by Gamulin et al., (2004). The similarity of the -10 regions indicated the possibility that the P1 promoters of \textit{recA} and \textit{ruvC} may be transcribed by the same sigma factor.

The LexA binding sites for both \textit{ruvC} and \textit{recA} differ from the consensus TCGAAC(N)$_4$GTTCGA (Davis \textit{et al}., 2002a) by 1bp, indicated in green (see figures 5.2b and 5.2c). The sequence of the spacer region between the palindromic sites are different for \textit{recA} and \textit{ruvC}, which may play a role in regulation/strength of repression. Interestingly the LexA
regulated promoters of recA (P2) and ruvC (P1) appear to be transcribed by different RNAP-σ complexes (according to sequence homology at the -10 and -35 sites). The P2 promoter of recA, regulated by LexA shows a high degree of homology to the σ32 promoter consensus (see figure 5.2c), whereas the P1 promoter of ruvC, regulated by LexA shows homology to the σ70 promoter consensus (see figure 5.2b).

5.2 Quantitative analysis of the transcriptional start sites of ruvC compared to recA by RNase protection

To address the differential expression observed from the P1 and P2 promoters of ruvC in the three different experimental strains of M. tuberculosis, a complementary method, RNase protection was utilised, which enables quantitation of the transcript from multiple promoters. RNase protection is a method by which the level of transcription can be quantified from a number of individual promoters of the same gene. The method involves in-vitro transcription of a complementary RNA probe, which is then incubated with the RNA sample of interest. The homologous regions of the probe and transcript bind to form dsRNA, any ssRNA probe not bound to the sample RNA is digested using RNaseA/T1, resulting in a protected fragment which can be run on a polyacrylamide gel and visualised by autorography.

The RNA used in the RNase protection assays was quantified using an Agilent bioanalyser (see figure 5.3), this accurate quantitation of the RNA prior to the RNase protection enabled greater precision in quantitation of the RNase protection assay, due to equal loading of RNA for all the different samples. As outlined in section 5.1, the RNA was extracted from exponential phase (OD 0.3) for H37Rv wild-type, ΔrecA strain and ΔsigG strain, under both induced (mitomycin C 0.02μg/ml) and uninduced conditions. RNA was also extracted for H37Rv wild-type and
Figure 5.3: Bioanalyser data from RNA extractions. RNA samples were quantified using the Bioanalyser (Agilent Technologies). Quantification of the RNA samples is indicated in the table above.
ΔsigG strain, under both induced (mitomycin C 0.02µg/ml) and uninduced conditions, at early exponential phase (OD 0.15) due to the preliminary Taqman data suggesting that sigG may be preferentially expressed early on in growth.

### 5.2.1 Design and optimisation of the RNase protection assay.

Constructs were designed for recA, sigG and ruvC containing a region of approximately 322-358bp, including approximately 151-271bp upstream of the translational start site, were cloned into pCR4 blunt, between T3 and T7 promoters (see table 2.2). Clones were then sequenced and the orientation was determined. The in-vitro transcription was carried out on a linearised template, in the reverse orientation to produce radiolabelled complementary RNA (section 2.9.7.1 and table 5.1).

<table>
<thead>
<tr>
<th>Gene</th>
<th>probe size (bp)</th>
<th>Probe Orientation</th>
<th>Test template</th>
</tr>
</thead>
<tbody>
<tr>
<td>recA</td>
<td>423</td>
<td>T3 for antisense RNA</td>
<td>NheI digest = 44bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agel digest = 265bp</td>
</tr>
<tr>
<td>ruvC</td>
<td>383</td>
<td>T3 for antisense RNA</td>
<td>ClaI digest = 152bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NheI digest = 231bp</td>
</tr>
<tr>
<td>sigG</td>
<td>404</td>
<td>T7 for antisense RNA</td>
<td>EagI digest = 68bp</td>
</tr>
</tbody>
</table>

Table 5.1: Probe size, orientation and test templates designed for the three genes of interest

The sizes were determined using Map Draw (DNAsStar) to predict the restriction sites in the cloned regions of interest. The orientation for probe transcription was derived from the sequencing data on the orientation of the insert of the gene in the plasmid. The test templates are linearized with the listed enzymes and transcribed from the opposite promoter to the probe. For ruvC, two separate digests were performed (ClaI and NheI), which were then pooled to determine if more than one in-vitro transcription product could be detected, and sized reasonably accurately.

The complementary RNA probe would then bind to the homologous transcript in the total RNA from the strain of interest. Digestion of single stranded RNA (ssRNA) was performed using.
RNase A/T1, the resultant protected dsRNA product was then separated on a polyacrylamide gel, and visualised by autoradiography. In order for the experiment to be quantitative, the probe must be in excess, and has to be a clear single band (see figure 5.4).

Prior to performing the RNase protection assay, a number of optimization steps were performed, including increasing the quantity of template in the in-vitro transcription reaction (data not included), altering the ratio of radiolabelled to unlabelled UTP, decreasing the temperature of the transcription reaction, as well as varying the type and quantity of RNase used in the digestion.

The quantity of unlabelled UTP was altered to limit premature termination of the probe, and therefore produce a pure probe for the RNase protection assay. An in-vitro transcription of the ruvC probe was performed using increasing concentration of unlabelled UTP, alongside the radiolabelled UTP (see figure 5.5). A size marker was also produced by linearizing the template construct at different sites along the upstream region of ruvC, these fragments were then pooled and used in an in-vitro transcription reaction, thus producing radiolabelled transcription products of known size (see figure 5.5). Figure 5.5 shows that whilst weighing up purity against quantity, the probe synthesised with 0.5mM unlabelled UTP yielded the optimum probe.

The effect of altering the temperature of the in-vitro transcription reaction was tested for ruvC and recA, to determine whether this had an effect on the purity of the probe; figure 5.6 shows that the in-vitro transcription reaction produced a probe of greater purity when the reaction was performed at 37°C rather than 15°C.
**Figure 5.4**

**Schematic of an RNase protection assay.** The region of interest, including a minimum of 150bp upstream of the predicted translational start site (green) was cloned into pCR4-blunt between the T7 and T3 promoters. The construct was linearised using a restriction site, and *in-vitro* transcription was performed in the opposite direction to transcription *in-vivo*. This resulted in the production of a complementary radiolabelled probe (purple), which was quantified using a scintillation counter, and combined with total RNA. The probe annealed to complementary regions of RNA, and following digestion using RNase A/T1, the resultant protected fragment was visualised by autoradiography.
Figure 5.5: *In-vitro* transcription of the *ruvC* probe with varying concentrations of unlabelled UTP. The transcription reaction was performed on the linearised (*NotI* digest) *ruvC* template DNA, using T3 RNA polymerase. The transcription reaction was set up using increasing concentrations of unlabelled UTP: 0.125mM (track 1), 0.25mM (track 2), 0.5mM (track 3), and 2.5mM (track 4). The *ruvC* template was digested with *NheI* and *ClaI*, these were then combined in an *in-vitro* transcription reaction to produce a marker of known size. These products were then run on a polyacrylamide gel and visualised by autoradiography.
Figure 5.6: In-vitro transcription of the *ruvC* and *recA* probes at altered temperature. The in-vitro transcription reactions to produce the probes were carried out using 0.5mM unlabelled UTP, at either 37°C or 15°C. The radiolabelled probes were then run on a polyacrylamide gel, and visualised using autoradiography.
Finally, it is important to determine the quantity and type of RNase used in the protection assay to optimise the signal-to-noise ratio, whereby the RNase will remove any background (non-specific hybridization), but will not degrade the protected fragment. Degradation of the protected fragment appears as a smear on the autoradiograph, and can occur when dsRNA breathes or locally denatures, usually at regions of high uracil content. In a protection assay, this could result in the degradation of the protected specific dsRNA. To minimise incorrect degradation, RNase T1 may be used alone, as it cleaves 3' of guanine residues, whereas RNase A cleaves 3' of both cytosine and uracil residues. Therefore a test RNase protection assay was designed to determine the optimum combination of probe, test template(s) and digestion enzyme(s) and ratio. A positive control was necessary to act as a template for the probe, so that a protected fragment could be visualised, hence test templates of known size were synthesised using \textit{in-vitro} transcription for each of the three genes, \textit{ruvC}, \textit{recA} and \textit{sigG} (see table 5.1). A similar \textit{in-vitro} transcription reaction was performed as previously described for the probe synthesis, however, radiolabelled UTP was omitted and the reactions were performed in the forward orientation as opposed to the reverse orientation used to synthesise the probes. This resulted in the production of ssRNA of known length, homologous to the \textit{in-vitro} transcribed cRNA probe. Figure 5.7 shows a test RNase protection assay, where the \textit{ruvC} probe, \textit{in-vitro} transcribed at 0.125mM, 0.25mM, 0.5mM and 2.5mM of unlabelled UTP, were used with the test templates for \textit{ruvC}, which clearly produced two bands at 152bp and 231bp, from the pooled \textit{ClaI} and \textit{NheI} digests used to form the test templates. The test RNase protection (figure 5.7) also shows the different digestions with a combination of RNase A/T1 mix, and T1 only. It is clear that probe 2 (0.5mM cold UTP) produced the most crisp band for both the probe and test template, when the digestion was performed with 1:100 ratio of RNaseA/T1 mix to reaction mix (see figure 5.7). Therefore the remaining RNase protection assays were performed with probe
Figure 5.7: Optimisation of the RNase protection. In-vitro transcription reactions were carried out as outlined in figure 5.4, with varying concentrations of unlabelled UTP; probe 1 was synthesised with 0.125mM, probe 2 with 0.25mM, probe 3 with 0.5mM and finally probe 4 with 5mM (track 4). These probes were then used in a test RNase protection assay with an in-vitro transcribed test template. The test templates and negative control were incubated with the probe and digested with 1:100 ratio of RNase A:T1 (track 1 test template, track 2, negative control). Non-specific yeast RNA was used as a negative control. Track 3 contains the undigested probe. A test RNase digestion was performed whereby different ratios of RNaseA/T1 mix were used; 1:50 (track A), 1:100 (track B), 1:500 (track C). RNase T1 was also used alone, 1:50 (track D), 1:100 (track E), 1:500 (track F).
transcribed with the additional 0.5mM cold UTP, and the RNase digestion was performed with a ratio of 1:100 RNaseA/T1 mix to reaction mix. To enable accurate quantitation to take place, the radiolabelled probe must be in excess, therefore the probes were synthesised fresh, prior to each RNase protection assay, and 1µl of probe was visualised using polyacrylamide gel electrophoresis and autoradiography. The radioactive incorporation and quantitation was determined using a scintillation counter, to ensure equal quantities of the relevant probe (5x10^4 CPM) were used in the subsequent reactions.

The final pilot experiment involved the synthesis of each of the three probes, *ruvC*, *recA* and *sigG* under the optimised conditions outlined above. These were then run on a polyacrylamide gel alongside two size markers, both of which were produced by *in-vitro* transcription, incorporating radiolabelled α^{32}P UTP; one was commercially available from Ambion, producing a ladder from 500bp to 100bp (see figure 5.8). The second was produced by two restriction digests (*Cla*I and *Nhe*I) of a *ruvC* clone, which, when transcribed from the T3 promoter results in a 166bp fragment (*Nhe*I) and a 245bp fragment (*Cla*I) (see figure 5.8). Figure 5.8 shows the probes and markers are working under the test conditions outlined above, and that the test-templates and the markers can be distinguished by size, thus indicating the power of resolution of the system.

### 5.2.2 RNase protection of *ruvC*

The RNase protection for *ruvC* was performed using 40µg of total RNA from H37Rv wild-type, Δ*recA* and Δ*sigG* strains, under both uninduced and induced (mitomycin C 0.02µg/ml) conditions at an OD of 0.3. The probe was used undigested as a positive control, and the *Nhe*I test template (see table 5.1) was used a positive control after RNase A/T1 digestion. These samples were run alongside the marker and the ladder produced by *in-vitro* transcription from
Figure 5.8: Synthesis and sizing of the ruvC, recA and sigG probes under optimised conditions. The in-vitro transcription reactions were performed with 0.5mM unlabelled UTP at 37°C. The probes were run alongside an in-vitro transcribed commercial ladder M1 (Ambion) and a constructed marker M2, sizes are indicated. The recA probe (track 1), the sigG probe (track 2) and the ruvC probe (track 3) all appear as clean bands.
the *ruvC* clone and a commercial ladder (Ambion) respectively. Only one clear band was present, corresponding to the P1 promoter of *ruvC*, which appears to be partially induced in the *ΔrecA* strain (see figure 5.9). It appears that the expression level of the second promoter P2 of *ruvC* is close to the limit of detection for the method (see figure 5.9), which may have been improved by using a greater quantity of RNA.

A phosphorimager was utilised to visualise the gel, so it was possible to quantify the levels of expression from each promoter using volume of intensity. Grids of equal sizes were placed around the uninduced and induced bands for the P1 and P2 promoters individually, enabling volume measurements to be obtained, for each sample. A background correction (background value) for each grid was applied to all the values so a direct comparison could be obtained between the levels of expression from the P1 and P2 promoters under uninduced and induced conditions (see figure 5.10a). The uninduced bands were barely detectable above the background correction and no difference was observed between strains (data not included), therefore the induced samples were used to analyse the difference in expression between the P1 and P2 promoters of *ruvC* in the different strains of *M. tuberculosis*. There was clearly no difference between the expression of the P1 promoter or the P2 promoter, under induced conditions for both H37Rv and Δ*sigG* strain (see figure 5.10b), suggesting that SigG does not play a role in expression from either of the two promoters. However, there was a decrease in the expression of the P1 promoter in the *ΔrecA* strain compared to the wild-type and Δ*sigG* strains (see figure 5.10b). This correlates with the presence of the SOS box and therefore implicates LexA binding in inhibition of expression from the P1 promoter of *ruvC* (see figure 5.10b). There also appeared to be a decrease in expression from the P2 promoter of *ruvC* in the *ΔrecA* strain compared to the wild-type and Δ*sigG* strains, but the low level of expression from
Figure 5.9: RNase protection assay to identify *ruvC* promoters in H37Rv, Δ*recA* and Δ*sigG* strains of *M. tuberculosis*. Assays were performed on 40μg of RNA, under uninduced and induced conditions (0.02μg/ml mitomycin C), at an OD of 0.3, tracks 1-6. Yeast RNA was used as the negative control (track 7), two positive controls were the undigested probe (track 8) and the test template (track 9). The *ruvC* test template was produced by non-radiolabelled *in-vitro* transcription in the forward orientation of the probe construct, and yields products of 152bp and 231bp, when protected by the *ruvC* probe. The hybridization reaction of the template/RNA and probe was carried out over night at 42°C. Digestion was carried out with a ratio of 1:100 RNase A:T1. Samples were run alongside *in-vitro* transcribed markers (track M₁ and M₂), sizes are indicated above. Samples were then visualised by polyacrylamide gel electrophoresis and autoradiography. The position of the two *ruvC* promoters and the test template are marked with arrows.
Figure 5.10a

<table>
<thead>
<tr>
<th>Grid Name</th>
<th>Grid Area</th>
<th>Volume</th>
<th>Background Value</th>
<th>Corrected volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>H37Rv Induced P1</td>
<td>600</td>
<td>8181.6</td>
<td>12.24</td>
<td>8169.36</td>
</tr>
<tr>
<td>ΔrecA Induced P1</td>
<td>600</td>
<td>4137.67</td>
<td>6.607</td>
<td>4131.063</td>
</tr>
<tr>
<td>ΔsigG Induced P1</td>
<td>600</td>
<td>7876.67</td>
<td>14.095</td>
<td>7862.575</td>
</tr>
<tr>
<td>H37Rv Induced P2</td>
<td>600</td>
<td>582.34</td>
<td>3.207</td>
<td>579.133</td>
</tr>
<tr>
<td>ΔrecA Induced P2</td>
<td>600</td>
<td>41.3</td>
<td>2.372</td>
<td>38.928</td>
</tr>
<tr>
<td>ΔsigG Induced P2</td>
<td>600</td>
<td>651.65</td>
<td>3.102</td>
<td>648.548</td>
</tr>
</tbody>
</table>

Figure 5.10b

Figure 5.10a: The expression level of P1 and P2 promoters of ruvC, under induced conditions in H37Rv wild type, ΔrecA and ΔsigG strains of M. tuberculosis. An RNase protection assay was performed on uninduced and induced RNA form H37Rv wild type, ΔrecA and ΔsigG strains to detect the ruvC promoters P1 and P2. Using the phosphorimagery software, grids of equal size were placed over the bands produced from an RNase protection assay, to determine the level of transcript for each promoter from each strain. The levels of expression from the uninduced samples were below the level of detection above background, therefore volume and background measurements were taken from only the induced bands. The area of each grid is listed, alongside the volume inside the grid, and the background correction (Local/average).

Figure 5.10b: A graph showing the differences in expression level of the P1 and P2 promoters of ruvC as determined using the phosphorimagery. The corrected volume data was used to create the graph in Graphpad programme, Prism 4.
the P2 promoter hindered accurate quantification of the transcript, so the reliability of this observation is uncertain.

### 5.2.3 RNase protection of recA

In view of the partial induction of ruvC observed by both RNase protection and primer extension in the ΔrecA strains of *M. tuberculosis*, it was of interest to assess the expression of recA. The RNase protection for recA was performed using 20μg of total RNA extracted at mid-exponential phase (OD 0.3), from H37Rv wild-type, ΔrecA and ΔsigG strains, under both uninduced and induced (mitomycin C 0.02μg/ml) conditions. The recA transcript is relatively abundant; therefore the quantity of total RNA required was decreased to 20μg. The RNase protection using the recA probe produced three bands, indicated by arrows 1, 2 and 3 in figure 5.11. However, Movahedzadeh *et al.*, (1997) published data indicating there were only two transcriptional start sites identified by primer extension. Due to the size and absence of band 2 (arrow 2, figure 5.11) in the ΔrecA strain, it is possible that the band is merely a partial degradation product of the P2 promoter. Analysis of the promoter region of recA indicated there was only one LexA binding site, situated within the P2 promoter (Brooks *et al.*, 2001). Arrow 3, figure 5.11 indicates that the P1 promoter of recA, appears to be induced in H37Rv wild-type, ΔrecA and ΔsigG strains. However, the level of induction appears to be decreased in the ΔsigG strain. In contrast, the P1 promoter in the ΔrecA strain gave an undiminished band, further confirming that the P1 promoter is DNA-damage inducible independently of RecA. The P2 promoter, indicated with arrow 3 (figure 5.11) appears to be induced in both the H37Rv and ΔsigG strains, but there is no induction of the P2 promoter in the ΔrecA strain (figure 5.11). The P2 promoter possesses a putative LexA binding site and therefore the LexA repressor
Figure 5.11: RNase protection assay to identify *recA* promoters in H37Rv, Δ*recA* and Δ*sigG* strains of *M. tuberculosis*. Assays were performed on 20 μg of RNA, under uninduced and induced conditions (0.02 μg/ml mitomycin C), for H37Rv, Δ*recA* and Δ*sigG* strains of *M. tuberculosis* induced at an OD of 0.3, tracks 1-6 respectively. The strains are marked under the tracks. Yeast RNA was used as the negative control (track 7), two positive controls were the undigested probe (track A) and the test template (track B). The *recA* test template was produced by non-radiolabelled *in-vitro* transcription in the forward orientation of the probe construct. The hybridization reaction of the template/RNA and probe was carried out overnight at 42°C. Digestion was carried out with a ratio of 1:100 RNase A:T1. Samples were run alongside *in-vitro* transcribed marker (track M), sizes are indicated. Samples were then visualised by polyacrylamide gel electrophoresis and autoradiography. The position of the three *recA* products are marked with arrows.
appears to abolish expression from the P2 promoter in the ΔrecA strain, indicating, as previously published, that RecA is required for de-repression of the P2 promoter.

The RNase protection assay was repeated using 20μg of RNA from H37Rv and ΔsigG strains. The RNA was harvested at an OD of 0.15, under both uninduced and induced (mitomycin C 0.02μg/ml) conditions. The recA probe was used undigested as a positive control (figure 5.12, track 6), and the Nhel test template (see table 5.1) was used a positive control after RNase A/T1 digestion (figure 5.12, track 7). The RNase protection samples were run alongside the size marker and ladder produced by in-vitro transcription from the ruvC clone and a commercial ladder (Ambion) respectively. Three protected bands were produced in the induced samples of H37Rv wild-type and the ΔsigG strain (see figure 5.12, arrows 1, 2 and 3), suggesting both the P1 and P2 promoters are DNA-damage inducible. The second band, indicated with arrow 2 (figure 5.12) is most likely a partial degradation product of the P2 transcript possibly due to local breathing of the RNA-probe hybrid, leading to degradation.

Intriguingly the level of expression of all three bands appeared weaker in the ΔsigG strain (figure 5.11 and 5.12). This may simply be due to differential loading of RNA; however, this is unlikely as the samples of RNA were quantified using a bioanalyser, which produces accurate quantitations, particularly in light of the fact that two biological replicates of the ΔsigG strain produced similar results. This might indicate that SigG may directly or indirectly partially regulate expression from both the P1 and P2 promoters of recA, however, this possibility would need to be tested, potentially using lacZ fusion assays in the ΔsigG strain compared to H37Rv.
Figure 5.12: RNase protection assay to identify recA promoters in H37Rv and sigG strains of M. tuberculosis. Assays were performed on 20μg of RNA, under uninduced and induced conditions (0.02μg/ml mitomycin C), for H37Rv and sigG strains of M. tuberculosis induced at an OD of 0.15, tracks 1-4 respectively. The strains are marked under the tracks. Yeast RNA was used as the negative control (track 5), two positive controls were the undigested probe (track 6) and the test template (track 7). The recA test template was produced by non-radiolabelled in-vitro transcription in the forward orientation of the probe construct. The hybridization reaction of the template/RNA and probe was carried out over night at 42°C. Digestion was carried out with a ratio of 1:100 RNase A:T1. Samples were run alongside in-vitro transcribed markers (track M1 and M2), sizes are indicated. Samples were then visualised by polyacrylamide gel electrophoresis and autoradiography. The position of the three recA promoters are marked with arrows 1, 2 and 3. The probe and test template are marked with arrows 5 and 4 respectively.
In order to quantify the difference in expression observed in figures 5.11 and 5.12, grids were placed around the bands of both RNase protection assays on the phosphorimager, and corrected volume calculations were derived to analyse the differences in expression (see figure 5.13); revealing that transcript levels under induced conditions for the P1 promoter of recA were approximately halved in the ΔsigG strain compared to the ΔrecA and H37Rv strains. The transcript levels for the P2 promoter in the ΔsigG strain were also approximately halved in comparison to the H37Rv strain (see figure 5.13 a and b). This was observed in both RNase protection assays using RNA from cells induced at the different optical densities (0.3 and 0.15). The level of transcription of the P2 promoter in the ΔrecA strain was almost completely abolished (see figure 5.13 a and b).

5.3 Validation of quantitation

The reduced expression of both the P1 and P2 promoters of recA in the ΔsigG strain was observed in both RNase protection assays. Therefore this data was compared to microarray data (see chapter 4), where RNA was extracted from uninduced and induced (0.02ug/ml mitomycin C) conditions for both H37Rv wild-type and ΔsigG strains. The microarrays were designed with DNA versus RNA hybridisations to enable both inter- and intra-strain comparisons. The normalisations carried out are outlined in chapter 4. There were no significant differences at the 1% confidence interval between the expression of recA or ruvC under uninduced and induced conditions for H37Rv wild-type and ΔsigG strain (see figure 5.14), although as expected the expression of sigG was clearly reduced in the ΔsigG strain.
Figure 5.13a

<table>
<thead>
<tr>
<th>Grid name</th>
<th>Area</th>
<th>Volume</th>
<th>Background Value</th>
<th>Corrected volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>H37Rv 0.35 Ind P2 (SOS Box)</td>
<td>975</td>
<td>34485.91</td>
<td>12.579</td>
<td>34473.331</td>
</tr>
<tr>
<td>ArecA 0.35 Ind P2 (SOS box)</td>
<td>975</td>
<td>3026.73</td>
<td>5.911</td>
<td>3020.819</td>
</tr>
<tr>
<td>AsigG 0.35 Ind P2 (SOS Box)</td>
<td>975</td>
<td>14365.75</td>
<td>10.099</td>
<td>14355.651</td>
</tr>
<tr>
<td>H37Rv 0.35 Ind P1</td>
<td>975</td>
<td>53332.96</td>
<td>19.834</td>
<td>53313.126</td>
</tr>
<tr>
<td>ArecA 0.35 Ind P1</td>
<td>975</td>
<td>54802.28</td>
<td>20.266</td>
<td>54782.014</td>
</tr>
<tr>
<td>AsigG 0.35 Ind P1</td>
<td>975</td>
<td>25271.22</td>
<td>20.354</td>
<td>25250.866</td>
</tr>
<tr>
<td>H37Rv 0.15 Ind P2 (SOS box)</td>
<td>975</td>
<td>118412.42</td>
<td>24.678</td>
<td>118387.742</td>
</tr>
<tr>
<td>AsigG 0.15 Ind P2 (SOS box)</td>
<td>975</td>
<td>63609.18</td>
<td>20.592</td>
<td>63588.588</td>
</tr>
<tr>
<td>H37Rv 0.15 Ind P1</td>
<td>975</td>
<td>112971.87</td>
<td>47.096</td>
<td>112924.774</td>
</tr>
<tr>
<td>AsigG 0.15 Ind P1</td>
<td>975</td>
<td>80674.5</td>
<td>39.962</td>
<td>80634.538</td>
</tr>
</tbody>
</table>

Figure 5.13b

**Figure 5.13a:** The expression level of P1 and P2 promoters of recA, under induced conditions in H37Rv wild type, ArecA and AsigG strains of *M. tuberculosis*. An RNase protection assay was performed on uninduced and induced RNA from H37Rv wild type, ArecA and AsigG strains to detect the *uvC* promoters P1 and P2. Using the phosphorimager software, grids of equal size were placed over the bands produced from an RNase protection assay, to determine the level of transcript for each promoter from each strain. The levels of expression from the uninduced samples were below the level of detection above background, therefore volume and background measurements were taken from only the induced bands. The area of each grid is listed, alongside the volume inside the grid, and the background correction (Local/average).

**Figure 5.13b:** A graph showing the differences in expression level of the P1 and P2 promoters of recA as determined using the phosphorimager. The corrected volume data was used to create the graph in the Graphpad programme, Prism 4.
Figure 5.14a: A comparison between the expression level of recA, ruvC and sigG in the wild-type H37Rv and ΔsigG mutant of M. tuberculosis. The data were obtained from micro-array experiments outlined in chapter 4. The graph shows a comparison between the gene expression level, under both uninduced and induced conditions (0.02μg/ml mitomycin C) at an OD of 0.15.

Figure 5.14b: A comparison between induction ratio of recA, ruvC and sigG in the wild-type H37Rv and ΔsigG mutant of M. tuberculosis. Induction ratios were calculated using the normalised values for uninduced and induced data outlined above.

