Citation for published version:
https://doi.org/10.1016/j.tube.2013.05.005

DOI:
10.1016/j.tube.2013.05.005

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Publisher Rights
CC BY
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

University of Bath

Alternative formats
If you require this document in an alternative format, please contact: openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 09. Mar. 2022
MOLECULAR ASPECTS

Characterisation of the Mycobacterium tuberculosis alternative sigma factor SigG: Its operon and regulon

Alison Gaudiona, Lisa Dawsonb, Elaine Davis, Katherine Smolletta

Division of Mycobacterial Research, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

ARTICLE INFO

Article history:
Received 26 February 2013
Received in revised form 15 May 2013
Accepted 19 May 2013

Keywords:
Mycobacterium tuberculosis
Sigma factors
SigG
DNA damage response

SUMMARY

A major step in the pathogenesis of Mycobacterium tuberculosis is the ability to survive inside macrophages, where it is exposed to a number of DNA damaging agents. The alternative sigma factor SigG has been shown to be upregulated by DNA damaging agents and by macrophage infection. SigG is induced by DNA damage and is part of the RecA independent DNA damage response, and mutation of sigG did not increase sensitivity to DNA damage. The RecA-independent DNA damage response was subsequently found to be regulated by the Clp protease regulator ClpR. We sought to determine a role for SigG both by looking at the control of its expression and by determining the genes that SigG itself regulates. In our previous work we attempted to compare gene expression in a sigG mutant strain to wild-type M. tuberculosis H37Rv but found no significant differences, possibly due to SigG being the lowest expressed of all sigma factors under normal growth conditions.

1. Introduction

Despite the availability of treatments and a vaccine Mycobacterium tuberculosis remains a major cause of human disease worldwide and was responsible for 1.4 million deaths in 2011. During the course of infection and transmission to new hosts M. tuberculosis is exposed to a number of stresses and its ability to adapt to these stresses is a key component of its survival. Not surprisingly the genome sequence of M. tuberculosis revealed over 100 genes encoding regulatory proteins involved in gene expression, including 13 sigma factors. Sigma factors are components of RNA polymerase that contain the promoter recognition domains. There are several different classes of sigma factor ranging from house-keeping sigma factors to the alternative sigma factors, which respond to specific external stimuli (reviewed in Helmann et al.).

The M. tuberculosis genome encodes 10 alternative sigma factors; here we examine the alternative sigma factor, SigG. Determining the conditions that induce expression of a particular alternative sigma factor can be useful in designating function. SigG is induced by DNA damage and is part of the RecA independent DNA damage response. There are two known pathways for induction of DNA repair genes in response to DNA damage in M. tuberculosis. One of these is the classical SOS response that is controlled by LexA and RecA controlling genes involved in DNA damage induced mutagenesis, cell cycle control and containing phage attachment sites. The other is the RecA-independent response that controls the majority of DNA damage inducible genes, including many of the genes involved in DNA repair. A DNA motif resembling a non-canonical promoter common to a number of genes regulated in this way has been identified leading to the suggestion that these genes were controlled by an alternative sigma factor. However, SigG was found not to control transcription of either the SOS or RecA-independent response genes, and mutation of sigG did not increase sensitivity to DNA damage. The RecA-independent DNA damage response was subsequently found to be regulated by the Clp protease regulator ClpR.

We sought to determine a role for SigG both by looking at the control of its expression and by determining the genes that SigG itself regulates. In our previous work we attempted to compare gene expression in a sigG mutant strain to wild-type M. tuberculosis H37Rv but found no significant differences, possibly due to SigG being the lowest expressed of all sigma factors under normal growth conditions.
growth conditions. Therefore, in this study we examined the regulon of sigG using an over-expression strain. We found that instead of controlling genes involved in DNA repair it controls genes with a potential role in detoxification.

2. Materials and methods

2.1. Bacterial strains and culture conditions

The mycobacterial strains used were Mycobacterium smegmatis mc²155 and *M. tuberculosis* wild-type strain H37Rv. sigG mutant strains in H37Rv ΔsigG1 and ΔsigG2, and a sigG operon deletion in H37Rv ΔsigGW0 (this study). Mycobacterial cultures grown in Dubos medium (Difco) supplemented with albumin and 0.2% glycerol or on Difco Middlebrook 7H11 agar plates (Beckton Dickinson) supplemented with 4% albumin and 0.5% glycerol. *M. tuberculosis* liquid cultures were grown at 37°C incubator at 2 rpm. All procedures with live *M. tuberculosis* were carried out under ACDP containment level 3 conditions. Antibiotics were added as appropriate: kanamycin was used at 25 μg ml⁻¹, hygromycin was used at 50 μg ml⁻¹. To induce DNA damage cultures were divided into two aliquots at an OD₆₀₀ 0.3–0.4 and one sample was treated with 0.02 μg ml⁻¹ mitomycin C for 24 h. The other sample was incubated in parallel without treatment to provide an uninduced control. The plasmids and primers used in this study are described in Supplementary Tables 1 and 2 respectively.

