

cDNA–RNA subtractive hybridization reveals increased expression of mycocerosic acid synthase in intracellular *Mycobacterium bovis* BCG

Ming-Shi Li,^{1†} Irene M. Monahan,¹ Simon J. Waddell,¹ Joseph A. Mangan,¹ Steve L. Martin,² Martin J. Everett² and Philip D. Butcher¹

Author for correspondence: Philip D. Butcher. Tel: +44 20 8725 5721. Fax: +44 20 8672 0234.
e-mail: butcherp@sghms.ac.uk

¹ Department of Medical Microbiology, St George's Hospital Medical School, University of London, Cranmer Terrace, London SW17 0RE, UK

² Glaxo Wellcome Research and Development, Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK

Identifying genes that are differentially expressed by *Mycobacterium bovis* BCG after phagocytosis by macrophages will facilitate the understanding of the molecular mechanisms of host cell–intracellular pathogen interactions. To identify such genes a cDNA–total RNA subtractive hybridization strategy has been used that circumvents the problems both of limited availability of bacterial RNA from models of infection and the high rRNA backgrounds in total bacterial RNA. The subtraction products were used to screen a high-density gridded *Mycobacterium tuberculosis* genomic library. Sequence data were obtained from 19 differential clones, five of which contained overlapping sequences for the gene encoding mycocerosic acid synthase (*mas*). *Mas* is an enzyme involved in the synthesis of multi-methylated long-chain fatty acids that are part of phthiocerol dimycocerosate, a major component of the complex mycobacterial cell wall. Northern blotting and primer extension data confirmed up-regulation of *mas* in intracellular mycobacteria and also revealed a putative extended –10 promoter structure and a long untranslated upstream region 5' of the *mas* transcripts, containing predicted double-stranded structures. Furthermore, clones containing overlapping sequences for *furB*, *groEL-2*, *rplE* and *fadD28* were identified and the up-regulation of these genes was confirmed by Northern blot analysis. The cDNA–RNA subtractive hybridization enrichment and high density gridded library screening, combined with selective extraction of bacterial mRNA represents a valuable approach to the identification of genes expressed during intra-macrophage residence for bacteria such as *M. bovis* BCG and the pathogenic mycobacterium, *M. tuberculosis*.

Keywords: high-density gridded genomic library, macrophage, mycobacterial mRNA microarray

INTRODUCTION

Tuberculosis, caused by the bacterium *Mycobacterium tuberculosis*, is an important cause of global morbidity and mortality (Dolin *et al.*, 1994) with 1·7 billion people

.....
† **Present address:** Department of Paediatrics, Imperial College School of Medicine, St Mary's Hospital, Norfolk Place, London W2 1PG, UK.

Abbreviation: MBN, mung-bean nuclease.

infected with dormant bacilli, 10 million new cases of active disease and 3 million deaths per year (Raviglione *et al.*, 1995). Pathogenicity of *M. tuberculosis* is thought to be a multifactorial process with both pathogen and host-response effector molecules contributing to the process of infection leading either to overt disease or to the control of infection and long-term persistence or dormancy (Rook & Bloom, 1994). It is increasingly recognized that bacterial virulence constitutes the correct temporal and spatial regulation of many genes that

may be necessary for a particular phase in infection in response to specific environmental cues (Mekalanos, 1992). The slow-growing mycobacteria, *M. tuberculosis*, *Mycobacterium bovis* (including *M. bovis* BCG) and *Mycobacterium leprae* (the causative agent of leprosy), are intracellular pathogens that can survive and grow within host macrophages. Little is known about this process at a genetic level, but the availability of the whole genome sequence of *M. tuberculosis* (Cole *et al.*, 1998) provides a powerful resource to explore the molecular basis of pathogenicity in *M. tuberculosis*. Comparative genome analysis using whole-genome *M. tuberculosis* microarrays have revealed multiple gene deletions in *M. bovis* compared with *M. tuberculosis*, with a disproportionate loss of regulatory genes in *M. bovis* BCG (Behr *et al.*, 1999). Thus *M. bovis* BCG, the vaccine strain for human tuberculosis, may be considered a regulatory mutant (Behr *et al.*, 1999) in which alterations in gene expression may account for its attenuation. How *M. bovis* BCG survives in and interacts with the host macrophage without causing disease whilst affording immunoprotection needs to be elucidated. In this study we used *M. bovis* BCG both as a model of host-pathogen interactions for the pathogenic mycobacteria as well as to define the patterns of gene expression that underpin the intracellular survival, attenuation and immunogenicity of BCG.

In an attempt to sample a large portion of the genome for differentially expressed genes we have modified a subtractive hybridization system (Hubank & Schatz, 1994) based on cDNA (tester)-RNA (driver) hybridization and selective PCR amplification, and linked this with screening of a high-density array of an *M. tuberculosis* genomic library with the subtracted products. We have identified a subset of genes as being differentially expressed, five of which were confirmed by Northern blotting as being up-regulated by *M. bovis* BCG while inside macrophages.

METHODS

Bacterial culture, macrophage infection and RNA extraction.

THP-1 cells stimulated with 20 nM phorbol 12-myristate-13-acetate (PMA; 12 ng ml⁻¹) for 24 h were infected with *M. bovis* BCG Pasteur strain at an m.o.i. of 10 and allowed to phagocytose for 24 h, as detailed by Monahan *et al.* (2001). Intracellular bacilli were recovered at the 24 h point post-infection by centrifugation after selective lysis of THP-1 cells using GTC solution (4 M guanidinium thiocyanate, 0.5% sodium *N*-lauroylsarcosine, 25 mM sodium citrate, pH 7.0, and 0.1 M 2-mercaptoethanol). Detailed protocols are published by Butcher *et al.* (1998). Total RNA was extracted from the recovered bacilli as previously described (Mangan *et al.*, 1997).

Preparation of tester and driver for subtractive hybridization.

The tester consisted of linker-ligated cDNA synthesized from total RNA extracted according to Mangan *et al.* (1997) from *M. bovis* BCG phagocytosed by THP-1 cells 24 h after infection (see above). The cDNA was synthesized using Copy Kit (Invitrogen) and random primers, according to the manufacturer's instructions, from 1 µg of total RNA. The molecular size of the cDNA products was confirmed by

including 10 µCi [α -³²P]dCTP (Amersham) in the cDNA synthesis reaction, alkaline agarose gel electrophoresis of the cDNA products and autoradiography (Sambrook *et al.*, 1989). The resulting cDNA was digested with *DpnII* (New England Biolabs) and ligated to annealed oligonucleotides P12 (5'-GATCCGTTTCATG-3') and P24 (5'-ACCGACGTCGACTATCCATGAACG-3') using concentrated T4 DNA ligase (New England Biolabs). The ligation mixture was purified using the QIAquick PCR Purification Kit (Qiagen) to remove excess oligonucleotides. The linker-ligated cDNA was initially incubated with *Taq* DNA Polymerase (Promega) at 68 °C for 5 min and then amplified by PCR (1 min at 95 °C, 2 min at 68 °C and 3 min at 72 °C) for 25 cycles with the same *Taq* Polymerase and the P24 primer. Hot start PCR was applied to all of the PCR amplifications in this study. The driver was total RNA (prepared by the same method as described above) from *M. bovis* BCG cultured in Dubos broth (Difco) for 10 d at 37 °C.

