Identification of *Mycobacterium paratuberculosis* gene expression signals

John P. Bannantine  
*Iowa State University, Ames, IA 50011, USA*

Raul G. Barletta  
*University of Nebraska - Lincoln, rbarletta1@unl.edu*

Charles O. Thoen  
*Iowa State University, Ames, IA 50011, USA*

Robert E. Andrews Jr.  
*Iowa State University, Ames, IA 50011, USA*

Follow this and additional works at: [http://digitalcommons.unl.edu/vetscipapers](http://digitalcommons.unl.edu/vetscipapers)

Part of the [Veterinary Medicine Commons](http://digitalcommons.unl.edu/vetscipapers/71)


[http://digitalcommons.unl.edu/vetscipapers/71](http://digitalcommons.unl.edu/vetscipapers/71)

This Article is brought to you for free and open access by the Veterinary and Biomedical Sciences, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Veterinary and Biomedical Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Identification of Mycobacterium paratuberculosis gene expression signals

John P. Bannantine,1† Raúl G. Barletta,2 Charles O. Thoen,1 and Robert E. Andrews Jr.1

1 Department of Microbiology, Immunology and Preventive Medicine, Iowa State University, Ames, IA 50011, USA
2 Department of Veterinary and Biomedical Sciences, Center for Biotechnology, University of Nebraska, Lincoln, NE 68583-0905, USA
Corresponding author: Robert E. Andrews Jr.: Email: randrews@iastate.edu
† Present address: Laboratory of Intracellular Parasites, Rocky Mountain Laboratories, NIH, NIAID 903 South 4th Street, Hamilton, MT 59840, USA.

The Genome Sequence Database accession numbers for the M. paratuberculosis sequences determined in this work are L76360, L76361, L76363, L76364, L76365, L76373, L76374 and L76378.

Abstract: Mycobacterium paratuberculosis promoter-containing clones were isolated from a genomic DNA library constructed in the transcriptional-translational fusion vector pYUB76. The promoter-containing DNA fragments were identified in the surrogate host Mycobacterium smegmatis by expression of the promoterless lacZ reporter gene of pYUB76. The expression signals exhibited a wide range of strengths, as indicated by their corresponding β-galactosidase activities. Eight clones were sequenced and characterized further. Predicted open reading frames and codon usage were identified by computer analysis. Database searching for related sequences using the BLAST method revealed no homologies. Transcriptional activity was measured by slot-blot hybridization with steady-state RNA isolated from lacZ M. smegmatis clones. Primer extension analysis identified the transcription start sites within the cloned fragments. The promoter regions characterized in this study were used to establish a consensus promoter sequence for M. paratuberculosis. M. paratuberculosis consensus hexanucleotide sequences of TGMCGT and CGGCCS centred approximately 35 and 10 bp upstream from the transcription startpoints do not correspond to the consensus hexanucleotides of Escherichia coli promoters.

Keywords: Mycobacterium, promoter, gene expression, codon usage, primer extension

INTRODUCTION: Mycobacterium, promoter, gene expression, codon usage, primer extension

Despite the medical importance of members in the genus Mycobacterium, characterization of mycobacterial gene expression signals has lagged behind other bacterial systems. Genetic manipulation is difficult because pathogenic mycobacteria are slow growing, have undeveloped methods of genetic exchange, and are refractile to chemical or enzymic lysis. For these reasons, Mycobacterium paratuberculosis (M. avium subsp. paratuberculosis), the aetiological agent of Johnne’s disease, has not been well characterized at the molecular level. Few genes have been cloned and sequenced (Stevenson et al., 1991; Gilot et al., 1993; Green et al., 1989; Hance et al., 1989; El-Zaatari et al., 1994) and only two promoters have been characterized in this species (Murray et al., 1992; Thomas et al., 1992). No composite class transposons have been discovered and a method of genetic exchange has only recently been established (Foley-Thomas et al., 1995).