2.2. Protein preparation and antibody production

Recombinant SigG was produced by expression of His-tagged *M. tuberculosis* SigG from plasmid pHH5 in *Escherichia coli* strain Tuner. Protein was purified using an AKTA prime (Amersham Biosciences) first using a nickel-loaded HiTrap chelating HP column (GE Healthcare), followed by purification by a HiLoad 26/60 Superdex 200 prep grade gel filtration column (GE Healthcare). Pure SigG was then used to immunise rabbits to produce polyclonal antibody at a 1:5000 dilution (Qiagen). The blot was developed using ECL Western blotting detection reagents (GE Healthcare).

2.3. Preparation of cell free extracts, Western blot and β-galactosidase assays

Cell free extracts were produced as described previously. The supernatants were filtered through a low-binding Durapore 0.22-μm-pore-size membrane filter (Ultrafree-MC; Millipore) to ensure complete removal of bacteria before removal from containment facilities.

Western blots were performed using rabbit anti-SigG at 1:1000 dilution and anti-rabbit conjugated to horseradish peroxidase (Dako), at a 1:5000 dilution, or using Penta-His conjugated antibody at a 1:5000 dilution (Qiagen). The blot was developed using ECL Western blotting detection reagents (GE Healthcare).

For β-galactosidase activity, protein levels of cell-free extracts were quantified using a BCA kit (Pierce) and normalisation to the corresponding value of the housekeeping gene 16s rRNA or sigA to give relative expression level. Where required the relative expression of each sample was divided by that of the untreated sample or empty vector control to give the induction ratio.

2.4. RNA extraction

RNA was prepared from mycobacteria using the FastRNA Pro Blue kit (Qbiogene). Contaminating DNA was removed using RNase-free DNase (Roche), and RNA purified using RNeasy Minikit (Qiagen), or with a TURBO DNA-free kit (Ambion). RNA quality was determined using an Agilent Bioanalyzer 2100 (Agilent Technologies).

2.5. Transcriptional start site mapping

Primer extension was carried out using Primer Extension System (Promega). Briefly, primer sigGPExt (Supplementary Table 2) was end-labelled with ³²P-ATP and extension reactions carried out using 40–100 μg RNA. Reactions were run on an 8% denaturing polyacrylamide gel, alongside sequencing reactions (T7 Sequenase Kit Amersham), and visualised using autoradiography. The transcriptional start site of Rv0887c was mapped using the GeneRacer kit (Invitrogen) for RNA ligase-mediated rapid amplification of 5′ cDNA ends. Briefly the GeneRacer Oligo (Invitrogen) was ligated to the 5′-ends of RNA from *M. tuberculosis* H37Rv. cDNA products for the genes of interest were produced by RT-PCR using a primer specific to the GeneRacer Oligo along with a gene specific primer (Supplementary Table 2). Amplified cDNA ends were cloned into pCR 4-TOPO (Invitrogen) for sequencing.

2.6. Microarray

Whole genome *M. tuberculosis* microarray slides were obtained from the Bacterial Microarray Group at St. George’s, London. Cy5-labelled RNA versus Cy3-labelled DNA hybridizations were performed and analysed as previously described. Significant differences were determined using two-way ANOVA with the Benjamini and Hockberg False Discovery Rate correction. Fully annotated microarray data have been deposited in BugBase (accession number TBC) and ArrayExpress (accession number TBC).

2.7. Quantitative RT-PCR

cDNA synthesis was performed using Superscript II (Invitrogen) and random hexamer primers. Real-time quantitative PCR was carried out using Fast SYBR green master mix (Applied Biosystems) on an Applied Biosystems 7500 Fast instrument and analysed with 7500 Fast SDS software version 1.4. Gene specific primers were designed using Primer Express version 3.0 (Applied Biosystems). For each gene cDNA samples and their RT negative controls were run along with a set of genomic DNA standards to give a quantity of specific cDNA per sample. Values for the RT negative controls were subtracted, followed by normalisation to the corresponding value of the housekeeping gene *rrs* encoding 16s rRNA or *sigA*, to give relative expression level. Where required the relative expression of each sample was divided by that of the untreated sample or empty vector control to give the induction ratio.

