The Mycobacterium tuberculosis katG promoter region contains a novel upstream activator

Michelle A. Mulder, Harold Zappe and Lafras M. Steyn

Author for correspondence: Lafras M. Steyn. Tel: +27 21 406 6363. Fax: +27 21 448 8153. e-mail: lsteyn@medmicro.uct.ac.za

Department of Medical Microbiology, University of Cape Town and Groote Schuur Hospital, Medical School, Observatory, 7925 Cape Town, South Africa

An Escherichia coli–mycobacterial shuttle vector, pJCluc, containing a luciferase reporter gene, was constructed and used to analyse the Mycobacterium tuberculosis katG promoter. A 1.9 kb region immediately upstream of katG promoted expression of the luciferase gene in E. coli and Mycobacterium smegmatis. A smaller promoter fragment (559 bp) promoted expression with equal efficiency, and was used in all further studies. Two transcription start sites were mapped by primer extension analysis to 47 and 56 bp upstream of the GTG initiation codon. Putative promoters associated with these show similarity to previously identified mycobacterial promoters. Deletions in the promoter fragment, introduced with BAL-31 nuclease and restriction endonucleases, revealed that a region between 559 and 448 bp upstream of the translation initiation codon, designated the upstream activator region (UAR), is essential for promoter activity in E. coli, and is required for optimal activity in M. smegmatis. The katG UAR was also able to increase expression from the Mycobacterium paratuberculosis P\textsubscript{An} promoter 15-fold in E. coli and 12-fold in M. smegmatis. An alternative promoter is active in deletion constructs in which either the UAR or the katG promoters identified here are absent. Expression from the katG promoter peaks during late exponential phase, and declines during stationary phase. The promoter is induced by ascorbic acid, and is repressed by oxygen limitation and growth at elevated temperatures. The promoter constructs exhibited similar activities in Mycobacterium bovis BCG as they did in M. smegmatis.

Keywords: Mycobacterium tuberculosis, katG, promoter, regulation

INTRODUCTION

The genus Mycobacterium is one of the largest bacterial genera and includes the pathogenic species Mycobacterium leprae and Mycobacterium tuberculosis. At present little is known about the control of gene expression in the mycobacteria. A number of mycobacterial promoters have been studied (reviewed by Mulder et al., 1997). Consensus promoter sequences have been proposed (Ramesh & Gopinathan, 1995; Bashyam et al., 1996); however, many of these promoters are specifically regulated in vivo and may not be typical. Some of the mycobacterial promoters resemble the typical Escherichia coli σ\textsuperscript{70} consensus promoter and function in this organism, but most have a higher G+C content and differ from the E. coli consensus. These function more efficiently in Streptomyces (Kieser et al., 1986). As yet, too few mycobacterial promoters have been studied to make accurate predictions on consensus sequences and promoter structures.

The isolation of plasmids which replicate in the mycobacteria, for example pAL5000 from Mycobacterium fortuitum (Labidi et al., 1985), has facilitated the construction of vectors for studying mycobacterial gene expression in a homologous host (Snapper et al., 1988; Aldovini & Young, 1991; Stover et al., 1991; Barletta et al., 1992). The surrogate host of choice is Mycobacterium smegmatis because of its non-pathogenicity and rapid growth. The M. smegmatis RNA polymerase has been purified and used for transcription initiation in vitro (Levin & Hatfull, 1993). The major form of this enzyme has marked conservation with that of E. coli (Predich et al., 1995), which suggests some common mechanisms of transcription initiation.

Abbreviations: ADC, albumin dextrose catalase; RLU, relative light units; tss, transcription start site; UAR, upstream activator region.
The *M. tuberculosis* katG gene encodes a dual-function catalase-peroxidase enzyme which protects the cell against excess hydrogen peroxide and, therefore, contributes to its survival in macrophages (Middlebrook & Cohn, 1953; Mitchison et al., 1960; Jackett et al., 1978; Wilson et al., 1995; Heym et al., 1997). Another important feature of the *M. tuberculosis* katG gene is its well-documented association with susceptibility to isoniazid (e.g. Winder, 1960; Zhang et al., 1992; Heym et al., 1993; Rouse & Morris, 1995). Examination of the mechanisms of regulation of this gene may, therefore, contribute to our knowledge of both the interaction of the organism with its host and the conditions which lead to resistance to isoniazid. We report here on the characterization of the *M. tuberculosis* katG promoter region and examination of the conditions under which katG is expressed in an *M. smegmatis* host.

**METHODS**

**Bacterial strains and plasmids.** All cloning steps were performed in *E. coli* LKIII (Zabeau & Stanley, 1982) grown in 2YT medium (Miller, 1992). Expression studies were performed in *M. smegmatis* LR222, a high-frequency transforming strain (Beggs et al., 1995), and *M. bovis* BCG Tokyo vaccine strain 172 (supplied by the State Vaccine Institute, Pinelands, South Africa). *M. tuberculosis* H37Rv was obtained from the American Type Culture Collection. All mycobacterial strains were cultivated in Middlebrook 7H9 medium (Difco) supplemented with catalase-free ABC (albumin and dextrose) and 0.05% (v/v) Tween 80, or on Middlebrook-ADC agar plates. pGEM-luc (Promega) and pJC86 (J.T. Crawford, Centers for Disease Control and Prevention, Atlanta, GA, USA) were used for construction of the plasmids.