Figure 5.14c: A comparison between the expression level of recA, ruvC and sigG in the wild-type H37Rv and ΔrecA mutant of M. tuberculosis. Data was modified from Rand et al., 2003.
To further examine the differences observed in the RNase protection between the expression of the different promoters from the three strains, quantitative PCR was used to look at the levels of expression of \textit{ruvC}, \textit{recA} and \textit{sigG} under uninduced and induced (0.02\mu g/ml mitomycin C) conditions in H37Rv wild-type, \textit{\Delta recA} and \textit{\Delta sigG} strains. The drawback of this method and microarray analysis are that one is unable to distinguish expression from individual promoters of a particular gene; therefore, the expression level is presented as combined expression from all of the promoters of a specific gene.

FAM and TAMRA fluorescently labelled probes and primers were designed in the coding region for \textit{recA}, \textit{ruvC} and \textit{sigG}. These were used along side an internal standard control \textit{sigA} and were also run alongside a standard curve of known concentrations of DNA, to enable quantitation. The expression level of \textit{ruvC} and \textit{sigA} were measured in H37Rv, \textit{\Delta sigG} and \textit{\Delta recA} strains of \textit{M. tuberculosis} (see figure 5.15a). The expression levels were calculated in pico grams (pg) using a standard curve. There appears to be a difference in the level of expression of \textit{sigA} in H37Rv induced conditions (see figure 5.15a). There is no difference however between the expression levels of \textit{sigA} under uninduced and induced conditions for \textit{\Delta sigG} and \textit{\Delta recA} strains. This indicates that either \textit{sigA} is induced in H37Rv wild-type under mitomycin C stress, or more likely, the apparent increase in expression of \textit{sigA} is merely a result of a more efficient reverse transcription reaction.

In figure 5.15b, the expression of \textit{ruvC} is expressed as a proportion of internal standard, \textit{sigA}. The expression level of \textit{ruvC} in uninduced conditions is higher in the \textit{\Delta sigG} strain than the H37Rv and \textit{\Delta recA} strains (figure 5.15b). Despite this, there is no difference in the induced level of expression of \textit{ruvC} between H37Rv and \textit{\Delta sigG} strains. A decrease in the level of induction of \textit{ruvC} in the \textit{\Delta recA} strain was observed, although this decrease was not as marked as the
Figure 5.15a: Taqman expression levels of ruvC in H37Rv wild-type, ΔrecA and ΔsigG strains of *M. tuberculosis*. 1μg of RNA for each strain under both uninduced and induced (0.02μg/ml mitomycinC) conditions were used in a cDNA synthesis reaction, using random primers (RT+), a negative control omitting reverse transcriptase was also used (RT-). The resultant cDNA was diluted 1:20 with DEPC dH₂O, and 5μl were used in the Taqman real time PCR. The sigA and ruvC probes used were FAM/TAMRA labelled. sigA was used as an internal standard and known concentrations of DNA were used to produce a standard curve to determine the concentration (pg). Both RT+ and RT- reactions were used in triplicate. The expression level obtained from each sample for both probes was used in Graph Pad Prism to create a graph.

Figure 5.15b: Taqman expression level of ruvC expressed as a proportion of sigA. The data obtained for ruvC from the above Taqman was expressed as a proportion of the internal control sigA.
decrease observed by microarray analysis (see figure 5.15c), or RNase protection (see figure 5.9).

5.4 Discussion

The data presented by both primer extension and RNase protection for \textit{ruvC} indicates the presence of two promoters P1 and P2. It appears that the P1 promoter of \textit{ruvC} is DNA-damage inducible, whereas the methods utilised have been unable to detect induction of the P2 promoter of \textit{ruvC}, in response to mitomycin C stress. RNase protection is a more quantitative measure of transcription. It was noted by Movahedzadeh \textit{et al.}, (1997) that primer extension assays performed on \textit{recA} showed the P1 promoter was DNA-damage inducible, but it did not appear that the P2 promoter was DNA-damage inducible. The P2 promoter of \textit{recA} contains a palindromic binding site for the LexA repressor and was predicted to be DNA-damage inducible. Therefore, LacZ fusion assays were performed with the P1 and P2 promoters of \textit{recA}, which showed that both promoters were in fact DNA-damage inducible (Davis \textit{et al.}, 2002b). The RNase protection assay performed on \textit{recA} showed both the P1 and P2 promoters were in fact DNA-damage inducible (figure 5.11 and 5.12). Therefore it appears that RNase protection is a better method to detect induction of promoters. Taking this into account, it indicates that the P1 promoter of \textit{ruvC} is DNA-damage inducible as observed by both primer extension and RNase protection, whereas the P2 promoter is probably not DNA-damage inducible. This could however be addressed by using LacZ fusion assays as described by (Davis \textit{et al.}, 2002b)

It is particularly interesting to note that the P1 promoter of \textit{ruvC}, remains partially DNA-damage inducible in the \textit{ΔrecA} strain of \textit{M. tuberculosis}, which indicates that the P1 promoter
may be dual regulated, especially as the P2 promoter of \textit{recA} in the $\Delta\text{recA}$ strain shows no induction. This indicates that the P2 promoter of \textit{recA} is regulated by the RecA/LexA system: ablation of expression of the \textit{recA} P2 promoter in the $\Delta\text{recA}$ strain indicates dependence of the P2 promoter on RecA for de-repression. Conversely the P1 promoter of \textit{ruvC} is able to partially overcome the repression by LexA, even in the absence of RecA. There are a number of different mechanisms by which the P1 promoter of \textit{ruvC} may be able to partially overcome repression by LexA: These include the promoter recognition site, which may affect sigma factor specificity and competition for the core RNAP, along with the possibility that transcription factors play a role by either affecting sigma factor binding or RNAP stability. Possible transcription factors include ppGpp, H-NS, Fis and other architectural proteins, which also affect local DNA topology and may therefore influence repression by LexA. The other possibility is that the relative position of the LexA binding site may enable binding of the RNAP to the $-10$ and $-35$ regions, which may destabilize and overcome the repression of the LexA, or the RNAP may initiate transcription during equilibrium maintenance, discussed below.

To address the possibility that these differences in repression were a result of sigma factor specificity, promoter consensi were analysed. There have been a number of promoter consensus sequences identified, including \textit{E. coli} sigma S, encoded by RpoS; the stress response sigma factor for \textit{E. coli} (Lee and Gralla, 2001). This sigma factor is particularly interesting as it drives transcription from approximately 100 genes, involved in the stress response regulon, including those involved in oxidative stress, such as \textit{sodC} (superoxide dismutase), \textit{katE} and \textit{katD} (catalases), \textit{xthA} (endonuclease) and \textit{ada} (methyl transferase), as well as virulence factors, \textit{csgBA} (curli) (Lacour and Landini, 2004). SigS is a member of the $\sigma^{38}$ family and the consensus sequence for $\sigma^{38}$ at $-10$ is CTAcacT (Lee and Gralla, 2001), which was refined
further by Lacour and Landini, (2004) to TGN_{o,2}C(T/A)ATA(C/A)T. The underline indicates where this sequence is nearly identical to the consensus of sigma 70 in *E. coli* (TATAAT) -10 and (TTGACA) -35 (Lisser and Margalit, 1993). The -35 site of RpoS is identical to the sigma 70 site. This is particularly interesting, as it suggests two sigma factors are able to recognise similar -10 and -35 promoter elements. It has been shown *in-vitro* that σ^70/σ^{38} is able to initiate transcription from some σ^70 promoters. The promoters all had similar -10 regions, and were divided into three categories, type I were recognised by both σ^70 and σ^{38}, type II were recognised by mainly σ^70 whereas type III were recognised by only σ^{38} (Tanaka et al., 1993). Therefore these sigma factors could compete to initiate transcription. However, RpoS is expressed at low levels in the cell, until stationary phase, when the level gradually increases. Conversely the levels of σ^70 appear to remain relatively constant (Tanaka et al., 1993). This is particularly relevant, as the P1 promoter region of *ruvC* appears to show homology to both the σ^70 and σ^{38} – 10 promoter consensus sequences, suggesting the possibility that transcription of *ruvC* from P1 may be regulated by two sigma factors.

It is noteworthy that *in-vitro* transcription levels do not always conform with the *in-vivo* findings, suggesting that different transcription factor(s) and/or specific transcription conditions may be necessary for transcription by σ^{38} in preference to σ^70 (Tanaka et al., 1995). Interestingly, the *csgA* promoter of *E. coli* is transcribed by both σ^{38} and σ^70, yet the histone like protein H-NS selectively inhibits transcription of the *csgA* promoter by σ^70, thus indicating the importance of DNA structure in transcriptional initiation (Tanaka et al., 1995). Other structural proteins including Lrp, CRP, IHF and Fis have been implicated in determining whether expression from a given promoter is driven by σ^{38} or σ^70 (Hengge-Aronis, 1999). Transcriptional analysis of genes transcribed in *E. coli* by σ^{38} revealed that the region
downstream of −17 (including the −10 hexamer) is important for recognition, by σ^38 (Tanaka et al., 1995). Transcription of a single promoter may be subject to differential regulation by two sigma factors. This has been observed with genes involved in osmoregulation, osmB and osmY in *E. coli*. Under normal conditions, transcription is driven equally by σ^38 and σ^70; however under elevated concentrations of glutamate or acetate the level of transcription by σ^38 increased by approximately 20 fold (Tanaka et al., 1995). It has been shown that the −10 regions for *E. coli* and *M. tuberculosis* housekeeping promoters are highly conserved, but there are differences at the −35 sites (Lee and Gralla, 2001).

Another factor affecting transcription of genes dual regulated by σ^5 and σ^70 is alarmone guanosine tetraphosphate (ppGpp), which can act as a positive or negative regulator of transcription. ppGpp can bind to the β or β′ subunits of RNA polymerase, which reduces the ability of the RNAP-σ complex to form an open complex, necessary for elongation during transcription. However ppGpp can also work as a positive regulator of transcription, by destabilising the RNAP-σ complex bound to *rrnP1* promoters, thus increasing the amount of free RNAP-σ complex, which is then able to drive transcription of promoters which are less well adapted to recruiting RNAP-σ complex (Jishage et al., 2002). Interestingly regulators of alternative sigma factors require elevated concentrations of ppGpp for transcription. Po and Pu inducers of σ^54 promoters require ppGpp for transcription. ppGpp has been shown to affect sigma factor competition for core RNAP, favouring interactions with alternative sigma factors rather than σ^70 (Gralla, 2005).

It has been shown that sigma factors compete to bind with core RNAP and initiate transcription of their regulon. As previously mentioned, σ^5 encoded by *rpoS* is a stress response sigma factor
in *E. coli*, responding to a number of environmental stresses including oxidative stress. In a σ^5^ mutant strain of *E. coli*, super induction of genes regulated by σ^70^ was observed. This was thought to be due to the quantity of σ^70^ binding to core RNAP, brought about by the absence of competition for the core RNAP by σ^5^ (Nystrom, 2004a). Interestingly, over-expression of σ^5^, results in a decrease in the expression of genes transcribed by other sigma factors, suggesting the levels of competing sigma factors provide an additional layer to the regulation of transcription at different stages during growth. However, it is not known whether this competition is a direct result of the relative concentrations of specific sigma factors, or their relative affinities for the core RNAP (Nystrom, 2004a). The current hypothesis states that there is a transcriptional trade off between stress defence and growth and proliferation. SpoT and RelA affect levels of ppGpp in the cell, by affecting production and hydrolysis of ppGpp, which in turn affects the affinity of the RNAP for the alternative sigma factors involved in the stress response. However, to complicate matters, DskA is a polymerase binding protein, which is attracted to the active site of the RNAP, like ppGpp and nucleotides. In the presence of the co-regulator, significantly less ppGpp is required to inhibit transcription (Gralla, 2005).

Interestingly Hengge-Aronis, (1999) suggested that σ^70^ and σ^5^ recognise almost identical promoter regions, and that the stress response genes transcribed by σ^5^, were actually transcribed by the interactions of the holoenzymes containing σ^70^ and σ^5^, along with other factors, such as H-HS, Lrp, CRP, IHF or Fis. These additional factors were key in determining whether transcription was initiated by σ^70^ or σ^5^ (Hengge-Aronis, 1999). It was also suggested that the local difference in DNA structure may play a role in the differential choice between σ^70^ and σ^5^ (Hengge-Aronis, 1999).
The proximity of the +1, -10 and -35 sites may also have an effect on transcription. Closer analysis of the promoter regions upstream of the *E. coli* rRNA P1 revealed that there is an unusually short spacer between the −10 and −35 sites (this is also observed with the P1 promoter of *M. tuberculosis recA*), as well as a CG rich sequence upstream of the +1, and a long spacer between the −10 and the +1. A reduction of supercoiling with the rRNA promoter has been shown to significantly decrease transcription (Lew and Gralla, 2004).

Another alternative mechanism for regulation of transcription of a particular gene has been observed with the heat shock induced *dnaK* in *Clostridium acetobutylicum*, whereby there is an 11bp inverted repeat (GCACTC) present between the transcriptional and translational start site, which forms a hairpin loop structure, and may be involved in regulation of expression (Narberhaus et al., 1992). Thus further indicating the importance of DNA topology in regulation of gene expression in response to environmental stress.

The sequence and the position of the LexA binding site may play a role in the efficacy of repression. However, taking into account that both *recA* and *ruvC* show the same mismatch to the SOS box consensus, it is more likely that the position of the SOS box plays a role in differential repression of *recA* and *ruvC*. The relative position of repressor binding sites has been shown to be of importance in regulation by TrpR, a repressor protein. Genes regulated by the TrpR repressor have either weak or strong TrpR boxes, which show weak or strong binding of the TrpR repressor respectively (Pittard *et al.*, 2005). The level of repression is dependent on the binding site (TrpR box) and the relative position of the TrpR box to the −10 and −35 promoter elements as well as to the transcriptional start site. The formation of the repressor is also important, as the TrpR repressor can form both dimers and hexamers, dimers bind to strong
TrpR boxes, whereas hexamers of TrpR bind to weak TrpR boxes. TrpR is a dimer in solution, but in the presence of ATP and tyrosine can form a hexamer (Pittard et al., 2005). It has also been demonstrated that weak TrpR box overlaps with the promoter. This is of interest as it has been shown that LexA in-vitro is present in dimers (Chattopadhyaya and Pal, 2004).

These weak TrpR boxes are recognised and repressed by the hexamers, which bind to the TrpR boxes with the low affinity. The dimers usually bind strong TrpR boxes, but can bind to weak TrpR boxes with low affinity. The binding of the different repressors if further complicated by the involvement of co-factors, such as phenylalanine, which either enables binding of the RNAP, or TrpR dimer to a weak adjacent TrpR Box. Therefore the TrpR repressor is able to not only repress, but also activate transcription of the Trp regulon, by coordination with aromatic amino acids and co-factors. An additional note is that although the sequence between the palindromic binding sites of the repressor is not thought to be significant, a particularly strong TrpR box, upstream of aroG, contains the palindromic binding consensus with a GC rich region between the palindromic sites (Pittard et al., 2005).

Under certain circumstances, when the binding site of the repressor is upstream of the promoter, it can actually initiate transcription, for example, when the TrpR repressor is bound upstream of the promoter, in the presence of co-factors such as phenylalanine, tyrosine or tryptophan the Trp protein interacts with the α subunit of RNAP to initiate transcription (Pittard et al., 2005). Therefore the positioning of the repressor is vital to the mechanism by which the repressor affects transcription.

The relative positions of the repressor binding sites and promoter recognition sites are important for the repression of transcription, as this can alter the mode of repression. One mechanism of
repression involves the exclusion of RNAP from the promoter, thus preventing the RNAP from forming a contact between the -10 and -35 sites of the promoter, while the other involves the repressor interfering with the bound RNAP, inhibiting isomerisation, which inhibits the formation of an open complex (Neidhardt, 1996); this prevents the RNAP from exiting the promoter and thereby entering elongation phase of transcription (Pittard et al., 2005). In the second case, the RNAP is able to bind to the -10 and -35 sites of the promoter, but the repressor prevents the RNAP from forming a stable open complex and moving into elongation phase of transcription. The other possibility is that a stable elongation complex is formed, but the elongation is halted by the presence of the repressor, which acts as a road block, as observed in the lac operon (Neidhardt, 1996). The DNA template sequence can also cause lateral oscillation of the ternary RNAP complex, either slowing down elongation of RNA, pausing, or even causing cessation of transcription. Cis-acting antisense RNA can act to stabilise the RNA polymerase when bound to a segment of the transcript behind the RNAP, thus preventing backward translocation of the ternary complex (Toulme et al., 1999). Therefore regulation, either activation or inhibition of transcription can take place during all three stages of transcription: initiation, elongation and termination.

The site of the operator, can result in competition, whereby, the operator/DNA binding regions overlap with promoters, therefore causing the repressor protein and the RNAP to compete for binding to the operator or promoter respectively. In this case, they are unable to bind simultaneously and therefore compete either kinetically or thermodynamically for the promoter-operator region. The bound repressor decreases the amount of free promoter, so decreasing the rate of formation of an open complex and thus decreasing transcription (Neidhardt, 1996).
The Lac operator overlaps with the transcriptional start site (+1); therefore the position of the repressor is thought to be important in regulation of transcription, possibly by preventing elongation phase of transcription. However, repression may still be overcome, as RNA polymerase has the ability to oscillate back or forward at each template position (Toulme et al., 1999), which could result in destabilisation of a repressor protein to enable read-through transcription. During transcription the RNAP covers approximately a 35bp region of DNA; 12-15bp of this region are unwound, thus forming the transcription bubble. Inside the melted region the template forms a constant heteroduplex with the 3' region of the transcript, covering approximately 8-9bp. Where the RNA-DNA hybrid is weak, the ternary RNAP complex moves backwards (backtracking) rather than forming the next phosphodiester bond, therefore stabilising the complex before re-initiating the elongation process. Site-specific repressor proteins and chromosomal proteins can block transcription elongation. However transcriptional fusions assays have implied that these road blocks are overcome to a certain extent by RNAP, thus enabling some transcription (Toulme et al., 1999). Although repressors form transcriptional road blocks, analysis of the lac repressor revealed that transcriptional read through was increased when the elongation complex was able to extend the RNA chain by an additional nucleotide, which destabilises the repressor-operator complex, and allows increased transcriptional read through (Mosrin-Huaman et al., 2004). The frequency of chain extension by an additional nucleotide is increased when intracellular NTP concentrations are elevated, thus suggesting that NTP concentrations may play a key role in regulation of transcriptional elongation.

Finally, it has been suggested that the lateral stability of the complex can be altered by cis and trans acting RNA sequences, which adds another method of regulation of gene expression into the complex pot of activator/repressor proteins, co-factors, DNA topology, Histone like
architectural proteins, NTP concentrations, and backtracking of RNAP. In order for transcription to be efficient, the RNAP has to overcome a number of different obstacles. Therefore one could hypothesise that the regulation of the \textit{ruvC} P1 promoter is far more complicated that the mere presence of the LexA repressor. The position of the LexA binding site, downstream of the promoter recognition site indicates the possibility that RNAP and LexA may be able to bind simultaneously; therefore the LexA repressor may act as a roadblock to the RNAP. The laws of equilibrium suggest that the LexA repressor is not constantly bound to the SOS box, so therefore during dissociation and re-association, the RNAP would no longer be blocked by the repressor and would therefore be able to transcribe \textit{ruvC}. The fact that no induction is observed from the P2 promoter of \textit{recA} in the \textit{ArecA} strain suggests the LexA repressor binding site obscures the promoter binding region as they have been shown to overlap (Movahedzadeh et al., 1997). This would most likely prevent binding of the RNAP, and therefore prevent transcription.

It has been shown that the P1 promoter of \textit{ruvC} is DNA-damage inducible (figure 5.1b, 5.2a, and 5.9). Therefore if the promoter were transcribed by two sigma factors, the level of the stress response sigma factor, would increase under stress conditions. This would result in elevated levels of the stress response sigma factor, which would compete with $\sigma^{70}$ for the core-RNAP. This competition would increase the concentrations of the relevant RNAP-$\sigma$ complex in the cell, and may therefore have a downstream effect on the competition balance between the RNAP and the repressor for the P1 promoter of \textit{ruvC}. Interestingly these two promoters, P1 of \textit{ruvC} and P2 of \textit{recA} appear to have different recognition sequences for $\sigma$ factors, and therefore are probably recognised by different $\sigma$ factors. One could hypothesise that sigma factor specificity and affinity for particular recognition sequences may explain some of the differential
expression observed. However as previously mentioned regulation of transcription is a multifactorial process, which would need to be further dissected.

The primer extension technique is useful for the ability to determine the exact transcriptional start site of a gene, due to the advantage of running a manual sequence alongside the primer extension reaction. However, this method is less quantitative than RNase protection, which does not allow exact identification, down to the nucleotide of the transcriptional start site(s), but does allow approximate identification of the position of the transcriptional start site(s) using RNA markers, and more importantly allows accurate quantitation of the level of transcription from multiple transcriptional starts sites. Problems can arise using primer extensions when relying on the predictions of translational start sites, from which to design primers to perform the primer extension assay. The primer has to be within the coding region of the gene of interest, to enable the extension reaction to be carried out to detect the transcriptional start site(s). The potential difficulties of primer extension were observed in chapter 4, whereby the primer extension reactions carried out to find the transcriptional start site of \( \text{sigG} \) were unsuccessful due to the in-correct prediction of the translational start site. This incorrect prediction meant the primer used in the primer extension reactions, was in actual fact, upstream of the transcriptional start site, therefore no product was detected. This problem was overcome using RNase protection, which allows the use of a large region as a probe, thus taking into account any discrepancies between the predicted translational start site and the actual transcriptional start site.

Taqman is a method by which transcription can be quantified; however the limitations of the procedure are such that one cannot distinguish expression levels from different promoters of a given gene, but rather measures the overall combined expression. Therefore it does not provide
the ability to dissect transcription from different start sites for a particular gene. If the Taqman results are accurate, then there is a significant increase in expression of ruvC under uninduced conditions in the ΔsigG strain compared to H37Rv and ΔrecA strains, and a slight decrease in induction of ruvC in ΔrecA strain. This does not correspond with the results obtained by primer extension, RNase protection, and microarray analysis, where all methods show a significant decrease in the expression of ruvC in the ΔrecA strain; indicating there may be problems with quantitation of the Taqman RT-PCR.

There were a number of concerns about the suitability of sigA as an internal control. Data available in the current literature is expressed as a proportion of the internal standard (sigA), therefore one cannot determine whether any differences have been observed in the expression levels of the internal standard sigA. Data from myself (data not shown), as well as from other members of the laboratory have raised concerns about the validity of using sigA as an internal control, due to differences observed in the expression level of sigA, particularly at different temporal levels, as seen when assessing the temporal expression levels of sigG (data not shown). The purpose of an internal control is to distinguish between differences in efficacy of the cDNA synthesis reaction; however, this cannot be achieved if the internal control varies under different conditions. For this reason, Taqman analysis was not performed for sigG or recA in the H37Rv ΔsigG and ΔrecA strains of M. tuberculosis.

The RNase protection assays show no difference in the induction of ruvC from the P1 promoter in the ΔsigG strain compared to wild-type strain. This observation was also made using microarray and Taqman; however there appeared to be a decrease in induction of the P1 promoter in the ΔsigG strain when analysed by primer extension. Due to the limited
quantitative ability of the primer extension assay, alongside the conflicting data obtained by the RNase protection, microarray and Taqman, the collective conclusion is that *ruvC* P1 is most likely not regulated by *sigG*.

In the case of the P2 promoter of *ruvC*, the RNase protection and primer extension assays both indicate that *ruvC* P2 is not induced in response to DNA-damage. However, this hypothesis should be confirmed, using promoter fusion assays with *lacZ*, under uninduced and induced conditions.

Closer analysis of the expression level of *ruvC* in the *ΔrecA* strain revealed that the P1 promoter, despite processing a LexA binding site, was partially induced. This was compared to data obtained by RNase protection for *recA*, which clearly showed the P2 promoter, regulated by LexA was not induced in the *ΔrecA* strain. This indicates that RecA is required for removal of the repressor LexA, and therefore required for the expression of the P2 promoter of *recA*. However, it appears that RecA is required to return the induction level of the P1 promoter of *ruvC* back to full level, although in the absence of RecA, the promoter remained induced by DNA-damage to approximately half the magnitude. Taqman data showed that *ruvC* was expressed to a lesser extent in the *ΔrecA* strain, but the expression level was not decreased by half, as was also observed in the microarray data (Rand et al., 2003). Therefore, as previously discussed, there appears to be another level of regulation, enabling transcription from the P1 promoter of *ruvC* in the absence of RecA. As mentioned, this difference in expression level may be due to the RNAP being able to read through the road-block produced by the LexA repressor, and the positions of the LexA binding sites for the P2 promoter of *recA* and the P1 promoter of *ruvC* differ. Alternatively as noted, the promoters appear to have different binding sites for different sigma factors, which may play a role in their regulation along with other
transcriptional regulatory proteins, architectural proteins and cellular concentrations of NTP. It is also of interest to note that recA expression is not completely abolished in ΔrecA strain, as shown by microarray (Rand et al., 2003) and LacZ promoter fusion assays (Davis et al., 2002b). Induction was observed from the P1 promoter of recA, to the same extent in the wild-type, and ΔrecA strains, thus agreeing with previously published data, that the expression level of P1 promoter of recA is not regulated by the RecA/LexA system. However, the RNase protection assay, microarray data and Taqman indicate that the sigma factor regulating expression of recA P1 is not SigG. It may be that the same mechanism regulates transcription from the P1 promoters of recA and ruvC; however the decreased induction of the P1 promoter in ruvC may be assigned to the fact that the alternative mechanism is only able to partially overcome repression of the promoter by LexA.