Subtractive hybridization. Prior to subtraction enrichment, mRNA accounted for only 2% of the tester; therefore, large amounts of tester cDNA (5 µg) were included in the hybridization to ensure representation of rare mRNA species. Thus, 5 µg PCR-amplified tester cDNA and 50 µg driver (total) RNA were mixed, extracted with Tris/HCl-saturated phenol/chloroform/isoamyl alcohol (25:24:1), pH 8.0, followed by chloroform/isoamyl alcohol (24:1). The resulting nucleic acid mix was dried using a Spin-Vacuum-dryer and dissolved in 8 µl 62.5 mM *N*-2-hydroxyethylpiperazine-*N'*-3-propanesulfonic acid (EPPS; Sigma), pH 7.6, 62.5% formamide, 12.5% dextran sulphate and overlaid with mineral oil. The mixture was heated at 98 °C for 2 min followed by the addition of 2 µl 5 M NaCl (giving a final concentration of 1 M) through the oil layer and incubated at 42 °C for 20 h. The hybridization mix was then diluted with 110 µl H₂O, followed by phenol/chloroform and chloroform extraction. The aqueous phase was heated at 65 °C for 10 min and 12 µl of the resulting mixture was subjected to a 20 µl mung bean nuclease (MBN) digestion in 1 × digestion buffer II (New England Biolabs) supplemented with 2 mM ZnSO₄ and 20 units MBN (New England Biolabs). After incubation at 30 °C for 30 min the MBN digestion was stopped by the addition of 53.2 µl 50 mM Tris/HCl, pH 9.0, and heating at 98 °C for 5 min. The resulting products were amplified in a 100 µl PCR reaction with P24 primer (1.2 µM) and 5 units *Taq* Polymerase (Promega) for 25 cycles (1 min at 95 °C, 2 min at 68 °C and 3 min at 72 °C).

Analysis of subtraction products. To analyse the subtraction results the PCR products were purified using the QIAquick PCR Purification Kit and nucleic acid concentrations were determined by A₂₆₀ readings. Equal amounts of each sample obtained from PCR amplification of linker-ligated cDNA before and after subtractions were separated on 1.0% agarose gels and visualized by ethidium bromide staining. DNA samples were transferred onto Hybond-N (Amersham) nylon membranes and analysed by Southern blotting (Sambrook *et al.*, 1989) using radiolabelled rDNA sequence generated by PCR amplification of the complete 16S-23S-5S rRNA operon from *M. tuberculosis* H37Rv genomic DNA. The resulting membranes were exposed to X-ray films.

Northern blot analysis. The concentration of RNA samples was determined by A₂₆₀ readings and confirmed by rehybridizing Northern blots to ³²P-labelled rRNA sequences. Equal amounts of RNA samples (5 µg) were separated on glyoxal gels (Sambrook *et al.*, 1989), transferred onto Hybond-N (Amersham) nylon membranes and hybridized with radiolabelled gene-specific probes. High stringency washing at

Table 1. Plasmid clones selected for sequencing and their overlapping regions on the *M. tuberculosis* genome

Genes selected for Northern blot analysis are shown in bold type. The Genbank accession number for the sequence is AL123456.

Plasmid clone	Overlapping region on the genome	Name(s) and Rv numbers of the gene(s) in the region
41L2	3 281 126–3 286 589	<i>mas</i> (Rv2940c), <i>fadD28</i> (Rv2941), <i>mmpL7</i> (Rv2942)
17H23	3 278 947–3 284 243	<i>mas</i> , <i>fadD28</i>
41M18	3 281 010–3 284 449	<i>mas</i> , <i>fadD28</i>
46H23	3 278 882–3 282 232	<i>mas</i>
22F7	3 280 629–3 284 479	<i>mas</i> , <i>fadD28</i>
12P7	812 163–815 938	<i>rplE</i> , <i>rpsN</i> , <i>rpsH</i> , <i>rplF</i> , <i>rplR</i> , <i>rpsE</i> , <i>rpmD</i> , <i>rplO</i> , <i>sppA</i> (Rv0716–Rv0724)
16M17	809 500–815 231	Rv0712, Rv0713, <i>rplN</i> , <i>rplX</i> , <i>rplE</i> , <i>rpsN</i> , <i>rpsH</i> , <i>rplF</i> , <i>rplR</i> , <i>rpsE</i> , <i>rpmD</i> (Rv0714–Rv0722)
36B8	810 376–815 341	Rv0713, <i>rplN</i> , <i>rplX</i> , <i>rplE</i> , <i>rpsN</i> , <i>rpsH</i> , <i>rplF</i> , <i>rplR</i> , <i>rpsE</i> , <i>rpmD</i> , <i>rplO</i> (Rv0714–Rv0723)
13E7	527 076–530 474	<i>moeA3</i> (Rv0438c), Rv0439c, <i>groEL2</i> (Rv0440), Rv0441c
29F13	526 134–530 053	<i>moeA3</i> (Rv0438c), Rv0439c, <i>groEL2</i> (Rv0440)
34N24	2 636 899–2 643 486	PPE family (Rv2356c), <i>glyS</i> (Rv2357c), Rv2358, <i>furB</i> (Rv2359), Rv2360c, Rv2361c
32I21	3 600 045–3 605 145	Rv3224, Rv3225c, Rv3226c, <i>aroA</i> (Rv3228), Rv3228
33H5	2 359 051–2 364 094	Rv2100, <i>helZ</i> (Rv2101), Rv2102
7M1	1 454 289–1 457 670	<i>rho</i> , <i>rpmE</i> , <i>prfA</i> , <i>hemK</i> (Rv1297–Rv1300), Rv1301
7O5	1 354 726–1 358 577	Rv1212c, <i>glcC</i> (Rv1213), PE family (Rv1214c), Rv1215c
12I13	1 022 252–1 027 295	<i>betP</i> (Rv0917), Rv0918, Rv0919, Rv0920c, Rv0921
38K15	2 540 16–2 59 468	<i>nadR</i> , Rv0213c, <i>fadD4</i> , <i>fadE3</i> (Rv0212c–Rv0215c)
42G7	1 630 181–1 637 063	PE_PGSR family (Rv0212c–Rv0216)
48H9	3 238 552–3 241 550	<i>Fpg</i> (Rv2924c), <i>rnc</i> (Rv2925c), Rv2926c

65 °C and 0.1 × SSC was performed for 45 min to minimize any cross-hybridization with homologous sequences. The membranes were exposed to a phosphorimage screen and analysed using a Storm-840 phosphorimage scanner and ImageQuant software (Molecular Dynamics). Intensities from the whole lanes were quantified and equal areas between lanes were used. For gel G (see Fig. 5) the intensities of the two bands were quantified separately. The probes were generated by PCR with *M. tuberculosis* genomic DNA template and the primer sequences (5'–3') were: GATTTCGGGCAAAGATAAAGG and CATGTTGACGACGACTTTTCG for *rplE*, TC-GAGACCAAGGAGCAGATT and CTCACGGTCGTAGT-CGGAGT for *groEL-2*, GGGATGTCCAGAAGTCCGTT and GAGCCTACGAGGTGGTCAA for *moaA3*, ATCC-CCGCCATCAATTATCG and TTCTTCGTGCCGCTAG-TTC for *fadD28*, CGGACACGATCATTCTCGAACC and TCGGGGAGCACTACGACGATTA for *cysK*, and GATT-TCGGGCAAAGATAAAGG and CATGTTGACGACGAC-TTTTCG for *furB*.

The mycocerosic acid synthase gene (*mas*) probe was generated by PCR using T3 and T7 primers (Stratagene) and the clone 46H23 (see Table 1 and Fig. 4) as template.