Recently, several mycobacterial promoters have been cloned (Sela & Clark-Curtiss, 1991; Barletta et al., 1992; DasGupta et al., 1993; Thomas et al., 1992; Timm et al., 1994). In these studies the cloned promoter fragments were not sequenced (Barletta et al., 1992; Timm et al., 1994) and transcription start site mapping was either performed in recombinant Escherichia coli clones (Sela & Clark-Curtiss, 1991) or not at all (Barletta et al., 1992; DasGupta et al., 1993). Very recently, a thorough analysis of Mycobacterium smegmatis and M. tuberculosis promoter elements has been performed (Bashyam et al., 1996).

The transcriptional-translational fusion vector pYUB76 has been used to clone expression signals from mycobacteriophages (Barletta et al., 1992). We used this vector to extend the current knowledge of promoter architecture and gene expression in M. paratuberculosis by cloning and characterizing chromosomal promoter elements.
METHODS

Bacterial strains and plasmids. M. paratuberculosis ATCC 19698 was provided by D. Whipple, NADC, Ames, IA, USA. Excherichia coli DH10B (Gibco-BRL) served as the strain for routine plasmid maintenance and isolation. M. smegmatis mc²155 (Snapper et al., 1990) was the surrogate host for the M. paratuberculosis gene expression experiments. The transcriptional-translational lacZ fusion vector pYUB76 (Barletta et al., 1992) replicates in E. coli and M. smegmatis.

Growth conditions and extraction of Tween-80, OADC enrichment (Difco) and 2 mg l⁻¹ coli formants were selected on Middlebrook 7H10 agar (Difco) with 50 μg ml⁻¹ kanamycin ml⁻¹. The cells were harvested after 5 weeks of growth in tissue-culture flasks (Corning) at 37 °C. M. paratuberculosis pellets (1–3 g) were frozen at −70 °C and genomic DNA was extracted by the method of Whipple et al. (1987).

DNA cloning. Genomic DNA from M. paratuberculosis was partially digested with Sau3AI and size-fractionated by agarose gel electrophoresis. Fragments in the 0.2–2 kb size range were excised from the gel and purified by electroelution (Sambrook et al., 1989). Approximately 0.5 μg purified genomic DNA fragments was ligated to 1 μg dephosphorylated, BamHI-cleaved pYUB76 in a final volume of 10 μl. The ligation mixture was used to transform M. smegmatis mc²155 or E. coli DH10B by electroporation. E. coli transformants were selected on Luria–Bertani (LB) medium containing 50 μg kanamycin ml⁻¹ and X-Gal (45 μg ml⁻¹). M. smegmatis electrotransformants were selected on Middlebrook 7H10 agar (Difco) with 50 μg kanamycin ml⁻¹ and X-Gal. Blue colonies on this medium were considered candidate M. paratuberculosis promoter sequencers.

Quantitative β-galactosidase assays. β-Galactosidase assays for M. smegmatis were performed as described by Barletta et al. (1992). For all experiments, β-galactosidase activity is expressed in Miller units (Miller, 1972) and normalized by dividing by the OD₆₀₀ readings of the respective cultures, which ranged from 0.2 to 0.3.

RNA extraction. Total cell RNA was extracted from M. smegmatis using the protocol described by Basham & Tyagi (1994).

DNA sequencing. Sequencing reactions were performed manually with the Sequenase version 2.0 sequencing kit (United States Biochemical) for generating size markers to resolve the length of the primer extension products. The lacZ+ M. paratuberculosis inserts in pYUB76 were sequenced with the Taq DyeDeoxy Terminator Cycle Sequencing Kit (Perkin Elmer/Applied Biosystems Division: PE/ABD). Each sequencing reaction included 100 ng template DNA and 10 pmol sequencing primer in a total volume of 20 μl. Each reaction underwent 25 cycles (98 °C for 15 s, 50 °C for 2 s, 60 °C for 4 min) in the GeneAmp PCR System 9600 thermal cycler (Perkin-Elmer). The sequencing products were purified through Centri-Sep columns (Princeton Separations) and were loaded onto 6% denaturing polyacrylamide gels in an automated sequencer (373A DNA Sequencer; PE/ABD). Template DNA for automated sequencing was prepared by one of two methods: PEG preparation (Tartof & Hobbs, 1987) or Qiagen miniprep extraction.