The recA data was not quantified using Taqman due to the difference observed when using sigA as a standard control. Therefore this could be performed using a different internal standard such as 16s rRNA; however the problem of using 16s rRNA is one of relative abundance. The level of sigG is very low and the level of 16s rRNA is considerably higher, so using 16s rRNA as the normaliser may be unsuitable. Another alternative is gnd, which has been previously used as an internal standard control (Brooks et al., 2001). It may be more beneficial to use a combination of these to determine if the levels of these alter consistently with the cDNA preparations, thus indicating that the differences observed are actually due to the efficiency of the cDNA synthesis rather than differences in levels of the normaliser. The difference in expression observed in the RNase protection assay could be addressed in a similar way, but using an additional probe in the reactions to quantify whether the differences in expression are due to an effect of the knockout strain, or due to different levels of RNA. It would therefore be vital to choose a gene, whose expression level remained constant in H37Rv as well as ΔrecA and ΔsigG strain.
In conclusion, \textit{ruvC} possesses two promoters, only one of which appears DNA-damage inducible. The P1 promoter, repressed by LexA, is induced, but to a decreased level in the absence of RecA. The P2 promoter of \textit{ruvC} is expressed to lower levels than the P1 promoter and does not appear under these conditions to be DNA-damage inducible. The P2 promoter of \textit{recA} is not expressed in the absence of RecA, however the P1 promoter remains DNA-damage inducible. It does not appear that SigG plays a role in regulating transcription of \textit{recA} or \textit{ruvC} in response to DNA-damage. It appears that there is an alternative mechanism regulating expression of both \textit{recA} and \textit{ruvC} in response to DNA-damage; however it is unclear what this mechanism is. There may be a clue in the DNA sequence analysis, as it appears that there is a high degree if homology between the \textit{–10} promoter elements of the P1 promoters of \textit{recA} and \textit{ruvC}. \textit{lacZ} fusion assays along side site directed mutagenesis could be used to identify the important residues in the \textit{–10} and \textit{–35} regions.
6 Regulation of gene expression from the P1 promoter of recA

As outlined in the previous chapters, the recA gene is part of the SOS regulon in *M. tuberculosis*. The mechanism of action of the SOS response has been well defined in *E. coli*, where recA has been shown to be under the control of LexA (Little, 1982) a transcriptional repressor, which binds to a specific region (SOS box) located in the promoter of the SOS response genes.

Some proteins have been identified that are DNA-damage inducible, but are regulated independently of the LexA/RecA system (Kleinsteuber and Quinones, 1995); which has also been observed in *M. tuberculosis*, where a number of genes have been shown to be DNA-damage inducible independently of the LexA/RecA system, including recA itself (Rand *et al.*, 2003). Microarray data showed that the DNA-damage regulon is comprised of 112 genes upregulated 3-fold or more in response to DNA-damage by mitomycin C (0.2μg/ml), whereby the vast majority of genes are regulated in a RecA independent manner, as they exhibit either full or partial induction of expression in response to DNA-damage in the ΔrecA strain (Rand *et al.*, 2003). This observation suggests that there may be an alternative mode of regulation of some DNA-damage inducible genes.

The possibility that SigG played a role in regulation of the DNA-damage regulon, in response to DNA-damage, was assessed in chapters 3 and 4. The data revealed that SigG did not play a major role in regulation of the DNA-damage response; therefore other options were explored to elucidate the alternative mode of regulation of the DNA-damage regulon. One possibility was that another sigma factor was responsible for regulation of the DNA-damage response.
Alternatively, a regulatory protein could bind to the promoter regions of some of these DNA-damage inducible genes and act to either repress or activate transcription. To address the possibility that regulation of the RecA independent DNA-damage response was controlled by a regulatory protein, the recA promoters were analysed by gel-shift assay to look for binding of potential regulatory proteins. The recA promoters were chosen for this experiment, as the P1 and P2 promoters of recA are DNA-damage inducible, the P2 promoter is regulated by LexA, and it has been shown by RNase protection assay that the P2 promoter is not expressed in the ΔrecA strain (see chapter 5), suggesting the expression of this promoter is entirely dependent on RecA for de-repression. The P1 promoter remains DNA-damage inducible in the ΔrecA strain, suggesting its regulation is independent of RecA.

### 6.1 Gel retardation assays

A method designed to detect DNA-protein interactions is the gel retardation or 'bandshift' assay. A radiolabelled DNA probe, is incubated with protein extracts, and electrophoresed through a non-denaturing polyacrylamide gel; any interactions between a protein from the cell free extracts (CFE) and the labelled probe result in the formation of a complex, which leads to a reduction in the electrophoretic mobility of the fragment in the non-denaturing polyacrylamide gel (Lane et al., 1992), resulting in a bandshift.

### 6.2 Bandshifts using large P1 and P2 probes.

Two double-stranded oligonucleotide probes were designed to each cover a 60bp region of the P1 or P2 promoter regions of recA (see figure 6.1A), as these were enough to show induction of P1 in response to DNA-damage in a lacZ fusion assay (Davis et al., 2002b). These probes were then used in bandshift experiments with induced (mitomycin C, 0.2μg/ml) and uninduced cell
Figure 6.1A: A schematic representation of P1 and P2 promoter region of *recA*. The 60mer oligonucleotide probes used in gel retardation cover either the P1 or P2 promoters, or part of the internal coding region of *recA*.

Figure 6.1B: A gel retardation using 60mer oligonucleotide probes. Two cultures of H37Rv were grown to an OD of 0.3-0.4. These were then pooled and split. One was untreated, the other was induced with 0.2mg/ml mitomycin C and both were incubated in a rolling incubator for 24hrs at 37°C. Cell free extract was obtained from both uninduced and induced cultures and used in a gel retardation assay (track U and I respectively) with either P1 oligo, P2 oligo or internal *recA* oligo D60. A negative control containing no protein was used for all of the oligos (track -).
free extracts (CFEs). Binding reactions using purified LexA protein with labelled P2 probe, were used as a positive control. The labelled P2 probe bound to LexA in both the purified LexA protein preparation and the CFE, resulting in a visible shift or retardation (data not shown and figure 6.1B). The labelled P1 probe showed a range of retarded bands (see figure 6.1B), demanding optimisation of the conditions to aid specific binding and maintenance of a stable complex during electrophoresis.

Optimisation steps were carried out to decrease non-specific binding, and increase specific binding particularly to the P1 region of recA. Different binding buffer recipes were used as outlined in section 2.10.3. The most effective buffer was buffer 11 (see figure 6.3B). Two different gel recipes were also tested, one containing glycerol, and the other without: the EMSA mix (with glycerol) proved the best for resolution and separation of retardation (data not shown).

Bandshift experiments were repeated using the radiolabelled P1 and P2 probes, under optimised conditions, with the addition of another labelled oligo D60, designed from the internal coding region of recA, which should not bind to a potential activator or repressor protein. The labelled D60 and P1 probes showed different retardation patterns, suggesting the binding to P1 was specific (see figure 6.1B). However, there were several faint retardations visible. Therefore to narrow down the potential binding site, and get a clearer single retardation, shorter 30mer oligonucleotide probes were designed (see figure 6.2A)
Figure 6.2A: A schematic representation of P1 and P2 promoter region of recA. The 30mer oligonucleotide probes used in gel retardation cover either the P1 or P2 promoters, or part of the internal coding region of recA.

Figure 6.2B: A gel retardation using the 30mer oligonucleotide probes. Uninduced H37Rv CFE was used in a gel retardation assay (track +) with either P1 oligo A, B, C, the P2 oligo E, or the internal recA oligo D. A negative control, omitting the protein was used for all of the oligos (track -). Arrow 1 indicates possible non-specific band for oligos A, B and C; whereas arrow 2 indicates a potential specific band for P1 oligos A, B and C. Arrow 3 indicates the non-specific band produced by the internal recA oligo D and arrow 4 indicates LexA bound to P2 oligo E.
Figure 6.3A: A schematic representation of P1 and P2 promoter region of recA. The 30mer oligonucleotide probes used in gel retardation cover either the P1 or P2 promoters, or part of the internal coding region of recA.

Figure 6.3B: A gel retardation comparing binding buffers using the 30mer oligonucleotide probes. Uninduced H37Rv CFE was used in a gel retardation assay (track +) with either P1 oligo A, or the P2 oligo E. A negative control, omitting the protein was used for all of the oligos (track -). Oligo A was used in reactions with uninduced CFE with different binding buffers, as indicated. Competition experiments were also carried out, where 100-fold excess of unlabelled P1 competition oligos, A, B and C, were included in the binding reaction along with labelled oligo A.

Figure 6.3C: A competitive gel retardation using the 30mer oligonucleotide probes. Uninduced CFE H37Rv CFE was used in a gel retardation assay (track +) with P1 oligo C. A negative control, omitting the protein was used (track -). Binding of labelled oligo C was competed with 100-fold excess of unlabelled P1 A, B and C, P2 E and recA internal oligo D. The arrow indicates the retardation product using the oligo C probe.
6.3 Bandshifts using small P1 and P2 probes.

Bandshifts were carried out using the optimised conditions outlined in section 6.2. The positive control oligo E, designed to contain the LexA binding site of the P2 promoter, produced a retardation with LexA (see figure 6.2B and 6.3B). Oligos A, B and C produced two distinct retarded bands (see figure 6.2B), with one potential due to an activator/repressor protein, and the other retardation potentially a non-specific band. Oligo probe D (see figure 6.3B), designed from the coding region of recA, showed no specific binding. The top retarded band present in both binding with the P1 specific oligo probes A, B and C, and the negative control oligo probe D, suggested it was the result of non-specific binding. The bottom retarded band visible with oligos A, B and C ran at a different level to the oligo D retarded band (see figure 6.2B), suggesting a specific retardation.

6.4 Competitive bandshifts

Competitive bandshifts were carried out with a 100-fold excess of unlabelled probe designed to compete out any non-specific binding. Oligos A, B and C were labelled and used in competitive binding assays with unlabelled A, B and C, where labelled oligo A was competed with Oligo A, B and C separately (see figure 6.3B). No binding was seen under these conditions, suggesting either the binding was non-specific, or there were two potential binding sites for a repressor/activator protein. This was repeated, using labelled oligo B, then labelled oligo C. Again, no binding was observed under these conditions (data not shown).

To determine if there were two potential binding sites, or the initial binding was non-specific, labelled oligo C was competed with oligos A, B, C, D (recA internal oligo) and E (LexA...
specific oligo). The binding observed with labelled oligo C was removed in all the competition reactions (see figure 6.3C), suggesting the binding observed under these conditions was non-specific.

One difficulty in detecting binding by gel retardation is the necessity for the protein-DNA interactions to form a stable complex during electrophoresis. Therefore, a method of cross-linking the complexes to aid stability was employed. Gluteraldehyde was added to the competitive bandshift reactions prior to electrophoresis, but still no specific shifts were visualised, i.e. the competitors A, B, C, D and E all removed binding observed with oligo C alone.

6.5 Discussion

Close analysis of the different DNA-damage repair systems in E. coli and B. subtilis have shown that they have a common regulatory system, which responds to DNA-damage, the SOS response (Cheo et al., 1993; Raivo and Silhavy, 2001). This regulatory system is also present in M. tuberculosis. However, unlike E. coli and B. subtilis, M. tuberculosis has two promoters regulating DNA-damage inducible gene expression of recA. Only one of those promoters, P2, is regulated by the SOS response, and the other, P1, remains DNA-damage inducible independently of LexA/RecA (Davis et al., 2002b). Therefore, an alternative mechanism of regulation probably regulates DNA-damage inducible gene expression of the P1 promoter.

The gel retardation experiments were carried out to determine if double-stranded oligonucleotides covering the P1 or P2 regions of recA could bind to a protein from a CFE of M. tuberculosis and produce a visible retardation when electrophoresed in a non-denaturing...
polyacrylamide gel. Detection of protein-DNA interactions by gel retardation depends on two critical factors, resolution from uncomplexed DNA, and maintenance of a stable protein-DNA complex (Lane et al., 1992). The P2 promoter region was used as a positive control, and the formation of a retarded band confirmed the integrity of the CFE. However, the ideal conditions for the maintenance of a stable complex between P2 and LexA will not necessarily be ideal for binding of a repressor/activator protein to the P1 region. Although binding was detected to the P1 probe, competition bandshifts revealed that this binding was not specific. A specific binding complex was not detected even following cross-linking of complexes prior to electrophoresis.

The lack of formation of a specific complex with CFE might have been due to there being insufficient quantities of the regulatory protein in the extract. However, this may also have been due to the conditions of the experiment. It is possible that too little probe was used in the experiment, therefore a titration of the probe with the CFE could enhance the possibility of detecting a stable interaction. The competition experiments may also have been too stringent, therefore alongside the titration of probe, a titration of competitor could also be used. Another potential pitfall of using the A, B and C oligos to compete for binding, is that the potential binding site may cover a large region of the DNA, which could be looped in the presence of the repressor/activator, as observed with dnaK, in Clostridium acetobutlicum, in which a hairpin loop structure which is thought to play a role in regulation of expression (Narberhaus et al., 1992). If this were the case, then the overlapping oligos would not be a good choice for the competitive oligo, as the binding site of a potential repressor/activator protein may overlap these oligos: hence the use of the non-specific oligo designed for and used in the assays (oligo D). This result does not indicate that a protein does not bind to the P1 region of recA, just that it was not detectable by this method. For this reason, an alternative approach was taken in which a gene inactivation knockout was produced in a predicted regulatory protein, which may
potentially be involved in regulation of the RecA independent DNA-damage regulon (see chapter 7).
7 Construction of a knockout in a predicted regulatory protein

7.1 Introduction

A wide variety of different environmental signals can alter gene expression, both temporally and spatially. Regulation of gene expression in response to these environmental cues requires different response regulators. As well as sigma factors, there are many families of response regulators, which either activate or repress expression of a specific regulon (Withey and DiRita, 2005). Activator or repressor proteins can also be used in conjunction with sigma factors, which form the specific interaction between RNA polymerase and the promoter (Gomez et al., 2000). Microarray experiments carried out by another member of the laboratory identified 5 predicted transcriptional regulatory proteins that were upregulated in response to DNA-damage (Rand et al., 2003): Rv0586, Rv1956, Rv1985c, Rv2017 and Rv2884 (see Table 7.1).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Induction ratio in Wild type</th>
<th>Induction ratio in recA mutant</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0586</td>
<td>2.16 ± 0.16</td>
<td>2.20 ± 0.42</td>
<td>Non-essential (Sassetti et al., 2003)</td>
</tr>
<tr>
<td>Rv1956</td>
<td>6.95 ± 0.34</td>
<td>9.43 ± 0.86</td>
<td>Non-essential (Sassetti et al., 2003), induced by heat shock (Stewart et al., 2002)</td>
</tr>
<tr>
<td>Rv1985c</td>
<td>2.86 ± 0.43</td>
<td>5.23 ± 1.01</td>
<td>Non-essential (Sassetti et al., 2003)</td>
</tr>
<tr>
<td>Rv2017</td>
<td>3.36 ± 0.65</td>
<td>4.37 ± 1.21</td>
<td>Induced by heat shock (Stewart et al., 2002)</td>
</tr>
<tr>
<td>Rv2884</td>
<td>4.62 ± 0.97</td>
<td>6.20 ± 1.43</td>
<td>Non-essential (Sassetti et al., 2003)</td>
</tr>
</tbody>
</table>

Table 7.1: Relative induction ratio in response to mitomycin C induction of 5 predicted regulatory genes in both wild-type and ΔrecA strains of M. tuberculosis

A recent transposon mutagenesis study performed by Sassetti et al., (2003) showed mutations in certain genes were recoverable in-vitro, and were therefore classified as non-essential. Data for induction ratios was obtained from supplementary data tables (Rand et al., 2003).
One or more of these transcriptional regulators could be involved in a novel mechanism regulating DNA-damage inducible gene expression.

### 7.2 Analysis of homology of the predicted regulatory proteins.

In order to assess the potential of the regulatory proteins to regulate gene expression of part of the DNA-damage regulon, and to further elucidate their possible relationships to other known regulatory proteins, homology studies were performed using BLAST, EMBOSS and PfAM, and domain searches were performed using NCBI conserved domain search (RPS-BLAST).

RPS-BLAST using the NCBI website revealed that Rv0586 showed homology to the GntR family of transcriptional regulators (see figure 7.1). An example is, *yhcF*, a repressor of the gluconate operon in *Bacillus cereus*. This homology covered a 60 residue region, which formed the helix-turn-helix motif, with which Rv0586 shared 100% homology (3e-11). At the N-terminal region, it shows homology to the GntR family of transcriptional regulator, pfam00392 *Bacillus cereus* (96.9% identity, 9e-12). The GntR family can be sub-divided depending on the C-terminal effector domains. In this case Rv0586 shows homology to a FadR transcriptional regulator, COG2186. BLASTP shows Rv0586 has homology to MAP4081 in *M. avium subsp. paratuberculosis*, and YP-116369 of *Nocardia farcinicia* with 67% and 60% identity respectively. Homology searches using FASTA (nucleotide search) showed Rv0586 shares homology (E=9.3e-08) with the regulatory protein P33233 of the L-lactate dehydrogenase operon from *Escherichia coli*, and to other GntR family of transcriptional regulators in *M. tuberculosis* e.g. Rv3060c and Rv0792c.
Figure 7.1: Conserved domains present in the 5 predicted regulatory proteins. Protein sequences available from the Tuberculist website were used in the NCBI conserved domain search to detect any homology to known domains. The light blue represents regions of low complexity and the homology to specific domains is indicated below each protein.
RPS-BLAST using the NCBI website revealed that Rv1956 shared similarity with the xenobiotic response XRE family of transcriptional regulators (see figure 7.1) (cd00093). The helix-turn-helix domain also showed homology to Cro and Ci family. BLASTP shows Rv1956 has homology with NP_855641 of *M. bovis* and CAF20199 of *Corynebacterium glutamicum* with 100% and 52% identity respectively.

RPS-BLAST using the NCBI website revealed that Rv1985c shared homology with LysR family of transcriptional regulators (see figure 7.1)(Pfam00126 gnl/CDD/25406)), with a 95% alignment 2e-10. BLASTP shows Rv1985c has homology with NP_855657 from *M. bovis* and CAB93745 of *Streptomyces coelicolor*, with 100% and 48% identity respectively.

Protein blast (BLASTP) search using Rv2017 also showed homology to GI: 29345183, a zinc binding transcriptional regulator of the Cro/CI family, from *Enterococcus faecalis* V583 (E-value = 3e⁻³), as well as homology to Rv2515c (E-value = 3e⁻⁸) from *M. tuberculosis* and Mb2040 (E-value = 5e⁻⁷) from *M. bovis*, a hypothetical regulatory protein to which, Rv2017 showed a 100% nucleotide identity (FASTA). Rv2515c is particularly interesting, as it is listed as a hypothetical protein in TubercuList, and was shown to be essential by transposon mutagenesis by Sassetti *et al.*, (1999); it too contains a neutral zinc metallopeptidase, zinc binding region signature. RPS-BLAST using the NCBI website revealed that Rv2017 contains a Helix-turn-helix domain which shares homology with both XRE family of xenobiotic response regulators, and pbsX family; it also contains a domain of unknown function (DUF955) (see figure 7.1).

As previously stated by (Earl *et al.*, 2002), *irrE* of *D. radiodurans* shows homology to Rv2017 of *M. tuberculosis*. However homology blast searches using both nucleotide (FASTA) and
protein (BLASTP) sequences of Rv2017 downloaded from TubercuList failed to detect homology to \textit{irrE} within the standard default cut-off. When the BLASTP was repeated using \textit{irrE} protein sequence as the query, homology was detected with Rv2017 (E-value of 5e^-7) (Figure 7.2a and b). An alignment of the regions of homology between IrrE and Rv2017 was also assessed using the protein alignment database EMBOSS (figure 7.2a). The alignments produced by both EMSOSS (figure 7.2a) and BLASTP (figure 7.2b) show a region of 31 residues of similarity, which is contained within the domain of unknown function (DUF 955). Pfam analysis showed that Rv2017 contains a predicted helix-turn-helix (HTH 3) DNA binding domain from residue 6 to 59, and a domain of unknown function (DUF 955), which contains a H-E-X-X-H motif. This motif is consistent with the presence of a catalytic active site, which has similarity to Peptidase M48, a metalloprotease.

RPS-BLAST using the NCBI website revealed that Rv2884 shared homology with the effector domain of response regulators (trans\_reg\_C) (see figure 7.1), present in two-component signal transduction systems. BLASTP for Rv2884 revealed homology to NP\_855641 of \textit{M. bovis} and CAF20199 of \textit{Corynebacterium glutamicum} with an identity of 100\% and 52\% respectively. Interestingly, Rv2884 exhibits homology (E=3e^-15) to Q55733, a regulatory component of a sensory transduction system, from \textit{Synechocystis} sp. strain PCC 6803.

Rv2017 and Rv2884 were selected for further study based on their levels of induction in response to mitomycin C (see table 7.1), taking into account the fact that a mutant of Rv1956 was already under construction by another member of the laboratory. Rv2017 was particularly interesting due to sequence homology with \textit{irrE} from \textit{Deinococcus radiodurans} (\textit{D. radiodurans}) (see figure 7.2a and b), a radio-resistant bacterium (Earl \textit{et al.}, 2002). IrrE has been shown to positively regulate gene expression of \textit{recA} (Earl \textit{et al.}, 2002). Although \textit{recA} is
**Figure 7.2a**

<table>
<thead>
<tr>
<th>Search method</th>
<th>EMBOSS-align water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aligned sequence 1</td>
<td>Rv2017 Mycobacterium tuberculosis</td>
</tr>
<tr>
<td>Aligned sequence 2</td>
<td>IrrE Deinococcus radiodurans</td>
</tr>
<tr>
<td>Matrix</td>
<td>EBLOSUM62</td>
</tr>
<tr>
<td>Gap penalty</td>
<td>10</td>
</tr>
<tr>
<td>Extended penalty</td>
<td>0.5</td>
</tr>
</tbody>
</table>

![Alignment Alignment](image)

Length: 280  
Identity: 68/280 (24.3%), Similarity: 98/280 (35.0%), Gaps: 104/280 (37.1%), Score: 134.5

**Figure 7.2b**

<table>
<thead>
<tr>
<th>Search method</th>
<th>BLASTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query</td>
<td>IrrE Deinococcus radiodurans (328 letters)</td>
</tr>
<tr>
<td>Database</td>
<td>UniProt/Swiss-Prot + DADB + PRF protein</td>
</tr>
<tr>
<td>Sequence database</td>
<td>[Last update May/03/2005]</td>
</tr>
<tr>
<td>Results</td>
<td>H37Rv</td>
</tr>
<tr>
<td>POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN protein. (Rv2017)</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>346</td>
</tr>
</tbody>
</table>

Query: 96  
Query: 156  
Query: 178  
Query: 224

Score: 57.4 bits (137), Expect = 5e-07  
Identities = 27/77 (35%), Positives = 37/77 (48%), Gaps = 14/77 (18%)

**Figure 7.2a:** An alignment of IrrE from *D. radiodurans* with Rv2017 from *M. tuberculosis*. Alignments were performed using EMBOSS-align, with water alignment, designed to align local regions of homology. Homologous amino acids are indicated with yellow shading, while similar amino acids are indicated with green shading.

**Figure 7.2b:** Results of a blast search using IrrE from *D. radiodurans*. The blast search was performed using IrrE as the query. The results included Rv2017 from *M. tuberculosis*. Regions of homology are indicated in yellow, regions of similarity are indicated in green.
DNA-damage inducible in *D. radiodurans*, it does not appear to be negatively regulated by LexA (Narumi *et al.*, 2001) as in *E. coli* (Little, 1982), despite the fact that LexA, present in *D. radiodurans*, negatively regulates gene expression and undergoes autocatalytic cleavage in the presence of activated RecA (Narumi *et al.*, 2001), as extensively described in *E. coli*.

In order to address the possibility that a repressor or activator protein was responsible for regulating gene expression of some DNA-damage inducible genes, the approach chosen was to construction *M. tuberculosis* knockouts for Rv2884 and Rv2017, to assess the phenotypic effects and the effects of the knockout on gene expression.

### 7.3 Construction of Gene inactivation knockouts

Gene inactivation knockouts were planned for Rv2017 and Rv2884, to elucidate the function of the predicted regulatory genes and to identify their regulons.

#### 7.3.1 Chromosomal location of Rv2017 and Rv2884

Firstly positional analysis of Rv2017 and Rv2884 was carried out using Tuberculist, to determine the orientation of the genes and the direction and orientation of surrounding genes (see figure 7.3a and b). This suggested that Rv2884 is mono-cistronic, whereas Rv2017 appears to be poly-cistronic. This hypothesis for Rv2017 was tested using RT-PCR. Forward and reverse primers were designed in adjacent genes to amplify the intergenic region separating the genes of interest (see figure 7.4a). If the genes were co-transcribed, forming a polycistron, these primers would produce a product from cDNA; however if the genes were mono-cistronic the intergenic regions would not be transcribed, and therefore a PCR would not produce any
Figure 7.3a

**LEGEND**
- virulence, detox, adapt
- Information pathway
- stable RNA
- intermediary metabolism
- conserved
- lipid metabolism
- cell wall, process
- IS/phage
- unknown
- conserved in *M. bovis*
- regulatory
- rRNA
- tRNA

Figure 7.3b

**LEGEND**
- virulence, detox, adapt
- Information pathway
- stable RNA
- intermediary metabolism
- conserved
- lipid metabolism
- cell wall, process
- IS/phage
- unknown
- conserved in *M. bovis*
- regulatory
- rRNA
- tRNA

Figure 7.3a: Chromosomal location of Rv2017. The data was downloaded from the *M. tuberculosis* genome website, Tuberculist. Arrows indicate the direction of transcription and are colour coded depending on their predicted functions.

Figure 7.3b: Chromosomal location of Rv2884. The data was downloaded from the *M. tuberculosis* genome website, Tuberculist. Arrows indicate the direction of transcription and are colour coded depending on their predicted functions.
Figure 7.4a: A schematic representation of the relative position of the predicted transcriptional regulatory protein, Rv2017. The orientation and relative positioning of Rv2017 is depicted with arrows and overlap/intergenic distances between upstream and downstream genes are marked. The primers used in co-transcription studies are indicated: A (Rv2016-Rv2017), B (Rv2017-Rv2018), C (Rv2018-Rv2019), and the internal control D (sigG).

Figure 7.4b: An agarose gel showing co-transcription of Rv2017 operon. PCR reactions were carried out using cDNA (track 1), a positive DNA control (track 2), and a negative control RT- (Track -). The A, B, C and D label underneath, are equivalent to the primer pairs described in figure 7.4a, and the sizes were as expected A (492bp), B (649bp), C (715bp) and D (571bp).
product (see figure 7.4b). RNA that had not been reverse transcribed to form cDNA was used as a negative control (RT-) to determine if there was any DNA contamination in the RNA sample. DNA was used as a positive control (see figure 7.4b). Primers were designed to PCR the intervening region between Rv2016 and Rv2017, Rv2017 and Rv2018, then also Rv2018 and Rv2019 (see figure 7.4a). Internal sigG primers were used as a positive control to check the integrity of the cDNA (see figure 7.4b). The size of the intergenic regions suggests Rv2016 and Rv2017 are co-transcribed (see figure 7.4a) and that Rv2018 and Rv2019 are co-transcribed (see figure 7.4a), however the intergenic distance between Rv2017 and Rv2018 make prediction difficult (see figure 7.4a). Initial RT-PCR results suggested that Rv2016, Rv2017 and Rv2018 are co-transcribed (see figure 7.4b), whereas Rv2019 is not part of the same operon. The lower level of signal from the Rv2018 to Rv2019 primers on genomic DNA may explain why RT-PCR appeared to produced no band. Due to the overlap between these genes, it is highly likely that they are co-transcribed.