2.8. Murine bone marrow-derived macrophage (BMDM) infection model

Infection of BMDM was carried out as described previously. Macrophages were seeded into 24-well plates at a density of 2 × 10⁵ cells per well. To activate macrophages 10 μg ml⁻¹ murine interferon-γ (IFN-γ; Roche) was added. Naïve and activated macrophages were infected with *M. tuberculosis* strains grown to mid-log (OD₆₀₀ ~ 0.5) and resuspended in PBS containing 0.05% Tween 80 at an MOI of 0.1:1. Survival and multiplication of *M. tuberculosis* was determined by calculating CFUs at time points post infection.
3. Results

3.1. Characterisation of the SigG operon

As a first step in the characterisation of SigG, we confirmed its genetic organisation. The gene immediately upstream of sigG is transcribed in the opposite orientation, and downstream there are three genes transcribed in the same orientation (Figure 1A). Examination of correlated gene expression predicts that an operon runs from sigG-Rv0180c, missing lprO (Pearson's correlation coefficient 0.37 sigG-Rv0180c, 0.08 sigG-lprO, TB database www.tbdb.org). We confirmed this to be the case by RT-PCR (Figure 1) using a series of primers spanning from the beginning of sigG to the end of lprO. Transcripts could be obtained from sigG to primer lprOR1 located 250 bp within lprO as well as from the end of Rv0181c to lprOR1, but not to the 3' end of lprO indicating that the operon terminates within lprO. A further transcript could be detected between the 3' end of lprO and 300 bp upstream of the Rv0180c stop codon. Therefore, the sigG operon consists of sigG, Rv0181c and Rv0180c, while the lprO transcript starts from a promoter internal to Rv0180c.

3.2. Determining sigG transcriptional start sites

To determine potential promoters for the sigG operon the transcript start sites were mapped by primer extension using RNA extracted from wild-type M. tuberculosis H37Rv and ΔsigG strains grown with and without mitomycin C treatment (Figure 2A). This revealed three potential start sites located at −12, −57 and +114 bp from the annotated start codon of sigG, labelled P3, P2 and P1 respectively. P1 shows induction by the DNA damaging agent, mitomycin C, and also gives the strongest band both with and without mitomycin C treatment. P2 and P3 give much weaker bands with P2 showing some induction by mitomycin C and P3 being potentially SigG dependent, being only identified from wild-type RNA.

3.3. Mapping SigG translational start site

As both P1 and P2 are located downstream of the annotated translational start site of sigG, this site must be incorrect. Located immediately downstream of P1 are 2 potential GTG start codons (Figure 2F). To determine if one of these is the start codon for sigG we performed a frame-shift mutagenesis translational start site assay as previously described.40 Due to the low expression of SigG during normal growth25 protein could not be detected. Instead a SigG over-expression construct, pKS09, was developed. This plasmid was derived from the moderate copy number plasmid pMV26144 and contains the entire sigG annotated coding region plus 500 bp upstream, to include its own regulatory elements (see Supplementary Table 1). This region also contains a significant proportion of the upstream gene Rv0183 (Figure 2B). Quantitative RT-PCR analysis found substantial over-expression of sigG when transformed into wild-type M. tuberculosis H37Rv. Expression of Rv0181c and Rv0180c was unaffected in this strain (Figure 2C), indicating that the genomic sigG operon was not over-expressed. Further RT-PCR analysis found that this transcript started within the vector sequence (data not shown) indicating that sigG over-expression was due to read-through from a vector encoded promoter and not auto-regulation by SigG. Western blot on cell free extract obtained from this strain, using polyclonal antibody against

![Figure 1](https://example.com/figure1.png)
M. tuberculosis SigG, found that the expressed protein was 4 kDa smaller than expected from the annotation (Figure 2D). Two individual single base pair deletions were made in this construct immediately up or downstream of the 1st potential GTG start codon downstream of P1, termed pKS09TTSmut1 and pKS09TTSmut2 respectively (Figure 2F). These constructs were expressed in M. smegmatis, which has been shown to be a suitable model for determining translational start codons of M. tuberculosis proteins.40 Western blots were performed on cell-free extracts obtained from these strains grown to mid-log phase, and equal loading was determined by Coomassie staining of an equivalent gel (Figure 2E). SigG was undetectable in wild-type M. smegmatis, whereas the over-expression construct gave high levels of SigG protein. SigG could be detected in pKS09TTSmut1 but was not detectable with pKS09TTSmut2. This indicates that the translational start site for SigG is between these two mutations and so must be located at the GTG 114 bp downstream of the annotated start codon resulting in a protein 38 amino acids shorter.
3.4. Promoter identification for sigG