**DNA manipulations.** Standard recombinant DNA methods were performed by previously described protocols (Ausubel et al., 1987; Sambrook et al., 1989). DNA fragments were sequenced by the dideoxy sequencing protocol (Sanger et al., 1977) using a T7 Polymerase Sequencing kit (Pharmacia), pUC19 forward and reverse sequencing primers (Promega), and a synthetic oligonucleotide complementary to the 5′ end of the luciferase gene, 5′-CTTTATGTTTGGCCTGTC-3′. 

**Isolation of genomic DNA and PCR.** Genomic DNA was purified from *M. tuberculosis* H37Rv as described by Jacobs et al. (1991). PCR amplification of the katG upstream sequences was performed using *M. tuberculosis* H37Rv genomic DNA (0.1 μg) as the template with the primers 5′-CTG-GTAAAGCrtGCGGGCAAAACAGC-3′ and 5′-CACAG-ggaTCCCTTCCAGGATTTGGT-3′ (based on the published sequence, GenBank accession no. X68081). The lower-case nucleotides indicate mismatches and the underlined nucleotides represent the restriction sites for HinIII and BamHI, respectively. The amplifications were performed using Tag polymerase according to the manufacturers’ specifications, except that DMSO was added to a final concentration of 10% (v/v).

**Plasmid constructions.** The plasmids constructed in this study are listed in Table 1. Plasmid pJC1uc was constructed by cloning the 1748 bp HinIII-Stul fragment from pGEM-luc (GenBank/EMBL accession no. X65316), containing the promoterless firefly luciferase gene, into the KpnI and HinIII sites of pJC86 (Fig. 1). A 1943 bp PCR product containing the region upstream of the *M. tuberculosis* katG gene was digested with BamHI and HindIII and cloned into the corresponding restriction sites of pJC1uc to yield plasmid pK10. A 559 bp SmaI–BamHI fragment of the PCR product, immediately upstream of the katG gene, was cloned upstream of the luciferase gene in pJC1uc to produce pK20. Various deletion derivatives of pK20 were constructed, making use of the restriction sites present in the insert. These are summarized in Table 1 and Fig. 2. The construct pANIIIW (a gift from Karen Kempsall, Glaxo-Wellcome), consisting of a 170 bp PCR fragment containing the P<sub>AN</sub> promoter cloned into the T-tailed EcoRV restriction site of the vector pT7Blue (Novagen), was used as a source of the *M. paratuberculosis* promoter (Murray et al., 1992). The 218 bp HindIII–BamHI fragment from pANIIIW was cloned into the corresponding restriction sites of pJC1uc to form pLPan. In this construct, the P<sub>AN</sub> promoter is in the correct orientation to promote expression of the luciferase gene. The 262 bp HindIII–SphI fragment of pK20 containing the katG upstream activator region (UAR) was cloned upstream of the P<sub>AN</sub> promoter in pLPan to form pLPur. The structures of all the above constructs were confirmed by DNA sequencing, using the pUC19 forward sequencing primer (Promega) and the luciferase primer described above.

**Nuclease BAL-31 deletions.** pK20 DNA (30 μg) was digested to completion with HindIII, extracted with phenol/chloroform/isomyl alcohol and precipitated with ethanol. The linear fragment was digested with nuclease BAL-31 according to a standard protocol (Ausubel et al., 1987). The digested DNA was then precipitated with ethanol, religated and used to transform competent *E. coli* LKIII cells.

**Electroporation.** *M. smegmatis* cells were prepared for electroporation as described previously (Jacobs et al., 1991). For electroporation, 60 μl cells were combined with 140 μl 10% (v/v) glycerol and 1 μg DNA (5 μl), and placed on ice for 5 min. Electroporations were performed using 0.1 cm gap Gene Pulser cuvettes (Bio-Rad) at 1 kV, 25 μF and 1000 Ω, using a Gene Pulser apparatus (Bio-Rad). After electroporation, 900 μl Middlebrook-ADC was added, and the cells were incubated at 37 °C for 3–4 h before plating. Plates were incubated at 37 °C for 4–5 d. Electrocompetent *M. bovis* BCG cells were prepared and electroporated using the method of Jacobs et al. (1991), with minor modifications. A stationary-phase culture of *M. bovis* BCG grown in Middlebrook-ADC-Tween was diluted 100-fold in freshly prepared Middlebrook-ADC-Tween, and grown to mid-exponential phase (8 d) in a roller culture at 37 °C. Cells were harvested by centrifugation at 4 °C and washed five times in sterile deionized water to remove all salts and residual medium. Washed cells were resuspended in a final volume of 1 ml ice-cold 10% (v/v) glycerol per 100 ml cells (original volume). Electroporations were performed as described for *M. smegmatis*. The cells were kept on ice for 5 min post-electroporation. Middlebrook-ADC (700 μl) was added and the cells were incubated for 3–4 h at 37 °C. The cells were plated on Middlebrook-ADC agar plates containing 30 μg kanamycin ml<sup>−1</sup> and 0.01% (w/v) cycloheximide (Fluka Biochemicals), and incubated at 37 °C for 3–4 weeks.