Primer extension of *mas*. RNA samples were treated with RNase-free DNaseI. The primer, 5'-ATTCCATACTCTT-CCAACCTCGTCT-3' (see Fig. 6b), which showed low homology across the genome by a BLASTN search and was therefore considered to be gene-specific, was end-labelled with ³²P using Ready-To-Go T4 Polynucleotide Kinase Reaction Kit (Pharmacia) and [³²P]ATP (Amersham). The radio-labelled primer was mixed with the RNA templates and heated at 70 °C for 10 min and then quickly chilled on ice.

Each primer extension reaction was performed in a solution of 50 mM Tris/HCl, pH 8.3, 3 mM MgCl₂, 75 mM KCl, 10 mM DTT, 0.5 mM dNTPs, 35 units RNAGuard (Pharmacia) and 200 units SuperScriptII (Life Technologies) at 42 °C for 1 h. The resulting products were electrophoresed on a 6% polyacrylamide/8 M urea gel. The gels were dried and exposed to Phosphorimage screens (Molecular Dynamics) or X-ray films. Each primer extension reaction contained 3 µg total RNA and 5 pmol radiolabelled primer and was loaded into one sample well. Experimental controls with RNA samples subjected to RNase digestion prior to the primer extension reaction did not result in any obvious bands on the autoradiography.

The DNA ladders for the analysis of primer extension products were generated by the dideoxynucleotide chain-termination method (Sanger *et al.*, 1977) by using the Sequenase Version II DNA Sequencing Kit (Amersham) and [³⁵S]dATP (Amersham). The sequencing templates were recombinant plasmids of clone 41M18 (see Table 1 and Fig. 4) and the primer was the same as used in the primer extension analysis.

Hybridization of gridded mycobacterial genomic DNA library filters. An *M. tuberculosis* genomic library was prepared by *Sau3A* partial digestion of strain H37Rv genomic DNA, ligation of the size-fractionated fragments (3–8 kb) into the *Bam*HI site of pBluescript IISK (Stratagene) and transformation of the recombinant plasmid into *Escherichia coli* XL-1 Blue (Stratagene). Recombinant clones, identified by blue/white screening on X-Gal medium, were picked into 384-well microtitre plates containing freezing medium using the Q Bot robot (Genetix). The library plates were incubated overnight,

replicated and stored at -80°C . Restriction analysis of 23 white clones showed 19 had inserts (82%) with a size range of 2.7–6.2 kb. One copy of the library was gridded onto 22×22 cm Hybond-N nylon membranes (Amersham) again using the Q Bot robot. Each membrane contained 18432 clones representing the entire genomic library (sufficient for 99.9% genome coverage) and each clone was duplicated in a pre-designed pattern within a 4×4 array block to allow easy identification. The inoculated membranes were incubated on LB agar (containing ampicillin at $100 \mu\text{g ml}^{-1}$) overnight, allowing each clone to grow into a microcolony. Plasmid DNA from the microcolonies was released and denatured by standard methods and fixed to the membrane by UV irradiation (Sambrook *et al.*, 1989). The grids were tested with two test-gene probes and positive clones were identified. Representation of this gridded library has been further confirmed by the subsequent use of these grids by ten different research laboratories, all of which have identified appropriate clones of interest. The membranes were hybridized with radiolabelled cDNA (tester) or subtracted products (100 ng) in 15 ml solution containing 2% dextran sulphate, $6 \times \text{SSC}$, $5 \times$ Denhardt's solution, 0.5% SDS, 50% formamide and $100 \mu\text{g}$ denatured salmon sperm DNA ml^{-1} at 42°C for 60 h in a hybridization oven (Hybaid) with two sheets of nylon net to sandwich the membrane. Stringent washing was carried out in a 2 l solution containing 0.5% SDS and $0.1 \times \text{SSC}$ at 65°C for 1 h. The results were quantitatively imaged by phosphor-imaging and analysed by ImageQuant software (Molecular Dynamics).

Radioactive labelling of nucleic acid probes. Nucleic acid probes were labelled by random priming using the Rediprime Kit (Amersham) and $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ (Amersham) and purified using MicroSpin columns (Pharmacia).

Sequence analysis of selected clones. The DNA sequence was obtained using a conventional dye-terminator cycle sequencing protocol (Perkin-Elmer) and an ABI373 or ABI377 automated DNA sequencer at GlaxoWellcome (Stevenage). For each clone, the start and end sequences of the inserts were generated using T3 and T7 primers, which anneal to either side of the MCS of pBluescript II SK (Stratagene). The BLAST programme was used to search for matches to *M. tuberculosis* sequences deposited in the EMBL/GenBank databases, allowing the size and content of each insert to be extrapolated. The genes that resulted from the homology searches are listed in Table 1.

RESULTS

Subtractive hybridization design

The subtractive hybridization strategy outlined in Fig. 1 was designed to enrich for mRNA sequences differentially expressed by intracellular mycobacteria. In the first step, tester cDNA was prepared by synthesis of ds-cDNA from intracellular *M. bovis* BCG total RNA, extracted free from macrophage RNA. This was followed by restriction digestion of the cDNA and ligation of linkers to the fragments to provide priming sites for further PCR amplification. The second step involved the hybridization of linker-ligated cDNA (tester) with total RNA from *in vitro*-cultured *M. bovis* BCG (driver) which was followed by MBN treatment coupled to PCR; this removed single-stranded tester and driver sequences and tester/driver hybrids, thus allowing

specific amplification of tester/tester hybrids. The resulting PCR products were then used to hybridize high-density gridded genomic DNA libraries of *M. tuberculosis* to identify the genes corresponding to the enriched mRNA sequences in the subtracted products.

Two major considerations were necessary at the outset of this study. First, difficulties in purification of mRNA from total RNA, due to the lack of poly(A) tails in prokaryotic mRNA, results in a high abundance of rRNA sequences present in both tester and driver. Second, the amount of RNA available from macrophage-recovered mycobacteria was very limited: about 1–2 μg total mycobacterial RNA was recovered from 10^8 THP-1 cells under the infection conditions used. In the subtraction strategy (Fig. 1) the use of total RNA as driver allowed both the enrichment of differences and removal of ribosomal sequences at the same time. It also provided better hybridization efficiency than a cDNA–cDNA hybridization since reannealing of double-stranded driver cDNA was avoided. The use of linker-ligated cDNA tester and MBN-coupled selective PCR, adapted and modified from the representational difference analysis (RDA) technology of Hubank & Schatz (1994), accommodated the possibility of using small amounts (1 μg) of tester RNA as starting material. In this study two rounds of subtraction were carried out.