SigG translation starts immediately downstream of P1, which corresponds with the previously identified RecA_NDp promoter\textsuperscript{21} (Figure 2F). To determine whether each of the potential promoters were active, transcriptional fusions for each promoter were constructed fusing approximately 60 bp upstream of each transcriptional start site to the reporter gene lacZ, giving plasmids pLDlac1, pLDlac2 and pLDlac3 (containing P1, P2 and P3 respectively, Figure 3E). The promoter activity for each construct was determined in wild-type \textit{M. tuberculosis} H37Rv with and without mitomycin C treatment (Figure 3A). Only P1 gave activity above background and this was increased approximately 3-fold upon DNA damage induction. An A to C mutation was made in the –10 region of RecA_NDp of P1 giving pLDlac1-mut. This mutation is known to inactivate this promoter family\textsuperscript{22} and abolished the activity from the P1 promoter. Therefore, promoter 1 is active and is responsible for DNA damage induction of the sigG operon. To determine whether the other potential promoters play a role in expression of sigG, pGA04 was constructed, containing all 3 potential promoters fused to the reporter gene lacZ (Figure 3E). This longer construct could also contain binding sites for other potential regulatory elements necessary for activity that may be missing in the smaller constructs. Mutations were made in the –10 region of P1 (as above) giving pAG04-mut1, and also in the –10 regions of P2 and P3 giving plasmids, pAG04-mut2 (P1 and P2 mutation) and pAG04-mut3 (P1, P2 and P3 mutation). The promoter activity for these plasmids was determined in wild-type \textit{M. tuberculosis} H37Rv (Figure 3B). Mutation of the RecA_NDp (pAG04-mut1) reduced but did not abolish activity from the wild-type promoter (pAG04); while this was not a significant reduction this may be due to the presence or absence of additional regulatory elements of P1 in the longer construct. Additional mutation of the –10 region of P2, (giving pAG04-mut2), reduced this activity further, to approximately 1 quarter that of the wild-type, whereas mutation of P3 had no additional effect. Therefore sigG is expressed from at least two promoters. To confirm this at the miRNA level qRT-PCR was performed on RNA isolated from \textit{M. tuberculosis} H37Rv with and without mitomycin C treatment (Figure 3C). Primers were located either downstream of promoter P1, therefore detecting transcripts originating from all three promoters, or between promoter P2 and P1, therefore detecting transcripts originating from promoters P3 and P2 only. Detection of transcripts downstream of P1 showed a high level of expression and was induced 3-fold upon mitomycin C induction. A similar level of transcript could be detected upstream using primers located between P1 and P2 but outside of the \textit{Rv0183} coding region. This transcript reduced upon DNA damage induction. The difference between the transcripts detected by the two

Figure 3. Promoter activity for sigG. (A) β-galactosidase activity for \textit{M. tuberculosis} containing pJ414 empty vector (EV), or vectors containing translational fusions of each promoter identified in Figure 2A to reporter gene lacZ (as indicated in E), pLDlac1 (P1), pLDlac2 (P2) and pLDlac3 (P3) and P1 containing a mutation of the –10 region (pLDlac1-mut, P1-mut) with (black bars) and without (white bars) treatment with 0.02 μg ml\textsuperscript{-1} mitomycin C. (B) Translational fusion of 339 bp region upstream of sigG containing all 3 putative promoters (pAG04, WT Pr) with mutations to –10 regions of P1 (pAG04-mut1), P1 and P2 (pAG04-mut2) and P1, P2 and P3 (pAG04-mut3). (C) qRT-PCR using primers sigGqRTF and sigGqRTR (located downstream from P1, detecting transcript from all promoters), and sigGP2qRTF and sigGP2qRTR (located between transcriptional starts sites P1 and P2, detecting transcript starting upstream of P1). Expression was assessed from wild-type \textit{M. tuberculosis} H37Rv with (black bars) and without (white bars) mitomycin C treatment and normalised to sigA. (D) β-galactosidase activity for pAG04, containing the full length sigG promoter in wild-type \textit{M. tuberculosis} H37Rv and ΔsigGWO. (E) 339 bp DNA sequence of promoter region of sigG used to make construct pAG04 including P1, P2 and P3 with potential –10 and –35 promoter regions shaded grey. Locations of the DNA fragments used in the individual pLDlac promoter constructs are indicated by boxes, together with the transcriptional start site for each promoter (arrow and bold) and the annotated (ATG) and experimentally determined (GTG) start codons \textbf{BOLD ITALIC CAPITALS}. Data represents mean ± standard deviation from 3 biological replicates, statistical significance by two-tailed \textit{T} test (P value * < 0.1, ** < 0.01).
under DNA damaging conditions P1 is the primary promoter for the sigG operon. One other promoter possibly also controls expression of sigG, however, due to the proximity of the Rv0183 and sigG coding regions, the sigGP2qRTF primer is located only 37 bp upstream of the annotated Rv0183 coding region. It is therefore possible that the transcript detected upstream of sigG promoter P1 is actually the Rv0183 transcript.