**Preparation of cell extracts and luciferase assays.** For the luciferase activity assays, 5 ml cultures were grown to late exponential phase at 37 °C with aeration (Orbital shaker, Hugger Designs, at 150 r.p.m.). *E. coli* cell extracts were prepared by harvesting cells from 1 ml of culture and resuspension of the cells in 100 μl ice-cold 1 x Cell Culture Lysis Reagent (Luciferase Assay System, Promega). The cells were lysed by sonication in ice water (three bursts of 40 s each)
Table 1. Plasmids used in this study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Size (bp)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJC86</td>
<td>7427</td>
<td>Promoterless firefly luciferase gene in pJC86</td>
</tr>
<tr>
<td>pK10</td>
<td>9370</td>
<td>1943 kb katG upstream region in pJC86 (HindIII–BamHI)</td>
</tr>
<tr>
<td>pK20</td>
<td>7956</td>
<td>559 bp katG upstream region in pJC86 (SmaI–BamHI)</td>
</tr>
<tr>
<td>pK20AEN</td>
<td>7943</td>
<td>pK20 with the 13 bp EcoRV–NrdI fragment deleted</td>
</tr>
<tr>
<td>pK20AH</td>
<td>7845</td>
<td>pK20 with the 111 bp HindIII–AluI fragment deleted</td>
</tr>
<tr>
<td>pK20AS</td>
<td>7801</td>
<td>pK20 with the 155 bp AluI–SphI fragment deleted</td>
</tr>
<tr>
<td>pK20AS2</td>
<td>7850</td>
<td>pK20 with the 106 bp SalI–EcoRV fragment deleted</td>
</tr>
<tr>
<td>pK21</td>
<td>7656</td>
<td>pK20 with 300 bp vector upstream of insert resected*</td>
</tr>
<tr>
<td>pK22</td>
<td>7588</td>
<td>pK20 with 18 bp insert resected*</td>
</tr>
<tr>
<td>pK23</td>
<td>7733</td>
<td>pK20 with 123 bp insert resected*</td>
</tr>
<tr>
<td>pK24</td>
<td>7521</td>
<td>pK20 with 167 bp insert resected*</td>
</tr>
<tr>
<td>pK25</td>
<td>7402</td>
<td>pK20 with 283 bp insert resected*</td>
</tr>
<tr>
<td>pK20ASB</td>
<td>7689</td>
<td>262 bp HindIII–SphI katG UAR in pJC86</td>
</tr>
<tr>
<td>pK20AH</td>
<td>7727</td>
<td>300 bp Sp6l–BamHI katG promoter fragment in pJC86</td>
</tr>
<tr>
<td>pLPan</td>
<td>7646</td>
<td>218 bp P\textsubscript{AN} promoter in pJC86</td>
</tr>
<tr>
<td>pLPuar</td>
<td>7913</td>
<td>218 bp P\textsubscript{AN} promoter downstream of the katG UAR in pJC86</td>
</tr>
</tbody>
</table>

*BAL-31 nuclease derivatives of pK20.

Fig. 1. Schematic representation of the construction of the luciferase reporter plasmid pJC86.
and cell debris were removed by centrifugation. The supernatants were assayed for luciferase activity as described above. The protein concentrations of the cell extracts were measured using the D$_c$ Protein Assay kit (Bio-Rad), and the luciferase activities are quoted as relative light units (RLU) per mg protein used in each assay.

**RNA extractions and primer extension analysis.** RNA was extracted from *E. coli* and *M. smegmatis* using the FastPrep system [FastPrep FP120 instrument (Savant Instruments) and FastRNA Kit-Blue (Bio101)], according to the manufacturers’ instructions. Potential transcription start sites of the gene were identified by primer extension analysis (Sambrook et al., 1989). A complementary 18-mer primer (5'-GCAGTTGCTCTCCAGCGG-3') was designed to anneal 58 bp downstream of the ATG initiation codon of the luciferase gene. Approximately 100 ng primer was end-labelled with $[^32]P$ dATP using polynucleotide kinase, in a reaction with the supplied buffer. After incubation at 30 °C for 30 min, the unincorporated nucleotides were removed using a Promega G25 spin column. RNA samples (100 µg) were precipitated with sodium acetate and ethanol, resuspended in 30 µl RNA hybridization buffer (Ambion), and added to 25 ng labelled primer. The nucleic acids were denatured at 85 °C for 10 min, and annealed overnight at 50 °C. The annealed RNA was precipitated with sodium acetate and ethanol, and resuspended in diethyl pyrocarbonate (DEPC)-treated water. The primer was extended on the RNA template at 37 °C. The products were denatured at 90 °C for 2 min and separated on a 6 % (w/v) polyacrylamide sequencing gel adjacent to dideoxy sequencing reactions primed with the same 18-mer primer. Gels were dried and the bands visualized by autoradiography.

**ATP assays.** ATP concentration was measured as an indication of cell mass using the Promega Enliten luciferase/luciferin bioluminescence detection reagent in a Bio-Orbit 1253 luminometer. The ATP levels were monitored by measuring the production of light when ATP, luciferin and oxygen were combined in the presence of luciferase. The assay system was calibrated using an ATP standard of known concentration (Prioli et al., 1985; Stanley et al., 1989). The *M. smegmatis* cells were treated with 1 % (w/v) (final concentration) trichloroacetic acid (TCA), and incubated on ice for 30 min. The precipitation was terminated by the addition of 50 µl neutralization buffer (1 M Tris-acetate, pH 7.75) and 622 µl distilled water. The ATP assays were performed according to the manufacturer’s instructions. At higher cell densities, the TCA-treated samples were diluted for the assay and the calculation adjusted accordingly.

