Isolation of strong expression signals of *Mycobacterium tuberculosis*

James A. Triccas, 1,2† Warwick J. Britton 2 and Brigitte Gicquel 1

Author for correspondence: James A. Triccas. Tel: +61 2 9565 6163. Fax: +61 2 9565 6101. e-mail: J.Triccas!centenary.usyd.edu.au

1 Unité de Génétique Mycobactérienne, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

2 Centenary Institute of Cancer Medicine and Cell Biology, Locked Bag No. 6, Newtown, NSW, 2042, Australia

The natural fluorescence of the *Aequoria victoria* green fluorescent protein was exploited to isolate strong expression signals of *Mycobacterium tuberculosis*. *Mycobacterium bovis* bacille Calmette–Guérin harbouring *M. tuberculosis* fragments driving high levels of gfp expression were isolated by fluorescence-activated cell sorting (FACS). DNA sequencing and subsequent comparison with the *M. tuberculosis* genome sequence revealed that a total of nine postulated promoters had been identified. The majority of the promoters displayed activity that was greater than or equal to the *Mycobacterium fortuitum* β-lactamase promoter, one of the strongest mycobacterial promoters characterized to date. Two of the promoters corresponded to proteins predicted to be involved in calcium and magnesium utilization, the importance of such functions for cell physiology suggesting why these two genes are controlled by strong transcription signals. The seven other promoters corresponded to genes encoding proteins of unknown function. Promoter activity was maintained after prolonged incubation within macrophages, implying that these promoters could be used to drive sustained foreign gene expression *in vivo*. The strength of these expression signals identified could be employed for the overexpression of foreign genes in mycobacteria to aid protein purification and vaccine vector development. Furthermore, this study demonstrated that FACS provides a sensitive and efficient technique to measure and select strong mycobacterial expression signals.

**Keywords:** strong promoters, green fluorescent protein, genome, macrophage

**INTRODUCTION**

Tuberculosis, caused by the intracellular pathogen *Mycobacterium tuberculosis*, remains a major public health problem of global importance. An understanding of the genetic elements controlling the expression of *M. tuberculosis* genes, especially those encoding virulence determinants, would allow the development of strategies to combat the disease. Many studies have attempted to identify promoters controlling the expression of genes within the host, in particular the macrophage, which is the preferred niche for *M. tuberculosis* during infection. These studies have relied on reporter genes such as the *Aequoria victoria* green fluorescent protein (GFP; Via *et al.*, 1996; Triccas *et al.*, 1999), or directly compared mRNA levels outside or within host cells (Graham & Clark-Curtiss, 1999). Alternatively, more generalized approaches have been used to identify a broad range of promoters, including those of a constitutive nature, through the use of reporter proteins and direct analysis of individual recombinant clones. These latter types of studies are hindered by the large number of recombinant colonies that need to be screened as the frequency of clones carrying true promoters is typically quite low (Timm *et al.*, 1994a; Das Gupta *et al.*, 1993).

While the majority of studies analysing mycobacterial transcription regulation have focused on promoters whose activity varies in response to defined environmental signals, it is of interest to selectively define strong constitutive expression signals of *M. tuberculosis*. This would provide important information on those sequences required for optimal gene expression within mycobacteria. A limiting factor in the development of versatile expression systems for the mycobacteria is the...
lack of promoters permitting consistent, high-level expression of foreign genes in recombinant mycobacteria, analogous to the lac promoter of *Escherichia coli* (de Boer et al., 1983). Thus the isolation of strong mycobacterial promoters would permit expression of foreign genes in mycobacteria to aid protein purification and recombinant vaccine vector development. This is of particular importance as recombinant proteins obtained from fast-growing mycobacterial hosts, such as *Mycobacterium smegmatis*, have demonstrated superiority over the same proteins purified from *E. coli* expression systems, as assessed by structural and immunological analysis (Garbe et al., 1993; Roche et al., 1996; Triccas et al., 1996), an important consideration in the development of specific diagnostic reagents and improved vaccines.

In this report we have identified strong expression signals of *M. tuberculosis*. A random library of *M. tuberculosis* fragments was constructed in a vector containing a promoterless copy of the gene encoding the A. **victoria** GFP. After transformation into *Mycobacterium bovis* bacille Calmette–Guérin (BCG), recombinant bacteria displaying strong fluorescence were separated by fluorescence-activated cell sorting (FACS), resulting in the identification of *M. tuberculosis* sequences demonstrating strong promoter activity.