To sequence lacZ+ inserts from pYUB76, synthetic oligonucleotides (5′-CATTATTATCATGACATTACC-3′ and 5′-GTGCTGCAAGGGCATTAAGTGTG-3′) that flank the cloning site and hybridize to the vector were employed. When needed, internal primers were designed to sequence further into the cloned inserts.

Computer analysis of DNA sequences. Sequence alignments were performed with the software package SeqEd (Version 1.0.3; PE/ABD). Basic Local Alignment Search Tool (blast; Altschul et al., 1990) was used for database homology searches. The ORFs were identified with either Gene Finder, Sequencher 3.0, or the Wisconsin package GCG (Devereux et al., 1984).

Slot-blot hybridization analysis. Aliquots (5 μg) of M. smegmatis total cell RNA were heated and chemically denatured prior to application to a Hybrid-N nylon membrane (Amersham). RNA samples were applied with a filtration manifold consisting of a luscine block containing a number of slots (slot-blotting apparatus; Gibco-BRL). Prehybridization, hybridization and washes were performed by routine methods (Sambrook et al., 1989). The radioactive probes used for hybridization consisted of [α-³²P]dATP-labelled 3.2kb EcoRI fragment from pCClac5 (Cupples & Miller, 1988) and 16S and 23S rRNA from M. smegmatis. Plasmid pCClacl5 contains the 3.2kb EcoRI fragment carrying lacZ. Unincorporated nucleotides were separated from labelled probe by two consecutive precipitations with ethanol in the presence of 7.5 M ammonium acetate. Slot-blot signals were quantified by scanning the autoradiogram into the shareware computer program NIH Image and performing an internal calibration. The levels of RNA were adjusted by first normalizing the relative amounts of RNA used in the rRNA slots then similarly adjusting the RNA levels in the lacZ slots.

Primer extension analysis. Total cell RNA from M. smegmatis strains harbouring the lacZ fusions was used as a template for the synthesis of cDNA products from 5′-end-labelled synthetic oligonucleotide primers. The primer (5′-GTGCTGCAAGGGCATTAAGTGTG-3′), which corresponds to nucleotides internal to lacZ in pYUB76, was end-labelled by T4 polynucleotide kinase (Promega) with [α-³²P]dATP (Amersham). The annealing and reverse transcription reactions were performed with the AMV reverse transcriptase primer extension system following the instructions of the manufacturer (Pro-mega). The products of the reactions were visualized by autoradiography after separation in 6% urea-polyacrylamide gels.

RESULTS

Cloning M. paratuberculosis promoters in pYUB76

To define M. paratuberculosis promoter elements, a genomic DNA library was constructed in the transcriptional-translational fusion vector pYUB76 (Barletta et al., 1992) and transformed into E. coli DH10B. This library was pooled, amplified and transformed by electroporation into M. smegmatis. In a representative experiment, 24 lacZ+ colonies were obtained from 1774 kanamycin-resistant transformants. Eight clones displaying varying intensities in the chromogenic assay (e.g. dark to light blue on X-Gal plates) were selected for further study.

Relative β-galactosidase activity and mRNA levels of the M. smegmatis lacZ-fusion constructs

Figure 1 shows quantitative β-galactosidase activities and steady-state mRNA levels of M. smegmatis trans-ormants harbouring the pYUB76 vector with and without promoter-containing inserts. Also shown are the schematic maps of the lacZ fusions in this study. The lacZ fusions pAJB300, pAJB301 and pAJB303 displayed high expression of the reporter gene; fusions pAJB86, pAJB125, and pAJB305 gave medium levels of β-galactosidase activity; and fusions pAJB73 and pAJB304 showed low levels. No detect-
able activity was observed with the strain carrying the vector pYUB76. The *M. smegmatis* (pAJB73) *lacZ* fusion exhibited a blue colour on X-Gal-containing media, but had little β-galactosidase activity when quantitatively determined using ONPG as the substrate. When these constructs were transformed into *E. coli* DH10B, only pAJB73, pAJB86 and pAJB125 expressed β-galactosidase, at moderate levels (data not shown).