7.4 Designing and testing of knockout constructs

In brief, the approach used employed the design and construction of a non-integrating plasmid containing a deletion in the coding region, replaced by a hygromycin antibiotic resistance cassette. This was electroporated into H37Rv, where homologous recombination resulted in the replacement of the functional chromosomal copy of the relevant gene with the inactivated copy present on the plasmid, as outlined in (Parish et al., 1999).

PCR primers were designed to amplify the 5’ and the 3’ coding region of Rv2017 and Rv2884, including some upstream and downstream flanking region to enable the homologous recombination to take place. These regions were cloned sequentially into pBackbone, with part
Chapter 7 Construction of a knockout in a predicted regulatory protein

of the coding region between the 5' and 3' regions forming an internal deletion of 208bp for Rv2017 and 214bp for Rv2884. A hygromycin resistance cassette was inserted between the 5' and 3' coding regions to enable selection for the knockout. During the cloning stage, numerous problems occurred with Rv2884, including, obtaining a PCR product of the correct sequence for the 3' region of Rv2884, and later on, insert the sacBlacZ cassette. As Rv2017 was the most interesting gene and time was a limiting factor, Rv2884 was abandoned and the study focused on Rv2017.

A sacBlacZ cassette was cloned into the PacI site of the targeting construct to enable blue-white screening from the lacZ gene and sucrose counter-selection from the sucrose sensitivity gene sacB in M. tuberculosis. This construct (pLD2) was then electroporated into M. tuberculosis, H37Rv strain.

The selection and counter selection process was carried out as outlined in section 2.11.3, to detect potential knockouts, exhibiting Km\(^{+}\) and Hyg\(^{+}\) phenotypes. Screening of potential knockout colonies proved particularly difficult; in the first round of selections 50 colonies were screened and 16 were Km\(^{+}\) and Hyg\(^{+}\). Of the 16 potential knockouts, 4 colonies were contaminated, the other 12 were DNA extracted (section 2.6.1) and screened by PCR using primers from the deleted region and Southern blot. The PCR using primers from the deletion gave bands for each of the 12 colonies screened (data not included), suggesting an intact copy of Rv2017 was present. The design of the Southern blot was such that 3' and 5' single crossovers were theoretically distinguishable from double crossovers by size difference (see figure 7.5). The PCR result was confirmed by the Southern blot, which produced two bands of different sizes, suggesting none of the colonies were double crossovers (see figure 7.6a).
Figure 7.5: A schematic representation of formation of single and double crossovers for the Rv2017 knockout. The sizes produced after a Clal digest, for a 3' or 5' single crossover and a double crossover event are indicated below the arrows. The positions of the probe in a Southern blot are marked, indicating both 3' and 5' single crossovers produce two bands in a Southern blot, whereas double crossover produces a single band.
Figure 7.6a: Southern blot to detect potential knockouts of Rv2017. DNA was extracted from potential knockout colonies, with the Km\(^R\), Hyg\(^R\) phenotype. *Clal* restriction digests were performed, separated using agarose gel electrophoresis, and Southern blot was performed.

Figure 7.6b: A schematic representation of the size of fragments produced in a Southern blot with *Clal* digest. Size calculations for 3' and 5' single crossover events, double crossover (mutant) events and wild-type were derived from figure 8.6a.
crossovers (see figure 7.6a and b). A further 60 white colonies were replica plated onto Km and Hyg to screen for potential knockouts; however, none were Km$^s$ and Hyg$^R$.

After going back to the original transformation, 6 blue colonies were picked and used to allow double crossovers to take place (see section 2.11.3). After serial dilutions, plates were left for 8 weeks before 100 colonies were taken and replica plated onto Km and Hyg + X-gal (to confirm loss of lacZ). A mixture of small and large colonies were tested as potential double crossovers on the Km$^s$ and Hyg$^R$ selection. Out of the 100 colonies screened only 5 were Km$^s$ and Hyg$^R$.

PCR using primers from the deletion was carried out to determine if there were any potential knockouts of Rv2017; the PCR showed 2 out of 5 colonies did not produce a product (data not included), therefore, Southern blots were performed using a Clal digest (see figures 7.7a and 7.7b) and an Xhol/Hpal digest (see figures 7.8a and 7.8b), with one of the single crossover colonies from the previous Southern used as a control along with Wild-type H37Rv (see figure 7.7b and 7.8b). Two knockouts were detected (see figure 7.7b and 7.8b).

7.5 *in-vitro* and *in-vivo* phenotyping of the Rv2017 knockout

To assess whether Rv2017, as a predicted regulatory gene, regulates expression of the DNA-damage regulon, and/or regulates virulence in a mouse model of infection, *in-vitro* and *in-vivo* phenotyping was employed to study the effects of the mutation on the response to DNA-damage and on infection in the mouse model compared to H37Rv wild-type.
Figure 7.7a: A schematic representation of the size of the product produced by a Southern blot for both Rv2017 knockout and wild-type, following a Clal digest. The positions of the Clal restriction sites are indicated with the relative size products. The probe binding site overlaps the 5' coding region of Rv2017 (yellow) and the 5' flanking region of the Rv2017 construct (green). Genomic region outside the construct is black.

Figure 7.7b: A Southern blot using Clal digest to detect potential Rv2017 knockout. Arrows indicate Rv2017 knockout, H37Rv wild-type and Rv2017 3' single crossover.
Figure 7.8a: A schematic representation of the size of the product produced by a Southern blot for both Rv2017 knockout and wild-type, following a Xhol/Hpal double digest. The positions of the Xhol and Hpal restriction sites are indicated with the relative size products. The probe binding site overlaps the 5' coding region of Rv2017 (yellow) and the 5' flanking region of the Rv2017 construct (green). Genomic region outside the construct is black.

Figure 7.8b: A Southern blot using Xhol/Hpal double digest to detect potential Rv2017 knockout. Arrows indicate Rv2017 knockout, H37Rv wild-type and Rv2017 3' single crossover.
7.5.1 Growth curve

Growth curves for the Rv2017 knockout and wild-type H37Rv were produced in-vitro under normal growth conditions, in a rolling incubator, as outlined in section (2.12.1). The Rv2017 knockout grew at the same rate as the wild-type H37Rv (see figure 7.9), suggesting the knockout has the same ability to grow and divide as the wild-type under normal growth conditions.

7.5.2 Susceptibility of the Rv2017 knockout and wild-type to DNA-damage

The DNA damaging agent mitomycin C was used to determine whether the Rv2017 knockout exhibited altered susceptibility to DNA-damage, compared with wild-type H37Rv. Viability assays were performed as outlined in methods section 2.12.2. Experiments were performed on two distinct biological replicates and duplicate colony counts were recorded for each. The viability was presented as percentage viable counts, with treated viable counts expressed as a proportion of mean untreated viable counts. The data for Rv2017 knockout was then compared with wild-type H37Rv. The Rv2017 knockout appeared to be less susceptible to mitomycin C than H37Rv wild-type, on each concentration of mitomycin C tested (see table 7.2 and figure 7.10).

Statistical analysis was performed on the data to determine whether the differences observed in viability following mitomycin C exposure were significant. An F-test was performed on the data to determine if the data was normally distributed. The F-test showed that both sets of data, Rv2017 knockout and H37Rv wild-type had significant differences in the variance at all
Figure 7.9: The *in-vitro* growth curve of ΔRv2017 strain compared to wild-type H37Rv. For each strain, single colonies were inoculated into 5ml of Dubos plus albumin in triplicate and incubated in static culture for 10 days at 37°C. Static cultures were then used to inoculate a 100ml rolling cultures in Dubos plus albumin, to an OD of 0.005 (approx 1ml). Triplicate cultures were returned to the rolling incubator for a total of 14 days. Optical density readings were taken at the same time daily, and dilutions were made in Dubos + albumin after an OD of 0.4 to enable accurate reading by the spectrophotometer.
Table 7.2

<table>
<thead>
<tr>
<th>Mitomycin C (μg/ml)</th>
<th>H37Rv</th>
<th>Rv2017 knockout</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard error</td>
</tr>
<tr>
<td>Untreated</td>
<td>100.00</td>
<td>± 12.74</td>
</tr>
<tr>
<td>0.02</td>
<td>41.16</td>
<td>± 6.89</td>
</tr>
<tr>
<td>0.05</td>
<td>7.70</td>
<td>± 0.98</td>
</tr>
<tr>
<td>0.1</td>
<td>1.74</td>
<td>± 0.46</td>
</tr>
<tr>
<td>0.2</td>
<td>0.23</td>
<td>± 0.03</td>
</tr>
</tbody>
</table>

Figure 7.10

Table 7.2: Percentage viability of ΔRv2017 strain compared to H37Rv wild-type in response to Mitomycin C stress. Duplicate cultures were grown to exponential phase (0.3 to 0.4 OD), 40ml aliquots were incubated for 24 hours untreated or with the relevant concentration of Mitomycin C. After 24 hours the cultures were serial diluted and plated on 7H11 plates in duplicate and incubated at 37°C for 16-18 days. Colony counts were taken of at least 2 dilutions. The viable CFU was then expressed as a percentage of the untreated for each sample, and the mean and standard error was calculated.

Figure 7.10: A graph comparing the of viability of the ΔRv2017 strain with H37Rv wild-type to Mitomycin C stress. In Graphpad Prism 4, the percentage viability data as outlined above was transformed using log10, and plotted on a linear scale. A non-linear regression sigmoidal dose response curve was fitted to the data.
concentrations of Mitomycin C. This therefore meant the data were not normally distributed. The data were therefore transformed to approximate a normal distribution by a log₁₀ transformation. The F-test was then performed again and this time showed no significant difference between the variances of the H37Rv data and the Rv2017 data, at the 95% confidence interval (see table 7.3), thus suggesting the data approximated a normal distribution. A student t-test was performed on the transformed data to determine whether there was a significant difference between the means when directly comparing data for H37Rv and Rv2017 knockout. Student T-tests were performed with the following hypotheses:

H₀: P value > 0.05. Accept H₀: There is no significant difference between the mean viable counts of Rv2017 knockout and H37Rv in response to mitomycin C stress at a given concentration.

H₁: P value < 0.05. Accept H₁: There is a significant difference between the mean viable counts of Rv2017 knockout and H37Rv in response to mitomycin C stress at a given concentration.

A two-tailed, unpaired T-test revealed there was a significant difference between the viability of H37Rv and Rv2017 knockout for mitomycin C stress at 0.05, 0.1 and 0.2µg/ml mitomycin C (95% confidence interval) see table 7.3. No significant difference was observed at the 95% confidence interval (student t-test) between wild-type and Rv2017 knockout for untreated and 0.02µg/ml mitomycin C.
Table 7.3: Statistical analysis of the response to Mitomycin C stress of the Rv2017 knockout compared to H37Rv wild-type. The percentage viability for ΔRv2017 and H37Rv was transformed using log10. An F-test was performed to determine if both samples had equal variance (H0) (a requirement for parametric tests), when the p-value >0.05, H0 is accepted, the populations have equal variance and a t-test is appropriate. Then unpaired, two-tailed t-tests were performed between each sample for both ΔRv2017 and H37Rv wild-type, under each stress condition individually, to determine if the means for each sample were the same (H0) or the means of ΔRv2017 and H37Rv were different (H1). Where the p-values <0.05, H0 is rejected and H1 is accepted: there is a significant difference between the means.

<table>
<thead>
<tr>
<th>Table Analyzed</th>
<th>Untreated</th>
<th>0.02μg/ml MitC</th>
<th>0.05μg/ml MitC</th>
<th>0.1μg/ml MitC</th>
<th>0.2μg/ml MitC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transformed log10</td>
<td>Transformed log10</td>
<td>Transformed log10</td>
<td>Transformed log10</td>
<td>Transformed log10</td>
</tr>
<tr>
<td>H37Rv wild-type</td>
<td>H37Rv U</td>
<td>H37Rv 0.02</td>
<td>H367Rv 0.05</td>
<td>H37Rv 0.1</td>
<td>H37Rv 0.2</td>
</tr>
<tr>
<td>vs</td>
<td>vs</td>
<td>vs</td>
<td>vs</td>
<td>vs</td>
<td>vs</td>
</tr>
<tr>
<td>Rv2017 Knockout</td>
<td>Rv2017 U</td>
<td>Rv2017 0.02</td>
<td>Rv2017 0.05</td>
<td>Rv2017 0.1</td>
<td>Rv2017 0.2</td>
</tr>
<tr>
<td><strong>Unpaired t test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td>0.8767</td>
<td>0.0913</td>
<td>0.0001</td>
<td>P&lt;0.0001</td>
<td>P&lt;0.0001</td>
</tr>
<tr>
<td>P value summary</td>
<td>ns</td>
<td>ns</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Significant difference at (P &lt; 0.05)</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>One- or two-tailed P value?</td>
<td>Two-tailed</td>
<td>Two-tailed</td>
<td>Two-tailed</td>
<td>Two-tailed</td>
<td>Two-tailed</td>
</tr>
<tr>
<td>t, df</td>
<td>t=0.1585 df=12</td>
<td>t=1.836 df=12</td>
<td>t=5.325 df=13</td>
<td>t=6.812 df=16</td>
<td>t=6.943 df=9</td>
</tr>
<tr>
<td><strong>F test to compare variances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F,DFn, Dfd</td>
<td>2.648, 5, 7</td>
<td>2.332, 7, 5</td>
<td>3.174, 6, 7</td>
<td>1.650, 7, 9</td>
<td>9.568, 7, 2</td>
</tr>
<tr>
<td>P value</td>
<td>0.2366</td>
<td>0.3682</td>
<td>0.1564</td>
<td>0.4749</td>
<td>0.1957</td>
</tr>
<tr>
<td>P value summary</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Are variances significantly different</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
7.5.3 *in-vivo* phenotyping of the Rv2017 knockout compared to wild-type H37Rv

A mouse model of infection was used to determine the *in-vivo* phenotype of the Rv2017 knockout in comparison to wild-type H37Rv. Three month old female mice were injected intravenously in the tail vain with 200μl of each strain of bacteria by a trained animal technician (see section 2.12.3). Initial infection colony counts (CFU) were determined from the inoculum, by serial dilution and plating of the inoculum (see table 7.4); the CFU varied slightly, even though each strain was grown to an optical density (OD) of 0.3-0.4, then diluted to an OD of 0.02 in DAG (Dubos + albumin + glycerol).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Viable counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>H37Rv Wild-type</td>
<td>3.18E+05</td>
</tr>
<tr>
<td>Rv2017 Knockout</td>
<td>2.38E+05</td>
</tr>
</tbody>
</table>

**Table 7.4: Viable CFU counts of H37Rv wild-type and Rv2017 knockout**

The infection inoculum was serial diluted in saline in duplicate, and plated onto 7H11 plates. Viable colony counts were taken after 13-18 days.

Pathology of the infection was monitored using bacterial load calculated from harvested lungs and spleen obtained from 4 mice at days 1, 34, 68, 103 and 146 (see figure 7.11a and 7.11b). There was no difference observed between the H37Rv wild-type and the ΔRv2017 strain at day 1 or day 34; however, after this there is a significant increase in CFU in the lung in the ΔRv2017 strain compared to H37Rv wild-type (see figure 7.11a). There were problems with the initial plating of the 103 day time point, in that there were no colonies observed; therefore
**Figure 7.11a:** *In-vivo* mouse data comparing ΔRv2017 strain with H37Rv. The data was obtained from intravenous infections of 3 month old BALBC female mice with the ΔRv2017 strain and H37Rv wild-type of *M. tuberculosis*. The lung and spleen were harvested from 4 mice at day 1, 34, 68, 103 and 146, after infection, and colony forming units (CFU) were counted, however there were problems with CFU counts for the wild-type, after day 34, and for the ΔRv2017 strain in the lung on day 148.

**Figure 7.11b:** *In-vivo* mouse data comparing mean spleen and lung weights for the ΔRv2017 strain and H37Rv, without re-spotted wild-type.
the organ stock for the H37Rv wild-type was re-plated and colony counts were obtained. However, there are potential problems with the viability of frozen and re-streaked samples. Therefore, in order to determine the magnitude of the loss of viability, the lung and spleen counts for day 103 and day 146, were re-plated for both ΔRv2017 strain and H37Rv, however CFU data was not available due to technical problems in the animal facility. Nevertheless, the organ weights were available, and therefore were plotted for both the lung and spleen at each time point (see figure 7.1b). From the CFU data from initial time points, and the organ weight data, it appears that the ΔRv2017 strain may be hypervirulent in the mouse model of infection.

7.6 Discussion

Five predicted regulatory proteins were up regulated in response to DNA-damage in the ΔrecA strain of M. tuberculosis, suggesting their regulation is independent of the RecA/LexA system, implicating the possibility that one or more of these predicted regulatory proteins may play a role in an alternative mechanism of regulation of the DNA-damage response.

7.6.1 Functional classification of the five predicted regulatory proteins

These five predicted regulatory proteins show similarities to different families of transcriptional regulators. The annotated genome sequence of M. tuberculosis (TubercuList), together with homology searches and homology based domain predictions (NCBI website), suggests that Rv0586 is a predicted regulatory protein of the GntR family of transcriptional regulators, possessing a probable GntR signature domain and a helix-turn-helix domain. For Rv1956, Tuberculist did not indicate homology to a specific family of transcriptional regulators, yet,
Rv1956 contains a probable helix-turn-helix domain, which shows homology to the XRE family of transcriptional regulators, that respond to xenobiotics and is reported to be upregulated in response to starvation. Rv1985c is predicted to be part of the LysR family of transcriptional regulators, possessing a LysR signature domain and a helix-turn-helix domain. Rv2017 is not assigned to a particular family of transcriptional regulator, but, shows N-terminal similarity to a number of transcriptional regulators belonging to the XRE family and shows C-terminal homology to a zinc binding domain, suggesting a role as a metalloprotease. The N-terminal region of Rv2884 shows similarity to transcriptional regulators of sensory transduction systems, which exhibits some homology to OmpR, a stress response regulator present in E. coli; interestingly M. tuberculosis lacks an OmpR homologue.

Although (Earl et al., 2002) suggested Rv2017 from M. tuberculosis and irrE D. radiodurans share homology, blast searches using Rv2017 protein as the query failed to detect IrrE, whereas blast searches using IrrE sequence as the query pulled out Rv2017 (E-value of 5e⁻⁷) (figure 7.2a and b). The alignments produced by both EMSOSS (Figure 7.2a) and BLASTP (Figure 7.2b) show a region of 31 residues of homology, which corresponds to the DUF955 conserved region. Protein blast (BLASTP) search using Rv2017 also showed homology to GI: 29345183, a zinc binding transcriptional regulator of the Cro/Cl family, from Enterococcus faecalis V583 (E-value = 3e⁻³), as well as homology to Rv2515c (E-value = 3e⁻⁸) from M. tuberculosis and Mb2040 (E-value = 5e⁻⁷) from M. bovis. Rv2515c is particularly interesting, it is listed as a hypothetical protein in TubercuList, and was shown to be essential by transposon mutagenesis by (Sassetti et al., 2001) it too contains a neutral zinc metallopeptidase, zinc binding region signature.
7.6.2 Analysis of the Rv2017 gene inactivation knockout

The gene inactivation knockout of the predicted regulatory proteins (Rv2017) was used to study the effect of the mutation on gene expression and phenotype, to determine whether it was involved in regulation of DNA-damage repair, or pathogenesis. Prior to the construction of the gene inactivation knockout in Rv2017, chromosomal locational and transcriptional analysis revealed that Rv2017 was part of a polycistron (Figure 7.3a and Figure 7.4a and b). This is important information in designing and constructing a knockout, as within polycistrons, inactivation via deletion and insertion may have polar effects on any downstream genes. In order to determine whether any phenotypic effects are solely due to the gene inactivation, a complementing construct should be analysed alongside the knockout. Complementation with a construct re-instating only the inactivated (Rv2017) gene will determine whether the effects observed are due to the gene inactivation, or as a result of the polar effects on the downstream genes. To assess any potential polar effects, RT-PCR can be performed using RNA from both the mutant and wild-type strains to determine if there are any differences in gene expression of other genes in the polycistron. The 3' region of Rv2017 was retained in the knockout construct to help limit the potential for polar effects: the reason being the post-transcriptional modification of the mRNA should remain intact if the regions recognized by the post-translational modification process remain intact. Due to time constraints, a complementing construct was not produced for ΔRv2017 strain, but is vital for checking the effects of the Rv2017 gene inactivation knockout.

The first round of screening for an Rv2017 knockout revealed only single crossover events (figure 7.6a). Successful knockout strains of Rv2017 were not detected until the length of incubation for the double crossover stage of selection had been allowed to proceed for 8 weeks.
This may suggest the knockout has attenuated growth; however, later growth curves produced suggest that in liquid culture there was no difference between the H37Rv wild-type and the \(\Delta Rv2017\) strain (Figure 7.9). This discrepancy is difficult to explain.

The design of the Southern blot used in detection of the \(\Delta Rv2017\) strain enabled one to distinguish between single crossover events, random integrations, and the required double crossover events. However, due to the nature of the Southern blot, there are slight discrepancies between predicted and observed sizes for both wild-type and knockout strains. The Southern blot size estimation relies on the DNA marker run along side digested DNA samples, on an agarose gel. The gel is photographed under UV illumination with a fluorescent ruler, which is used to estimate sizes after autoradiography. Therefore the slight discrepancies in size are most likely due to human error when measuring the distance on the autoradiograph and comparing this to the gel picture. Nevertheless, two different digests show a continuous pattern for the Rv2017 knockout strains marked 1 and 2 on figures 7.7b and 7.8b, indicating that a knockout has been successfully obtained.

As previously stated there appears to be no difference between the in-vitro growth of the \(\Delta Rv2017\) strain compared to H37Rv wild-type, suggesting the deletion of Rv2017 does not affect the ability to divide and replicate under normal in-vitro growth conditions. However the \(\Delta Rv2017\) strain is hyposensitive to treatment with the chemical DNA damaging agent mitomycin C, compared to wild-type H37Rv. This difference in sensitivity is significant with treatment of greater than or equal to 0.05\(\mu\)g/ml mitomycin C. This may suggest that Rv2017 acts as a repressor protein of genes involved in the repair of interstrand cross links, such as those produced by mitomycin C. Possible candidates would include genes involved in NER, as NER mutants (\textit{uvrA}, \textit{uvrD}) show increased susceptibility to mitomycin C (Rand, 2003),
therefore in the absence of the repressor, these genes may be constitutively produced and so may act more quickly to repair such interstrand cross links between alkylated bases.

An alternative hypothesis is that Rv2017 may regulate genes involved in permeability, such as protein channels. Thus, knocking out Rv2017 could make the mutant strain less permeable to exogenous protein/chemicals. The hyposensitivity of Rv2017 to mitomycin C could be due to the regulon of Rv2017.

Rand, (2003) observed that mutants in base excision repair (BER) pathways showed decreased susceptibility to mitomycin C; however, there was no significant difference between the susceptibility of the BER knockouts and the wild-type H37Rv, suggesting that inter-strand cross links were not repaired via BER. Mutants in nucleotide excision repair were more susceptible to mitomycin C than wild-type H37Rv. There are three main processes by which damaged DNA is repaired, depending on the type of DNA-damage. BER and NER removed damaged bases directly, using a different mechanism, and DNA-damage across DNA strands is repaired by homologous recombination. BER removes oxidised and alkylated bases (Mizrahi and Andersen 1998), whereas NER, studied in E. coli is performed by the exinucleases uvrABC and D, which function to excise pyrimidine dimers in damaged DNA (Freidberg et al., 1994).

Due to the apparent insusceptibility of ΔRv2017 strain to mitomycin C stress, it would be interesting to look at the effects of other DNA damaging agents, which target specific DNA repair pathways in the ΔRv2017 strain. During the infection process, M. tuberculosis is phagocytosed by macrophages, which develop into granulomas, producing a hostile environment for the bacterium. The macrophages produce reactive oxygen and nitrogen intermediates, known to damage DNA and the granuloma provides an hypoxic environment,
lacking in nutrients. Hydrogen peroxide and cumene hydroperoxide are reduced to superoxide radicals, known to damage DNA. Therefore, these could be used to determine the viability of ΔRv2017 strain. The bacteria are also subject to nitrosative stress in macrophages, which could be mimicked using sodium nitrite at acidic pH, to determine susceptibility of the ΔRv2017 strain.

An alternative approach to looking at the \textit{in-vitro} susceptibility of the ΔRv2017 strain to stresses in liquid culture would be to look at the viability of the ΔRv2017 strain in \textit{in-vitro} macrophage infections. Graham et al., (1999) showed that genes involved in DNA repair and virulence are induced in macrophages, to potentially cope with the hostile environment. The heat shock induced protein α-crystallin is upregulated under low oxygen tension, and is required for growth in macrophages. Therefore, activated bone marrow derived macrophages produce a number of DNA damaging agents, which would better mimic the intracellular environment that \textit{M. tuberculosis} is exposed to during infection.

Animal models have been developed to mimic the effects of \textit{in-vivo} infection. After 3 time points there appears to be no difference between H37Rv wild-type and ΔRv2017 strain in the spleen, yet there was a significant difference in CFU in the lung at day 68 in ΔRv2017 strain compared to H37Rv (t-test p<0.05). The log difference in CFU observed between the lung and the spleen can be accounted for due to the method of infection: intravenous injection results in an initially higher CFU in the spleen than the lungs. In a human infection, the aerosol route of infection determines that the CFU is highest in the lungs; therefore the only way to mimic this would be to use an aerosol route of infection, which is currently not available at our animal facility.
8 Identification of the Rv2017 regulon by microarray analysis

8.1 Introduction

Microarray analysis performed on ΔrecA strain of M. tuberculosis revealed that the majority of genes in the DNA-damage regulon were regulated in a RecA/LexA independent manner (Rand et al., 2003). It was hypothesised that another regulatory protein, such as an activator/repressor may be involved in regulation of expression of some of the genes in the DNA-damage regulon. Rv2017 appeared to be a potential candidate, due to the higher levels of induction observed in ΔrecA strain compared to H37Rv (Rand et al., 2003) and the homology of Rv2017 to IrrE of Deinococcus radiodurans (D. radiodurans), shown to positively regulate recA (Earl et al., 2002).

Microarray experiments were designed to address the possibility that Rv2017 was involved in regulation of genes in the DNA-damage regulon. The experimental design was such that inter- and intra-strain comparisons could be made between the level of gene expression under uninduced and induced (0.02μg/ml mitomycin C) conditions. The advantages of using RNA versus DNA arrays are outlined in chapter 4.

The microarray slides used, were PCR spotted whole genome M. tuberculosis arrays, produced by the BµGS microarray unit at St. George’s Hospital Medical School. Triplicate RNA samples were harvested from H37Rv and ΔRv2017 strains of M tuberculosis, that were either uninduced or induced with 0.02μg/ml mitomycin C in exponential phase (OD 0.15). Both the uninduced and induced cultures were incubated for a further 24 hours after the addition of the chemical
DNA damaging agent mitomycin C. DNA contamination was removed by DNase treatment for the H37Rv strain; however problems occurred with the DNase treatment of the ΔRv2017 strain, resulting in degradation of the RNA. Therefore the small amount of sample retained, that was not DNase treated was analysed for the presence of DNA contamination using PCR. A PCR was performed on all the ΔRv2017 strain RNA samples alongside a DNA control to detect any potential DNA contamination. The PCR showed that the ΔRv2017 RNA samples were not contaminated with DNA (see figure 8.1). The RNA samples for both H37Rv wild-type and ΔRv2017 strain were then analysed using a Bioanalyser (Agilent Technologies) to determine the quantity and quality of the RNA preparation (data not shown).