Primers extension analysis indicated that P3 was potentially SigG dependent (Figure 2A); however, over-expression of SigG did not affect expression from the rest of the sigG operon, suggesting that expression of the genomic operon is not SigG dependent (Figure 2C). Therefore we examined activity of the SigG promoter region in ΔsigGW0, in which the SigG-Rv0180c operon was deleted and replaced with a gentamicin resistance cassette by homologous recombination (Supplementary Figure 1A). The first 32 bp of sigG and last 380 bp of Rv0183 were included in the targeting construct to prevent effects on the promoters of adjacent genes. The absence of expression of sigG, Rv0181c and Rv0180c in ΔsigGW0 was confirmed by qRT-PCR (Supplementary Figure 1B). The sigG promoter construct pMG04, containing all three potential promoters was expressed in ΔsigGW0, but showed no difference in activity compared to expression in wild-type M. tuberculosis (Figure 3D). The dominant P1 promoter could be masking the effects of a sigG dependent promoter. However, it is possible that lack of detection of the P3 transcriptional start site in ΔsigGW1 is due to a defect in this strain, which also contains a spontaneous mutation affecting the cell wall.39

3.5. Investigation of the role of SigG in vivo

The potential role of SigG during in vivo infection was assessed in a macrophage infection model. Wild-type H37Rv, ΔsigG2 and ΔsigGW0, were used to infect naïve and activated murine bone marrow derived macrophages (BMDM). Neither the ΔsigG2 or ΔsigGW0 strains were attenuated compared to wild-type (Figure 4).

3.6. Identification of the SigG regulon

The possible function of SigG was further assessed by determining its regulon. Previous investigations examining the gene expression profiles of a sigG mutant strain with and without DNA damage induction found no significant differences from the wild-type H37Rv strain.39 Therefore, in this investigation we compared the expression profiles of M. tuberculosis H37Rv containing pKS09, a SigG over-expression construct, to that of H37Rv containing an empty vector (pKS12) by microarray (Table 1). This identified 13 genes significantly up-regulated in the over-expression strain by more than 1.75 fold (P value < 0.05). The two most upregulated genes were sigG and Rv0183. A significant portion of Rv0183 is located within the pKS09 construct used for over-expression (see Figure 2B). Quantitative-PCR was performed using two sets of primers, one located either within Rv0183 and contained in pKS09 (primer set 2), the other within Rv0183 but outside of the region contained in pKS09 (primer set 1). This confirmed that the higher level of Rv0183 RNA observed in the microarray analysis is an artefact of the over-expression strain and not specific regulation by SigG (Figure 2C). Rv0181c and Rv0180c were not identified as being upregulated in pKS09, which confirmed previous qRT-PCR analysis (Figure 2C) and shows further shows that SigG is not autoregulatory. Quantitative PCR confirmed that 9 of the remaining 11 genes were upregulated more than 1.5-fold compared to an empty vector strain (Table 1). To confirm whether these genes identified by microarray analysis were upregulated due specifically to the SigG protein, or artefacts of the over-expression construct, induction relative to the empty vector containing strain was assessed in H37Rv containing pKS09FS, in which a single base pair deletion was made within the coding region of sigG. This resulted in a strain which gave the same level of RNA but no functional SigG protein due to a frame-shift (Table 1). Induction of Rv0942, vapB15, vapC15, gbd1 and Rv2004 occurred in both the SigG overexpression (pKS09) and frame-shift (pKS09FS) strains and so most likely represent false positives. The vapB15 operon encodes a toxin-antitoxin system, the toxin of which degrades mRNA and has been shown to be induced during stress such as hypoxia.33 This operon may be instead responding to the abnormal pattern of RNA expression rather than to SigG itself. Induction of jfrB, Rv0887c, Rv0911, Rv0912, prpD and inhA was significantly reduced in the SigG frame-shift (pKS09FS) compared to functional SigG overexpression (pKS09) indicating that their induction was dependent on functional SigG protein.

The two genes most highly induced by SigG were Rv0887c (16-fold induction) and Rv0911 (11-fold induction). These are both annotated as conserved hypothetical proteins of unknown function, however, analysis of their sequence structure identified that they both contain Glo_EDL_BRP_like superfamilies (Figure 5A). These domains are characteristic of glyoxalases, type I extradiol dioxygenases and bleomycin resistance proteins, which are all involved in inactivation of toxic compounds, many of which damage DNA.18 Expression of Rv0887c in response to DNA damage by mitomycin C was assessed in wild-type H37Rv and ΔsigG2, in which sigG is mutated, without the spontaneous cell wall mutation39 and ΔsigG2 containing the pLDL-8T complementing vector, containing the entire sigG operon (Figure 5B). Expression of Rv0887c was found to increase upon mitomycin C treatment in the wild-type H37Rv strain (induction ratio 5.7, p-value < 0.05). Expression in the ΔsigG2 mutant was 3.6-fold lower than wild-type without induction, and no increase in expression was seen after treatment with mitomycin C (induction ratio 1.5, p-value > 0.1). Expression in the complementing strain was similar to wild-type (induction ratio 3.9, p-value < 0.05). There was also a slight, but not significant, increase

Figure 4. SigG does not affect survival of M. tuberculosis during infection of macrophages. Survival and multiplication of M. tuberculosis H37Rv, mutants ΔsigG2 and ΔsigGW0 in naïve and activated BMDMs. Data represents the mean ± standard deviation of 3 biological replicates.
in expression of \( \text{Rv0911} \) with mitomycin C in wild-type \( H37Rv \), which was reduced in the mutant (data not shown). This indicates that expression and DNA damage induction of \( \text{Rv0887c} \) is dependent on \( \text{sigG} \).