Stepwise reduction of rRNA sequences in the tester by multiple rounds of cDNA–RNA subtractions

It is known that rRNA sequences are the most abundant sequences common to tester cDNA and driver total RNA. To measure the stepwise reduction of rRNA-derived sequences in the tester cDNA upon multiple rounds of cDNA–RNA subtraction, equal amounts of tester cDNA, before subtraction (Fig. 2a, lane 1) and after first- and second-round subtraction (Fig. 2a, lanes 2 and 3, respectively) were analysed by Southern blotting and hybridization with a whole-operon rDNA probe. The autoradiograph (Fig. 2b) shows a reduction of rRNA-derived sequences in the first-round subtraction products (lane 2) compared to unsubtracted tester cDNA (lane 1), with further reduction in the second-round subtraction sample (lane 3). Quantitative analysis revealed a reduction of up to 70% of rRNA-derived sequences in the tester cDNA after two rounds of subtraction (mean 57.3%; range 50–70%; $n = 3$). Equal amounts of cDNA sample (as determined by A_{260}) were added to each lane, confirmed by ethidium bromide staining (Fig. 2a). The size distribution of *DpnII*-digested, linker-ligated, PCR-amplified cDNA up to ~ 700 bp correlates well with the predicted restriction fragment sizes for the ribosomal operon, which would be 4–740 bp with 79 cut sites. The origin and distribution of the major bands on the ethidium-stained gel within the randomly primed cDNA smear (Fig. 2a) was not determined. However, the banding pattern, which changed during rounds of subtraction, was probably only in part derived from ribosomal cDNA sequences,

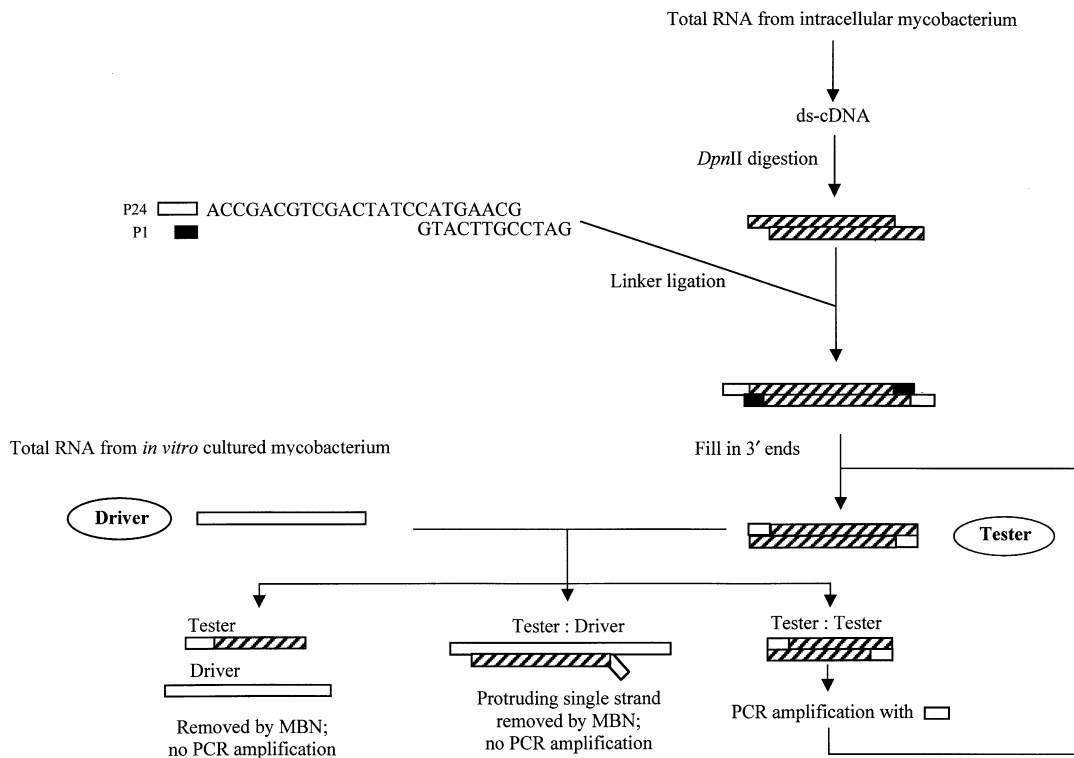


Fig. 1. Schematic diagram of the linker-ligated cDNA-RNA subtractive hybridization procedures.

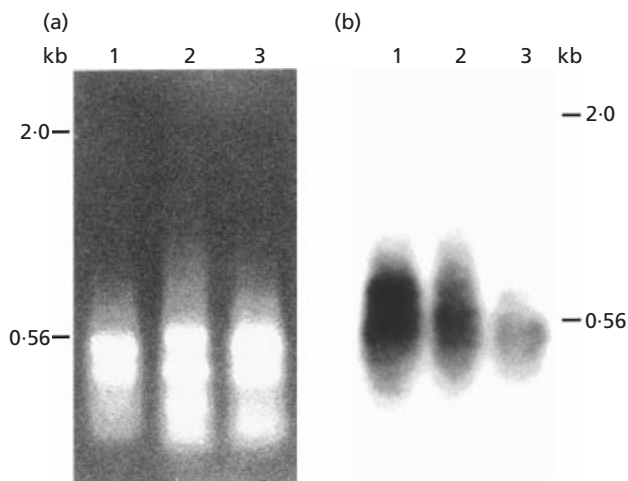


Fig. 2. Southern blot analysis of the tester cDNA and subtractive hybridization products to show stepwise reduction of rDNA sequences during subtractions. (a) Equal amounts of *DpnII*-digested, linker-ligated, PCR-amplified cDNA (tester) before subtraction (lane 1) and after first (lane 2) and second (lane 3) rounds of subtraction were separated by agarose gel electrophoresis and stained with ethidium bromide. (b) Autoradiograph of a Southern blot of (a), hybridized to a ^{32}P -labelled mycobacterial rRNA probe. DNA molecular sizes are indicated.

as the stained bands did not correspond completely with the bands revealed by rDNA-probed Southern blots (Fig. 2b).

Enrichment of specific gene sequences during subtraction: hybridization analysis of an *M. tuberculosis* high-density gridded genomic library

Subtracted products were radiolabelled with ^{32}P to high specific activity ($> 10^9$ d.p.m. μg^{-1}) and used to screen a high-density *M. tuberculosis* genomic DNA library consisting of ~ 18000 clones on 22×22 cm nylon membranes. Each clone was duplicated within a 4×4 array so as to give a pre-designed pattern for easy location of the clones and to eliminate false positives. Fig. 3 shows representative phosphorimages after hybridization with macrophage-induced mycobacterial double-stranded, PCR-amplified cDNA (tester) before (a) and after (b) two rounds of subtractive hybridization.

In Fig. 3(a) the image shows a range of spot intensities which in part represents the relative abundance of cDNAs in the unsubtracted tester probe. The clones with low hybridization intensities, at or a little above background, probably correspond to the low abundance mRNA pool as well as representing clones either with no insert ($\sim 17\%$; see Methods) or that failed to grow well. A relatively high background from the vector DNA on the clone arrays precluded a quantitative analysis of the range of transcript abundances in complex cDNA probes. The two pairs of spots with very strong intensities (boxed) were determined by sequencing to correspond to rRNA sequence-containing clones. It is known that slow-growing mycobacteria have only one copy of the rRNA operon per genome (Bercovier *et al.*, 1986; Cole *et al.*, 1998), so the number of such clones