**Conditions of expression of katG.** Stationary-phase cultures of *M. smegmatis/pK20* were diluted 100-fold into 500 ml Middlebrook-ADC-Tween, and incubated at 37 °C with shaking. Growth was monitored using OD$_{600}$ readings to measure cell density and ATP concentration to measure cell mass. Samples were removed at various times and assayed for luciferase activity. To test the effect of various stresses on promoter activity, an *M. smegmatis/pK20* culture (800 ml) was grown to mid-exponential phase (24 h), and divided into eight 100 ml cultures. The cells were harvested by centrifugation and resuspended in Middlebrook-ADC-Tween with the following additions per individual culture: (1) 10 mM hydrogen peroxide; (2) 10 mM ascorbic acid; (3) 60 mM
RESULTS

Construction of pJCluc

The E. coli–mycobacterial shuttle vector pJCluc was constructed as shown in Fig. 1. The plasmid pJC86 contains the EcoRV–HpaI pAL5000 backbone for replication in mycobacteria (Snapper et al., 1990), the lacZ gene and oriE from pUC18, and the kanamycin-resistance gene from Tn5. pJCluc contains the promoterless luc gene in the opposite orientation to the lacZ gene, and has no Shine–Dalgarno sequence. pJCluc and pJC86 were assayed for luciferase activity in E. coli and M. smegmatis. Cells containing pJC86 had no detectable luciferase activity, indicating no endogenous luminescence in either E. coli or M. smegmatis. E. coli/pJCluc exhibited a low level of activity [45 ± 18 RLU (mg protein)⁻¹] but a higher level of activity was detected in M. smegmatis/pJCluc [143 ± 28 RLU (mg protein)⁻¹]. This activity may be due to read-through from an upstream promoter which appears to be more active in M. smegmatis and, therefore, may be located in the pAL5000 fragment of the plasmid.

Amplification and expression analysis of the M. tuberculosis katG promoter

Plasmids pK10 and pK20, containing different-sized fragments of the katG upstream region (Table 1), were tested for luciferase activity in M. smegmatis and E. coli. For M. smegmatis/pK10 and M. smegmatis/pK20, the luciferase activities were 41800 ± 2400 and 38500 ± 10700 RLU (mg protein)⁻¹, respectively. Those for E. coli/pK10 and E. coli/pK20 were 381 ± 100 and 474 ± 13 RLU (mg protein)⁻¹, respectively. These results indicate that the katG promoter lies within the amplified fragment and is recognized by the E. coli and M. smegmatis transcription apparatus. They also suggest that the promoter region lies within the smaller Smal–BamHI fragment, since the luciferase activities of cells harbouring pK10 and pK20 do not differ significantly for either E. coli or M. smegmatis. Thus it appears that no sequences required for optimal expression are located in the 1374 bp region upstream of the Smal restriction site. The smaller construct, pK20, was therefore used for all further studies.

Mapping of the transcription start site by primer extension analysis

Primer extension analysis was performed using an antisense strand primer complementary to sequences near the 5’ end of the luciferase gene, and total RNA isolated from late-exponential-phase M. smegmatis/pK20 and E. coli/pK20 cells. In both E. coli and M. smegmatis, transcription was initiated at two sites located 47 bp (transcription start site A, tssA) and 56 bp (tssB) upstream of the translation initiation codon (Fig. 3). The possibility that the bands are due to a
Deletion analysis of the katG promoter region

A number of deletions were made, using the restriction sites present on the insert of pK20, in order to determine whether other regions of the M. tuberculosis katG upstream region are required for promoter activity. The constructs were tested for luciferase activity in E. coli and M. smegmatis. The regions present in the various deletions and the resulting decreases in luciferase activity in M. smegmatis are summarized in Fig. 2. Removal of 111 bp from the 5’ end of the insert (pK20AHA) resulted in an 82- and 55-fold reduction in activity in E. coli and M. smegmatis, respectively, suggesting that this region is essential for optimal promoter activity. Removal of the 13 bp between the EcoRV and NruI restriction sites (pK20AEN) in the promoter region resulted in a threefold decrease in promoter activity in E. coli (results not shown). This deletion removes the putative −10 hexamer associated with tssA, as well as tssB and a region of its putative −10 hexamer. The deletion should, therefore, abolish promoter activity completely. In M. smegmatis, deletion of this region had little effect (1.3-fold reduction) on expression. Removal of sequences upstream of the EcoRV site (pK20ASE) gave a similar result, i.e., a threefold decrease in activity in E. coli, and no significant change in M. smegmatis. Deletion of the Sall–EcoRV fragment results in the removal of half of the −10 hexamer of tssB and the −35 sequences of both tssA and tssB. Transcription initiates at a different promoter in both pK20AEN and pK20ASE (see below). The absence of the 155 bp Alul–SphI fragment had no significant effect on promoter activity in either E. coli or M. smegmatis. Thus, based on these deletion constructs, the sequences necessary for maximal promoter activity lie upstream of the Alul site. This region may act as a regulatory region which enhances transcription initiation and it was designated an upstream activator region (UAR). The results also indicate that the distance between the UAR and the promoters is not critical.