The transcriptional start site of \( \text{Rv0887c} \) was mapped using 5’ RACE in wild-type \( H37Rv \) expressing \( \text{pKS09} \). This identified a potential start site 52 bp upstream of the annotated start codon (Figure 5C). This region showed significant homology to the promoter region of \( \text{Rv0911} \) and was used to identify a potential promoter motif for \( \text{sigG} \) as being CGATGA(N18)GTCNNTA.

### Table 1

<table>
<thead>
<tr>
<th>Rv number</th>
<th>Gene name</th>
<th>Microarray(^1)</th>
<th>p-value</th>
<th>qRT-PCR(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Rv0182} )</td>
<td>( \text{sigG} )</td>
<td>110.22</td>
<td>4.11 ( \times ) 10(^{-08} )</td>
<td>473.1</td>
</tr>
<tr>
<td>( \text{Rv0183} )</td>
<td></td>
<td>52.52</td>
<td>2.81 ( \times ) 10(^{-04} )</td>
<td>ND</td>
</tr>
<tr>
<td>( \text{Rv0887c} )</td>
<td></td>
<td>15.96</td>
<td>2.36 ( \times ) 10(^{-06} )</td>
<td>229.9</td>
</tr>
<tr>
<td>( \text{Rv0911} )</td>
<td></td>
<td>11.37</td>
<td>3.13 ( \times ) 10(^{-07} )</td>
<td>12.13</td>
</tr>
<tr>
<td>( \text{Rv0886} )</td>
<td>( \text{fprB} )</td>
<td>10.89</td>
<td>3.13 ( \times ) 10(^{-07} )</td>
<td>6.72</td>
</tr>
<tr>
<td>( \text{Rv0912} )</td>
<td></td>
<td>6.45</td>
<td>4.63 ( \times ) 10(^{-04} )</td>
<td>2.22</td>
</tr>
<tr>
<td>( \text{Rv1130} )</td>
<td>( \text{prpD} )</td>
<td>1.93</td>
<td>0.042</td>
<td>2.44</td>
</tr>
<tr>
<td>( \text{Rv1484} )</td>
<td>( \text{inhA} )</td>
<td>1.78</td>
<td>0.0073</td>
<td>1.96</td>
</tr>
<tr>
<td>( \text{Rv0942} )</td>
<td></td>
<td>4.93</td>
<td>0.0014</td>
<td>1.22</td>
</tr>
<tr>
<td>( \text{Rv0234c} )</td>
<td>( \text{gabD1} )</td>
<td>2.05</td>
<td>0.019</td>
<td>1.64</td>
</tr>
<tr>
<td>( \text{Rv0204c} )</td>
<td></td>
<td>1.98</td>
<td>0.027</td>
<td>0.46</td>
</tr>
<tr>
<td>( \text{Rv0209} )</td>
<td>( \text{vapB15} )</td>
<td>1.98</td>
<td>0.027</td>
<td>2.81</td>
</tr>
<tr>
<td>( \text{Rv2010} )</td>
<td>( \text{vapC15} )</td>
<td>1.93</td>
<td>0.022</td>
<td>2.32</td>
</tr>
</tbody>
</table>

\(^1\) Microarray analysis performed with three biological and two technical replicates; fold change compared to empty vector containing wild-type \( M. \text{tuberculosis} \) \( H37Rv \); genes upregulated >1.75 fold with Benjamin Hochberg corrected \( p \)-value \(< 0.05.\)

\(^1\) qRT-PCR from at least three biological replicates; relative expression normalised to that of \( \text{sigG} \); fold change for \( \text{SigG} \) over-expression (\( \text{pKS09} \), or \( \text{SigG} \) frameshift mutant (\( \text{pKS09FS} \)) compared to empty vector containing wild-type \( M. \text{tuberculosis} \); significant differences between expression in \( \text{pKS09} \) compared to \( \text{pKS09FS} \) containing strains determined by \( T \)-test (\( p < 0.1 \); \( p < 0.01 \); \( p < 0.001 \); \( p < 0.0001 \)).