8.2 Analysis and input of the microarray slides

For the microarray experiments, each triplicate biological sample was used in duplicate for both the uninduced and induced (mitomycin C 0.02μg/ml) cultures of H37Rv wild-type and the ΔRv2017 strain, resulting in a minimum of 6 slides per strain, per treatment. (Slides were repeated where either problems with scratching of the slide or background fluorescence occurred).

The slides were scanned and analysed as outlined in the methods. The spread of the data was determined using Genespring to produce a graph of the normalised data plotted against the strain, and induction (+/− mitC), to enable a comparison to be made between H37Rv and the ΔRv2017 strain (see figure 8.2). The spread of the data determines the type of statistical analysis performed on the data. Therefore the unequal spread of the data suggests the variances are not equal for H37Rv and ΔRv2017 strains, therefore the assumption was used that the
Figure 8.1: PCR to detect DNA contamination in ΔRv2017 RNA extractions. 1μg of RNA was used in the PCR reaction, for each of the biological replicates (1-3) of uninduced (tracks 1,3 and 5) and induced (0.02μg/ml mitomycin C) (tracks 2,4 and 6) samples from ΔRv2017 strain of *M. tuberculosis*. Genomic *M. tuberculosis* DNA was used as a positive control (track 7), negative control (track 8). The samples were run alongside commercial markers M1 (1kb ladder) and M2 (100bp ladder).
Figure 8.2: Schematic representation of the spread of microarray data for H37Rv, ΔsigG and ΔRv2017 strains of *M. tuberculosis*. PCR spotted *M. tuberculosis* whole genome microarrays (Btgs Unit, St. Georges Hospital) were performed to compare expression of genes in H37Rv, ΔsigG and ΔRv2017 strains of *M. tuberculosis*. Arrays were performed as two colour experiments cy3/cy5 arrays, using DNA versus RNA respectively. The strain is marked under each of the three figures, and the data goes from uninduced (-) on the left, to the induced (+MitC) 0.02μg/ml mitomycin C, on the right. The data shown is comprised of three biological replicates each with two technical replicates for each strain under each condition.
variances of the data were unequal, which required the use of the Welsh’s approximation to the Student T-test.

8.2.1 Statistical analysis of microarray data

The microarray data was analysed in two ways, to answer two different questions. The first was to identify whether Rv2017 played a role in the regulation of genes induced by DNA-damage, identified in the H37Rv wild-type as genes that were 2-fold or more up-regulated in response to DNA-damage (mitomycin C stress). The second strategy was to look globally at genes regulated by Rv2017 under either uninduced or induced conditions, and determine whether there was any overlap between the two groups.

8.2.1.1 Rv2017 and the genes involved in the DNA-damage response

To address the possibility that Rv2017 was involved in gene regulation following DNA-damage, a list of genes induced two fold or more in the H37Rv wild-type in response to DNA-damage, was collated using the ‘filtering on fold change’ option in Genespring, with the minimum cut off value of 2-fold induction. The resultant gene list, as outlined in chapter 4, contained 118 genes, whose induction ratio was greater than or equal to 2, and for which expression in the induced conditions compared to uninduced conditions in H37Rv was significantly different, with a p value P<0.01 (Student T-test).

The normalised data for both H37Rv and ΔRv2017 from the list of 118 DNA-damage inducible genes were exported into Excel, where calculations were performed to determine the mean, standard deviation and standard error for both the uninduced and induced values for each strain
individually. Any standard errors greater than 3 were highlighted, and the individual values were assessed. In some cases, spurious data spots had not been effectively removed in Gene Pix, which could be seen as huge numbers, usually in one out of the 6 replicates. When the data for 5 out of the 6 biological and technical replicates were within one standard deviation of the mean, and the last replicate was greater than 3 standard deviations away from the mean, these data points were removed as errors.

After the removal of the outliers, the induction ratios were calculated for all 118 genes, using induced/uninduced for all the individual biological and technical replicates. The mean, standard deviation and standard error were then calculated for the induction ratios of both H37Rv and ΔRv2017 strains. The induction ratios were transformed using log10, before a two-tailed Welsh’s approximation to a T-test was used with a p value p<0.01, to calculate whether there was a significant difference between the induction ratio of H37Rv compared to ΔRv2017 strain. A total of 59 genes were identified with significantly different induction ratios, all of which were decreased in ΔRv2017 compared to H37Rv (see figure 8.3); therefore 59 genes were also identified whereby there was no significant difference in induction ratio between ΔRv2017 strain and H37Rv. However, potential pitfalls of using the induction ratio alone were identified in chapter 4. Therefore the data were further analysed to determine which genes were significantly different in ΔRv2017 strain compared to H37Rv under uninduced conditions and then induced conditions, using a two-tailed Welsh’s approximation to a T-test, with a p value p<0.01, performed using transformed uninduced and induced data (log10).

There are 30 genes with a significantly different induction ratio, which are also significantly different under induced conditions (see figure 8.3 and 8.4). The induction ratios are all significantly lower (p<0.01), as are the induced values; the normalised data in the bottom graphs
Figure 8.3: Schematic representation of the analysis of genes in the DNA damage regulon. The DNA damage regulon was defined by those genes that were regulated 2-fold or more in response to DNA damage (mitomycin C stress 0.02µg/ml) in H37Rv wild-type, with a p≤0.01 (Students T-test). This data was then compared to the ΔRv2017 strain. The induction ratio was calculated (induced/uninduced) for both H37Rv and ΔRv2017 strains, a T-test was then performed p≤0.01 to determine which genes had a significantly different induction ratio in ΔRv2017 compared to H37Rv. These were then further analyzed to determine which genes were significantly different under induced conditions between H37Rv and ΔRv2017, then under uninduced conditions using a Student T-test p≤0.01. The overlap between the induced and uninduced comparisons is also shown.
Figure 8.4: 30 genes with significantly different induction ratios, also significantly different under induced conditions. The data were obtained from DNA versus RNS microarrays with H37Rv and ΔRv2017 strain of M. tuberculosis under both uninduced and induced (0.02μg/ml mitomycin C) conditions. The induction ratios were calculated for both H37Rv and ΔRv2017, as induced/uninduced, a Student T-test was performed p<0.01, top graphs. The induced values for each strain were also significantly different using Student T-test p<0.01, bottom graphs.
shows the induced levels are significantly lower in ΔRv2017 strain compared to H37Rv. In this case, the induction ratios and the normalised data tie in together, and show a significant difference in expression under induced conditions, suggesting that these genes are directly or indirectly regulated by Rv2017, in response to the DNA damaging agent mitomycin C (0.02μg/ml).

Of the 59 genes that have a significantly decreased induction ratio, 11 of these are significantly different under uninduced conditions (see figure 8.3 and 8.4). Although the induction ratios are significantly different (p<0.01), when analysing the normalised data, it appears that these genes are not induced in ΔRv2017 strain (figure 8.5, bottom graphs). Rather, the uninduced level of expression appears to be significantly higher in ΔRv2017 strain than in H37Rv. Therefore the overall expression level in the induced is similar for Rv0861c, ligB, Rv2413c, Rv3517 and Rv0094c, but the induction ratio is significantly lower due a higher level of expression under uninduced conditions (see figure 8.5). The high level of expression of all the 11 genes under uninduced conditions could be directly or indirectly due to Rv2017, whereby Rv2017 may act as a repressor, or may regulate a repressor protein, which, under uninduced conditions would normally suppress transcription. Therefore, if Rv2017 regulates expression of a repressor protein, this protein would not be expressed in ΔRv2017 strain, therefore, the repressors’ regulon would no longer be suppressed, and the level of expression would be similar under both uninduced and induced conditions.

Surprisingly, 5 genes have a significantly decreased induction ratio, but neither the uninduced, nor the induced values are significantly different (see figure 8.3 and 8.6). Further analysis of the induced conditions for the ΔRv2017 and H37Rv strains, revealed that although the induced values are not significantly different at p<0.01, they are significantly different at p<0.05 (see
Figure 8.5: 11 genes with significantly different induction ratios, also significantly different under uninduced conditions. The data were obtained from DNA versus RNA microarrays with H37Rv and ARv2017 strain of M. tuberculosis under both uninduced and induced (0.02μg/ml mitomycin C) conditions. The induction ratios were calculated for both H37Rv and ARv2017, as induced/uninduced, a Student T-test was performed p<0.01, top graph. The uninduced values for each strain were also significantly different using Student T-test p<0.01, bottom graph.
Figure 8.6: 5 genes with significantly different induction ratios, which are not significantly different under uninduced or induced conditions. The data were obtained from DNA versus RNA microarrays with H37Rv and ΔRv2017 strain of *M. tuberculosis* under both uninduced and induced (0.02μg/ml mitomycin C) conditions. The induction ratios were calculated for both H37Rv and ΔRv2017, as induced/uninduced, a Student T-test was performed p<0.01, top graph. The uninduced values for each strain were not significantly different using Student T-test p>0.01, bottom graph. Asterisks mark the induced values that are significantly different in ΔRv2017 compared to H37Rv, with p<0.05. The uninduced values are all significantly different at p<0.05.
asterisks figure 8.6); therefore this may be due to the stringency of the statistical test. The normalised data shows an apparent decrease in the level of expression under induced conditions for all of the 5 genes (see figure 8.6). Therefore, the differences observed in the induction ratio and the normalised data appear to tally, in that, expression level is decreased in ΔRv2017 strain, suggesting that Rv2017 may directly or indirectly regulate expression of these genes, in a similar manner to the first group outlined above.

Finally, of the 59 genes with a significantly different induction ratio, there are 13 genes, which are significantly different under both uninduced and induced conditions (see figure 8.3 and 8.7). In the majority of cases, the decrease in induction ratio reflects the decrease in expression under induced conditions, except for Rv3467 and Rv0765c, where the level of expression is increased under both uninduced and induced conditions. This results in an induction ratio that appears to be decreased in ΔRv2017 strain, where the actual expression levels, given by the normalised data (see figure 8.7, bottom graph), shows an increase in expression. This is an example of where it is advisable to look at both the uninduced and induced conditions separately. The majority of the 13 genes also have a significantly increased expression level under uninduced conditions (except Rv2979c and Rv2791c). The increased level of expression under uninduced conditions may skew that data towards a smaller induction ratio; however, looking at the normalised data alongside the induction ratio shows that the decrease in expression in ΔRv2017 strain is consistent with the decreased induction ratio, in every case, except Rv3467 and Rv0765c.

Interestingly, there are a large proportion of genes that do not have a significantly different induction ratio, but are significantly different under induced and uninduced conditions (see figure 8.3). Of the 59 genes without a significantly different induction ratio, 26 are significantly
Figure 8.7: 13 genes with significantly different induction ratios, also significantly different under both induced and uninduced conditions. The data were obtained from DNA versus RNA microarrays with H37Rv and ΔRv2017 strain of *M. tuberculosis* under both uninduced and induced (0.02μg/ml mitomycin C) conditions. The induction ratios were calculated for both H37Rv and ΔRv2017, as induced/uninduced, a Student T-test was performed *p*<0.01, top graph. The induced and uninduced values for each strain were also significantly different using Student T-test *p*<0.01, bottom graph. The + symbol indicates genes expressed to a higher level in the ΔRv2017 strain.
different under induced conditions (see figure 8.8). All of these genes have a significantly lower expression level under induced conditions in ΔRv2017 strain compared to H37Rv. It seems odd that these genes are expressed to a significantly different level under induced conditions; yet do not have a significantly different induction ratio. However, when the Welsh’s approximation to the Student T-test is performed with p<0.05, 11 of these genes do in fact have a significantly different induction ratio, with the less stringent test. Therefore, it is possible that some of the important differences in expression level would be missed purely using the induction ratio. Even though the induction ratios are not significantly different at p<0.01, it appears that the induced values are significantly lower at p<0.01, which suggests that Rv2017 has a direct or indirect effect on expression of these genes in response to the DNA damaging agent mitomycin C (0.02µg/ml). Either Rv2017 is an activator protein, that positively regulates gene expression in response to the DNA damaging agent mitomycin C (0.02µg/ml), or Rv2017 positively regulates one or more activator proteins responsible for transcription of these DNA-damage inducible genes.

There are only 4 genes, which are significantly different under uninduced conditions that do not have a significantly different induction ratio (see figure 8.3 and 8.9). The levels of expression appear to be the same for three of the genes (Rv3836, Rv1084 and Rv2974c), under both induced and uninduced conditions in the ΔRv2017 strain, with the uninduced expression levels significantly higher under in ΔRv2017 strain compared to H37Rv (see figure 8.9), suggesting that these genes may be directly or indirectly regulated by Rv2017, whereby Rv2017 either acts as a repressor protein, to suppress transcription, or activates one or more repressor proteins, to suppress transcription under uninduced conditions.
Figure 8.8: 26 genes without significantly different induction ratios, but which were significantly different under induced conditions. The data were obtained from DNA versus RNA microarrays with H37Rv and ΔRv2017 strain of *M. tuberculosis* under both uninduced and induced (0.02 mg/ml mitomycin C) conditions. The induction ratios were calculated for both H37Rv and ΔRv2017, as induced/uninduced, a Student T-test was performed p<0.01 and none were significantly different, top graph. However the induced values for each strain were significantly different using Student T-test p<0.01, bottom graph. Asterisks mark the induction ratios that are significantly different at p<0.05.
Figure 8.9: 4 genes without significantly different induction ratios, but which were significantly different under uninduced conditions. The data were obtained from DNA versus RNA microarrays with H37Rv and ΔRv2017 strain of M. tuberculosis under both uninduced and induced (0.02μg/ml mitomycin C) conditions. The induction ratios were calculated for both H37Rv and ΔRv2017, as induced/uninduced, a Student T-test was performed p<0.01, which was not significantly different, top graph. However, the uninduced values for each strain were significantly different using Student T-test p<0.01, bottom graph. The asterisks mark the induction ratios that are significantly different at p<0.05.
It was anticipated that if the induction ratios were not significantly different, then the individual uninduced and induced conditions would not be significantly different. However, this was the case for only a proportion of the 59 genes without significantly different induction ratios: 20 genes fell into this category (see figure 8.3 and 8.10). The majority of these genes have normalised expression levels for uninduced and induced conditions to reflect the non-significantly different induction ratios. However, there appeared by eye, to be a few exceptions; therefore, the Welsh’s approximation to the T-test was performed with cut off p<0.05, the induction ratios that were significantly different with the less stringent cut off value are marked with an asterisk (see figure 8.10).

Finally for the genes without a significantly different induction ratio, there are a total of 9 genes which are significantly different under both induced and uninduced conditions, even though the induction ratio is not significantly different for these genes, the overall expression level under both uninduced and induced conditions is significantly decreased p<0.01 for all the genes (see figure 8.11). Therefore, these genes are still being induced to the same extent proportionally, but the levels are significantly lower in ΔRv2017 strain. This suggests that Rv2017 directly or indirectly affects the intrinsic ability of these genes to be expressed. This means that there is proportionally less transcript produced for recA, ruvC and other 7 DNA-damage inducible genes. This highlights the importance of analysing both the normalised expression levels and the induction ratios simultaneously.

### 8.2.1.2 Analysis of the Rv2017 regulon

To determine whether Rv2017 directly or indirectly regulates expression of genes besides those in the DNA-damage regulon, statistical analysis in Genespring was used to generate genes lists
Figure 8.10: 20 genes without significantly different induction ratios, which are not significantly different under uninduced or induced conditions. The data were obtained from DNA versus RNA microarrays with H37Rv and ΔRv2017 strain of *M. tuberculosis* under both uninduced and induced (0.02μg/ml mitomycin C) conditions. In the top figure the large standard errors required a different scale. The induction ratios were calculated for both H37Rv and ΔRv2017, as induced/uninduced, a Student T-test was performed p<0.01, which was not significantly different, top graph. When a student T-test was performed for both the uninduced and induced conditions separately, there was no significant difference. Asterisks mark the induction ratios were the t-test was significant at p<0.05.
Figure 8.11: 9 genes without significantly different induction ratios, yet significantly different under both induced and uninduced conditions. The data were obtained from DNA versus RNA microarrays with H37Rv and ARv2017 strain of M. tuberculosis under both uninduced and induced (0.02μg/ml mitomycin C) conditions. The induction ratios were calculated for both H37Rv and ARv2017, as induced/uninduced, a Student T-test was performed p<0.01, were not significantly different, top graph. The induced and uninduced values for each strain were significantly different using Student T-test p<0.01, bottom graph. Asterisks indicate the induction ratios that are significantly different at p<0.05.
to determine whether there was any significant difference between the expression level of the genes in ΔRv2017 strain compared to H37Rv under uninduced conditions and induced conditions separately. The Welch's approximation to a T-test was performed using $p<0.01$, and the Benjamini and Hochberg false discovery rate (FDR) correction was applied to the data. There were a total of 1347 genes that were significantly different under uninduced conditions between ΔRv2017 strain compared to H37Rv, and 740 genes that were significantly different under induced conditions in ΔRv2017 strain compared to H37Rv. An overlap of 595 genes were significantly different $p<0.01$ with FDR correction in ΔRv2017 strain compared to H37Rv (see figure 8.12). Of these 595 genes, 224 were significantly decreased ($p<0.01 + \text{FDR}$) in both uninduced and induced conditions in ΔRv2017 strain compared to H37Rv, 366 were significantly increased ($p<0.01 + \text{FDR}$) in ΔRv2017 strain compared to H37Rv under both uninduced and induced conditions, and 5 were significantly decreased ($p<0.01 + \text{FDR}$) under only induced conditions in ΔRv2017 strain compared to H37Rv.

The 595 genes that were significantly different under both uninduced and induced conditions were classified according to functional group (see table 8.1). They covered a broad range of categories, with the majority of genes involved in small molecule metabolism. Interestingly, there were 31 broad regulatory genes in this group, comprised of 24 predicted activator/repressor proteins, 5 two-component regulatory proteins, and 2 serine-threonine protein kinases (see table 8.1). These are potentially the most interesting, therefore, these were further grouped into those that were decreased in the ΔRv2017 strain compared to H37Rv (see figure 8.13a), and those that were increased in ΔRv2017 strain compared to H37Rv (see figure 8.13b). The level of expression of these regulatory genes appears to be the same in uninduced
Figure 8.12: Venn diagram showing the genes significantly different in ΔRv2017 strain compared to H37Rv wild-type. The diagram was produced in Genespring using the microarray data for the DNA versus RNA arrays from ΔRv2017 and H37Rv strains of *M. tuberculosis*, under uninduced and induced (0.02μg/ml mitomycin C) conditions. Gene lists were produced to identify genes that were significantly different in ΔRv2017 compared to H37Rv under uninduced conditions, then separately under induced conditions. The significant differences were determined using a Student’s T-test, performed with cut off p<0.01, with the Benjamini and Hochberg false discovery rate correction.
Table 8.1

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Small molecule metabolism</strong></td>
<td></td>
</tr>
<tr>
<td>Degradation</td>
<td>27</td>
</tr>
<tr>
<td>Energy metabolism</td>
<td>75</td>
</tr>
<tr>
<td>Central intermediary metabolism</td>
<td>8</td>
</tr>
<tr>
<td>Amino acid biosynthesis</td>
<td>16</td>
</tr>
<tr>
<td>Polyamine synthesis</td>
<td>0</td>
</tr>
<tr>
<td>Purines/pyrimidines/nucleotides</td>
<td>9</td>
</tr>
<tr>
<td>Biosynthetic co-factors</td>
<td>14</td>
</tr>
<tr>
<td>Lipid biosynthesis</td>
<td>9</td>
</tr>
<tr>
<td>Polyketide</td>
<td>8</td>
</tr>
<tr>
<td>Broad regulatory function</td>
<td>31</td>
</tr>
<tr>
<td>Activator/repressors (24)</td>
<td></td>
</tr>
<tr>
<td>Two component systems (5)</td>
<td></td>
</tr>
<tr>
<td>Serine-threonine protein kinases (2)</td>
<td></td>
</tr>
<tr>
<td><strong>Macromolecular metabolism</strong></td>
<td>119</td>
</tr>
<tr>
<td>Synthesis and modification of macromolecules</td>
<td>51</td>
</tr>
<tr>
<td>Degredation of macromolecules</td>
<td>9</td>
</tr>
<tr>
<td>Cell envelope</td>
<td>59</td>
</tr>
<tr>
<td><strong>Cell processes</strong></td>
<td>44</td>
</tr>
<tr>
<td>Transport(binding sites)</td>
<td>27</td>
</tr>
<tr>
<td>Chaperones/heat shock</td>
<td>4</td>
</tr>
<tr>
<td>Cell division</td>
<td>4</td>
</tr>
<tr>
<td>Protein and peptide synthesis</td>
<td>1</td>
</tr>
<tr>
<td>Adaptation and atypical conditions</td>
<td>2</td>
</tr>
<tr>
<td>Detoxification</td>
<td>6</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>71</td>
</tr>
<tr>
<td>Virulence</td>
<td>9</td>
</tr>
<tr>
<td>IS elements</td>
<td>22</td>
</tr>
<tr>
<td>PE and PPE</td>
<td>12</td>
</tr>
<tr>
<td>Antibiotic production/resistance</td>
<td>0</td>
</tr>
<tr>
<td>Bacteriocin</td>
<td>1</td>
</tr>
<tr>
<td>Cytochrome P450</td>
<td>6</td>
</tr>
<tr>
<td>Miscellaneous transferases</td>
<td>12</td>
</tr>
<tr>
<td>Miscellaneous phosphatases</td>
<td>3</td>
</tr>
<tr>
<td><strong>Conserved hypothetical proteins</strong></td>
<td>114</td>
</tr>
<tr>
<td><strong>Unknown</strong></td>
<td>50</td>
</tr>
</tbody>
</table>

Table 8.1: Functional classification of 595 genes significantly different under uninduced and induced conditions in ΔRv2017 compared to H37Rv. The gene list of those genes significantly different under both uninduced and induced (0.02μg/ml mitomycin C) conditions from the Rv2017 strain compared to H37Rv wild-type, were classified using the classification genes lists in Genespring, obtained from the annotated genome sequence of *M. tuberculosis*, Cole et al., 1998.
Significant increased in expression in Rv2017 compared to H37Rv under both Uninduced and induced conditions

Decreased expression of regulators, under both uninduced and induced conditions

Figure 8.13: Transcriptional regulators with altered expression in ΔRv2017 compared to H37Rv under both uninduced and induced conditions. The data was obtained from the gene list of 595 genes significantly different under both induced and uninduced conditions in ΔRv2017 strain compared to H37Rv, at p<0.01 with False discovery rate correction (Benjamini and Hochberg). The top graph indictates genes with increased expression in ΔRv2017 strain and the bottom graph indictaes genes with decreased expression in ΔRv2017 strain compared to H37Rv.
and induced conditions, suggesting they are not induced in response to mitomycin C stress in either H37Rv or ΔRv2017 strain.

In order to determine which genes were the most severely affected in the ΔRv2017 strain, all 1347 genes that were significantly different in ΔRv2017 strain compared to H37Rv under uninduced conditions p<0.01 + FDR correction were imported into Excel, and the fold change was calculated with Rv2017 expression level as a proportion of H37Rv. A list was derived whereby genes with differential expression of greater than or equal to 5 fold were collated (cut-off 0.2 for genes expressed to a lesser extent in ΔRv2017 strain and 5 for those genes expressed to a greater extent in ΔRv2017 strain). A total of 100 genes were identified with cut-off of less than or equal to 0.2, the reciprocal was then taken, and displayed as a negative number (see appendix III). There were 90 genes identified that were upregulated in the ΔRv2017 strain, i.e. their fold change was greater than or equal to 5 (see appendix III). The top 15 genes with the greatest fold are listed (see table 8.2).

<table>
<thead>
<tr>
<th>Table 8.2: (continued over-page)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Function</th>
<th>Mean Fold change</th>
<th>SEM</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Virulence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>groEL2</td>
<td>Prevents mis-folding/aids refolding of proteins</td>
<td>-13.3119</td>
<td>1.14E-02</td>
<td>Rv0440</td>
</tr>
<tr>
<td><strong>Lipid metabolism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kasA</td>
<td>Fatty acid biosynthesis</td>
<td>-9.9907</td>
<td>5.43E-03</td>
<td>Rv2245</td>
</tr>
<tr>
<td><strong>Cell wall and cell processes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1892</td>
<td>Probable membrane protein</td>
<td>-9.3952</td>
<td>2.95E-02</td>
<td>Rv1892</td>
</tr>
<tr>
<td>Rv0584</td>
<td>Conserved export protein</td>
<td>8.7677</td>
<td>1.72E+00</td>
<td>Rv0584</td>
</tr>
<tr>
<td><strong>IS and phage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv3827c</td>
<td>Possible transposase for IS1537</td>
<td>-10.4465</td>
<td>1.99E-02</td>
<td>Rv3827c</td>
</tr>
<tr>
<td>Rv2648</td>
<td>Probable transposase IS6110</td>
<td>9.9949</td>
<td>3.21E+00</td>
<td>Rv2648</td>
</tr>
<tr>
<td><strong>PE/PPE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>PE</td>
<td>-15.7004</td>
<td>1.25E-02</td>
<td>Rv3477</td>
</tr>
<tr>
<td><strong>Intermediary metabolism and respiration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 8.2: Genes with the highest fold change in ΔRv2017 strain compared to H37Rv

The top 15 genes with the highest and lowest fold changes were classified according to function. The fold changes indicated were calculated using ΔRv2017/H37Rv, the reciprocal was taken and expressed as a - , indicating the gene expression was less in the ΔRv2017 strain. The standard error (SEM) for the fold change is indicated along with the gene identification.

GroEL2 is involved in the general stress response, and is upregulated in response to heat shock, however these uninduced conditions were standard growth conditions at 37°C, without the addition of chemical or physiological stresses to either the ΔRv2017 strain or the H37Rv wild-type, which suggests the deletion of Rv2017 may affect the general stress level of the bacterium. Interestingly a two component sensor kinase (Rv0600c) and a two-component transcriptional regulator (Rv0602c) were both upregulated by 12 and 11 fold respectively, and
are located in close proximity, separated by Rv0601c, the first part of a probable two component sensor kinase. Rv0600c and Rv0602c are significantly different under both uninduced and induced conditions in ΔRv2017 strain compared to H37Rv. Surprisingly, Rv0601c is not significantly different in ΔRv2017 strain compared to H37Rv under either uninduced or induced conditions at $P<0.01$, but this may be due to the stringency of the statistical test.