To quantify transcriptional activity from the pYUB76::*lacZ* fusions, slot-blot hybridization experiments were performed. Steady-state RNA from *lacZ*-positive *M. smegmatis* strains was applied to nylon membranes and hybridized (Figure 2). Comparing the relative amounts of *lacZ* fusion message, *M. smegmatis* cells harbouring pAJB300, pAJB301, pAJB303 and pAJB304 showed the highest level of transcriptional activity. No transcriptional activity was detected in *M. smegmatis* harbouring pYUB76 by either Northern hybridization analysis (data not shown) or slot-blot analysis (Figure 2). As indicated by the normalized ratio of mRNA levels and β-galactosidase activities, the correlation between these two measurements is not complete. The pAJB303 construct shows the most steady-state mRNA, but β-galactosidase activity ranks it third strongest. Similarly, pAJB300 has the highest β-galactosidase activity and is second to pAJB303 in transcriptional activity. pAJB125 shows the least correlation, displaying medium levels of β-galactosidase activity, but very low *lacZ* mRNA levels. The reverse situation occurs with pAJB73. This lack of correlation is probably attributable to several factors, such as mRNA stability and translational efficiency.
DNA sequence analysis

The nucleotide sequence was determined for each of the lacZ+ M. paratuberculosis inserts. Insert sizes, mol% G+C, and number of amino acids derived from each mycobacterial fusion, are listed in Table 1. The mol% G+C of sequenced inserts agrees well with the published figure of 66–67 mol% for M. paratuberculosis DNA (McFadden et al., 1987). The putative ORFs were identified by computer analysis and confirmed for conditions specific to lacZ expression in pYUB76 [i.e. all cloned sequences must contain a promoter, ribosome-binding site (RBS), and a start codon that must be in-frame with the ninth codon (GUC) of lacZ]. The pAJB303 construct has two possible in-frame start codons (AUG and GUG). Since the GUG codon is separated from the RBS by only two nucleotides, we chose the AUG start codon which is separated by 8 nucleotides, a more prototypical spacing seen in prokaryotes. blast searches of nucleic acid databases showed no significant homology to any of the M. paratuberculosis sequences in this study. However, pAJB73 and pA JB304 have a M. paratuberculosis nucleotide or protein sequence of insufficient length to reveal meaningful homologies in the database.

Primer extension analysis

To localize the promoter region within the cloned M. paratuberculosis inserts, the transcription start sites were mapped. The mRNA initiation sites of the cloned lacZ-fusion transcripts were determined by primer extension analysis. Primer extension products were generated for the pA JB73 (data not shown), pA JB86, pA JB125, pA JB300, pA JB301, pA JB303, pA JB304, and pA JB305 constructs (Figure 3). The start-point of the pA JB303 transcript was not determined to single base-pair accuracy due to the large size of the cDNA product. Therefore, an estimate of the extension product was made using radiolabelled φX174 DNA markers digested with Hinfl (data not shown).

Table 1. Sequence data from lacZ+ fusion constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>Insert size (bp)</th>
<th>G+C content (mol%)</th>
<th>No. of aa derived from the mycobacterial gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>pA JB303</td>
<td>610</td>
<td>67.2</td>
<td>95</td>
</tr>
<tr>
<td>pA JB300</td>
<td>412</td>
<td>66.1</td>
<td>24</td>
</tr>
<tr>
<td>pA JB301</td>
<td>292</td>
<td>70.9</td>
<td>61</td>
</tr>
<tr>
<td>pA JB304</td>
<td>663</td>
<td>68.2</td>
<td>7</td>
</tr>
<tr>
<td>pA JB73</td>
<td>104</td>
<td>69.2</td>
<td>15</td>
</tr>
<tr>
<td>pA JB305</td>
<td>559</td>
<td>71.0</td>
<td>10</td>
</tr>
<tr>
<td>pA JB86</td>
<td>&gt;1800</td>
<td>66.3</td>
<td>20</td>
</tr>
<tr>
<td>pA JB125</td>
<td>1660</td>
<td>60.8</td>
<td>81</td>
</tr>
</tbody>
</table>