### 4. Discussion

We examined the regulation of \( \text{sigG} \) at the transcriptional level, as well as determining the regulon for \( \text{SigG} \). The dominant promoter for \( \text{sigG} \) was found to coincide with the \( \text{RecA}_{\text{NDp}} \) promoter. This promoter is located downstream of the annotated start site for \( \text{sigG} \) indicating that this site must be incorrect. Due to the GC rich genome incorrect annotation of translational start codons is common in \( \text{Mycobacteria} \), which has implications for predictions of promoters and gene function.\(^40\) Translational start site mapping of \( \text{sigG} \) revealed that the start codon is in fact located 114 bp downstream of the annotated start codon, producing a protein 38 amino acids shorter. This region does not contain any of the functional domains for the sigma factor. Codon usage database analysis\(^30\) found that this region contained 7 (18.4%) codons which occur at a frequency less than 5 per 1000 for the entire \( M. \text{tuberculosis} \) protein coding genes, compared to 15 (4.5%) in the rest of the \( \text{sigG} \) gene. This may mean that this region is much less able to be translated. The alternative sigma factor, \( \text{SigE} \), was shown to start from three different codons depending on the promoter used,\(^36\) it is therefore possible that \( \text{sigG} \) is expressed from the annotated start codon under different growth conditions. The DNA damage inducible promoter for \( \text{SigG} \) is the furthest downstream, therefore if \( \text{SigG} \) was also able to start from two translational start codons you would expect the protein expressed under DNA damaging conditions to be smaller than that expressed under normal, exponential growth. In fact there was no evidence of any larger isoforms from over-expression of \( \text{SigG} \) in either \( M. \text{smegmatis} \) or \( M. \text{tuberculosis} \) (Figure 2). It was also noted that expression of soluble recombinant \( \text{SigG} \) in \( E. \text{coli} \) could only be obtained when this region was omitted from the expression construct (data not shown). This indicates that \( \text{SigG} \) likely starts from this site under all conditions.

This study identified three potential promoters upstream of the experimentally determined \( \text{SigG} \) translational start site, including the previously characterised DNA damage inducible \( \text{RecA}_{\text{NDp}} \), recognised by \( \text{ClpR} \).\(^{21,45} \) Expression of sigma factors from multiple promoters has been shown previously for \( \text{SigB} \), \( \text{SigC} \), \( \text{SigE} \) and \( \text{SigL} \).\(^{11,13,38} \) We were able to demonstrate the activity and DNA-damage inducible nature of promoter \( \text{P1} \); however, when included in the longer \( \text{lacz} \) construct, mutation of the \( \text{P1} \) –10 region did not significantly reduce expression levels. This is likely due to the presence of regulatory elements in this longer construct that were not present in the individual promoter constructs. We were unable to ascertain whether promoters 2 and 3 were genuine; both promoters showed no activity when expressed individually, and the longer construct, containing potential binding sites for other regulatory elements, showed only a slight reduction in activity when promoter 2 was mutated. It’s possible that the mutations used did not completely eliminate activity or an additional, uncharacterised promoter is responsible for the residual activity. Initial primer extension analysis indicated that promoter 3 was \( \text{SigG} \) dependent, however this is unlikely to be the case as over-expression of \( \text{SigG} \) had no effect on expression from its own promoter (as expression of the rest of the operon was unaffected); also promoter activity was unaffected in a \( \text{SigG} \) mutant strain (Figure 3D). Promoter 3 does, however, bear some similarity to the \( \text{SigM} \) consensus, which has been shown to be induced during temperature stress and stationary phase and regulate genes involved in oxidative stress and secretion.\(^3,4,35\)

\( \text{sigG} \) was found to be transcribed in a three gene operon with a long 3’ untranslated region overlapping the downstream gene \( \text{lprO} \). Whole genome sequencing has revealed that long 3’ UTR’s are common in \( M. \text{tuberculosis} \).\(^3 \) In most bacteria termination is tightly regulated by a hairpin loop structure followed by a poly-U-stretch, conversely \( \text{mycobacterial terminators} \) lack the poly-U-stretch which may result in less efficient termination and have implications for the regulation of downstream genes.\(^3,28\) Interestingly, when the entire \( \text{sigG} \) operon, up to 380 bp before the end of \( \text{Rv0180c} \), was deleted an increase in the expression of \( \text{lprO} \) was observed (Supplementary Figure 1), indicating that during normal growth either expression of the \( \text{sigG} \) operon inhibits \( \text{lprO} \) expression or a negative regulatory element for \( \text{lprO} \) is present within the deleted section of \( \text{Rv0180c} \).

\( \text{SigG} \) is regulated as part of the DNA damage response of \( M. \text{tuberculosis} \). However, mutation of \( \text{sigG} \) caused no increase in

\( A. \text{Gaudion et al. / Tuberculosis 93 (2013) 482–491} \)
sensitivity to DNA damaging agents. SigG was predicted to regulate the RecA-dependent SOS response, in particular the SOS response regulator lexA.