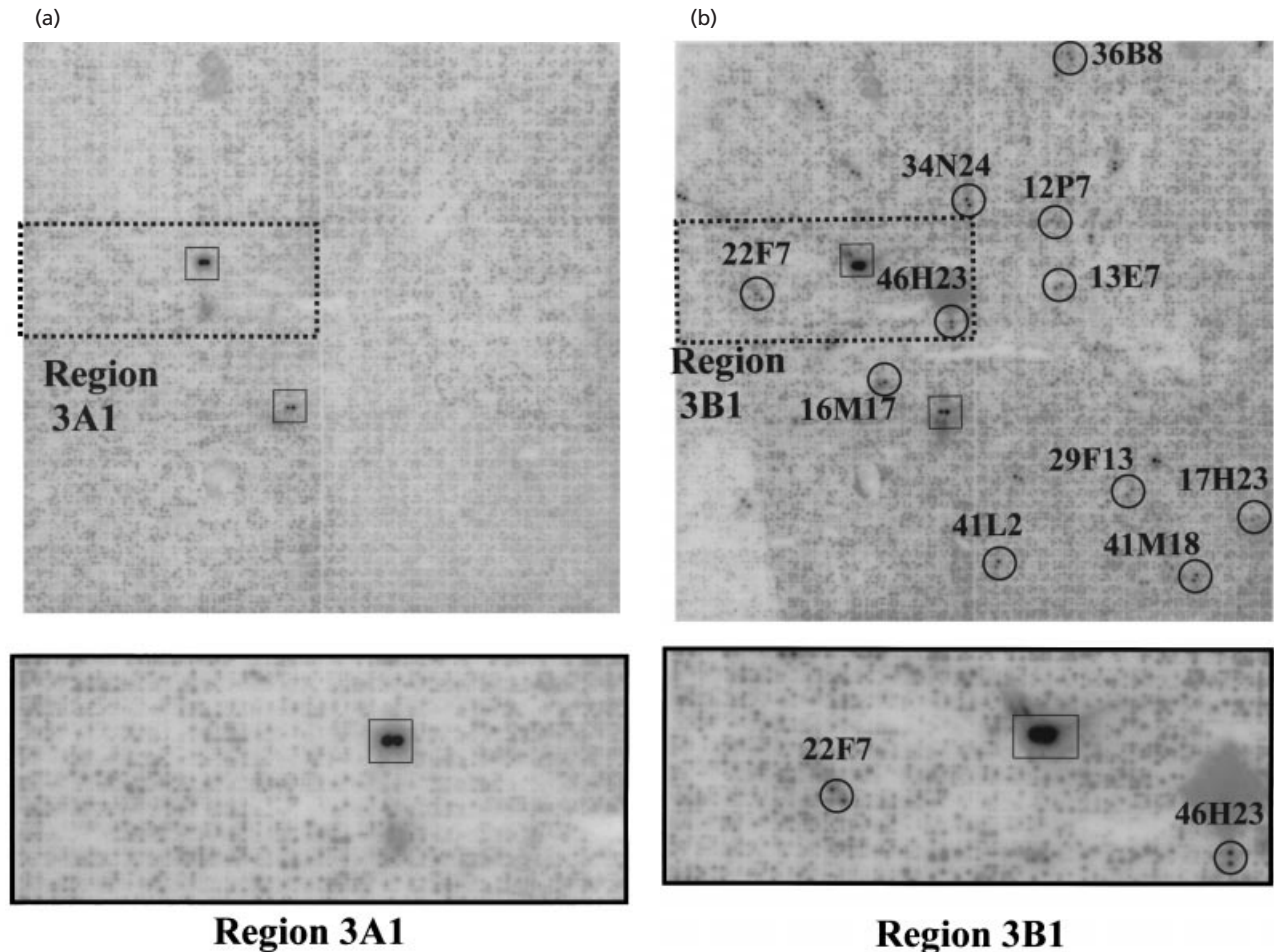


Fig. 3. Hybridization of high-density gridded *M. tuberculosis* genomic library filters with mycobacterial cDNA probes before and after subtractive hybridization showing enrichment of differentially expressed sequences. The probes used for hybridizing the membranes are: (a) ds-cDNA, linker-ligated and PCR-amplified from macrophage-phagocytosed *M. bovis* BCG before subtractive hybridization; (b) cDNA as in (a) after two rounds of subtractive hybridization. Parts of the image from (a) and (b), namely Region 3A1 and Region 3B1, respectively, are shown as enlarged images below the corresponding panels. Clones selected for sequencing and confirmed as differentially expressed by Northern blot analysis (gene name in bold in Table 1) are indicated by circles and clone number. The intense doublets of rRNA sequence containing clones are indicated (boxes).

observed was predictably low. In Fig. 3(b), where the subtraction products after two rounds of enrichment were used as probes, the hybridization profile was altered compared to Fig. 3(a) in that the number and intensity of distinct spots were increased and low intensity spots were reduced. The results indicated that the differentially expressed gene sequences in the tester cDNA were enriched and common sequences (in both tester and driver) were reduced or removed during subtractive hybridization. The increased contrast between spot intensities for the two probes (unsubtracted and subtracted tester cDNA), as a result of the subtractive process, allowed the identification of individual high intensity spots and confirmation of correct pairs by eye. To select for further characterization those clones showing the greatest enrichment, spot intensities from Fig. 3(b) were quantified using ImageQuant software (Molecular Dynamics). A mean intensity of all spots was

taken as the background and each spot was expressed as a value above that background. Ninety-five pairs of spots were identified with a greater than 20-fold intensity above the mean spot intensity, with a maximum of 50-fold (Fig. 3b). The hybridizing clones from Fig. 3(b) were ranked according to their intensities and the top 33 clones were selected for sequence analysis. Of the 33 clones, 24 were recovered from frozen stocks and 19 were sequenced.

Sequence identification and confirmation of gene up-regulation

Our analysis focused on the sequence data obtained from the 19 clones selected from our mycobacterial genomic library. The results of the homology searches for each of the sequenced clones against the *M. tuberculosis* genome sequence are listed in Table 1.

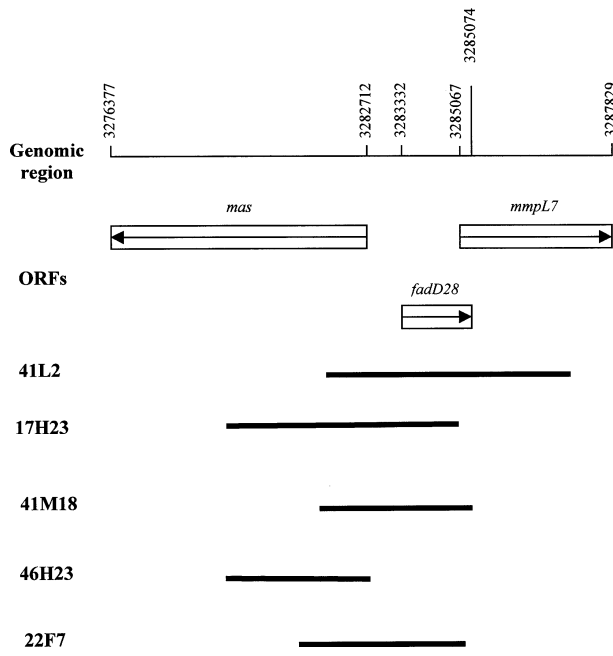


Fig. 4. Schematic representation of the overlapping nature of the genomic fragments of the five clones containing *mas* and their sequence homology relation to the corresponding mycobacterial genomic region. The names of the clones obtained in this study are indicated on the left. ORFs are annotated according to Cole *et al.* (1998).

Because of the limited availability of intracellular mycobacterial RNA, only a few genes (Table 1, gene names in bold) were selected for Northern blot analysis (requiring > 5 µg) to confirm their increased expression. This selection was based on the highest number of clones containing overlapping sequences for a particular gene.