There appear to be a large number of genes, which are present in operons where the expression levels are altered in the ΔRv2017 strain compared to H37Rv. An example is gyrB (Rv0005) and gyrA (Rv0006), DNA gyrase subunits, which form an operon; their fold changes are -4.3 and -3.8 respectively (down regulated in ΔRv2017 strain). There are a lot of ribosomal proteins which appear to have significantly decreased expression in ΔRv2017 strain compared to H37Rv, including those located in close proximity: rpsF (Rv0053) a conserved 30S ribosomal protein, ssb (Rv0054) – a single stranded binding protein, rpsR1 (Rv0055) – a probable 30S ribosomal protein and rplL (Rv0056) – a probable 50S ribosomal protein as well as Rv0700 through to Rv0709, which are also ribosomal proteins. There are also genes involved in heat shock that are down-regulated in ΔRv2017 strain compared to H37Rv, which appear to form an operon; dnaK (Rv0350) a chaperone, grpE (Rv0351) which stimulates dnaK with DnaJ, dnaJ (Rv0352) a co-chaperone and hspR (Rv0353) a heat shock transcriptional repressor which negatively regulates dnaJ.

### 8.3 Discussion

The analysis of the genes involved in DNA-damage repair revealed many of the same genes that were identified as up-regulated in response to DNA-damage in Rand et al (2003) were also
identified in the ΔRv2017 and H37Rv comparison, even though there were differences in the experimental design. Rand et al. (2003) used a 3-fold induction ratio cut off for the DNA-damage regulon genes, following induction with the mitomycin C at a higher concentration (0.2 μg/ml) and also used RNA versus RNA arrays. The difference in experimental design explains the subtle differences observed in the gene lists of the DNA-damage repair regulon. However, the majority of genes identified in this study correspond with the published DNA-damage regulon (Rand et al., 2003).

Analysis of genes involved in DNA-damage repair revealed that 59 genes were significantly decreased (p<0.01) in their induction ratio, in ΔRv2017 strain compared to H37Rv. However, analysis of the individual values for the induced and uninduced conditions revealed that additional genes that were significantly different under one or both conditions, that were not detected by the induction ratio analysis, suggesting that this data need to be used in conjunction with the induction ratio to determine whether genes are being missed or overlooked that were actually significantly different.

A significantly decreased induction ratio was observed for just over half of the genes that were significantly different under induced conditions (figures 8.4 and 8.8) in ΔRv2017 strain compared to H37Rv. This draws our attention to 26 genes that were significantly decreased (p<0.01) under induced conditions in ΔRv2017 strain compared to H37Rv, yet the induction ratios were not significantly different (figure 8.8). Repeating the statistical test with a less stringent p value (p<0.05) still does not account for all the genes that appear to be induced to a lesser extent in ΔRv2017 strain, but do not have a significantly different induction ratio. Consequently, it is important to look at both the induction ratio, and the separate conditions, otherwise genes would be overlooked that are not induced to the same level in the ΔRv2017
strain compared to H37Rv. In some cases the standard errors may account for the lack of significant difference in induction ratio, when the genes are significantly different under induced conditions. In calculating the induction ratio, one has to compound standard errors for both the uninduced and induced values, making this data less reliable overall.

Examining that data for the uninduced and induced conditions separately can also give clues as to whether genes are being regulated by a repressor or an activator, which the induction ratio alone does not reveal. For example, if the level of expression in the uninduced condition is similar to that under induced conditions, i.e. the uninduced value is significantly greater in one of the strains, the induction ratio will appear very small. Where this is the case, it suggests that Rv2017 directly or indirectly regulates a repressor protein, which under uninduced conditions suppresses transcription. When the repressor is no longer present, the level of transcription is the same under both uninduced and induced conditions.

One would expect that when the individual components, i.e. the uninduced and induced conditions are not significantly different, the induction ratio would also not be significantly different. This was observed in the majority of cases with a few exceptions, that were in fact significantly different under induced conditions, when a less stringent p value was used (p<0.05).

Perhaps some of the most interesting observations were seen with genes that were significantly different under both uninduced and induced condition. The initial hypothesis was that these genes would all have had a significantly different induction ratio; however, this was not the case, just under half of the genes did not have a significantly different induction ratio. Of particular interest was recA, which fell into this category. The induction ratio appears to be
equal in ΔRv2017 strain compared to H37Rv; however it was clearly apparent from the
uninduced and induced values that these were significantly lower in ΔRv2017 strain, by
approximately 5-fold. This 5-fold decrease was consistent throughout the uninduced and
induced values, thus resulting in an identical induction ratio. There are 9 genes (figure 8.11)
that have similar induction ratios in ΔRv2017 strain compared to H37Rv, but show a significant
decrease in expression under both uninduced and induced conditions. One possible explanation
is that these genes are regulated by the same sigma factor, which recognises and controls
transcription from their promoters. They are still induced, as the mechanism of induction has
not been interrupted, yet the overall level of transcription is much less; this could be explained
if the expression or activity of the responsible sigma factor is reduced or if another sigma factor
were compensating for the reduction in the normal sigma factor. Interestingly, the expression of
two sigma factors is significantly decreased under both uninduced and induced conditions in
ΔRv2017 compared to H37Rv: $\text{sigB}$ and $\text{sigE}$ (see table 8.3). The decrease in expression of
these sigma factors may have a role in the differential expression observed in others of the 118
genes involved in DNA-damage repair.

<table>
<thead>
<tr>
<th></th>
<th>H37Rv Uninduced</th>
<th>H37Rv Induced</th>
<th>ΔRv2017 Uninduced</th>
<th>ΔRv2017 Induced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>sigB</strong></td>
<td>4.670 ± 0.482</td>
<td>4.865 ± 0.604</td>
<td>0.843 ± 0.073</td>
<td>1.259 ± 0.138</td>
</tr>
<tr>
<td><strong>sigE</strong></td>
<td>2.723 ± 0.438</td>
<td>3.138 ± 0.412</td>
<td>1.082 ± 0.075</td>
<td>1.086 ± 0.131</td>
</tr>
</tbody>
</table>

Table 8.3: Expression level of two sigma factors with decreased expression in ΔRv2017
strain compared to H37Rv

The genes were identified in the 595 genes that were significantly different under both
uninduced and induced conditions in the ΔRv2017 strain compared to H37Rv, using a Welsh’s
approximation to a T-test $p<0.01$ with the Benjamini and Hochberg FDR correction.
The other important question to address is: how does this data relate to the data obtained from the \(\Delta recA\) strain of \textit{M. tuberculosis}. Does Rv2017 regulate expression of genes in the DNA-damage regulon that are not regulated by the classical RecA/LexA system? In order to determine whether this is the case, the 59 genes with significantly different induction ratios were compared to data from the \(\Delta recA\) strain (Rand \textit{et al.}, 2003), classified in accordance with the expression patterns seen in the \(\Delta recA\) strain (see table 8.4): group 1 genes were not expressed in \(\Delta recA\) strain (21 genes, most with an SOS box) and therefore thought to be solely regulated by the RecA/LexA system; group 2 genes were partially expressed in \(\Delta recA\) strain (28 genes, including \textit{recA}), suggesting these gene were subject to dual regulation; group 3 genes were expressed to the same extent in the \(\Delta recA\) strain (50 genes), suggesting they were not regulated by the classical RecA/LexA system and group 4 genes were expressed to a greater extent in \(\Delta recA\) strain (13 genes), suggesting their expression may be negatively regulated indirectly by RecA/LexA. In this analysis, an additional group was identified; group 5 containing genes that were not identified as upregulated in Rand \textit{et al.}, (2003) in response to DNA-damage, however, the conditions differed between this study and this Rand \textit{et al.}, (2003).

Perhaps surprisingly, a large proportion of the genes (approximately 22\%) fall into group 1, which are genes predicted to be regulated solely by the RecA/LexA system, interestingly 13 out of the 21 genes in this group showed reduced induction in the \(\Delta Rv2017\) strain. This might be accounted for by the decreased overall expression of \textit{recA}. Even though the induction ratio is the same in the \(\Delta Rv2017\) strain compared to H37Rv, the actual expression level under both induced and uninduced conditions is approximately 5-fold lower in Rv2017 compared to H37Rv. This decrease in the level of expression of \textit{recA}, may have a direct effect on the amount of RecA produced, which in turn would results in less RecA filament present in the cell,
Table 8.4
Group 1

Group 2

Group 3

Group 4

Group 5

System atic
Rv0336
Rv0515
Rv0516c
RvlOOO
R vl378c
R vl702c
Rv2578c
linB
lexA
Rv3074
dnaE2
Rv3395c
Rv3777
Rv0094c
Rv1148c
R vl588c
R vl945
Rv2717c
Rv2718c
Rv2719c
Rv2979c
Rv3466
Rv3467
Rv3828c
Rv0922
R vl277
fmt
uvrB
R vl765c
R vl833c
Rv2014
Rv2015c
Rv2191
Rv2791c
Rv2792c
Rv2884
Rv2885c
thiL
Rv2978c
Rv3201c
Rv3202c
Ihr
nei
Rv3517
fdxB
Rvl956

Rv2119
Rv2734
Rv0765c
Rv0861c
Rv0997
R v ll4 7
Rv2362c
Rv2413c
Rv2559c
recX
ligB

Rv3075c
Rv3645

Common
Rv0336
Rv0515
Rv0516c
RvlOOO
R vl378c
R vl702c
Rv2578c
Rv2579
Rv2720
Rv3074
Rv3370c
Rv3395c
Rv3777
Rv0094c
R v ll4 8 c
R vl588c
R vl945
Rv2717c
Rv2718c
Rv2719c
Rv2979c
Rv3466
Rv3467
Rv3828c
Rv0922
RV1277
R vl406
R vl633
R vl765c
R vl833c
Rv2014
Rv2015c
Rv2191
Rv2791c
Rv2792c
Rv2884
Rv2885c
Rv2977c
Rv2978c
Rv3201c
Rv3202c
Rv3296
Rv3297
Rv3517
Rv3554
R vl956
Rv2119
Rv2734
Rv0765c
Rv0861c
Rv0997
R v l147
Rv2362c
Rv2413c
Rv2559c
Rv2736c
Rv3062
Rv3075c
Rv3645

|H 37R v in d u ction ratio
S error
Mean
2.81
12.07
2.83
12.89
0.76
4.7 6
1.11
6.95
4.30
12.47
3.37
1.01
5.80
1.18
4.17
0.51
6.48
0.61
2.88
15.63
14.16
3.55
4.42
12.80
0.57
5.35
1.38
6.80
1.07
8.39
2.99
10.06
8.34
2.03
3.98
0.65
6.16
0 .76
0.70
6.66
1.31
4.75
7.85
2.55
5.56
1.40
5.07
1.16
4.18
0.89
5.48
0.82
3.93
0.48
0.52
3.20
6.47
0.48
4.5 6
1.17
1.41
6.18
1.74
9.39
8.90
1.56
7.01
1.68
1.83
6.13
3.10
0.53
8.16
1.47
5.64
1.34
8.48
1.30
20.05
5.26
16.74
4.43
5.98
0.91
4.10
0.65
4.92
1.24
3.52
0.71
3.09
0.37
3.19
0.38
5.18
1.16
1.93
0.23
2.37
0.48
2.02
0.30
3.6 4
0.90
2.85
0.35
2.31
0.40
2.92
0.43
9.12
1.82
2.55
0.43
2.75
0.22
2.32
0.25

R v 2 0 1 7 in d u ctio n ratio |
Mean
S error
T-test
1.73
0.18
5E-05
1.52
4E-05
0.23
1.62
0.0012
0.10
0.1 4
1.32
0.0003
2.28
0.39
0.0088
1.09
0.06
0.0061
0.0002
1.10
0.08
1.35
0.0003
0.20
2.07
0.13
6E-06
2.87
0.58
0.0038
1.59
0.25
0.0003
1.10
0.006
0.0 5
1.86
0.3 5
0.01
1.29
0.17
0.0005
1.45
0.22
IE-05
1.38
0.10
0.0006
0.17
1.56
0.0004
1.48
0.19
0.0043
1.68
0.37
0.0017
1.45
0.2 5
0.0003
1.46
0.17
0.0099
1.27
0.07
0.0036
1.28
0 .18
0.0026
1.69
0.18
0.0059
1.33
0.17
0.0028
1.42
0.21
0.0001
1.01
0.08
3E-05
1.20
0.12
0.003
2.01
2E-05
0.23
1.36
0.17
0.0096
1.18
0.14
0.0011
2 .0 4
0.28
7E-05
1.25
0.11
5E-06
1.70
0.13
0.0008
1.29
0.15
0.0042
1.46
0.0064
0.16
1.52
0.13
0.0001
1.02
0.14
0.0002
1.59
0.13
0.0001
2.1 3
0.74
0.0003
1.69
0.20
0.0001
1.10
0.19
6E-05
1.39
0.0021
0.19
1.15
0.08
0.0062
1.24
0.11
0.0054
1.58
0.21
0.0095
0.14
1.39
0.0015
1.59
0.1 7
0.009?
1.10
0.09
1.00
0.0 0 6 2
0.09
0.97
0.11
I 0.0073
1.15
0.09
1.03
0.15
0.0008
0.92
0.13
0.0066
1.30
0.15
0.0026
2.78
0.22
0.001
0.99
0.09
0.0019
1.17
0.14
1.23
0 .04
| 0.0022 |

Table 8.4: C lassification o f 59 gen es with significantly different induction ratios according to ex­
pression patterns seen in the ArecA strain. The genes were grouped into 5 categories; group 1 genes
were not expressed in ArecA strain, group 2 were partially expressed in ArecA strain, group 3 were ex­
pressed to the same extent in ArecA strain, group 4 were expressed to a greater extent in ArecA strain,
and group 5 were not listed in the comparison between the ArecA strain and H37Rv wild-type in Rand
et al., 2003.
297


thus resulting in reduced autocatalytic cleavage of LexA. This may result in a decrease in the levels of de-repression of the RecA/LexA operon, reflected a decreased level of induction of genes regulated solely by the RecA/LexA system. This could therefore account for the decrease in induction ratio of this sub-set of DNA-damage repair genes. However, this phenomenon could cause problems when addressing the possibility of whether the dual regulated genes (i.e. those partially induced in ΔrecA strain) are partially regulated by Rv2017. The decrease in induction ratio observed in ΔRv2017 strain for 11 of the 28 genes in group 2 may also be due to the decreased level of transcription of RecA, and therefore decreased de-repression via LexA.

A large proportion of genes also fell into category 3, a total of 21 out of 50 genes showed a decreased induction ratio compared in the ΔRv2017 strain compared to H37Rv: these genes were classified as those expressed to the same extent in H37Rv wild-type and the ΔrecA strain. These genes are therefore thought to be regulated independently of the RecA/LexA system, however, Rv2017 does not appear to directly or indirectly regulate expression of all the genes in this category, therefore, although it appears that Rv2017 plays a role in regulation of genes in the DNA-damage response, it does not solely regulate expression of genes in a RecA dependent or RecA independent manner, but appears to regulate a subset of both categories. The genes in group 4 were expressed to a greater extent in ΔrecA strain, however, none of these were expressed to a higher level in ΔRv2017 strain than H37Rv, and in fact, the induction ratio of 3 genes were decreased in the ΔRv2017 strain. The group 5 genes were not present in the microarray data from (Rand et al., 2003), this is most likely due to the different induction conditions (optical density of induction and concentration of mitomycin C).

A total of 595 genes appear to be regulated by Rv2017 directly or indirectly, as they are significantly different under both uninduced and induced conditions. Consequently, it appears that determination of the Rv2017 regulon is more complicated that initially predicted. To
complicate matters further, Rv2017 appears to positively or negatively regulate the expression of 31 genes involved in transcriptional regulation via direct or indirect means. Consequently, it is impossible to determine with the data collated, whether the effects on transcription are as a direct result of the Rv2017 deletion, or as an indirect result, due to the sheer number of transcriptional regulators that appear to have their expression levels modified in ARv2017 strain.

The analysis of fold change of the genes significantly different under uninduced conditions \( p<0.01 \) with the Benjamini and Hochberg false discovery rate correction, has revealed that 100 genes are downregulated 5-fold or more in ARv2017 strain compared to H37Rv, and 90 genes are upregulated 5-fold or more in ARv2017 strain compared to H37Rv. The list of genes upregulated 5-fold or more in ARv2017 strain includes transcriptional regulators, and two component systems, along with PE and PPE genes. Interestingly two genes involved in homologous recombination were also upregulated in the ARv2017 strain; \( \text{recC} \), upregulated by 6.58±1.27 fold and \( \text{recD} \), upregulated by 2.18±0.22 fold. These genes are co-transcribed and form a complex with \( \text{recB} \), but on checking the arrays it was found that \( \text{recB} \) was a bad spot, consistently throughout all the samples, so its expression could not be determined.

Genes that were 5-fold or more down regulated in ARv2017 strain include genes involved in rRNA synthesis, which may indicate that Rv2017 is a master regulator, as the conditions of treatment were the same for the H37Rv wild-type and ARv2017 strain. The decrease in expression of these genes, including other house keeping genes could be a direct result of the decrease in expression of the sigma factor \( \text{sigB} \). Due to the nature of the data, it is impossible to determine which genes are directly regulated by Rv2017, and which are indirectly regulated, potentially through a transcriptional regulatory protein whose expression is controlled by...
Rv2017; for example, WhiB1 is a transcriptional regulatory protein down-regulated in ΔRv2017 strain, which is thought to bind with *M. tuberculosis* CRP (Roger Buxton - personal communication). The repression of such a protein would therefore potentially affect all the genes regulated by CRP, thus indicating the difficulty of dissecting the Rv2017 regulon, from the regulons of other transcriptional regulators whose expression is altered in the ΔRv2017 strain. Nevertheless, this does raise the possibility that Rv2017 may in fact be a master regulatory protein, which may function as both an activator or repressor protein to positively and negatively regulate transcription at different genes.
Chapter 9 Discussion

*M. tuberculosis* is an intracellular pathogen, which resides and replicates within macrophages (Graham and Clark-Curtiss, 1999; Mariani *et al.*, 2000). After receptor-mediated phagocytosis, *M. tuberculosis* prevents phago-lysosome fusion. However, this arrest of maturation of the phagosome is not universal, and in some cases phagosome maturation is enhanced by activation with IFN-γ (Kaufmann, 2001). Activated macrophage produces an array of DNA damaging agents, in the form of reactive oxygen (ROI) and nitrogen intermediates (RNI), which are known to damage DNA (Kaufmann, 2001). Therefore the repair of DNA-damage is thought to be of utmost importance in survival of *M. tuberculosis*. In a number of bacteria the expression of genes in the DNA-damage regulon is controlled by RecA/LexA; termed the SOS response, this system has been extensively studied in both *E. coli* and *B. subtilis*. It was recently discovered that the majority of genes in the DNA-damage regulon of *M. tuberculosis* were in fact regulated in a RecA independent fashion (Rand *et al.*, 2003), which lead to the question of what mode of regulation governed these DNA-damage repair genes. This project was designed to probe the alternative mechanism that was responsible for regulation of genes in the DNA-damage regulon that were regulated independently of the RecA/LexA system. A number of different possibilities were addressed regarding the mode of regulation of this subset of DNA-damage repair genes, and two different avenues were pursued to unravel the mode of regulation.

It has been suggested that the majority of regulation in bacteria takes place at the transcriptional level (Raman *et al.*, 2001). There are many different types of transcriptional regulation, that are controlled both temporally and spatially and culminate in a multi-factorial process of regulation. The gene expression of the sub-set of DNA-damage repair genes, defined by Rand *et al.*, (2003) as RecA independent, may be regulated by a sigma factor, which, by its specificity regulates
transcription from a certain regulon; the other alternative considered was that a regulatory protein, analogous to LexA, was responsible for the activation or suppression of genes in the RecA independent DNA-damage regulon. However, the microarray data published by Rand et al., (2003) indicated that the RecA dependent and the RecA independent DNA-damage regulons are not mutually exclusive. Certain key genes, including recA itself, were partially induced in response to DNA-damage in the ΔrecA strain of M. tuberculosis, suggesting that LexA was only partially responsible for regulation of these genes. The possibilities that a sigma factor or a predicted regulatory protein were responsible for regulation of transcription of a subset of the genes in the DNA-damage regulon were assessed.

9.1 The role of SigG

SigG was the most highly upregulated sigma factor in response to DNA-damage (Rand et al., 2003). SigG was also upregulated in the ΔrecA strain of M. tuberculosis suggesting that the expression and regulation of SigG were RecA independent.

A gene inactivation knockout of sigG was constructed in a laboratory strain of M. tuberculosis, H37Rv, using a deletion and insertion method, in which, a functional SigG was no longer produced. Complementation constructs were also produced for the ΔsigG strain to ascertain whether the phenotype observed was a direct result of the gene inactivation, or due to the indirect effects of the deletion on the two downstream co-transcribed genes Rv0181c and Rv0180c, or due to a secondary mutation elsewhere in the genome.

No differences were observed in the in-vitro growth rate for the ΔsigG strain and the full operon complement, compared to H37Rv. The ΔsigG strain was significantly more susceptible to
mitomycin C stress than the wild-type at a range of 0.02 to 0.2μg/ml. The full operon complement did not restore the viability back to wild-type, but was significantly less susceptible to mitomycin C stress in 4 out of the 5 tested concentrations than the ΔsigG strain. Mitomycin C causes interstrand cross links in the DNA (Paz et al., 2004), which are generally repaired by homologous recombination; therefore one would expect upregulation of the recA and other genes involved in homologous recombination, and genes in the RecA/LexA regulon. Interestingly, these genes do not appear to be differentially expressed in the ΔsigG strain, despite the increased susceptibility to mitomycin C of the ΔsigG strain. It is possible that the increased susceptibility of ΔsigG strain to mitomycin C could be as a result of the significantly decreased expression of some of the cell wall/cell process genes and lipid metabolism genes. These decreases in expression may result in problems with cell wall formation in ΔsigG strain, which in turn, could make the mutant more susceptible to mitomycin C, as a higher proportion may be entering the cell.

Interestingly, preliminary mouse in-vivo data revealed that the ΔsigG strain was attenuated in comparison to H37Rv, which suggests that SigG or the SigG regulon is involved in virulence of M. tuberculosis in the mouse model of infection. Other ECF family sigma factor mutants have only shown attenuation in the mouse model of infection when time to death studies were performed: SigF (Chen et al., 2000), SigE (Ando et al., 2003), SigC (Sun et al., 2004), and SigH (Kaushal et al., 2002). No differences were observed in the CFU counts for the lung and spleen over a 20 day period for SigE (Ando et al., 2003), whereas a SigF mutant showed decreased CFU compared to wild-type (Geiman et al., 2004). The role of other sigma factors in M. tuberculosis had been identified using gene inactivation mutants. Analysis of these mutants revealed that SigH was responsible for regulation of other sigma factors, regulating inducible
gene expression of \textit{sigE}, along with basal and inducible gene expression of \textit{sigB}, which is subject to dual regulation by SigH and SigE (Raman \textit{et al.}, 2001).

Analysis of the SigG locus revealed that the two downstream genes, Rv0181c and Rv0180c, are co-transcribed with SigG and may possibly function in the role of regulation of the sigma factors activity, as the majority of sigma factors are co-transcribed with their cognate anti-sigma factors (Raivio and Silhavy, 2001; Yoshimura \textit{et al.}, 2004). Analysis of genes downstream of sigma factors, revealed that they are usually comprised of a transmembrane protein (Hughes and Mathee, 1998), with an extracytoplasmic sensory domain and an intracellular cytoplasmic domain that binds to its cognate sigma factor to prevent transcription of the sigma factors regulon (Yoshimura \textit{et al.}, 2004) and a periplasmic domain which may act to detect environmental signals (Helmann, 1999). There are a wide variety of regulatory mechanisms governing sigma factor activity: regulation can take place at transcriptional, translational and post-translational levels (Helmann, 1999). Anti-sigma factors also operate in various ways to inhibit sigma factor-RNAP binding, which can be broadly grouped into three categories: export from the cell, as seen with the flagella biosynthesis anti-sigma factor FlgM of \textit{Salmonella typhimurium}, partner switching modules as outlined with SigF of \textit{B. subtilis} (i.e. regulation by an anti-anti sigma factor) and interactions with small molecules or protein ligands (Helmann, 1999). Sigma factors can also be synthesised as inactive pro-proteins, which undergo cleavage to form an active sigma factor (Haldenwang, 1995). Homology searches revealed the \textit{sigG} locus is also present in \textit{M. bovis} and shares a high degree of homology. The domain predictions and homology searches suggest that Rv0181c is a cytoplasmic protein, whereas Rv0180c appears to be a transmembrane protein with multiple transmembrane domains. Based on homology and domain predictions along with evidence from the regulation of other sigma
Chapter 9 Discussion

factors, it is possible that Rv0181c is an anti-sigma factor, and Rv0180c is an anti-anti sigma factor, which detect and respond to environmental signals to regulate the function of SigG.

The microarray data from the \( \Delta sigG \) strain showed that the two downstream genes appear to have slightly reduced transcriptional in the \( \Delta sigG \) strain compared to H37Rv. However, the limited decrease, particularly in translation, potentially observed in the Western blot suggests that the polar effects of the \( \Delta sigG \) strain are minimal, and indicates that the phenotype of the \( \Delta sigG \) strain is most likely a direct result of the gene inactivation of SigG. Although the deletion and insertion were not in-frame, this inclusion of the 3' region of sigG in the knockout construct resulted in the proper formation of the stop codon at the end of sigG, so as not to interfere with any post-translational modification/processing.

Microarray data and potential Western data, indicated that the vast majority of expression of the sigG operon was independent of SigG, suggesting that SigG does not autoregulate, as is the case for a proportion of sigma factors (Raman et al., 2004). However, primer extension and RNase protection assays of sigG revealed three separate transcriptional start sites, two of which are DNA-damage inducible, while the third promoter (P3) is not upregulated in response to DNA-damage, but appears to be autoregulated. The P1 and P2 promoters are regulated independently of SigG. Analysis of the promoter –10 and –35 regions supports the hypothesis that the P3 promoter is regulated by SigG, as they bore no similarity to the promoter motifs described for the ECF sigma factors or the primary sigma factor. Further analysis of the upstream regions of 5 genes upregulated in the \( \Delta sigG \) strain, revealed that a potential SigG consensus of CGACC(R)(t/c)C-N_{13-22}-TGTCCG (where R is a/g/t) was present within 100bp upstream and downstream of the predicted translational start sites. There appears to be a greater degree of homology at the –35 site than at the –10 site, which has been observed before with other ECF
family sigma factors (Raman et al., 2004). Interestingly, nuoB, Rv0655 and Rv0312 show a high degree of homology at both the -10 and -35 sites. Confirmation of these observations by either RNase protection or primer extension would enable clarification and further refinement a consensus motif, which could then be used in conjunction with the microarray findings to search for other genes regulated by SigG.