Figure 3. Primer extension analysis of lacZ fusion transcripts from pA JB300 (lane 1), pA JB301 (lane 2), pA JB303 (lane 3), pA JB86 (lane 4), pA JB305 (lane 5), pA JB125 (lane 6), and pA JB304 (lane 7). The positions of the primer extension products are indicated by arrows along with the nucleotide (nt) length of the product. An M13 sequence ladder was used to size the extension products with the potential limitation of a 1–2 nucleotide inaccuracy. Details of the experiment are described in Methods.
Table 2. Nucleotide sequences of predicted \textit{M. paratuberculosis} ribosome-binding sites

Predicted RBSs are underlined, with nucleotides that may base-pair with rRNA shown in bold. Potential AUG and GUG initiation codons are also underlined.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Nucleotide sequence</th>
<th>Δ\text{G} (kcal mol(^{-1}))</th>
<th>Position of start codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAJB303</td>
<td>AGGGCUGGCGGGCGCUAGCACGGAUGAAGUGAAAGAU</td>
<td>-6.0</td>
<td>+94</td>
</tr>
<tr>
<td>pAJB300</td>
<td>GCCAAGUGCUCAAAACGGCGGUAGACAGUAAGUGAA</td>
<td>-4.8</td>
<td>+96</td>
</tr>
<tr>
<td>pAJB301</td>
<td>GCCGGCGGCGAAGCGCGGCAAGCGCGGCAAG</td>
<td>-5.6</td>
<td>+59</td>
</tr>
<tr>
<td>pAJB304</td>
<td>GCCGGCGGCGAAGCGCGGCAAGCGCGGCAAG</td>
<td>-15</td>
<td>+22</td>
</tr>
<tr>
<td>pAJB73</td>
<td>GCCGGCGGCGAAGCGCGGCAAGCGCGGCAAG</td>
<td>-3.6</td>
<td>+15</td>
</tr>
<tr>
<td>pAJB305</td>
<td>GCCGGCGGCGAAGCGCGGCAAGCGCGGCAAG</td>
<td>-10</td>
<td>+79</td>
</tr>
<tr>
<td>pAJB86</td>
<td>GCCGGCGGCGAAGCGCGGCAAGCGCGGCAAG</td>
<td>-8.0</td>
<td>+118</td>
</tr>
<tr>
<td>pAJB125</td>
<td>GCCGGCGGCGAAGCGCGGCAAGCGCGGCAAG</td>
<td>-19</td>
<td>+76</td>
</tr>
</tbody>
</table>