Deletion analysis of the katG promoter fragment using nuclease BAL-31

In a further attempt to locate the sequences required for promoter activity, the pK20 insert was progressively deleted from the HindIII restriction site at the 5’ end, using nuclease BAL-31. Five recombinants were selected and sequenced to determine the size of the deletions. Since the activity of BAL-31 is bidirectional, the construct pK20AHS, containing the 300 bp SphI–BamHI katG promoter region in pJCluc, was used to determine whether the deletion of vector sequences was responsible for the observed differences in luciferase activity. The plasmid pK20ASB was constructed to determine whether the UAR on its own was able to promote expression of the luciferase gene. E. coli and M. smegmatis cells harbouring the deletion plasmids, as well as the undeleted and vector controls, pK20 and pJCluc, respectively, were tested for luciferase activity (Fig. 2). Removal of the first 18 bp (pK22) of the insert resulted in an approximately 40% reduction in promoter activity, while the removal of 123 bp from the 5’ end (pK23) resulted in a sixfold reduction in activity in M. smegmatis. Removal of a further 160 bp (pK25) resulted in an 11-fold reduction in activity in this host. Similar results were obtained for E. coli. These results show that the region up to 123 bp from the 5’ end of the insert is

Fig. 4. Nucleotide sequence of the M. tuberculosis katG promoter region including the first three codons, labelled with respect to the published sequence (GenBank accession no. X68081). The transcription start sites (tss) are indicated with arrows. The putative −10 and −35 hexamers associated with each transcription start site are underlined, and the putative ribosome-binding site (RBS) is shown in bold. The lower-case nucleotides represent the restriction endonuclease sites EcoRV and NruI. (a) Promoters associated with tssA and tssB. (b) Promoters associated with tssC, which becomes functional in deletion derivatives pK20∆HS, pK20AEN and pK20ASE. (c) Summary of the potential katG promoter sequences.

Premature stop site of the reverse transcription reaction can be excluded, as sequencing of this region revealed no signs of band compressions, suggesting no tendency for secondary structure formation. The putative −10 and −35 sequences relative to these transcription start sites are represented in Fig. 4. The putative −10 hexamer for tssA, GACACT, lies 8 bp upstream of the transcription start site and shows a high degree of similarity to other mycobacterial −10 sequences. The first G of the hexamer is the least conserved. Associated with this −10 hexamer are three possible −35 sequences: TCTATG 19 bp upstream, TGT CCT 15 bp upstream, and TCTCTGA 13 bp upstream. The putative −10 hexamer for tssB, GATATC, lies 6 bp upstream of the transcription start site. Again, the first G of the hexamer is the least conserved. Associated with this −10 region are two possible −35 sequences, TCTACT 22 bp upstream, and TACTTG 20 bp upstream.
required for maximal promoter activity in both hosts. The region between 123 and 283 bp from the 5' end is not additionally important. Deletion plasmid pK21 has no insert sequences deleted, but has 300 bp of vector sequence removed. The construct exhibited similar luciferase activities to the parent plasmid, pK20, suggesting that deletion of at least the first 300 bp of vector sequence upstream of the katG promoter fragment has no effect on promoter activity. In addition the construct pK20∆HS, in which no vector sequences are deleted but 262 bp of insert are removed, exhibited a similar decrease in luciferase activity to pK23, pK24 and pK25. These results indicate that the decreases in luciferase activity are due to removal of essential insert fragments, rather than vector sequences. The background luciferase activity of cells harbouring pK20∆SB suggests that the 262 bp HindIII–SphI UAR contains no promoter sequences capable of promoting expression of the katG gene. The MIC of kanamycin for M. smegmatis and E. coli cells harbouring a selection of the restriction-endonuclease- and BAL-31-generated deletion constructs was determined as a measure of the relative copy number of the relevant plasmids (Stolt & Stoker, 1996) (results not shown). There were no significant differences in the MICs of the cells harbouring different deletion constructs in either of the two species. Thus it is unlikely that the effects of the deletions on the luciferase activities observed were due to differences in the plasmid copy number.

**Primer extension analysis of selected deletions**

Primer extension analysis was performed on total RNA extracted from M. smegmatis/pK20∆HS, M. smegmatis/pK20ΔEN and M. smegmatis/pK20ΔSE, in order to identify alternative promoter sequences that allowed expression of the luciferase gene. In all three cases, transcription initiated at a C residue located 23 bp upstream of the translation initiation codon (Fig. 5). Therefore, in the absence of either the UAR or the usual promoter sequences, a third region, designated P_c, is recognized as a promoter. Located 7 bp upstream of transcription start site C is the putative −10 hexamer, CACAGC (Fig. 4b). Associated with this are two putative −35 regions: TTCGCG, 14 bp upstream, and TCCGAC, 22 bp upstream.

**Enhancement of the pAN promoter by the katG UAR**

The functionality of the katG UAR with other mycobacterial promoters was tested by cloning of the 262 bp HindIII–SphI fragment from pK20 upstream of the M. paratuberculosis pAN promoter in pJCluc (pLPuar). The P_an promoter was also cloned into pJCluc on its own (pLPan) as a control. The constructs were electroporated into M. smegmatis and cell extracts from the resulting transformants were tested for luciferase activity. The presence of the katG UAR resulted in a 15-fold increase in expression of the P_an promoter in E. coli [278±72 to 4100±1400 RLU (mg protein)^−1] and a 12-fold increase in M. smegmatis (9900±1900 to 115600±24200 RLU (mg protein)^−1). This result confirms that the region acts as an enhancer of transcription initiation.