9.1.1 Identification of the SigG regulon.

The preliminary microarray analysis of the ΔsigG strain revealed that ruvC was partially induced in the ΔsigG strain, suggesting that ruvC may be partially regulated by SigG. This possibility was further analysed using primer extension and RNase protection assays to determine the transcriptional start sites of ruvC. As previously mentioned, ruvC, like recA was partially upregulated in response to DNA-damage in the ΔrecA strain, and it has been shown that recA processes two transcriptional start sites, both of which are DNA-damage inducible, but only one of which is regulated by the repressor LexA (Davis et al., 2002b; Movahedzadeh et al., 1997). RNase protection assay confirmed that the P2 promoter of recA (containing the SOS box) showed complete abolition of expression in the ΔrecA strain, suggesting that expression from the P2 promoter was completely dependent on RecA for derepression. In accordance with the findings (Movahedzadeh et al., 1997), the P1 promoter was DNA-damage inducible even in the ΔrecA strain, thus suggesting that induction of expression from this promoter is independent of RecA. There appeared to be a decrease by approximately 2-fold, in the level of expression in the ΔsigG strain from both the P1 and P2 promoters of recA, however, theses findings were not validated with the microarray data, which revealed there was no significant difference in the expression level of recA in the ΔsigG strain compared to H37Rv at the 1% level. The drawback
to this method is that one is not able to distinguish between the expression levels from individual promoters.

The primer extension assays revealed that $ruvC$ possesses two transcriptional start sites, only one of which appeared to be DNA-damage inducible (P1). The P1 promoter, possesses a LexA binding site, and was only partially decreased in the $\Delta recA$ strain, whereas the P2 promoter appeared to be transcribed to a lesser extent in the $\Delta sigG$ strain. The RNase protection confirmed the observation that the P2 promoter does not appear to be DNA-damage inducible, however, neither the P1 nor the P2 promoter show differential expression by RNase protection in the $\Delta sigG$ strain. The RNase protection clearly shows that expression from the P1 promoter is not completely abolished in the $\Delta recA$ strain, thus suggesting there is another mechanism present, which is partially able to overcome the derepression of the P1 promoter of $ruvC$ by LexA. This is particularly interesting when one looks at the potential promoter motifs identified upstream of the transcriptional start sites of the P1 promoters of both $recA$ and $ruvC$. These promoters are both DNA-damage inducible and appear to be regulated at least partially by an alternative mechanism to the RecA/LexA system. Gamulin et al., (2004) revealed the presence of a potential promoter motif, which they defined as the RecA-ND for $recA$ non-dependent promoter. Analysis of the transcriptional start sites for $ruvC$ reveal this motif is actually positioned in close proximity to the P1 promoter, in the potential -10 location. The -35 region is located more proximally in $recA$, and there is an alternative -35 sites for the P1 promoter of $ruvC$ other than the RecA-ND, defined by Gamulin et al., (2004). Nevertheless, it suggests that the P1 promoters of both $recA$ and $ruvC$ may be recognised by the same sigma factor. Interestingly, the region 2 of SigG shows the greatest degree of homology to the $\sigma^{70}$, and the -10 sites of $ruvC$ show homology to both the $\sigma^{70}$ and $\sigma^{38}$ of $E. coli$. However the regulation of which sigma factor transcribes from a dual regulated promoter is more complicated than pure
competition, as in the case with $\sigma^5$ and $\sigma^7$, architectural proteins such as Lrp, CRP, IHF and Fis have been implicated in determining which sigma factor initiates transcription (Hengge-Aronis, 1999), and the choice of sigma factor can also be modified by the pleiotropic regulator ppGpp, which has been shown to favour interactions of RNAP with alternative sigma factors at the expense of $\sigma^7$ (Gralla, 2005). This is of particular interest, as it has also been alluded to that genes under the negative regulation of the LexA repressor have been de-repressed by ppGpp in the absence of RecA (Kvint et al., 2000).

The differences observed in repression of the P2 promoter of recA and the P1 promoter of ruvC by LexA could be explained by the relative location of the LexA binding site. The sequences of the binding sites are identical for both recA and ruvC and vary from the consensus by 1bp. They also have a different spacer region, between the palindromic sites. The LexA binding site for the P2 promoter of recA is located over the RNAP binding site, and therefore most likely prevents binding of the RNAP-$\sigma$ complex by obscuring the $-10$ and the $-35$ sites. However, the location of the LexA binding site for the P1 promoter of ruvC is downstream of the $-10$ promoter site, overlapping the transcriptional start site. This is interesting as the TrpR operon has both strong and weak TrpR boxes, depending on the location of the promoter $-10$ and $-35$ sites relative to the position of the TrpR box (Pittard et al., 2005). It is possible that the location of the LexA repressor binding site enables transcriptional read through. In ruvC, the position of the LexA binding site would allow binding of the RNAP-$\sigma$ complex, but the LexA repressor may possibly act as a road-block, preventing transcriptional elongation, rather than transcriptional initiation as is potentially the case for recA. It has been demonstrated that oscillation of the RNAP-$\sigma$ complex at the site of a road-block can lead to transcriptional read through. Analysis of the location of the operator of the lac operon revealed that the Lac repressor was able to bind to the operator and did not prevent binding of the RNAP to the
promoter, yet the repressor blocked the clearance of the RNAP-σ complex from the promoter, and hence prevented transcription (Gralla, 1996). This repression was overcome in-vitro through oscillation of the RNAP-σ complex. Therefore it is possible that the LexA repressor only partially inhibits transcription via the position of the binding site rather than the affinity of the repressor for the SOS box.

Microarray data indicated that SigG was not involved in the regulation of genes involved in DNA-damage repair, but does regulate expression of a number of genes. Transcriptional analysis in the ΔsigG strain of recA and ruvC, genes which are partially regulated in a RecA independent manner, revealed that although preliminary microarray data and primer extension assays eluded to the possibility that ruvC was regulated by SigG, RNase protection assays revealed that SigG was not involved in regulation of expression of ruvC from either of the two promoters. The RNase protection result was confirmed by subsequent microarray analysis, which showed that the overall expression levels of ruvC were unaffected by the absence of SigG. It is noteworthy, that the expression level of ruvC from the P2 promoter was increased in the ΔsigG strain, which may have resulted from increased availability of RNAP in the absence of SigG. Sigma factor competition for the RNAP in the absence of SigG may have resulted in an increase in basal level of transcription.

RNase protection assay for recA, also partially regulated by RecA, lead to the tantalising possibility that recA P1 and P2 promoters were regulated partially by SigG. Two predicted regulatory proteins, Rv0232 and Rv3050c of the TetR/AcrR and AsnR family of transcriptional regulators respectively were downregulated in the ΔsigG strain, suggesting this decrease in expression of recA could be attributed directly to SigG, or indirectly via one of the two predicted regulatory proteins. Alternatively, SigE was upregulated under uninduced conditions.
in the ΔsigG strain, although the level of SigH remained constant; however, SigE is only regulated by SigH under induced conditions (Raman et al., 2001). SigE may regulate transcription of a repressor protein, therefore if this were the case, then the repressor would be upregulated in the ΔsigG strain, and could therefore result in decreased transcription of recA from both promoters. A common motif was identified upstream of the translational start sites of both sigG and recA, although the importance of this motif is unclear, as it is present with 1bp mismatch in close proximity to the predicted translational start sites (-150bp to +100bp) in 24 other mycobacterial genes.

The promoter similarity suggests the possibility that a sigma factor is responsible for regulation of expression of the RecA independently regulated genes. Analysis of the published sigma factor consensus sequences revealed that these do not appear to have homology to the P1 promoters of recA and ruvC. It is possible that ruvC could be transcribed from the same promoter by different sigma factors, which are able to elicit transcription from the same promoter. This type of compensation has been observed with over-expression of SigA which resulted in a decrease in expression of genes regulated by alternative sigma factors (Nystrom, 2004a). It has been observed with E. coli, that there is an overlap between the recognition of some promoter elements, there are cases where a single promoter can be regulated by more than one sigma factor, which is the case with σ^5 and σ^70 in E. coli (Tanaka et al., 1993). Dual regulated promoters have similar -10 regions, which form the specific interactions with region 2 of the sigma factor.

Interestingly some genes repressed by LexA in M. tuberculosis have multiple LexA binding sites, as is the case for Rv2719c, which is divergently transcribed with LexA. Rv2719c possesses three LexA binding sites, which overlap the promoter recognition -10 and -35 sites.
or overlap with the predicted transcriptional start site. These sites however contain mismatches to the LexA consensus, therefore it was suggested that multiple LexA binding sites were present when there were two or more mismatches in the SOS box, this may allow regulation in the overall level of transcription depending on the degree of DNA-damage (Dullaghan et al., 2002).

Regulation of transcription can be controlled by a dual mechanism, as observed with the uspA gene in *E. coli*, whereby the gene is subject to both positive and negative regulation. The repression by the FadR repressor can be overcome by the activator alarmone guanosine 3', 5'-bisphosphate (ppGpp), for certain genes in the operon, whereas *fad* itself is regulated by two activator molecules, ppGpp and the cAMP-CAP protein complex (Kvint et al., 2000).

### 9.2 The role of an activator or repressor in DNA-damage

The second possibility was addressed, that a regulatory protein was potentially responsible for regulation of transcription of those DNA-damage inducible genes regulated independently of RecA. *In-vitro* binding assays, in the form of bandshift assays were unable to detect binding of a specific protein to the P1 region of *recA*, although binding of the LexA repressor to the P2 promoter was readily detectable. These regulatory proteins chosen for the *in-vitro* transcription covered a wide range of the transcriptional regulatory families, however, Rv2017 proved potentially the most interesting, due to homology with IrrR of *D. radiodurans*, which positively regulates *recA*, despite the presence of the LexA repressor (Narumi et al., 2001).

The bandshift may not have been able to detect binding of an activator, as this process can be complicated as transcriptional regulation can take place at either proximal or distal sites. Distal activation occurs at enhancer sites located upstream of transcriptional start sites. Sig54 (σ^{54}) in
*E. coli* is part of the distal regulation, whereby, unlike the proximal enhancers, which enhance binding of the RNAP to the promoter recognition site by specific interactions with the alpha subunit of RNAP, the distal enhancers require the presence of an activator, and $\sigma^{54}$, which enables the RNAP-$\sigma^{54}$ complex to bind to the promoter and remain inactive until the enhancer, causes the assembly of a large complex with ATPase activity via a signal transduction cascade (Merrick, 1993). The enhancer bound ATPase then makes contact with the RNAP most likely with the sigma subunit, and loops out the DNA region (Gralla, 1996), enabling initiation of transcription.

Analysis of the ΔRv2017 strain produced by deletion and insertion proved more fruitful. *In-vitro* growth curves revealed there was no difference between the growth of the ΔRv2017 strain and H37Rv under the conditions utilised. Interestingly the ΔRv2017 strain was hyposensitive to mitomycin C stress compared to H37Rv, in concentrations of mitomycin C greater than 0.02μg/ml, and preliminary time points for a mouse model of infection suggested that the ΔRv2017 strain was hypervirulent in comparison to H37Rv.

Time constraints did not allow the construction of a complement of ΔRv2017 strain. However, the RT-PCR performed to determine whether Rv2017 was mono or polycistronic revealed that Rv2017 was the second gene in an operon with Rv2016 located 5’ of Rv2017. RT-PCR also suggested that Rv2018 was also co-transcribed with Rv2017; Rv2018 is annotated in Tuberculist to overlap with Rv2019, suggesting they also form a polycistron. The microarray data produced for the ΔRv2017 strain compared to H37Rv contradicts the RT-PCR data, as the expression levels of Rv2018 and Rv2019 are not significantly different under either uninduced or induced (mitomycin C 0.02 μg/ml) conditions. The microarray data for ΔsigG strain revealed
a significant decrease in expression of the two downstream co-transcribed genes, which suggests a potential problem with the RT-PCR. One could speculate that the promoter for Rv2018 overlaps with the 3' region of Rv2017, which would make it appear as though the genes were co-transcribed, where in actual fact this was probably not the case.

Microarray analysis of the ΔRv2017 strain in comparison to H37Rv suggested that Rv2017 was partially involved in regulation of the DNA-damage repair regulon. Problems were encountered as with the ΔsigG strain microarray, when attempting to compare the data to that of ΔrecA strain, as there was a 10-fold difference in the concentration of mitomycin C used for induction, but more importantly, only the induction ratio was available for the previous microarray data for the ΔrecA strain (Rand et al., 2003), limiting the information that could be obtained as compared to the availability of individual uninduced and induced results in this study. It did appear however, that Rv2017 was directly or indirectly responsible for regulation approximately half of the genes in the DNA-damage regulon. These genes were not limited to those regulated partially by RecA and independently of RecA, overlapping with the genes whose expression was dependent on RecA for derepression. Interestingly, part of the explanation may lie in the expression level of recA itself. At first glance using the induction ratio data, it appears that the level of induction of recA in the ΔRv2017 strain is the same as in the wild-type; however when the uninduced and induced values are assessed individually, there is a significant decrease in the expression level of recA under uninduced (t-test p<0.01) and induced (t-test p<0.01 + FDR) conditions, which, equates to an approximate 3.5-fold decrease in the expression level under induced conditions of recA in the ΔRv2017 strain. This decreased level of recA expression could result in a decreased quantity of RecA, which in turn would affect those genes partially and completely dependent on RecA for derepression.
With regard to the role of Rv2017 in regulation of the DNA-damage regulon, Rv2017 does not regulate all of the genes identified by Gamulin et al., (2004) as possessing a RecA-ND promoter motif, nor is the role of Rv2017 limited to RecA independent gene expression. Out of the 21 genes classified as DNA repair genes possessing the RecA-ND promoter motif, only \textit{recA}, \textit{ruvC}, \textit{lhr}, \textit{ssb}, \textit{dnaB} and Rv0184 were significantly decreased in the \textit{ΔRv2017} strain in comparison to H37Rv. This indicates that Rv2017 and its regulon alone, are not fully responsible for regulation of genes induced in a RecA independent fashion.

\textit{M. tuberculosis} has 196 predicted broad spectrum regulatory proteins (Cole \textit{et al.}, 1998), of which 152 are predicted activator/repressor proteins, 30 are 2 component regulators, and 14 are serine-threonine kinases. In the \textit{ΔRv2017} strain, 7 of these broad-spectrum regulatory proteins were expressed to a lesser degree in \textit{ΔRv2017} strain than H37Rv, while 24 were upregulated in the \textit{ΔRv2017} strain. A comparison between the uninduced and induced expression levels revealed that these transcriptional regulators do not appear to be induced in response to mitomycin C stress. This suggests that the Rv2017 deletion has resulted in the differential expression of all of the broad spectrum transcriptional regulators, which may help explain why the expression of a huge proportion of genes was altered in the \textit{ΔRv2017} strain compared to H37Rv. This suggests that Rv2017 may act as a master regulator, affecting expression of a large proportion of genes in a possible cascade.

If Rv2017 acts as an activator, then potentially the most interesting genes are those that are downregulated in \textit{ΔRv2107} strain, as these may be regulated directly by Rv2017. Rv3477 is one of the genes most highly downregulated in the \textit{ΔRv2017} strain, and encodes a PE protein. This is of particular interest as the PE/PPE family are extracellular surface proteins, which act as potential antigens for the host immune response (Plotkin \textit{et al.}, 2004), and form about 10% of
the mycobacterial genome (Chaitra et al., 2005). Other genes downregulated include \textit{ctaB} a cytochrome oxidase and \textit{groEL2}, which is involved in heat shock response. Interestingly, both SigF of \textit{M. tuberculosis} and SigB of \textit{Streptomyces coelicolor} appear to have large regulons, including other transcriptional regulators, suggesting that although their specific regulons are small, they regulate a large number of other genes indirectly (Geiman et al., 2004; Lee et al., 2005). Both \textit{sigB} and \textit{sigE} are significantly decreased in the $\Delta$Rv2017 strain, as are certain operons involved in homologous recombination; \textit{recC}, \textit{D} and \textit{G}, with no data available for \textit{recB}. The general stress response chaperones DnaJ and DnaK are also downregulated in the $\Delta$Rv2017 strain, which suggests Rv2017 may play a direct or indirect role in stress response. However, it is noteworthy that the genes up/down regulated in the $\Delta$Rv2017 strain, appear to cover a wide range of functional groups, which further suggests that Rv2017 may be a master regulator.

In conclusion, it appears that the control of DNA-damage repair is complex, particularly as control of transcription appears to be a multifactorial process, which differs depending on the regulon, and often requires a combination of complex transcriptional regulation. A prime example of the complexity of transcriptional regulation is observed in the heat shock response in \textit{B. subtilis}, which is regulated by at least 4 different mechanisms controlling different genes in the heat shock response. Class I genes, such as those that encode the molecular chaperones \textit{gorESL} and \textit{dnaK}, are regulated by the HecA repressor, which binds to a palindromic site in the operator, termed the CIRCE site (controlling inverted repeat of chaperone expression). The class II heat shock genes are regulated by the alternative sigma factor SigB (Grandvalet et al., 2005), which account for a small proportion of the 709 genes with 4-fold differential expression in the $\Delta$sigB strain (Price et al., 2001). The class III genes are regulated by the CtsR repressor, which recognises and binds as a dimer to a heptanucleotide tandem repeat in the operator (Derre
et al., 2000). The class IV heat shock genes are regulated by an unknown mechanism (Grandvalet et al., 2005). This is particularly relevant, as the control of the transcriptional repressors differs. HrcA is synthesised as an inactive protein, which is activated upon binding to GroEL; the active HrcA is then able to repress transcription by binding to the palindromic site of the operator. The repression is transiently lost under heat shock, modulated by GroE. Regulation of the CtsR repressor is somewhat different and involves degradation via proteases (ClpX and ClpP), while, ClpC is thought to protect CtsR from degradation under normal conditions (Derre et al., 2000).

Further research is therefore required to dissect the complex network of regulation of the DNA-damage regulon, of which, I suspect only the tip of the iceberg has been discovered.

9.3 Where to begin

The research thus far has identified a number of tantalising possibilities for the alternative mechanism of regulation of the DNA-damage regulon. In order to further elucidate the possible mode of regulation, there are a number of avenues that it would be beneficial to pursue:

To determine a consensus promoter recognition site for SigG, primer extension/RNase protection assays could be performed using the 5 genes potentially directly regulated by SigG, to detect their transcriptional start site(s). This data could then these be used to design site directed mutagenesis and reporter fusion assays to define the SigG recognition consensus. Taqman quantitative RT-PCR would be used to confirm this SigG dependent expression, by looking for decreased levels of expression in the \( \Delta \text{sigG} \) strain. This could also be used to
determine whether recA is partially regulated by SigG, as suggested by the primer extension and RNase protection assays.

The role of SigG in infection could be assessed by repeating the mouse in-vivo infections; alongside in-vitro infections of steady state and IFN-γ activated bone marrow derived macrophages. However, prior to this, both the full and partial operon complement should be re-tested for in-vitro viability.

The role of the two downstream co-transcribed genes in the regulation of SigG could be further addressed by re-synthesis of the SigG antibody, with a view to using the SigG, Rv0181c and Rv0180c antibodies in co-immunoprecipitations to determine if the two downstream genes are involved in regulation of SigG. Alongside this approach yeast-2-hybrid assays could be used to detect any interactions between SigG, Rv0181c and Rv0180c.

Determination of the Rv2017 regulon could be addressed using primer extension/RNase protection assays of the genes most highly downregulated in the ΔRv2017 strain, to identify a potential Rv2017 binding site and expression/purification of Rv2017 using a pET expression system could be performed; with a view to using purified Rv2017 in bandshift experiments with the P1 promoters of both recA and ruvC to determine if the decrease in expression of these genes was directly dependent on Rv2017.

The potential hypervirulence of the ΔRv2017 strain during infection could be assessed by using established in-vivo and in-vitro models (as outlined for ΔsigG strain).
10 References


Duncan, L., and Losick, R. (1993) SpoIIAB is an Anti-{sigma} Factor that Binds to and Inhibits Transcription by Regulatory Protein {sigma}F from Bacillus subtilis. PNAS 90: 2325-2329.


References


References


References


References


determinants for Escherichia coli RNA polymerase holoenzyme containing sigma

for back and forth oscillations of the transcription elongation complex. Embo
Journal 18: 5052-5060.

Nature 222: 537-540.

Opinion in Microbiology 4: 178-185.


Walker, G.C. (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in


Withey, J.H., and DiRita, V.J. (2005) Activation of both acfA and acfD transcription by Vibrio
cholerae ToxT requires binding to two centrally located DNA sites in an inverted

sigma factor sigA mediates enhanced growth of Mycobacterium tuberculosis in

Yan, S., Myler, P.J., and Stuart, K. (2001) Tetracycline regulated gene expression in
Leishmania donovani. Molecular and Biochemical Parasitology 112: 61-69.

extracytoplasmic function (ECF) sigma factors with the N-terminal regions of their
potential anti-sigma factors. Microbiology 150: 591-599.

Zahrt, T.C., Song, J., Siple, J., and Deretic, V. (2001) Mycobacterial FurA is a negative

Zahrt, T.C., and Deretic, V. (2002) Reactive nitrogen and oxygen intermediates and bacterial
defenses: unusual adaptations in Mycobacterium tuberculosis. Antioxid Redox
Signal 4: 141-159.

Appendix I: Media and solutions

Modified Dubos broth

- $K_2HPO_4$: 1g
- $Na_2HPO_4.12H_2O$: 6.25g
- $Na_3$ citrate: 1.25g
- $MgSO_4.7H_2O$: 0.6g
- Asparagine: 2g
- 10% Tween 80: 5mls
- Casamino acids (Difco): 2g
- pH 7.2 (2M NaOH):
- $dH_2O$ up to 960mls

Autoclave 121°C for 15 mins

Freezing medium

- $K_2HPO_4$: 12.6g
- $Na_3$ citrate: 0.9g
- $MgSO_4.7H_2O$: 0.18g
- $(NH_4)_2SO_4$: 1.8g
- $KH_2PO_4$: 3.6g
- Glycerol: 96g
- $dH_2O$ to 1 litre

Autoclave 121°C for 15 mins

Lauria-Bertani agar (L-agar)

- Tryptone (Difco): 10g
- Yeast extract (Difco): 5g
- NaCl: 10g
- Agar (Difco): 15g
- $dH_2O$ to 1 litre

Lauria-Bertani broth (L-broth)

- Tryptone (Difco): 10g
- Yeast extract (Difco): 5g
- NaCl: 10g
- pH 7.5 (1M NaOH):
- $dH_2O$ to 1 litre

Autoclave 121°C for 15 mins
Middlebrook 7H9 broth

Glycerol 2mls
7H9 medium powder (Difco) 4.7g
dH2O to 900mls
Autoclave 121°C for 10 mins

Middlebrook 7H11 agar

Glycerol 5mls
7H11 medium powder (Difco) 21g
dH2O to 900mls
Autoclave 121°C for 10 mins

Denhardt's reagent (50X)

Ficoll 5g
Polyvinylpyrrolidone 5g
BSA (Fraction V, Sigma) 5g
dH2O to 500mls
Filter sterilise and store at -20°C

Z-buffer

Na2HPO4.7H2O 0.06M
NaH2PO4.H2O 0.04M
KCl 0.01M
MgSO4.7H2O 0.001M
β-merceptoethanol 0.05M
pH7.0
### Appendix II: Primers and probes

<table>
<thead>
<tr>
<th>Primers and probes</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>co-transcription primers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD sigG Af/r</td>
<td>accaacagtgccccgcctatt</td>
<td>gccgcagccccctctcgc</td>
</tr>
<tr>
<td>LD sigG Bf/r</td>
<td>cctacggccgctcactctgga</td>
<td>gctgcgctcggcggatgcta</td>
</tr>
<tr>
<td>LD sigG Cf/r</td>
<td>agccgtaccacgtccttcc</td>
<td>tggccgcctttgcctccttc</td>
</tr>
<tr>
<td>LD sigG Df/r</td>
<td>ccctaccggcgtgaactgct</td>
<td>ggtcttcgcctgcgggtgaaatc</td>
</tr>
<tr>
<td>LD2016f/r</td>
<td>ccctgcggccggcgtgaagat</td>
<td>cgccttcggggttgggaact</td>
</tr>
<tr>
<td>LD2017f/r</td>
<td>ccaccaaccccacggagaacat</td>
<td>actgcgggtcgggtggggaagag</td>
</tr>
<tr>
<td>LD2018f/r</td>
<td>gcgccttgccccaggattttgtc</td>
<td>ataccgtggccgactgctctgc</td>
</tr>
</tbody>
</table>
## Knockout construction and detection

| LD sigG up/low | gctctagagaggtggtcggttgtaagctggac | ccaagctttataggatgacgcggccggaatgtgta |
| LDsigG 5' DC | gttgccggtggccacctggccaa | ggcacctgacgggaccaacatct |
| pLR11c 5'R | LDsigG low | gcaggtccggtaggaatcact |
| sigG D f/r | ttcgccccgcaagctc | gcagcagtcgccgcagcagta |
| sigG probe f/r | ggaattacatagctatgccccccccggtgtcaag | atggcatccccgcacaagttccgctcgtctatct |
| Rv2017 3'F/R | ccaactgacaccaaccacccagggaca | gggcagcaagaccgccacct |
| Rv2017 5'F/R | tgtactagtcactgggccacggcaactac | cgctgggccccctagttgcttcct |
| sigG compd1 f/r | gcagcgtgagcctacagcgcgcgcagcgcagtc | gacggtgcgccgcgatagcgtcagttc |
| Rv2017 probe f/r | ccatgcatacggcggtctgtcctcccc | cgctcctgcgccggtgaacttc |
| Rv2017 D f/r | cgtggcacaaggttaacgcttttc | cgatgggcccgggtgtaacct |
| Hyg f/r | gcactgcacgggacacacatccc | cgcggccaggtcacgaagatggt |

## RNase protection/primer extension

### primers

| LD sigG mid H | cccagcttttcacgtcaatgcgctacgcagagttcct |
| LD sigG mid2 | cgctccccgttaatgact |
| ruvC small f/r | aatctagaggtcgggcccatttcgaacttc | gagaagttagctgcgacccgacccgacat |
| recA RNase F/R1 | gatccgccccctcgcgcggttggagca | gagcgcccatcaccagaaacctgccgtaac |
| sigG RNase F/R1 | atgcggccccgttgaagtc | tctggtgcggcagagat |

## sequencing

### primers

| pKP186 seq f/r | ccttaatacggcctttgaggtgac | gcataaaaaggaagggccagttct |
| comp 186 seq f/r | tagcccgcccccatgcgcagtc | ctagcggctgsgcagac |
| M13 f/r | commercial (invitorgen) | commercial (invitorgen) |
**Bandshift probes**