16S rRNA 3′

AUUCUCACACUA

Table 3. −35 and −10 promoter regions and their spacings in mycobacteria

<table>
<thead>
<tr>
<th>Construct</th>
<th>−35 region</th>
<th>−10 region</th>
<th>+1</th>
<th>Spacing*</th>
<th>Initial codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{M. paratuberculosis} promoters from this study†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAJB303</td>
<td>GGTTCCGTTTGCCTTGATGCGCAA</td>
<td>CGGCCAGGCGCGTGC</td>
<td>T</td>
<td>92</td>
<td>ATG</td>
</tr>
<tr>
<td>pAJB300</td>
<td>CAGTGGACGTGCGGCGCCCGCAG</td>
<td>CGGCCAGGCGCGTGC</td>
<td>A</td>
<td>195</td>
<td>ATG</td>
</tr>
<tr>
<td>pAJB301</td>
<td>GCAGGTACAGGCGGCGTGCCTGC</td>
<td>CGGCCAGGCGCGTGC</td>
<td>G</td>
<td>21</td>
<td>ATG</td>
</tr>
<tr>
<td>pAJB304</td>
<td>CGACGGATGGGCGGCGTGCCTGC</td>
<td>CGGCCAGGCGCGTGC</td>
<td>T</td>
<td>13</td>
<td>ATG</td>
</tr>
<tr>
<td>pAJB73</td>
<td>GGTTCCGTTTGCCTTGATGCGCAA</td>
<td>CGGCCAGGCGCGTGC</td>
<td>T</td>
<td>77</td>
<td>ATG</td>
</tr>
<tr>
<td>pAJB305</td>
<td>CGTTGGACGTGCGGCGTCGGTT</td>
<td>CGGCCAGGCGCGTGC</td>
<td>T</td>
<td>116</td>
<td>ATG</td>
</tr>
<tr>
<td>pAJB86</td>
<td>CGTTGGACGTGCGGCGTCGGTT</td>
<td>CGGCCAGGCGCGTGC</td>
<td>C</td>
<td>74</td>
<td>ATG</td>
</tr>
<tr>
<td>pAJB125</td>
<td>CGTTGGACGTGCGGCGTCGGTT</td>
<td>CGGCCAGGCGCGTGC</td>
<td>C</td>
<td>63</td>
<td>ATG</td>
</tr>
</tbody>
</table>

Consensus‡

TGGCGTT

\(-\text{N}_{16–20}^{−}\) CGGCCS

Other mycobacterial promoters§

\(\text{bsp60 BCG}\)

<table>
<thead>
<tr>
<th>Construct</th>
<th>−35 region</th>
<th>−10 region</th>
<th>+1</th>
<th>Spacing*</th>
<th>Initial codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>bsp60 BCG</td>
<td>TGGAC</td>
<td>17*</td>
<td>T</td>
<td>183</td>
<td>ATG</td>
</tr>
<tr>
<td>(\text{P}_m\ \text{M. tb.})</td>
<td>TGGAC</td>
<td>17</td>
<td>A</td>
<td>40</td>
<td>ATG</td>
</tr>
<tr>
<td>16S rRNA \textit{lepra}</td>
<td>TGGAC</td>
<td>16</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16S rRNA \textit{M. tub.}\</td>
<td>TGGAC</td>
<td>18</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{cpn60 M. tub.}\</td>
<td>TGGAC</td>
<td>17</td>
<td>A</td>
<td>29</td>
<td>ATG</td>
</tr>
<tr>
<td>\textit{mpn70 BCG}\</td>
<td>TGGAC</td>
<td>17</td>
<td>G</td>
<td>176</td>
<td>ATG</td>
</tr>
<tr>
<td>\textit{bta fibroin}\</td>
<td>TGGAC</td>
<td>19</td>
<td>A</td>
<td>0</td>
<td>ATG</td>
</tr>
<tr>
<td>\textit{ask smegmatis}\</td>
<td>TGGAC</td>
<td>17</td>
<td>G</td>
<td>250</td>
<td>ATG</td>
</tr>
<tr>
<td>\textit{85A M. tub.}\</td>
<td>TGGAC</td>
<td>22</td>
<td>A</td>
<td>63</td>
<td>ATG</td>
</tr>
</tbody>
</table>

*Number of nucleotides between indicated regions.
†Nucleotides in bold represent the −35 and −10 consensus regions. Underlined nucleotides match the consensus at that position.
‡M represents C or A; S represents G or C. pAJB125 and pAJB304 were excluded when calculating the consensus.
§\textit{Mycobacterium} sp. promoter regions as compiled by Kremer et al. (1995). \textit{M. pht}, \textit{M. paratuberculosis}; \textit{M. tub.}, \textit{M. tuberculosis}. Only promoters for which the transcription start sites are known were included. Underlined nucleotides indicate a match to the \textit{M. paratuberculosis} consensus sequence.