The largest group of overlapping clones consisted of five clones (41L2, 17H23, 41M18, 46H23 and 22F7) containing varying lengths of the *mas* gene. The overlapping relationship of the five clones is shown schematically in Fig. 4. As the most overlapping ORF in the group, *mas* was selected for further investigation, initially by Northern blot analysis. The insert of clone 46H23 consisted of a 3.3 kb internal sequence of *mas* and was used as probe for the Northern blot analysis of RNA from *in vitro*-cultured and intracellular *M. bovis* BCG. Similar hybridization patterns were obtained using an *M. tuberculosis* genomic DNA PCR generated probe. The results were visualized by phosphorimaging and quantitatively analysed by ImageQuant software. Three independent Northern blot analyses confirmed a mean 2.7-fold increase (2.5–3.0) of the *mas* message in intracellular *M. bovis* BCG 24 h post-phagocytosis (Fig. 5b, lane 1) compared to *in vitro*-cultured *M. bovis* BCG (Fig. 5b, lane 2). Equal loading of the RNA samples was adjusted by measuring A_{260} and confirmed by rehybridizing and quantifying the same blot with an rRNA probe (Fig. 5a). The size distribution of the hybridizing *mas* sequences in Fig. 5(b) appeared as a

smear with a mean size of ~ 1 kb. This was much smaller than the predicted size of *mas* mRNA based on the genome sequence and was seen on each of three separate gels from two individual RNA preparations. Since intact rRNA bands are clearly visible in Fig. 5(a), and probing the same RNA with other gene-specific probes resulted in sharp bands of the correct size (see below and Fig. 5), the smears observed for *mas* appeared not to be the result of general degradation of the RNA samples during preparation or storage, but rather may represent some specific property of *mas* mRNA stability (see Discussion). A second gene in this region, *fadD28*, which overlapped with four clones (Fig. 4), was also studied by Northern blotting and the expression of *fadD28* demonstrated a 2.5-fold increase in intracellular *M. bovis* BCG compared with *in vitro*-cultured *M. bovis* BCG (Fig. 5d).

A second group of clones was composed of three clones, 12P7, 16M17 and 36B8 (Table 1). The sequences of these clones overlapped in a genomic region encoding ribosomal proteins equivalent to the *spc* operon in *E. coli* (Lindahl & Zengel, 1986) and *Bacillus subtilis* (Henkin *et al.*, 1989). To confirm differential expression in this region by Northern blot analysis, we selected the *rplE* gene as a probe on the basis of size and gene location within the proposed operon (see Discussion). Northern blotting with a 420 bp PCR product probe spanning most of the *rplE* ORF sequence (Fig. 5h) showed two discrete hybridization bands, designated *rplE*-(H) (high) for the higher molecular mass band and *rplE*-(L) (low) for the lower molecular mass band. After adjustment for unequal loading (lane2/lane1 = twofold) of total RNA on this gel as measured by rDNA reprobing (Fig. 5g) and quantitative phosphorimaging, the intensities of *rplE*-(H) and *rplE*-(L) bands were quantified as 4.1- and 1.2-fold higher in the intracellular (Fig. 5h, lane 1) than in the *in vitro*-cultured (Fig. 5h, lane 2) sample. The result demonstrated increased expression of the transcripts containing *rplE* by intracellular mycobacteria. The multiple-banded nature of the Northern blot analysis is unusual but consistent with studies in *E. coli* (see Discussion).

The sequence of clones 13E7 and 29F13 in Table 1 overlapped in the region of the genome containing *moaA3*, *Rv0439c* and *groEL-2*. Because of their potential biological interest, *moaA3* and *groEL-2* were selected for Northern blot analysis with PCR-generated, gene-specific probes. Confirmation of equal loading and Northern blot analysis were conducted in the same way as described above for *mas*. Fig. 5(c) which shows a single sharp band of the expected size for *groEL-2* mRNA in both RNA samples demonstrates a fourfold increase in intracellular *M. bovis* BCG compared to *in vitro*-grown bacteria (lanes 1 and 2, respectively). However, using the *moaA3* gene sequence probe, the same Northern blot did not show any visible RNA bands in either sample, suggesting either that *moaA3* is a low-abundance mRNA (not detectable by Northern hybridization) or that it is not expressed under these conditions. The *furB* gene corresponded to part of the

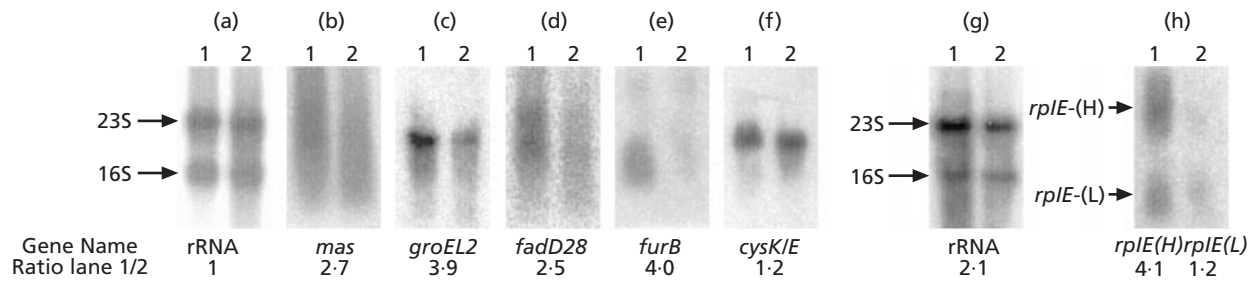


Fig. 5. Confirmation of intracellular differential expression of selected genes by Northern blot analysis of RNA (5 µg) from *in vitro*-cultured (lane 2) and macrophage-phagocytosed (lane 1) *M. bovis* BCG. Hybridization with gene-specific probes for *mas*, *groEL-2*, *fadD28*, *furB*, *cysK* and *rplE* on individual Northern blots are shown in panels (b), (c), (d), (e), (f) and (h), respectively. The RNA used was from two separate macrophage infection experiments and the Northern blots were repeated and confirmed three times for *mas* and twice for *fadD28*; only single experiments were performed for *groEL2*, *furB*, *cysK* and *rplE*. For each gel, equal loading of RNA samples was confirmed by stripping and rehybridizing with rRNA probes (a and g). Equal loading in gel panels (b), (c), (d), (e) and (f) was observed and is shown by a representative Northern rRNA analysis (a). Unequal loading in gel (h) is shown by rRNA probing of the same gel (g). The molecular mass of rRNA bands is indicated. At the bottom of the figure the intensities of hybridized bands in each Northern blot analysis are given, expressed as ratios of the relative intensities of lane 2/lane 1, as determined by quantitative phosphorimaging of a single hybridization (except for *mas*, where $n = 3$; range 2.3–3.0). Two discrete bands in (h) are indicated as *rplE*-(H) and *rplE*-(L). The band intensity ratios for (h) were readjusted according to the band intensity ratios for (g) (see text).

clone 34N24 sequence that was selected on the basis of spot intensity. A PCR-derived *furB* gene probe hybridized to a low-molecular-mass band in both RNA samples (Fig. 5e), quantified as a fourfold increase in the intracellular compared to the *in vitro*-cultured mycobacteria RNA sample. Thus, the five genes corresponding to overlapping clones in Table 1, selected on the basis of being the most enriched sequences by subtractive hybridization (highlighted in Fig. 3b), have been confirmed as up-regulated in intracellular mycobacteria by subsequent Northern blot analysis.

As further validation of the selection of differentially expressed genes, a clone that was identified as not containing differentially expressed sequences from the library screen with subtractively hybridized probes (i.e. not enriched during subtraction) was used as a control for Northern blot analysis with the same RNA samples as above. This clone contained three ORFs: *cysK*, *cysE* and Rv2336. Using a PCR-amplified *cysK* gene sequence as probe the Northern blot (Fig. 5f) showed a single band in both RNA samples of about 1.6 kb, correlating to a *cysK*–*cysE* coupled transcript, and the ratio of band intensities was almost equal (ratio = 1.2; see Fig. 5f).