**RESULTS AND DISCUSSION**

**Use of GFP to select M. bovis BCG clones harbouring strong expression signals of M. tuberculosis**

In a first step to isolate strong promoter sequences of *M. tuberculosis*, a library of *M. tuberculosis* DNA (0.2–1.5 kb) was constructed in the vector pJFX2 (Triccas et al., 1999), which contains a promoterless copy of the gene encoding a strongly fluorescent version of GFP.
Strong expression signals of *M. tuberculosis*

**Fig. 2.** Determination of relative *M. tuberculosis* promoter strength harboured by the isolated rBCG clones. Thirty-four colonies selected as strongly fluorescent by microscopy were analysed individually by flow cytometry. Results are represented as the fold difference in fluorescence of each clone compared to the fluorescence of BCG containing plasmid pJFX4 (gfp controlled by the *M. fortuitum* pblaF* promoter). The results are representative of two separate experiments.

**Table 1.** Analyses of clones harbouring strong *M. tuberculosis* expression signals

<table>
<thead>
<tr>
<th>Clone group (clone number)</th>
<th>Fold increase over pblaF* (± SEM)†</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (1, 2, F19, F20)</td>
<td>1.87 ± 0.38</td>
<td>Upstream of Rv0867c; unknown Ala-Pro-rich protein; signal sequence</td>
</tr>
<tr>
<td>B (5, 8, 20, F3, F5, F7, F8, F12, F17, F21, W23)</td>
<td>1.35 ± 0.09</td>
<td>Upstream of Rv1209; function unknown; in operon with tagA (DNA glycosidase)</td>
</tr>
<tr>
<td>C (9, F1, F14, F22)</td>
<td>2.52 ± 0.38</td>
<td>Upstream of Rv1233c; Pro-rich N terminus similar to calcium-binding protein A from <em>Dictyostelium discoideum</em></td>
</tr>
<tr>
<td>D (15, F6, F10)</td>
<td>2.76 ± 0.30</td>
<td>Upstream of Rv3108; function unknown</td>
</tr>
<tr>
<td>E (21)</td>
<td>0.91 ± 0.17</td>
<td>Upstream of Rv0203; function unknown; hydrophobic stretch near N terminus</td>
</tr>
<tr>
<td>F (23, W5)</td>
<td>2.00 ± 0.33</td>
<td>Within Rv2850; function unknown</td>
</tr>
<tr>
<td>G (F11, 13)</td>
<td>1.86 ± 0.06</td>
<td>Upstream of Rv2142c; function unknown</td>
</tr>
<tr>
<td>H (F18, F24)</td>
<td>3.44 ± 0.06</td>
<td>Upstream of Rv2778c; function unknown</td>
</tr>
<tr>
<td>I (F23)</td>
<td>1.75 ± 0.11</td>
<td>Upstream of Rv1811; probable magnesium transport ATPase protein C</td>
</tr>
</tbody>
</table>

† Fold increase over pblaF* represents the fluorescence value of bacteria harbouring the indicated clones divided by the fluorescence level of bacteria containing the pJFX4 plasmid (gfp controlled by the *M. fortuitum* pblaF* promoter). Values represent the mean fluorescence level for clones in a particular group ± SEM (in groups with less than three members (E–I) each individual clone was analysed three times).

(Cormack et al., 1996). After transformation of this library into *M. bovis* BCG and subsequent growth in 7H9 medium, a pool of recombinant (r) BCG clones was obtained that displayed varying levels of GFP expression as assessed by flow cytometry (Fig. 1a). Bacteria exhibiting medium to high levels of GFP fluorescence were collected by cell sorting, resulting in a pool of rBCG enriched for fluorescent bacteria (Fig. 1b). After plating of the pool onto Middlebrook 7H10, visual inspection by fluorescence microscopy revealed that approximately 50% of the colonies were fluorescent. Forty-four colonies identified as strongly fluorescent were selected for further analysis.

We next compared by flow cytometry the fluorescence of the 44 selected clones with that of BCG harbouring the pJFX4 plasmid (Triccas et al., 1999), which contains the gfp gene under the control of the β-lactamase promoter (pblaF*) of *M. fortuitum*, one of the strongest mycobacterial promoters characterized to date (Timm et al., 1994b). The strength of this promoter has been used to permit purification of important antigens of *M. tuberculosis* (Roche et al., 1996), *Mycobacterium leprae* (Triccas et al., 1996) and *Mycobacterium avium* (Triccas et al., 1998) from recombinant *M. smegmatis*. Systems based on pblaF* have also allowed overexpression of antigens in *M. bovis* BCG and subsequent detection of strong immune responses against the expressed products in animal models (Abdelhak et al., 1995). Of the 44 clones tested, 34 represented pure cultures showing fluorescence levels equal to or greater than that of BCG/pJFX4 (Fig. 2). The levels of fluorescence ranged...
from 0.91 to 3.50 times the level of gfp expression by virtue of pblaF*.