Putative ribosome-binding sites in \textit{M. paratuberculosis} mRNA

A region of homology to the 3′ terminus of \textit{M. lepra} 16S rRNA followed after a spacing of 4–37 bp by an initiation codon and an ORF for translation was observed for all the \textit{lacZ} fusions. Nucleotides that may base-pair with rRNA are shown in bold in Table 2. The predicted free energy of base-pairing (\(\Delta\text{G}\)) of these RBSs with 16S rRNA, calculated according to the rules of Tinoco \textit{et al.} (1973), ranged from −3.6 to −19 kcal mol\(^{-1}\) (−15.5 to −79.5 kJ mol\(^{-1}\)).
**Codon usage**

A compilation of the codons employed in the N-terminal mycobacterial ORFs which participated in the *lacZ* fusions was performed (data not shown). The 309 codons compiled for *M. paratuberculosis* in this study show a characteristic GC bias, with 73% of the codons containing a G or C in the third position, whereas codons ending in T and A are represented at frequencies of 15 and 12%, respectively. This codon usage is remarkably similar to that obtained for *M. leprae* (Honore et al., 1993) and the *M. tuberculosis* complex (Andersson & Sharp, 1996).

**Determination of a consensus promoter sequence**

The eight *M. paratuberculosis* promoters that had their transcription start sites identified by primer extension analysis were aligned as shown in Table 3. The nucleotides shown in bold in Table 3 represent the putative –35 and –10 regions. Based on these data, a *M. paratuberculosis* consensus sequence for the –35 region was established as TGMCGT. No strong consensus was established for any of the positions in the –10 region. However, for the aligned sequences in Table 3, a cytosine nucleotide was present in the first position of the highlighted –10 region in 87% of cases. Additionally, the occurrence at the other positions was 37% G, 62% G, 37% C, 50% C and 100% S. The putative promoter regions from pAJB125 and pAJB304 were excluded from the calculation of the consensus since those two promoter regions differed significantly from the other six. For comparison, the –35 and –10 regions of all mycobacterial promoters that contain an experimentally determined transcription start site were also compiled; these are listed at the bottom of Table 3. Nucleotides that match the consensus are underlined. None of these promoter regions display significant homology with the *M. paratuberculosis* consensus.

**DISCUSSION**

Eight authentic *M. paratuberculosis* promoter elements were identified in this study based on transcriptional/translational *lacZ* fusions and primer extension analysis using *M. smegmatis* as a surrogate host. Ideally, gene expression studies should be performed in the native host. However, a genetic system for *M. paratuberculosis* has only recently been developed (Foley-Thomas et al., 1995) and the expression of the *lacZ* reporter gene in *M. paratuberculosis* has not been documented. Nevertheless, a homologous mycobacterial system such as *M. smegmatis* offers clear advantages to the use of an *E. coli* system that is unable to recognize many mycobacterial expression sequences (Barletta et al., 1992; Clark-Curtiss et al., 1985; Thole et al., 1985).

Strong β-galactosidase activity was observed in the *lacZ* fusion constructs of pAJB300, pAJB301 and pAJB303. The Miller units obtained in the quantitative β-galactosidase assay indicated that pAJB300 is stronger than most of the mycobacteriophage promoters cloned in pYUB76 (Barletta et al., 1992). In contrast, pAJB301 and pAJB303 were slightly weaker than the cloned mycobacteriophage promoters. The cloned pAJB300 or pAJB303 inserts may be good candidates for expression of heterologous antigens in *M. bovis* BCG for recombinant vaccine development (Stover et al., 1991, 1993) because of their high level of expression.

The slot-blot hybridization of *M. smegmatis* RNA was performed to quantify steady-state levels of the *lacZ* fusion message. Thus, pAJB303 might contain the strongest promoter. Alternatively, the pAJB303 message might be more stable. The 5′ leader sequence of some stable transcripts such as the T4 bacteriophage gene 32 or the *E. coli* ompA can stabilize other sequences fused downstream from them (Ehretsmann et al., 1992). For example, the *lacZ* transcript has a half-life of 90 s, but a gene 32::*lacZ* fusion has a half-life of 20 min.