Primer extension analysis of *mas* transcripts

Primer extension analysis of *mas* was performed on total RNA from *in vitro*-cultured and macrophage-phagocytosed *M. bovis* BCG. A radiolabelled oligonucleotide that was complementary to the mRNA sequence in a region around the translation start site (see Fig. 6b and Methods) was hybridized to the RNA samples and the primer was extended using reverse transcriptase. The resulting products were analysed on denaturing polyacrylamide gels. The resulting autoradiograph (Fig. 6a) showed that single bands were clearly visible in both samples corresponding to a 185 bp

extension product, as compared with a DNA sequence ladder generated using the same primer on a recombinant plasmid 41M18 (see Fig. 4) template containing the *mas* DNA sequence. The result shows that transcription is initiated at a purine, 178 bases upstream of the translation start codon ATG (Fig. 6b). However, this result could not exclude the possibility of multiple transcription start sites for *mas*, since at least two smaller-sized primer extension products were noticed in several independent experiments, but appeared at substantially lower levels than the major 185 bp product. The up-regulation of *mas*, as seen by Northern analysis, was confirmed by quantitative phosphorimaging of the gene-specific primer extension data. This showed that with equal input RNA, a two- to threefold increase in the intensity of the extension product was observed in the intracellular RNA sample compared to *in vitro*-grown bacteria. The comparability between Northern and primer extension data on the levels of *mas* up-regulation indicated that any potential contribution to *mas* intensities on Northern blots (Fig. 5b), that might have resulted from cross-hybridization of the gene probe with closely related homologues of *mas* (such as *pks2*), was minimal.

Scrutiny of the upstream region of transcription start site in genomic DNA (Fig. 6b) shows a putative –10 promoter box with a TGN motif present immediately upstream of the –10 region. Such structures, defined as extended –10 promoters, were initially identified in *E. coli* (Ponnambalam *et al.*, 1986) and more recently in mycobacteria (Bashyam & Tyagi, 1998). This region lacks a typical –35 consensus sequence. A putative ribosome-binding site, GAGGT, is located 10 nt upstream of the ATG start codon and a perfect inverted repeat was identified between –28 and –13 with respect to the translation start site (Fig. 6b). RNA secondary structure prediction of this 5' untranslated

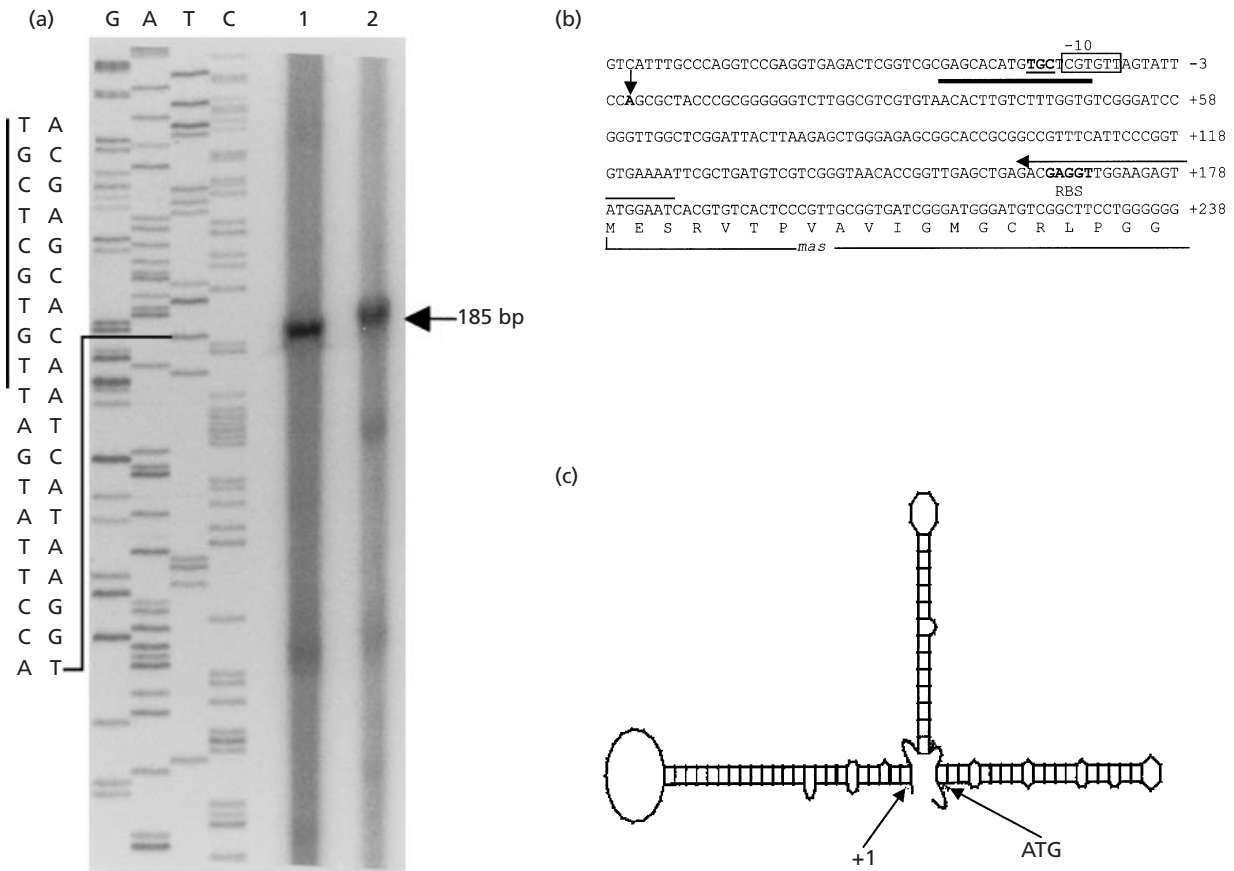


Fig. 6. Primer extension analysis and promoter region of *mas*. (a) Primer extension analysis for macrophage-phagocytosed (lane 1) and *in vitro*-cultured (lane 2) *M. bovis* BCG is shown. Extension products were sized at 185 bp using the DNA sequencing products (lanes marked G, A, T and C) generated on the *mas* gene containing recombinant DNA templates by the same primer as used in the primer extension analysis. The sequence up to -18 from transcription start site (A) is shown and the putative extended -10 region is indicated with a line on the right of the sense strand sequence. (b) Partial nucleotide sequence of the *mas* ORF and upstream region. A vertical arrow indicates a transcription start site. A horizontal arrow denotes the position of a reverse primer used in primer extension. A putative -10 box (boxed) and an extended TGN structure (in bold and underlined) are indicated. A direct invert repeat is also indicated by a wide bar underneath the sequence. Part of the *mas* coding region is indicated with the amino acid sequence underneath the corresponding codons. A putative ribosome-binding site (in bold) is indicated by RBS. Nucleotide positions with respect to the transcription start site as $+1$ are numbered on the right of the figure. (c) RNA secondary structure prediction for the *mas* upstream sequence. The sequence subjected to analysis spanned a 187 bp region from the transcription start site, $+1$, to 10 bp downstream of the translation start codon ATG. The structure prediction programme MFOLD used was based on Zuker's algorithm to determine the minimum free energy secondary structure (Zuker, 1989).

region revealed extensive hairpin-loop structures (Fig. 6c).