Subsequent investigations found no difference in the regulation either the SOS response, or RecA-independent DNA damage response genes in a sigG deletion mutant, and lexA was found to be controlled by a SigA consensus promoter. Our previous analysis of the transcriptional profile of a sigG mutant strain in fact showed no significant differences to the wild-type M. tuberculosis. We, therefore, looked at genes upregulated with a sigG over-expression strain. This gave five genes that were upregulated dependent on functional SigG, as their expression was no longer increased when a frameshift was introduced into the SigG coding region: fprB, Rv0887c, Rv0911, Rv0912, inhA and prpD. In addition, Rv0887c expression was confirmed to be regulated by SigG as its expression and DNA damage induction was reduced in a sigG mutant strain. No genes previously predicted to be part of the SOS response or the RecA-independent DNA damage response were found to be upregulated in the SigG over-expression strain, further confirming that SigG does not regulate either of these pathways.

The two genes most highly upregulated in the SigG over-expression strain were Rv0887c and Rv0911. These encode conserved hypothetical proteins of unknown function that both contained glyoxylase-like domains. Glyoxylase is involved in detoxifying reactive 2-oxoaldehydes, mainly methylglyoxal, a by-product of glycolysis thought to control the rate of carbon flux when moving between environments. Methylglyoxal can also cause DNA damage. Rv0911 was predicted to be involved in the methylglyoxal detoxification pathway of M. tuberculosis, as it was able to interact with pyrimidine-imidazole compounds, which
target the methyglyoxal detoxification pathway. Proline metabolism has also been predicted to play a role in methyglyoxal detoxification in M. smegmatis. Accordingly deletion of PruC, a transcriptional activator of proline metabolism, resulted in an increase in expression of two potential glyoxalases, including MSMEG_5680 the homologue of Rv0887c, as well as sigG and other genes belonging to the RecA-independent DNA damage induction pathway. Increased levels of methyglyoxal were found in M. tuberculosis infected macrophages and granulomas, suggesting they were part of the SigG regulon, have subsequently been detoxified and shown to not be SigG regulated:

The transcriptional start site of Rv0887c was mapped to a G residue, 52 base pairs upstream of the annotated start codon. Alignment of the upstream regions of Rv0887c and Rv0911 showed significant homology between the two and a putative SigG promoter sequence was identified as CGATAG(N18)GTCNTA. Previously a SigG consensus sequence has been hypothesised to be GCCGNT(N15–18)CGANCA however, at least two of the genes, whose decreased expression in a sigG deletion mutant was used to suggest they were part of the SigG regulon, have subsequently been shown to not be SigG regulated: lexA and Rv0654 (data not shown). It is therefore unlikely that this consensus is correct.

Rv0181c, which is cotranscribed with SigG, may also have a detoxifying role. BLAST analysis revealed that this protein had significant homology to the E. coli YhhW pirin protein (data not shown). YhhW has been shown to break down the antioxidant, quercetin, and is therefore a detoxifying enzyme known as a quercetinase. Purified Rv0181c protein has been shown to have some quercetinase activity thus supporting its potential role as a detoxifying enzyme (Supplementary Figure 2). Quercetin is a member of a group of compounds called flavonoids, which have been shown to inhibit fatty acid and mycolic acid biosynthesis in M. tuberculosis. In addition the strain ΔsigGWO showed increased susceptibility to the presence of Tween 80 compared to the wild-type strain. This phenotype could be complemented by all three genes of the sigG operon but not by sigG on its own (Supplementary Figure 3). This suggests that the susceptibility of the ΔsigGWO strain to Tween 80 is due to the absence of either Rv0181c and/or Rv0180c. Due to the fact that Tween 80 acts as a surfactant and Rv0180c is an integral membrane protein we hypothesise that the Tween 80 phenotype may be due to the absence of the Rv0180c gene.

We have shown that sigG is expressed as part of a three gene operon and its expression is controlled by at least two promoters. SigG is induced as part of the RecA independent DNA damage response but is not the sigma factor which controls this response. Instead SigG induces an alternative regulon, possibly involved in detoxification of methyglyoxal. SigG is co-transcribed with a putative quercetinase and a transmembrane protein. The absence of one or both of these two members of the sigG operon causes M. tuberculosis to become susceptible to the presence of Tween 80. Taken together, the potential functions of genes co-transcribed with sigG and members of the SigG regulon, outline a novel aspect of the DNA damage and stress response in M. tuberculosis, which, as well as inducing genes involved in DNA repair, also induces, via SigG, genes possibly involved in detoxifying factors that could be responsible for causing DNA damage and cell stress.

Ethical approval: Not required.

Funding: This research was supported by the UK Medical Research Council (programme number U1175 32056).

Competing interests: None declared.

Acknowledgements

We thank Dr Alan Williams for help with protein purification, and Joanna Houghton for making pH105.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.tube.2013.05.005.

References