Although we have not identified initiation codons directly by amino acid sequencing, all of the *lacZ* fusion mRNAs display a putative RBS complementary to the 3′ region of *M. leprae* 16S RNA followed by a possible initiation codon and an open reading frame. The free energies of interaction (ΔG) for the proposed RBSs vary from –3.6 to –19 kcal mol⁻¹ (–15.5 to –79.5 kJ mol⁻¹) as compared to –9.4 kcal mol⁻¹ (39.3 kJ mol⁻¹) for the prototype *E. coli* RBS AGGA (McLaughlin et al., 1981).

Six of the promoters transcriptionally mapped in this study display some degree of conformity in their –35 regions and, less so, in the –10 regions, with the canonical sequence including matches in at least half of the positions for each region. The intervening sequence between the –35 and –10 regions (16–20 bp) corresponds to the preferred interval for prokaryotic promoters. The –35 promoter regions from pAJB304 (no position conserved) and pAJB125 (two positions conserved) differ significantly. pAJB125 also differs significantly at the –10 box (two positions conserved). Significant differences are observed when we compare the *M. paratuberculosis* consensus sequence with nine previously studied promoters (Table 3). These deviations suggest that these promoters may be under the control of different sigma factors, a well-known situation in other microbial systems (Hawley & McClure, 1983). Since *M. paratuberculosis* promoters were identified in vitro under standard growing conditions, it is likely that the consensus sequence established in this study corresponds to the principal sigma factor involved in the transcription of housekeeping genes. In addition, since *M. smegmatis* was the surrogate host for these promoter studies, there may exist a bias toward *M. paratuberculosis* promoters that are highly active in that host. Conversely, some of the previously studied pro-
motors may require special sigma factors related to stress conditions, as exemplified by the heat-shock promoters hsp60 and mph70.

Recently, Bashyam et al. (1996) reported a thorough study on the consensus sequences of M. smegmatis and M. tuberculosis promoters. In that study, a fairly well-conserved –10 box was found which resembled the E. coli Pribrnow box. In contrast, we failed to find any conserved sequence of this kind. These differences may be attributable to the use of different reporter genes and selection markers, which may have led to the isolation of a different set of promoters. The observation that a bacterial sigma factor can recognize promoters with diverse –10 regions is unusual but not unique. Studies showing diverse –10 hexamers have been described for Caulobacter crescentus (Malakooti et al., 1995) and Chlamydia trachomatis (Douglas & Hatch, 1996).

Additional promoter sequences must be compared to test and refine the consensus regions. The –35 region of the paratuberculosis consensus bears some resemblance to E. coli whereas the –10 region does not. Upstream regions of mycobacterial genes have been shown to have a higher G+C content than their E. coli counterparts (Dale & Patki, 1990). This observation was confirmed in the present study, as the conserved regions have a G+C content of 66.7 mol% compared with 16.7 mol% for the E. coli consensus. As was pointed out by Dale & Patki (1990), analysis of the sequence of many mycobacterial genes does not reveal any regions upstream from the putative translational start position that resemble an E. coli-like consensus promoter. The M. paratuberculosis promoters defined in this study are more closely related to each other than they are to the other compiled mycobacterial promoters in Table 3. The results of this study have, for the first time, revealed a potential paratuberculosis consensus promoter.

ACKNOWLEDGEMENTS

We thank the DNA sequencing and synthesis facility (1184 Molecular Biology Building, Iowa State University) for performing the automated sequencing of the pYUB76::lacZ fusion inserts. We also thank Greg Phillips for providing plasmids and thoughtful discussions.

This is journal paper no. J-16741 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. The work is part of Project no. 3222, and was supported by Hatch Act and State of Iowa Funds. R.G.B. was supported by Cooperative State Research Service Project NEB 14-077.

REFERENCES