Identification of genes under the control of the isolated M. tuberculosis expression signals

As the complete genome sequence of M. tuberculosis is available (Cole et al., 1998), we next determined the genomic location of the expression signals isolated. PCR amplification was performed on the 34 rBCG clones after DNA extraction by freeze-thawing. PCR products were obtained for 30 of the clones, which after sequencing were shown to represent nine different M. tuberculosis genomic fragments (Table 1). Clones within the same group most likely represent sibling strains, as all have identical inserts within the vector. By analysing the genomic location of the sequenced fragments and determining the direction of transcription with respect to the gfp gene, we were able to determine which sequences most likely corresponded to a predicted M. tuberculosis gene (Table 1). The majority of the promoters corresponded to genes encoding proteins of unknown function (clone groups A, B, D, E, F, G and H). The DNA fragment of clone group F was located within the Rv2850 ORF of M. tuberculosis and thus may represent a cryptic promoter with strong activity in M. bovis BCG. Clone group C corresponded to a gene encoding a possible calcium-binding protein, while clone group I corresponded to a protein predicted to be involved in magnesium transport. The importance of calcium and magnesium in cell physiology may indicate why these two genes appear to be strongly expressed in mycobacteria. We cannot exclude the possibility that activity of these promoters is different in M. tuberculosis itself compared to BCG, as genomic regions present in M. tuberculosis yet absent from BCG may encode products exerting enhancing or deleterious effects on promoter activity. Furthermore, while this plasmid-based system allows comparative analysis of promoter strength, it may not truly reflect the promoter activity in M. tuberculosis, where promoter GENES would most likely be present as a single copy.

The cloned regions preceding the predicted translation start codon for the genes described in Table 1 were analysed using the BDGP Neural Network Promoter Prediction Program (Reese et al., 1996; www.fruitfly.org/seq-tools/promoter.html). Although we were able to identify — 10 and — 35 hexamers homologous to the consensus sequences of both E. coli (Harley & Reynolds, 1987) and mycobacterial promoters (Mulder et al., 1997), we did not identify patterns suggestive of a common sequence amongst promoters displaying elevated activity (seehttp://www.centenary.usyd.edu.au/research/tbres.html). The precise mapping of transcription start sites of these and other strong mycobacterial promoters would be required before a consensus pattern for strong promoters could be developed.

Activity of strong M. tuberculosis expression signals in M. smegmatis

We next determined if these M. tuberculosis expression signals that functioned strongly in M. bovis BCG also showed strong activity in M. smegmatis, the most widely used host for recombinant mycobacterial gene expression. Plasmids were recovered from rBCG, electroporated into M. smegmatis and fluorescence levels of bacteria analysed by FACS. Restriction analysis of plasmids revealed no apparent deletions (data not shown). While essentially all promoters showed greater activity than the pblaF* promoter in BCG (Fig. 3a), only the promoter of Rv3108 showed an enhanced level of activity in M. smegmatis compared to pblaF* (1.73-fold greater fluorescence; Fig. 3b). All other clones showed fluorescence levels between 0.15 and 0.69 of that of M. smegmatis harbouring pJFX4. Re-isolation of the plasmids and introduction into BCG gave similar GFP levels as in Fig. 3(a) (data not shown), suggesting that the reduced promoter activity was not due to vector deletion or rearrangement events within M. smegmatis. The reason for the reduced activity of M. tuberculosis promoters within M. smegmatis is unclear, but may reflect differences in transcription machinery amongst.
**Activity of the strong *M. tuberculosis* promoters within macrophages**

Any strong mycobacterial promoter useful for recombinant vaccine construction would ideally maintain its activity within the *in vivo* environment. This was assessed by infecting murine bone-marrow-derived macrophages with rBCG clones of each promoter group and analysis of fluorescence after 6 d infection. All clones displayed fluorescence that was not significantly different from that achieved in normal culture medium (Fig. 4). This suggests that the activity of these expression signals is not affected after prolonged exposure to the *in vivo* environment.

In this report we have exploited the fluorescent properties of the *A. victoria* GFP to permit the isolation of strong expression signals of *M. tuberculosis*. By using FACS to isolate recombinant bacteria displaying strong fluorescence, we significantly limit the number of colonies required to be screened by initially eliminating the majority of bacteria not displaying promoter activity, and can also place a limit on the level of promoter expression preferred during the cell sorting process. Such advantages are not seen with systems based on the β-galactosidase and chloramphenicol acetyltransferase reporter proteins which have previously been employed to isolate promoters of *M. tuberculosis* (Timm et al., 1994a; Das Gupta et al., 1993). Apart from providing information on the genes of *M. tuberculosis* that are strongly expressed by the bacterium, the promoters identified in this study may prove useful in the overexpression of foreign genes in mycobacteria to aid protein purification and vaccine vector development.

**ACKNOWLEDGEMENTS**

We thank Jean Rauzier for DNA sequencing, Dr Philippe Metezeau and Helene Kiefer for FACS sorting, Dr Brendan Cormack for the pKEN-GFPmut2 vector and Daniel Roach for helpful discussions. J. A. T. was the recipient of an Institut Pasteur Cantarini Fellowship. This work was supported by the European Community (grant BMH4 CT92167) and a Pasteur Merieux Connaught Research and Development grant.

**REFERENCES**