**Transformation of M. bovis BCG with the pJCluc constructs and luciferase activity assays**

The following constructs were electroporated into M. bovis BCG and tested for luciferase activity: pJCluc, pK20, pK20ΔHS, pK20ΔSB, pK24, pLPan and pLPuar (Table 2, Fig. 2). The background luciferase activity resulting from transcriptional readthrough was low in M. bovis BCG/pJCluc. The activity of the katG promoter in this host was, however, comparable to that in M. smegmatis. Removal of the katG UAR (pK20ΔHS) resulted in a 78-fold decrease in luciferase activity, indicating that the region also plays an important role in promoter activity in this host. The nuclease BAL-31 deletion plasmid, pK24, exhibited an eightfold decrease in luciferase activity relative to pK20. This plasmid.

---

**Fig. 5.** Primer extension analysis of the M. tuberculosis katG upstream region using a primer complementary to sequences downstream of the luciferase ATG and RNA isolated during late exponential phase from M. smegmatis cells harbouring (a) pK20∆HS, (b) pK20ΔEN, and (c) pK20ΔSE. The potential transcription start site, tssC, is indicated with an arrow. This corresponds to a C residue, 23 bp upstream of the GTG translation initiation codon. Lanes G, A, T, C show the sequencing reaction. Px, primer extension products.
Table 2. Luciferase activities of the pJCluc constructs in M. bovis BCG

<table>
<thead>
<tr>
<th>Construct</th>
<th>RLU (mg protein)</th>
<th>Relative activity†</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJCluc</td>
<td>12 ± 3</td>
<td>0.001</td>
</tr>
<tr>
<td>pK20</td>
<td>25 600 ± 2400</td>
<td>1.000</td>
</tr>
<tr>
<td>pK20AH5</td>
<td>300 ± 53</td>
<td>0.012</td>
</tr>
<tr>
<td>pK20ASB</td>
<td>46 ± 8</td>
<td>0.002</td>
</tr>
<tr>
<td>pK24</td>
<td>3 300 ± 1800</td>
<td>0.129</td>
</tr>
<tr>
<td>pLPan</td>
<td>4 800 ± 1300</td>
<td>0.188</td>
</tr>
<tr>
<td>pLPuar</td>
<td>21 420 ± 600</td>
<td>0.837</td>
</tr>
</tbody>
</table>

*Mean of three independent experiments †Activity normalized to that of the construct pK20.

contains an extra 93 bp of insert sequence relative to pK20AH5, which explains the smaller reduction in luciferase activity. The low background levels of luciferase activity resulting from pK20ASB indicates, as shown before, that the HindIII–SphI UAR of katG contains no active promoter sequences. The activity of the M. paratuberculosis P_An promoter increased fourfold in the presence of the katG UAR. This further confirms the importance of the region in enhancing transcription in this host and indicates that similar sequences are required for expression of the promoter in a slow-growing host.

Growth-phase-dependent expression from the katG promoter in M. smegmatis

Expression of the katG promoter in M. smegmatis was monitored over a 68 h period of growth by measurement of luciferase activity. The values used to plot the curves in Fig. 6 are taken from one experiment, but are representative of three independent experiments. The growth curves generated by optical density readings and ATP concentration showed similar trends (Fig. 6). The cells reach late exponential phase at approximately 30 h. Expression of luciferase from the katG promoter is low in early exponential phase, but increases steadily and reaches a transient peak in the late exponential phase, followed by a rapid decline as the cells enter stationary phase.

Expression from the katG promoter in cells exposed to various stresses

Some of the factors that are known to modulate katG expression in other bacteria were tested for their ability to influence expression from the M. tuberculosis katG promoter. Preliminary experiments revealed that a 4 h period was suitable for measurable response to stress in M. smegmatis. The experiments were performed in triplicate, and the luciferase activities of one representative experiment are given in Table 3. Treatment with 10 mM ascorbic acid resulted in an almost twofold (86%) increase in luciferase activity. Growth under

Table 3. Luciferase activities of M. smegmatis/pK20 exposed to various stresses

<table>
<thead>
<tr>
<th>Stress</th>
<th>Luciferase activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RLU (µg protein)−1</td>
</tr>
<tr>
<td>Control</td>
<td>88.60</td>
</tr>
<tr>
<td>Hydrogen peroxide</td>
<td>87.82</td>
</tr>
<tr>
<td>Sodium acetate</td>
<td>80.02</td>
</tr>
<tr>
<td>Ascorbic acid</td>
<td>164.48</td>
</tr>
<tr>
<td>Oxygen limitation</td>
<td>8.25</td>
</tr>
<tr>
<td>Low iron</td>
<td>78.26</td>
</tr>
<tr>
<td>pH 4</td>
<td>79.52</td>
</tr>
<tr>
<td>45 °C</td>
<td>2.12</td>
</tr>
</tbody>
</table>

*Activity normalized to that of the control.
oxygen-limiting conditions and at 45 °C resulted in an 11- and 42-fold decrease in luciferase activity, respectively. The other changes in the culture conditions did not significantly affect luciferase activity.

