Microbiology (2001), **147**, 1253–1258 Printed in Great Britain

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

†**Present address:** Centenary Institute of Cancer Medicine and Cell Biology, Locked Bag No 6, Newtown, NSW, 2042, Australia.

Abbreviations: BCG, bacille Calmette–Guérin; FACS, fluorescence-activated cell sorting; GFP, green fluorescent protein.

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.

METHODS

Bacterial strains and growth conditions. *E. coli* DH5α was grown on liquid or solid Luria–Bertani medium. Mycobacterial strains (*M. tuberculosis* strain 103 and *M. bovis* BCG Pasteur) were grown in liquid Middlebrook 7H9 medium (Difco) supplemented with ADC enrichment (Difco) or solid Middlebrook 7H10 medium (Difco) supplemented with OADC enrichment (Difco). When required, the antibiotic kanamycin was added at a concentration of 25 μg ml^{-1} for both *E. coli* and mycobacteria.

***M. tuberculosis* genome library construction.** Construction of the *M. tuberculosis* promoter library in the *E. coli*/mycobacterial shuttle vector pJFX2 is described elsewhere (Triccas *et al*., 1999). The pJFX4 vector was constructed by placing the strong β-lactamase promoter of *Mycobacterium fortuitum* (Timm *et al*., 1994b) into pJFX2. Preparation of competent cells and electroporation of mycobacteria was carried out as described previously (Pelicic *et al*., 1997).

FACS analysis and sorting of fluorescent bacteria. Bacteria were analysed with a FACScan (Becton-Dickinson Immunocytometry Systems) and sorted with an ELITE cell-sorting system (Beckman Coulter). After growth in 7H9 medium, BCG cells were resuspended in 1 ml PBS (approx. 1×10^6 bacteria), analysed by FACS and bacteria exhibiting medium to high levels of fluorescence were collected. Bacteria were plated directly onto 7H10 solid medium to recover individual clones.

Macrophage preparation and infection. Bone-marrow-derived macrophages were prepared as described previously (Triccas *et al*., 1999). Macrophage monolayers were infected with bacteria at an m.o.i. of 1:1. After 4 h infection, extracellular bacteria were removed by washing four times with PBS and incubation continued at 37 °C in 5% CO_2. After 6 d infection, BCG-infected macrophages were washed three times in PBS, centrifuged and lysed in water plus 0·1 % Tween 80. Recovered bacteria were analysed directly by flow cytometry.

Selection and analysis of clones harbouring strong *M. tuberculosis* expression signals. Individual BCG colonies were analysed directly using an Axioscop fluorescence microscope (Zeiss) and those colonies exhibiting strong fluorescence inoculated into 7H9 medium in 24-well culture plates. After 7 d growth, the geometric mean of fluorescence for bacterial populations was determined by flow cytometry using the Lysis II program (Becton-Dickinson). The identity of the *M. tuberculosis* insert was determined by PCR, DNA sequencing and database analysis as described previously (Triccas *et al*., 1999).

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

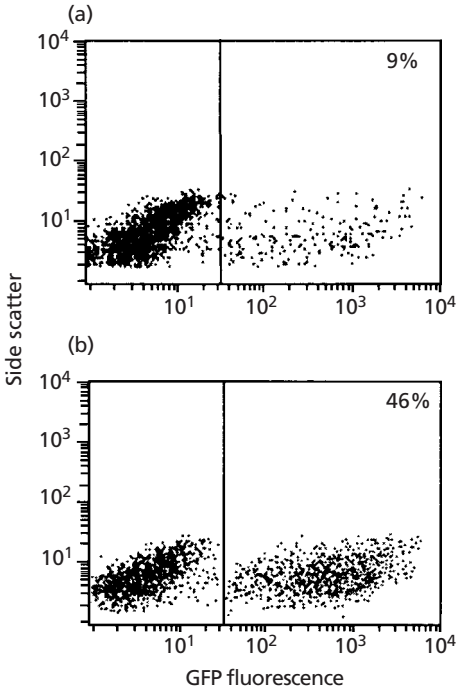

Fig. 1. Selection by FACS of rBCG clones harbouring strong *M. tuberculosis* expression signals. A library of *M. tuberculosis* 103 DNA fragments was constructed in the pJFX2 vector and transformed into *M. bovis* BCG. The proportion of bacteria displaying GFP fluorescence after growth in 7H9 culture medium was determined by flow cytometry (a). Bacteria exhibiting medium to high levels of GFP fluorescence were collected resulting in a pool of rBCG enriched for fluorescent bacteria (b).