<table>
<thead>
<tr>
<th>Probe Type</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>RecA P1 forward</td>
<td>ctagcattcggagcacgccgacattgcagtggtgggtctctagtggtcagcggccaaccga</td>
<td></td>
</tr>
<tr>
<td>RecA P1 reverse</td>
<td>ctagtcggtgcccgtgacactgacagacgccacggaagtgccgtgcctccgaatgat</td>
<td></td>
</tr>
<tr>
<td>RecA P2 forward</td>
<td>ctagacgcggcgtgtcacacttgacaggtgttcggctactgtggtgatcattcgga</td>
<td></td>
</tr>
<tr>
<td>RecA P2 reverse</td>
<td>ctagtcgaatgacacaccagctacccgggaacacttgttgcatttcgaagtcagtgacagcgcct</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 A forward</td>
<td>ctagatcattcggagcagccgactgtcagtgga</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 A reverse</td>
<td>ctagacactgacacagtgctgacactggaatgat</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 B forward</td>
<td>ctagagccgacttgacaggtgctgctctctagtga</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 B reverse</td>
<td>ctagtcagcagacgccagccacagtccgaagtcgct</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 C forward</td>
<td>ctagaggtgctctctagtgtcacggccaaccga</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 C reverse</td>
<td>ctagtggtgcccgtgacactgacagcct</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 D forward</td>
<td>ctagaatcgtgtgtgatcgtcagtggtgctggcggcct</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 D reverse</td>
<td>ctagtcagcgcgcacccgagctgagtctacccgct</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 E forward</td>
<td>tgaatcgaacaggtgctccgct</td>
<td></td>
</tr>
<tr>
<td>Oligo P1 E reverse</td>
<td>ttagccgacacacctgtgctgattca</td>
<td></td>
</tr>
</tbody>
</table>

**Taqman primers and probes**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Probe</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>sigA</td>
<td>ttgagccgactcctgccgg</td>
<td>tcgtcagcgcctcct</td>
<td>tggctagcgcacctctccct</td>
</tr>
<tr>
<td>sigG</td>
<td>cgcttgccatagtcgcttc</td>
<td>caacagtgcgcacctcct</td>
<td>cgttgccgatcacaaga</td>
</tr>
<tr>
<td>rucV</td>
<td>cacccaaatctctgccctgaagaatct</td>
<td>gcctcagtgccacggccttc</td>
<td>cggccggttctggtt</td>
</tr>
<tr>
<td>recA</td>
<td>tcgggggcaaccgggcat</td>
<td>accgggcggtgacttac</td>
<td>cgccgagctggtggatg</td>
</tr>
</tbody>
</table>
**Appendix III: Microarray fold change Rv2017**

### Higher in H37Rv

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Mean Fold change</th>
<th>SEM</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv3489</td>
<td>-16.38</td>
<td>1.84E-02</td>
<td>Rv3489</td>
</tr>
<tr>
<td>PE</td>
<td>-15.70</td>
<td>1.25E-02</td>
<td>Rv3477</td>
</tr>
<tr>
<td>ctaB</td>
<td>-15.01</td>
<td>6.24E-03</td>
<td>Rv1451</td>
</tr>
<tr>
<td>groEL2</td>
<td>-13.31</td>
<td>1.14E-02</td>
<td>Rv0440</td>
</tr>
<tr>
<td>Rv1516c</td>
<td>-13.02</td>
<td>2.73E-02</td>
<td>Rv1516c</td>
</tr>
<tr>
<td>Rv0893c</td>
<td>-12.18</td>
<td>2.64E-02</td>
<td>Rv0893c</td>
</tr>
<tr>
<td>Rv0289</td>
<td>-11.70</td>
<td>1.14E-02</td>
<td>Rv0289</td>
</tr>
<tr>
<td>Rv3384c</td>
<td>-11.54</td>
<td>2.10E-02</td>
<td>Rv3384c</td>
</tr>
<tr>
<td>Rv3827c</td>
<td>-10.45</td>
<td>1.99E-02</td>
<td>Rv3827c</td>
</tr>
<tr>
<td>Rv3745c</td>
<td>-10.43</td>
<td>3.12E-02</td>
<td>Rv3745c</td>
</tr>
<tr>
<td>Rv3583c</td>
<td>-10.32</td>
<td>2.55E-02</td>
<td>Rv3583c</td>
</tr>
<tr>
<td>kasA</td>
<td>-9.99</td>
<td>5.43E-03</td>
<td>Rv2245</td>
</tr>
<tr>
<td>Rv1892</td>
<td>-9.40</td>
<td>2.95E-02</td>
<td>Rv1892</td>
</tr>
<tr>
<td>Rv1519</td>
<td>-9.37</td>
<td>4.70E-02</td>
<td>Rv1519</td>
</tr>
<tr>
<td>Rv1245c</td>
<td>-9.33</td>
<td>3.84E-02</td>
<td>Rv0287</td>
</tr>
<tr>
<td>Rv0287</td>
<td>-9.31</td>
<td>3.19E-02</td>
<td>Rv3418c</td>
</tr>
<tr>
<td>groES</td>
<td>-9.28</td>
<td>1.83E-02</td>
<td>Rv1133c</td>
</tr>
<tr>
<td>rplO</td>
<td>-9.10</td>
<td>2.77E-02</td>
<td>Rv0722</td>
</tr>
<tr>
<td>metE</td>
<td>-9.09</td>
<td>2.02E-02</td>
<td>Rv0280</td>
</tr>
<tr>
<td>rpmH</td>
<td>-8.51</td>
<td>3.63E-02</td>
<td>Rv0167</td>
</tr>
<tr>
<td>rpmD</td>
<td>-8.24</td>
<td>4.58E-02</td>
<td>Rv0291</td>
</tr>
<tr>
<td>PPE</td>
<td>-8.12</td>
<td>2.50E-02</td>
<td>Rv2840c</td>
</tr>
<tr>
<td>Rv0167</td>
<td>-8.09</td>
<td>2.24E-02</td>
<td>Rv3872</td>
</tr>
<tr>
<td>Rv0291</td>
<td>-7.84</td>
<td>1.97E-02</td>
<td>Rv1245c</td>
</tr>
<tr>
<td>Rv2840c</td>
<td>-7.81</td>
<td>1.87E-02</td>
<td>Rv0682</td>
</tr>
<tr>
<td>PE</td>
<td>-7.78</td>
<td>3.73E-02</td>
<td>Rv0850</td>
</tr>
<tr>
<td>rpsL</td>
<td>-7.77</td>
<td>2.47E-02</td>
<td>Rv2412</td>
</tr>
<tr>
<td>Rv0850</td>
<td>-7.70</td>
<td>3.29E-02</td>
<td>Rv2778c</td>
</tr>
<tr>
<td>rpsT</td>
<td>-7.70</td>
<td>2.99E-02</td>
<td>Rv1307</td>
</tr>
<tr>
<td>Rv2778c</td>
<td>-7.61</td>
<td>2.58E-02</td>
<td>Rv0723</td>
</tr>
<tr>
<td>atpH</td>
<td>-7.60</td>
<td>2.21E-02</td>
<td>Rv3136</td>
</tr>
<tr>
<td>PPE</td>
<td>-7.54</td>
<td>3.92E-02</td>
<td>Rv1872c</td>
</tr>
<tr>
<td>lldD2</td>
<td>-7.54</td>
<td>2.44E-02</td>
<td>Rv3443c</td>
</tr>
<tr>
<td>rplM</td>
<td>-7.53</td>
<td>1.95E-02</td>
<td>Rv3615c</td>
</tr>
<tr>
<td>Rv3615c</td>
<td>-7.51</td>
<td>2.83E-02</td>
<td>Rv0636</td>
</tr>
</tbody>
</table>

### Higher in Rv2017

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Mean Fold change</th>
<th>SEM</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0648</td>
<td>15.04</td>
<td>5.86E+00</td>
<td>Rv0648</td>
</tr>
<tr>
<td>Rv0600c</td>
<td>12.01</td>
<td>1.71E+00</td>
<td>Rv0600c</td>
</tr>
<tr>
<td>tcrA</td>
<td>11.45</td>
<td>3.00E+00</td>
<td>Rv0602c</td>
</tr>
<tr>
<td>Rv0842</td>
<td>10.79</td>
<td>2.31E+00</td>
<td>Rv0842</td>
</tr>
<tr>
<td>Rv2624c</td>
<td>10.53</td>
<td>2.13E+00</td>
<td>Rv2624c</td>
</tr>
<tr>
<td>Rv2407</td>
<td>10.42</td>
<td>4.68E+00</td>
<td>Rv2407</td>
</tr>
<tr>
<td>Rv2897c</td>
<td>10.38</td>
<td>3.48E+00</td>
<td>Rv2897c</td>
</tr>
<tr>
<td>Rv2648</td>
<td>9.99</td>
<td>3.21E+00</td>
<td>Rv2648</td>
</tr>
<tr>
<td>Rv3467</td>
<td>9.25</td>
<td>1.45E+00</td>
<td>Rv3467</td>
</tr>
<tr>
<td>Rv2205c</td>
<td>9.14</td>
<td>2.86E+00</td>
<td>Rv2205c</td>
</tr>
<tr>
<td>Rv2415c</td>
<td>9.08</td>
<td>2.65E+00</td>
<td>Rv2415c</td>
</tr>
<tr>
<td>adhE</td>
<td>9.03</td>
<td>5.02E+00</td>
<td>Rv0162c</td>
</tr>
<tr>
<td>Rv1259</td>
<td>8.86</td>
<td>1.72E+00</td>
<td>Rv1259</td>
</tr>
<tr>
<td>Rv0584</td>
<td>8.77</td>
<td>1.72E+00</td>
<td>Rv0584</td>
</tr>
<tr>
<td>Rv0094c</td>
<td>8.65</td>
<td>7.49E-01</td>
<td>Rv0094c</td>
</tr>
<tr>
<td>Rv0934c</td>
<td>8.58</td>
<td>1.89E+00</td>
<td>Rv0934c</td>
</tr>
<tr>
<td>Rv1965</td>
<td>8.48</td>
<td>2.70E+00</td>
<td>Rv1965</td>
</tr>
<tr>
<td>Rv1726</td>
<td>8.47</td>
<td>1.19E+00</td>
<td>Rv1726</td>
</tr>
<tr>
<td>Rv3446c</td>
<td>8.33</td>
<td>2.74E+00</td>
<td>Rv3446c</td>
</tr>
<tr>
<td>Rv0541c</td>
<td>8.32</td>
<td>1.36E+00</td>
<td>Rv0541c</td>
</tr>
<tr>
<td>Rv3529c</td>
<td>8.19</td>
<td>2.54E+00</td>
<td>Rv3529c</td>
</tr>
<tr>
<td>Rv0621</td>
<td>7.92</td>
<td>7.84E-01</td>
<td>Rv0621</td>
</tr>
<tr>
<td>Rv0614</td>
<td>7.70</td>
<td>1.40E+00</td>
<td>Rv0614</td>
</tr>
<tr>
<td>Rv0368c</td>
<td>7.63</td>
<td>1.16E+00</td>
<td>Rv0368c</td>
</tr>
<tr>
<td>Rv2278</td>
<td>7.60</td>
<td>3.10E+00</td>
<td>Rv2278</td>
</tr>
<tr>
<td>uspA</td>
<td>7.56</td>
<td>2.07E+00</td>
<td>Rv2316</td>
</tr>
<tr>
<td>Rv0622</td>
<td>7.33</td>
<td>9.86E-01</td>
<td>Rv0622</td>
</tr>
<tr>
<td>Rv1188</td>
<td>7.26</td>
<td>1.50E+00</td>
<td>Rv1188</td>
</tr>
<tr>
<td>Rv0953c</td>
<td>7.09</td>
<td>1.52E+00</td>
<td>Rv0953c</td>
</tr>
<tr>
<td>Rv3448</td>
<td>6.88</td>
<td>1.70E+00</td>
<td>Rv3448</td>
</tr>
<tr>
<td>Rv0077c</td>
<td>6.79</td>
<td>1.66E+00</td>
<td>Rv0077c</td>
</tr>
<tr>
<td>Rv2813</td>
<td>6.78</td>
<td>2.05E+00</td>
<td>Rv2813</td>
</tr>
<tr>
<td>PE</td>
<td>6.66</td>
<td>7.50E-01</td>
<td>Rv2340c</td>
</tr>
<tr>
<td>recC</td>
<td>6.58</td>
<td>1.27E+00</td>
<td>Rv0631c</td>
</tr>
<tr>
<td>Rv1999c</td>
<td>6.52</td>
<td>1.79E+00</td>
<td>Rv1999c</td>
</tr>
<tr>
<td>Gene</td>
<td>Log2 Fold Change</td>
<td>p-Value</td>
<td>Gene</td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>Rv0636</td>
<td>-7.47</td>
<td>2.32E-02</td>
<td>Rv3633</td>
</tr>
<tr>
<td>Rv3633</td>
<td>-7.44</td>
<td>1.33E-02</td>
<td>Rv1981c</td>
</tr>
<tr>
<td>nrdF</td>
<td>-7.34</td>
<td>4.11E-02</td>
<td>Rv0702</td>
</tr>
<tr>
<td>rplD</td>
<td>-7.32</td>
<td>3.17E-02</td>
<td>Rv0282</td>
</tr>
<tr>
<td>Rv0282</td>
<td>-7.28</td>
<td>1.96E-02</td>
<td>Rv0719</td>
</tr>
<tr>
<td>rplF</td>
<td>-7.15</td>
<td>2.02E-02</td>
<td>Rv0637</td>
</tr>
<tr>
<td>Rv0637</td>
<td>-7.13</td>
<td>2.77E-02</td>
<td>Rv1013</td>
</tr>
<tr>
<td>Rv1013</td>
<td>-7.11</td>
<td>5.46E-02</td>
<td>Rv0667</td>
</tr>
<tr>
<td>pk516</td>
<td>-6.87</td>
<td>1.18E-02</td>
<td>Rv2160c</td>
</tr>
<tr>
<td>rpsB</td>
<td>-6.89</td>
<td>1.85E-02</td>
<td>Rv3633</td>
</tr>
<tr>
<td>Rv2160c</td>
<td>-6.88</td>
<td>4.43E-02</td>
<td>Rv3924c</td>
</tr>
<tr>
<td>whaB1</td>
<td>-6.74</td>
<td>1.25E-02</td>
<td>Rv0701</td>
</tr>
<tr>
<td>rplC</td>
<td>-6.69</td>
<td>1.60E-02</td>
<td>Rv1066</td>
</tr>
<tr>
<td>Rv1066</td>
<td>-6.66</td>
<td>3.60E-02</td>
<td>Rv1993c</td>
</tr>
<tr>
<td>Rv1993c</td>
<td>-6.66</td>
<td>1.59E-02</td>
<td>Rv2816c</td>
</tr>
<tr>
<td>PPE</td>
<td>-6.38</td>
<td>1.70E-02</td>
<td>Rv1518</td>
</tr>
<tr>
<td>Rv1518</td>
<td>-6.34</td>
<td>4.18E-02</td>
<td>Rv1195</td>
</tr>
<tr>
<td>Rv1195</td>
<td>-6.31</td>
<td>1.81E-02</td>
<td>Rv3048c</td>
</tr>
<tr>
<td>nrdG</td>
<td>-6.25</td>
<td>5.07E-02</td>
<td>Rv1641</td>
</tr>
<tr>
<td>infC</td>
<td>-6.25</td>
<td>2.69E-02</td>
<td>Rv1322</td>
</tr>
<tr>
<td>Rv1322</td>
<td>-6.20</td>
<td>8.26E-02</td>
<td>Rv0665</td>
</tr>
<tr>
<td>Rv0665</td>
<td>-6.20</td>
<td>4.00E-02</td>
<td>Rv3019c</td>
</tr>
<tr>
<td>Rv3019c</td>
<td>-6.19</td>
<td>2.66E-02</td>
<td>Rv3135</td>
</tr>
<tr>
<td>sigB</td>
<td>-6.15</td>
<td>3.45E-02</td>
<td>Rv1388</td>
</tr>
<tr>
<td>PPE</td>
<td>-6.13</td>
<td>4.92E-02</td>
<td>Rv3093</td>
</tr>
<tr>
<td>mlHf</td>
<td>-6.12</td>
<td>2.43E-02</td>
<td>Rv1630</td>
</tr>
<tr>
<td>Rv0309</td>
<td>-6.12</td>
<td>3.18E-02</td>
<td>Rv0653c</td>
</tr>
<tr>
<td>rpsA</td>
<td>-6.10</td>
<td>1.92E-02</td>
<td>Rv0608</td>
</tr>
<tr>
<td>Rv0608</td>
<td>-6.04</td>
<td>1.91E-02</td>
<td>Rv1308</td>
</tr>
<tr>
<td>Rv1308</td>
<td>-6.04</td>
<td>1.91E-02</td>
<td>Rv1383</td>
</tr>
<tr>
<td>Rv1383</td>
<td>-5.96</td>
<td>3.16E-02</td>
<td>Rv1054</td>
</tr>
<tr>
<td>Rv2186c</td>
<td>-5.95</td>
<td>4.06E-02</td>
<td>Rv2909c</td>
</tr>
<tr>
<td>Rv2909c</td>
<td>-5.95</td>
<td>4.06E-02</td>
<td>Rv1247c</td>
</tr>
<tr>
<td>Rv1247c</td>
<td>-5.92</td>
<td>4.63E-02</td>
<td>Rv3183</td>
</tr>
<tr>
<td>Rv3183</td>
<td>-5.86</td>
<td>7.11E-02</td>
<td>Rv3456c</td>
</tr>
<tr>
<td>Rv3456c</td>
<td>-5.86</td>
<td>7.11E-02</td>
<td>Rv3277</td>
</tr>
<tr>
<td>Rv3277</td>
<td>-5.85</td>
<td>6.30E-02</td>
<td>Rv0701</td>
</tr>
<tr>
<td>dnaK</td>
<td>-5.69</td>
<td>3.90E-02</td>
<td>Rv0701</td>
</tr>
</tbody>
</table>

---

### Appendix III

<table>
<thead>
<tr>
<th>Gene</th>
<th>Log2 Fold Change</th>
<th>p-Value</th>
<th>Gene</th>
<th>Log2 Fold Change</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0636</td>
<td></td>
<td></td>
<td>Rv3633</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv3633</td>
<td></td>
<td></td>
<td>Rv1981c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nrdF</td>
<td></td>
<td></td>
<td>Rv0702</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rplD</td>
<td></td>
<td></td>
<td>Rv0282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0282</td>
<td></td>
<td></td>
<td>Rv0719</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rplF</td>
<td></td>
<td></td>
<td>Rv0637</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0637</td>
<td></td>
<td></td>
<td>Rv1013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1013</td>
<td></td>
<td></td>
<td>Rv0667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pk516</td>
<td></td>
<td></td>
<td>Rv2160c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rpsB</td>
<td></td>
<td></td>
<td>Rv3633</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv2160c</td>
<td></td>
<td></td>
<td>Rv3924c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>whaB1</td>
<td></td>
<td></td>
<td>Rv0701</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rplC</td>
<td></td>
<td></td>
<td>Rv1066</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1066</td>
<td></td>
<td></td>
<td>Rv1993c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1993c</td>
<td></td>
<td></td>
<td>Rv2816c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPE</td>
<td></td>
<td></td>
<td>Rv1518</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1518</td>
<td></td>
<td></td>
<td>Rv1195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1195</td>
<td></td>
<td></td>
<td>Rv3048c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nrdG</td>
<td></td>
<td></td>
<td>Rv1641</td>
<td></td>
<td></td>
</tr>
<tr>
<td>infC</td>
<td></td>
<td></td>
<td>Rv1322</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1322</td>
<td></td>
<td></td>
<td>Rv0665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0665</td>
<td></td>
<td></td>
<td>Rv3019c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv3019c</td>
<td></td>
<td></td>
<td>Rv3135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sigB</td>
<td></td>
<td></td>
<td>Rv1388</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPE</td>
<td></td>
<td></td>
<td>Rv3093</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mlHf</td>
<td></td>
<td></td>
<td>Rv1630</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0309</td>
<td></td>
<td></td>
<td>Rv0653c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rpsA</td>
<td></td>
<td></td>
<td>Rv0608</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv0608</td>
<td></td>
<td></td>
<td>Rv1308</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1308</td>
<td></td>
<td></td>
<td>Rv1383</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1383</td>
<td></td>
<td></td>
<td>Rv1054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv2186c</td>
<td></td>
<td></td>
<td>Rv2909c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv2909c</td>
<td></td>
<td></td>
<td>Rv1247c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1247c</td>
<td></td>
<td></td>
<td>Rv3183</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv3183</td>
<td></td>
<td></td>
<td>Rv3456c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv3456c</td>
<td></td>
<td></td>
<td>Rv3277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dnaK</td>
<td></td>
<td></td>
<td>Rv0701</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix III: Gene with a 5-fold of more differential gene expression in ΔRv2017 strain compared to H37Rv.

Fold change was calculated as ΔRv2017 strain expressed as a proportion of H37Rv, genes were identified that were 5-fold different in expression level between ΔRv2017 strain compared to H37Rv. The left column represents the reciprocal fold change (indicated with -) therefore genes are expressed to a lower level in ΔRv2017 strain compared to H37Rv. Whereas the right hand column represents genes which are upregulated in ΔRv2017 strain by more than 5-fold compared to H37Rv.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold Change</th>
<th>Expression Level</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>rpsJ</td>
<td>-5.58</td>
<td>2.90E-02</td>
<td>Rv1344</td>
</tr>
<tr>
<td>Rv1344</td>
<td>-5.58</td>
<td>5.46E-02</td>
<td>Rv1464</td>
</tr>
<tr>
<td>Rv1464</td>
<td>-5.53</td>
<td>1.29E-02</td>
<td>Rv3856c</td>
</tr>
<tr>
<td>Rv3856c</td>
<td>-5.52</td>
<td>2.91E-02</td>
<td>Rv1361c</td>
</tr>
<tr>
<td>PPE</td>
<td>-5.50</td>
<td>2.45E-02</td>
<td>Rv2927c</td>
</tr>
<tr>
<td>Rv2927c</td>
<td>-5.49</td>
<td>4.01E-02</td>
<td>Rv1072</td>
</tr>
<tr>
<td>Rv1072</td>
<td>-5.44</td>
<td>2.50E-02</td>
<td>Rv0639</td>
</tr>
<tr>
<td>nusG</td>
<td>-5.43</td>
<td>6.46E-02</td>
<td>Rv3866</td>
</tr>
<tr>
<td>Rv3866</td>
<td>-5.43</td>
<td>1.91E-02</td>
<td>Rv3487c</td>
</tr>
<tr>
<td>lipF</td>
<td>-5.41</td>
<td>4.02E-02</td>
<td>Rv2247</td>
</tr>
<tr>
<td>accD6</td>
<td>-5.36</td>
<td>1.81E-02</td>
<td>Rv0685</td>
</tr>
<tr>
<td>tuf</td>
<td>-5.34</td>
<td>2.89E-02</td>
<td>Rv1080c</td>
</tr>
<tr>
<td>greA</td>
<td>-5.31</td>
<td>2.19E-02</td>
<td>Rv0283</td>
</tr>
<tr>
<td>Rv0283</td>
<td>-5.25</td>
<td>2.04E-02</td>
<td>Rv2381c</td>
</tr>
<tr>
<td>mbtD</td>
<td>-5.22</td>
<td>2.45E-02</td>
<td>Rv3407</td>
</tr>
<tr>
<td>Rv3407</td>
<td>-5.18</td>
<td>6.49E-02</td>
<td>Rv3316</td>
</tr>
<tr>
<td>sdhC</td>
<td>-5.16</td>
<td>4.59E-02</td>
<td>Rv2710</td>
</tr>
<tr>
<td>Rv0292</td>
<td>-5.12</td>
<td>2.84E-02</td>
<td>Rv0292</td>
</tr>
<tr>
<td>groEL1</td>
<td>-5.11</td>
<td>2.94E-02</td>
<td>Rv3417c</td>
</tr>
<tr>
<td>Rv3300c</td>
<td>-5.10</td>
<td>8.23E-02</td>
<td>Rv3300c</td>
</tr>
<tr>
<td>rpsG</td>
<td>-5.10</td>
<td>4.26E-02</td>
<td>Rv0683</td>
</tr>
<tr>
<td>atpD</td>
<td>-5.07</td>
<td>4.03E-02</td>
<td>Rv1310</td>
</tr>
<tr>
<td>Rv2376c</td>
<td>-5.06</td>
<td>2.60E-02</td>
<td>Rv2376c</td>
</tr>
<tr>
<td>cysM3</td>
<td>5.22</td>
<td>8.93E-01</td>
<td>Rv0848</td>
</tr>
<tr>
<td>PPE</td>
<td>5.06</td>
<td>1.15E+00</td>
<td>Rv3621c</td>
</tr>
<tr>
<td>PPE</td>
<td>5.12</td>
<td>5.03E-01</td>
<td>Rv2621c</td>
</tr>
<tr>
<td>Rv3186</td>
<td>5.05</td>
<td>1.55E+00</td>
<td>Rv3186</td>
</tr>
<tr>
<td>xylB</td>
<td>5.02</td>
<td>6.48E-01</td>
<td>Rv0729</td>
</tr>
<tr>
<td>Rv2488c</td>
<td>5.01</td>
<td>1.19E+00</td>
<td>Rv2488c</td>
</tr>
</tbody>
</table>
Appendix IV - Keystone abstract

A Novel Mechanism Regulating DNA-Damage Inducible Gene Expression in *Mycobacterium tuberculosis*

Lisa Dawson and Elaine Davis, Division of Mycobacterial Research, National Institute for Medical Research, Mill Hill, London, UK.

*Mycobacterium tuberculosis* (*Mtb*) is an intracellular pathogen, which causes human tuberculosis. Infected macrophage produce reactive oxygen and nitrogen intermediates, known to damage DNA; therefore DNA damage repair is thought to be important in survival of *Mtb* in the host. RecA is an integral part of the DNA damage repair system, being highly conserved and ubiquitous. In *Mtb* the *recA* gene is expressed from two promoters. One promoter (P2) is regulated by LexA, while the other promoter (P1) remains DNA damage inducible independently of the classical LexA/RecA system. Studies using gel shift assays have failed to detect a repressor or activator protein which binds the PI promoter from total cell free extracts of *Mtb*, unlike the P2 promoter, which clearly binds LexA. This may be due to low levels of expression. Therefore, five proteins with predicted regulatory functions, which are induced following DNA damage in both wild type and Δ*recA* strains of *Mtb*, are being expressed to examine binding individually.

Alternatively, regulation of gene expression from the P1 promoter could be controlled by a sigma factor. Sigma factors are protein subunits of RNA polymerase, which confer specificity of binding to the promoter of a specific gene. The function and/or expression of sigma factors can be regulated, resulting in regulation of expression of the genes they transcribe. The sigma factors SigG, SigE and SigH are induced following DNA damage in both wild type and Δ*recA* strains of *Mtb*. A *sigG* knockout is being constructed and will be used alongside *sigE* and *sigH* knockouts (kindly provided by R.N.Husson) to determine if any of these sigma factors regulate gene expression of the P1 promoter of *recA*.