**DISCUSSION**

In this study, the sequences required for expression from the *M. tuberculosis* katG promoter have been identified. The 1943 bp and 359 bp *M. tuberculosis* katG promoter fragments promoted expression of the luciferase gene with equal efficiency under the conditions used in this study, showing that sequences upstream of the Smal site are not essential for optimal promoter efficiency during normal growth. Deletion analysis revealed that the first 111 bp at the 5’ end of the Smal–BamHI *M. tuberculosis* katG promoter fragment are required for optimal expression in both *E. coli* and *M. smegmatis*. The absence of the 155 bp Alul–SphI fragment, on the other hand, had no significant effect on promoter activity in either *E. coli* or *M. smegmatis*, indicating that this region is not important for expression. The sequences necessary for maximal promoter activity, therefore, lie upstream of the Alul site. This region does not contain a functional promoter but may act as a regulatory region which enhances transcription initiation. The distance between the UAR and the promoter sequences does not appear to be critical. Within the UAR lies a 24 bp sequence (−523 to −499 with respect to the GTG translation initiation codon), with an A + T content of 66.67 mol%. This A + T content is high in comparison to normal *M. tuberculosis* promoter regions (Bashyam et al., 1996), and may be analogous to the UP elements found upstream of some *E. coli* promoters. These elements are usually AT-rich and can increase promoter activity up to 30-fold (Ross et al., 1993). They are also a frequent feature of Gram-positive promoters (Moran et al., 1982; Graves & Rabinowitz, 1986). Movahedzadeh et al. (1997) have recently identified a similar region, 269–310 bp upstream of the *M. tuberculosis* recA gene, which is essential for expression from the promoter. A database search with the *M. tuberculosis* katG UAR sequence revealed that it is not dispersed in the *M. tuberculosis* genome, but that it is similar to the region upstream of the *M. fortuitum* katG gene. The 24 bp AT-rich sequence within the katG UAR, described above, is 79.2% homologous to a region located 489 bp upstream of the *M. fortuitum* katG ATG translation initiation codon. This suggests possible similarities in the mode of regulation of the gene in the two mycobacterial species.

Two transcription start sites were identified for the *M. tuberculosis* katG promoter in both *E. coli/pK20* and *M. smegmatis/pK20*. This suggests that the gene is transcribed from two promoters and that the same promoter sequences are recognized by both hosts. The presence of two promoters upstream of mycobacterial genes is not uncommon and has been reported previously (Stover et al., 1991; Suzuki et al., 1991; Dhandayuthapani et al., 1997; Movahedzadeh et al., 1997). It is possible that the promoters are recognized by different RNA polymerase holoenzymes, and that they are utilized to different extents during growth. Removal of the sequence between the ribosome-binding site and tssB of the katG promoter results in a 45–50% reduction in luciferase activity (unpublished data, this laboratory). This suggests that the promoters contribute equally to katG expression under the conditions of the above experiment. In the absence of either the UAR, or the putative katG promoters, P_A and P_B, *M. smegmatis* utilizes a third promoter, P_C. It is possible that a similar situation occurs in *E. coli*. This promoter functions as efficiently as the other two in *M. smegmatis*; however, the presence of the UAR is also essential for optimal activity. If promoter P_C is used to promote expression of the luciferase gene in *E. coli/pK20AEN* and *E. coli/pK20ΔSE*, then it is evidently not as efficient as promoters P_A and P_B in this host, since these strains exhibit a threefold lower luciferase activity than *E. coli/pK20*.

The putative −10 and −35 sequences identified for the katG promoters show significant homology to many mycobacterial promoters characterized thus far, and may, therefore, be ‘typical mycobacterial promoters’ (Mulder et al., 1997). These sequences also all match the consensus *E. coli* sequences in at least one of the three highly conserved positions. The promoters are therefore probably recognized and bound by Eσ in *E. coli*. The spacing between the −10 and −35 regions, however, differs from that found in most *E. coli* promoters (17 ± 1) (Harley & Reynolds, 1987). This, together with the divergence from the consensus sequences, may be responsible for the poor efficiency of the promoters in *E. coli*. It has been noted that the distance between the −10 and −35 region is not critical in mycobacterial promoters (Kremer et al., 1995), and varies between 13 and 24 bp (Ramesh & Gopinathan, 1995). All three of the katG −10 hexamers are least conserved in the first position, where the T is replaced by a G or C. It has previously been reported that less conserved bases in *M. tuberculosis* promoters tend towards G and C substitutions (Bashyam et al., 1996).

Primer extension analysis using total RNA isolated from *M. tuberculosis* was unsuccessful, possibly due to low levels of the katG mRNA. The experiment was therefore performed using RNA isolated from *M. smegmatis* cells harbouring various katG promoter clones. The determination of transcription start sites for *M. tuberculosis* genes, using this heterologous host, has been shown to correlate well with results obtained from native mRNA transcripts (Levin & Hatfull, 1993; Dhandayuthapani et al., 1997; Movahedzadeh et al., 1997), and has successfully been used as a substitute in many cases (Murray et al., 1992; Kremer et al., 1995; Nesbit et al., 1995).