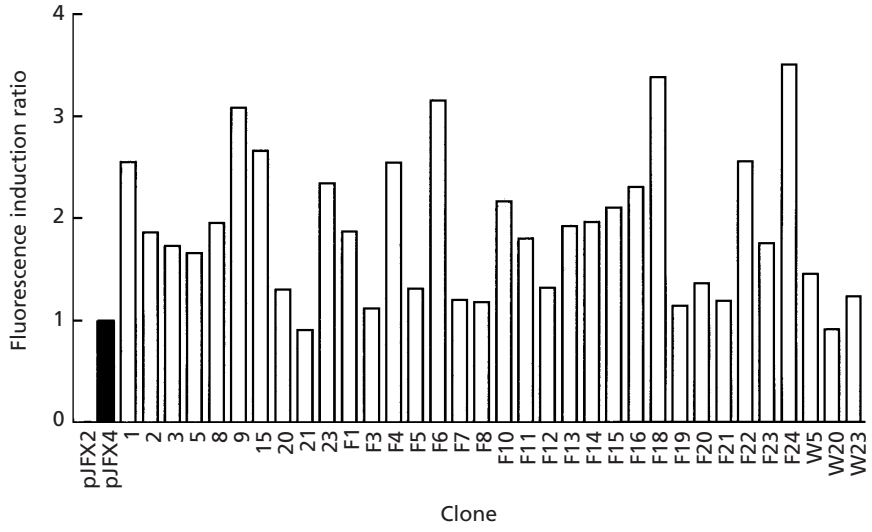

Fig. 2. Determination of relative *M. tuberculosis* promoter strength harboured by the isolated rBCG clones. Thirty-four colonies selected as strongly fluorescent by microscopy were analysed individually by flow cytometry. Results are represented as the fold difference in fluorescence of each clone compared 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

Clone group (clone number)	Fold increase over *pblaF** (±SEM)†	Description
A (1, 2, F19, F20)	1·87±0·38	Upstream of Rv0867c; unknown Ala-Pro-rich protein; signal sequence
B (5, 8, 20, F3, F5, F7, F8, F12, F17, F21, W23)	1·35±0.09	Upstream of Rv1209; function unknown; in operon with *tagA* (DNA glycosidase)
C (9, F1, F14, F22)	2·52± 0·38	Upstream of Rv1233c; Pro-rich N terminus similar to calcium-binding protein A from *Dictyostelium discoideum*
D (15, F6, F10)	2·76± 0·30	Upstream of Rv3108; function unknown
E (21)	0·91± 0·17	Upstream of Rv0203; function unknown; hydrophobic stretch near N terminus
F (23, W5)	2·00±0·33	Within Rv2850; function unknown
G (F11, 13)	1·86±0·06	Upstream of Rv2142c; function unknown
H (F18, F24)	3·44±0·06	Upstream of Rv2778c; function unknown
I (F23)	1·75±0·11	Upstream of Rv1811; probable magnesium transport ATPase protein C

† 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

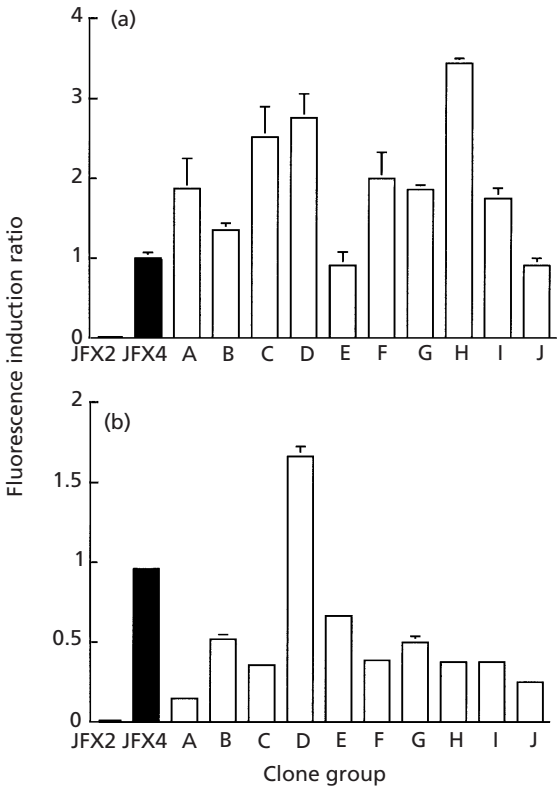

Fig. 3. Comparison of the strength of *M. tuberculosis* promoters in *M. bovis* BCG (a) and *M. smegmatis* (b). Promoter strength was assessed by using flow cytometry to determine the fluorescence of each clone. Results are represented as the mean fold difference in fluorescence for the clones (±SEM; three experiments) compared to the fluorescence of either BCG (a) or *M. smegmatis* (b) containing plasmid pJFX4 (*gfp* controlled by the *M. fortuitum pblaF** promoter).