As mentioned above, the first nucleotide of all three putative katG −10 hexamers deviates from the consensus *E. coli* sequence. *E. coli* promoters which deviate in this position generally require activators for efficient transcription initiation. These activators normally bind
to sites at various distances upstream of the —35 region (Raibaud & Schwartz, 1984; Zhou et al., 1994a, b).

J. Song & V. Deretic (GenBank accession no. AF002194) and Pagan-Ramos et al. (1998) have identified an open reading frame with homology to the Fur protein upstream of the M. tuberculosis and M. marinum katG genes, respectively. The FurA translation start codon lies immediately downstream of the 24 bp AT-rich sequence described above, while the stop codon lies immediately downstream of the katG tssA. This overlap of the FurA ORF with the katG promoter region and UAR may have implications for the regulation of both genes and suggests a possible coupling of the regulation of oxidative stress and iron metabolism genes in this organism.

The growth-phase-dependent expression of the M. tuberculosis katG promoter in M. smegmatis is similar to that in other bacteria. Many bacterial catalases are produced at low levels during exponential growth and are induced either during late exponential phase or during the transition to stationary phase (Mukhopadhyay & Schellhorn, 1994; Rocha & Smith, 1995; Schnell & Steinman, 1995). In E. coli, the induction of katG as the cells enter stationary phase is dependent on the stationary-phase or starvation response σ factor, KatF or RpoS (Loewen et al., 1985; Mulvey et al., 1988; Mukhopadhyay & Schellhorn, 1994). The effects of host cell stress on expression from the katG promoter were also investigated in this study. Mycobacterial cells were stressed during late exponential phase, at the time corresponding to maximum expression. The repression of the katG promoter under anaerobic conditions has been noted for other catalases (Morgan et al., 1986), and is probably due to a reduced requirement for the enzyme. The factor(s) responsible for this repression, and the repression of the promoter at elevated temperatures, have not been identified. In E. coli, treatment with 5–7 mM ascorbic acid results in an eightfold increase in catalase activity (Richer & Loewen, 1981), while treatment with 0–80 mM sodium acetate (pH 7·0) induces catalase activity sevenfold (Mukhopadhyay & Schellhorn, 1994). The treatment of M. smegmatis/pK20 with 10 mM ascorbic acid, however, resulted in less than twofold induction of the M. tuberculosis katG promoter, while no induction was observed in response to sodium acetate. These results suggest that there are differences in the regulation of the gene in the two organisms. The lack of induction of the M. tuberculosis katG promoter in the presence of hydrogen peroxide agrees with the results of Deretic et al. (1995), but not those of Sherman et al. (1996). The latter authors detected a sevenfold increase in expression from the M. tuberculosis katG promoter in an M. bovis BCG host in response to hydrogen peroxide. The differences in the response of the katG promoter to hydrogen peroxide may have been due to the use of different mycobacterial hosts. M. tuberculosis does not have a functional OxyR, which is responsible for the hydrogen-peroxide-dependent induction of the katG gene in most bacteria (Christman et al., 1985; VanBogelen et al., 1987; Tartaglia et al., 1989; Storz et al., 1990; Farr & Kogoma, 1991; Altuvia et al., 1994; Toledano et al., 1994). In addition, no homologous oxyR sequences have been detected in M. smegmatis. These observations suggest that the katG gene is regulated differently in these organisms, and that induction of the gene in response to hydrogen peroxide is unlikely, unless it occurs through an alternative mechanism.

IdeR-deficient mutants of M. smegmatis exhibit reduced catalase and peroxidase activities, associated with KatG (Dussurget et al., 1996). This suggests that the protein is either directly or indirectly responsible for the regulation of the katG gene in response to iron limitation in this organism. In this study, no changes in the expression from the katG promoter were noted under iron-limiting conditions. It is possible that the M. tuberculosis katG promoter is not recognized by the M. smegmatis IdeR or the factor(s) induced by the protein, or that regulation in response to iron limitation occurs at a post-transcriptional level in this organism. Alternatively, the M. smegmatis IdeR may not be induced or active under the conditions used in this study.

That the pJCluc promoter constructs in this study exhibited similar luciferase activities in M. bovis BCG and M. smegmatis was not unexpected. There were no significant differences between the activities of the M. bovis BCG hsp60, M. leprae 28 kDa and M. leprae 18 kDa promoters in M. smegmatis and M. bovis BCG (Dellagostin et al., 1995). In addition, the efficiency and specificity of transcriptional recognition is conserved in M. tuberculosis, M. smegmatis and M. bovis BCG (Bashyam et al., 1996). M. smegmatis is, therefore, a suitable host for the characterization of M. tuberculosis promoters.

A novel observation reported here is the identification of the UAR sequence upstream of the katG promoter. As mentioned earlier, these elements do occur in other bacteria, but there is a large degree of sequence divergence. No identical UAR sequences were found elsewhere in the M. tuberculosis genome and it would be intriguing to determine if other mycobacterial genes are controlled, in part, by divergent UAR elements. Furthermore, are these elements particularly common to the mycobacteria? Generally, the UAR or UP regions are only identified by functional analysis of the regions upstream of promoters, indicating the value of this kind of study.

ACKNOWLEDGEMENTS

This work was supported by the University of Cape Town and MRC and was conducted as part of the Glaxo-Wellcome Action TB initiative.

REFERENCES