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 (see http://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 myobacterial 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

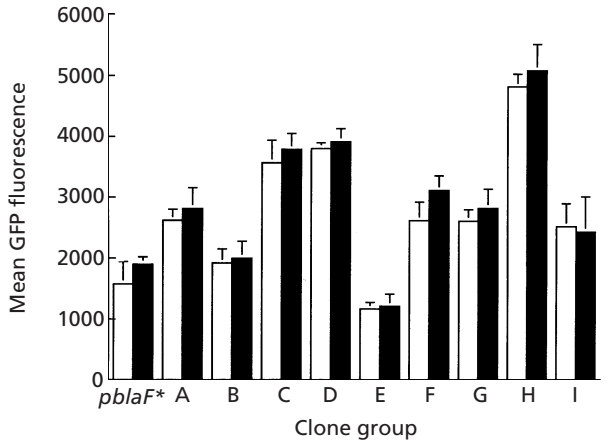

Fig. 4. Activity of strong *M. tuberculosis* promoters within macrophages. The relative GFP fluorescence intensities of bacteria recovered from murine bone-marrow-derived macrophages after 6 d (solid bars) or grown in 7H9 medium for 7 d (white bars) were compared by flow cytometry. Results are expressed as the mean fluorescence level ± SEM for three experiments.

mycobacterial species, or the existence of repressive elements within *M. smegmatis*.

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 strength 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 Métézeau and Hélène 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

Abdelhak, S., Louzir, H., Timm, J., Blel, L., Benlasfer, Z., Lagranderie, M., Gheorghiu, M., Dellagi, K. & Gicquel, B. (1995). Recombinant BCG expressing the leishmania surface antigen Gp63 induces protective immunity against *Leishmania major* infection in BALB/c mice. *Microbiology* **141**, 1585–1592.

de Boer, H. A., Comstock, L. J. & Vasser, M. (1983). The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci U S A* **80**, 21–25.

Cole, S. T., Brosch, R., Parkhill, J. & 39 other authors (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.

Cormack, B. P., Valdivia, R. H. & Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38.

Das Gupta, S. K., Bashyam, M. D. & Tyagi, A. K. (1993). Cloning and assessment of mycobacterial promoters by using a plasmid shuttle vector. *J Bacteriol* **175**, 5186–5192.

Garbe, T., Harris, D., Vordermeier, M., Lathigra, R., Ivanyi, J. & Young, D. B. (1993). Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation. *Infect Immun* **61**, 260–267.

Graham, J. E. & Clark-Curtiss, J. E. (1999). Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proc Natl Acad Sci U S A* **96**, 11554–11559.

Harley, C. B. & Reynolds, R. P. (1987). Analysis of *E. coli* promoter sequences. *Nucleic Acids Res* **15**, 2343–2361.

Mulder, M. A., Zappe, H. & Steyn, L. M. (1997). Mycobacterial promoters. *Tuber Lung Dis* **78**, 211–223.

Pelicic, V., Jackson, M., Reyrat, J.-M., Jacobs, W. R., Gicquel, B. & Guilhot, C. (1997). Efficient allelic exchange and transposition mutagenesis in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **94**, 10955–10960.

Reese, M. G., Harris, N. L. & Eeckman, F. H. (1996). Large scale sequencing specific neural networks for promoter and splice site recognition. In *Biocomputing: Proceedings of the 1996 Pacific Symposium*. Edited by Lawrence Hunter & Terri E. Klein. Singapore: World Scientific Publishing.

Roche, P. W., Winter, N., Triccas, J. A., Feng, C. & Britton, W. J. (1996). Expression of *Mycobacterium tuberculosis* MPT64 in recombinant *M. smegmatis*: purification, immunogenicity and application to skin tests for tuberculosis. *Clin Exp Immunol* **103**, 226–232.

Timm, J., Lim, E. M. & Gicquel, B. (1994a). *Escherichia coli*-mycobacteria shuttle vectors for operon and gene fusions to *lacZ*: the pJEM series. *J Bacteriol* **176**, 6749–6753.

Timm, J., Perilli, M. G., Duez, C. & 9 other authors (1994b). Transcription and expression analysis, using *lacZ* and *phoA* gene

fusions, of *Mycobacterium fortuitum* β-lactamase genes cloned from a natural isolate and a high-level β-lactamase producer. *Mol Microbiol* **12**, 491–504.

Triccas, J. A., Roche, P. W., Winter, N., Feng, C. G., Butlin, C. R. & Britton, W. J. (1996). A 35 kDa protein is a major target of the human immune response to *Mycobacterium leprae*. *Infect Immun* **64**, 5171–5177.

Triccas, J. A., Winter, N., Roche, P. W., Gilpin, A. & Britton, W. J. (1998). Molecular and immunological analyses of the *Mycobacterium avium* homologue of the highly antigenic *Mycobacterium leprae* 35 kDa protein. *Infect Immun* **66**, 2684–2690.

Triccas, J. A., Berthet, F.-X., Pelicic, V. & Gicquel, B. (1999). Use of fluorescence induction and sucrose counterselection to identify *Mycobacterium tuberculosis* genes expressed within host-cells. *Microbiology* **145**, 2923–2930.

Via, L. E., Curcic, R., Mudd, M. H., Dhandayuthapani, S., Ulmer, R. J. & Deretic, V. (1996). Elements of signal transduction in *Mycobacterium tuberculosis*: *in vitro* phosphorylation and *in vivo* expression of the response regulator MtrA. *J Bacteriol* **178**, 3314–3321.

Received 11 September 2000; revised 12 December 2000; accepted 2 February 2001.