DISCUSSION

In this report we have used cDNA-RNA subtractive hybridization to enrich for differentially expressed mRNA sequences from complex RNA populations coupled with gridded-genomic DNA library screening to identify genes differentially expressed by *M. bovis* BCG at a 24 h time-point after phagocytosis by the THP-1 macrophage-like cell line. The library screening initially identified 95 clone pairs that appeared to be enriched after subtractive hybridization by > 20 -fold for intracellular bacteria, each of which was likely to

contain genes that were up-regulated by *M. bovis* BCG inside macrophages compared to growth *in vitro*. Sequence data were obtained from 19 clones and several groups of clones contained sequences that overlapped in the same genomic region. Subsequent Northern blot analysis of a limited number of genes from these sets of overlapping clones confirmed the differential expression of five genes: *mas*, *fadD28*, *furB*, *groEL* and *rplE*, thus validating the subtractive hybridization approach to selecting up-regulated genes. The disadvantages of our experimental strategy include: (i) the limited number of clones that can be realistically selected for further characterization by sequencing and validation by Northern blotting (which requires more total RNA than can be readily recovered from the macrophage infection

model), thus sampling only a relatively small portion of the genome for differentially expressed genes; (ii) multiple time points cannot be easily studied and genes that are differentially expressed transiently will be missed when studying a single time point, in this case 24 h post-phagocytosis; (iii) the non-quantitative nature of the library screens with unsubtracted cDNA, which did not show a large dynamic range of mRNA transcripts (as seen in Fig. 3a); and (iv) the potential redundancy in the genomic library, limiting a total genome-wide screen. Clearly, the future application of whole-genome microarray hybridization technology will overcome many of these limitations (Wilson *et al.*, 1999).

A range of techniques have been described to study differential gene expression in other bacteria *in vitro*, but very few have been applied to mycobacterial models of infection (for review see Clark-Curtiss, 1998). The protein profiles of intracellular *M. bovis* BCG (Monahan *et al.*, 2001), *M. avium* (Sturgill-Koszycki *et al.*, 1997) and *M. tuberculosis* (Alavi & Affronti, 1994; Lee & Horowitz, 1995) have previously been studied by two-dimensional PAGE and changes in protein synthesis have been revealed, but only a few macrophage-induced *M. tuberculosis* proteins have been specifically identified. Mycobacterial gene expression has also been studied at the mRNA level using RT-PCR for individual genes both *in vitro* (Patel *et al.*, 1991) and during macrophage interaction (Butcher *et al.*, 1998), as well as by RAP-PCR (Mangan & Butcher, 1998; Rivera-Marrero *et al.*, 1998).

Several technical considerations of our method are warranted. Our subtractive hybridization strategy differed slightly from previously reported methods applied to mycobacteria (Kinger & Tyagi, 1993; Plum & Clark-Curtiss, 1994; Robinson *et al.*, 1994; Utt *et al.*, 1995) in that separate pre-removal of rRNA sequences from either tester or driver was not performed. The combination of single-stranded RNA (driver) to double-stranded (tester) cDNA hybridization, together with MBN removal of common hybrids and PCR amplification to selectively enrich the differentially expressed sequences, resulted in a robust technique for small amounts of starting material (1 µg RNA) with significant removal of ribosomal sequences (up to 70%). Our hybridization conditions included a low incubation temperature at 42 °C in a buffer containing 50% formamide, as it has been demonstrated that RNA can be damaged by prolonged incubations at high temperatures, typically 68 °C for 20 h (Wu & Bird, 1994). PCR amplification of a pool of different sequences such as total cDNA may be considered to alter the sequence representation within the pool. A recent study proposed a method that maintains proportional representation of individual mRNA sequence obtained from *M. tuberculosis* exposed to isoniazid *in vitro* (Alland *et al.*, 1998). This was achieved by restricting the size of selected sequences within a 400–1500 bp range and amplification only of the captured DNA sequences. In our study, the subtracted PCR products were also limited within the range 200–1000 bp (see Fig. 2a). In

our protocol, however, PCR amplification was involved in cDNA generation and the subsequent rounds of subtractive hybridization. We cannot therefore exclude changes in representation of subpopulations of mRNA caused by multiple amplifications. Nevertheless, selective PCR coupled to subtractive hybridization significantly increased the proportional differences between the up-regulated genes and background genes, thus magnifying the hybridization intensities for differentially enriched sequences by library array screening. The validity of our approach is confirmed since it clearly selects for mRNAs that are differentially expressed in the tester population, as confirmed by Northern blot analysis. Another recent study described a positive cDNA selection approach (Graham & Clark-Curtiss, 1999) to enrich for differentially expressed mRNA sequences from intra-macrophage *M. tuberculosis* by employing a novel normalization strategy which was enhanced by multiple rounds of amplification. The resulting enriched cDNAs were used as probes to analyse Southern blots of selected gene fragments, including a polyketide synthase gene (*pks2*), but the differences were not validated by Northern blotting.

With only limited amounts of RNA available from intracellular mycobacteria, we were limited to only a few genes for further validation by Northern analysis. Consideration of these genes is instructive.

***mas* and *fad28*.** One gene, *mas*, had the most sequence hits (5 clones) amongst the 19 clones selected and Northern analysis confirmed its up-regulated expression. Mycocerosic acid synthase is a multifunctional enzyme that is involved in the synthesis of the branched and multi-methylated long-chain mycocerosic acids (Rainwater & Kolattukudy, 1985). Mycocerosic acids and their esterified derivatives, such as phenolphthiocerols, are found exclusively in the cell wall of pathogenic mycobacteria (Minnikin, 1982). The up-regulation of the *mas* gene in macrophage-phagocytosed mycobacteria has not been reported previously. This result is particularly significant since it has been suggested that a *mas* knockout mutant of *M. bovis* BCG has increased susceptibility to host defences (for review see Kolattukudy *et al.*, 1997). Similarly, mycobacterial cell wall thickening (Cunningham & Spreadbury, 1998) and changes in mycolic acid composition (Yuan *et al.*, 1998) have been identified under low oxygen conditions. We suggest that changes in cell-wall-associated lipids are induced by the intra-macrophage environments based on our observations of the up-regulation of *mas* and *fad28*.

The importance of the *fadD28-mas* locus for mycobacterial survival during murine infection has also been recently highlighted in a study using signature-tagged transposon mutagenesis (STM) technology (Camacho *et al.*, 1999; Cox *et al.*, 1999), in which *fadD28* and *mmpL7* were shown as important genes for replication of *M. tuberculosis* in mouse lungs. *FadD28* (a probable acyl CoA synthase; Cole *et al.*, 1998) was suggested to be involved in the transfer of mycocerosic acids onto

phthiocerol and phenolphthiocerol (Cox *et al.*, 1999; Fitzmaurice & Kolattukudy, 1998) and MmpL7 to be involved in mycoserosic-acid-containing lipid transport across the bacterial cell wall (Cox *et al.*, 1999). The report (Cox *et al.*, 1999) also suggested the transport of phthiocerol dimycoserolate (PDIM) lipids across the cell wall into the cytoplasm of the cell and proposed a virulence role for released lipids. That *mas* and *fadD28* are shown in our study to be up-regulated inside macrophages reinforces the hypothesis that mycoserosic-acid-containing lipids play important roles in host–pathogen interaction in addition to being structural components of the mycobacterial cell wall.

Primer extension analysis has revealed a long untranslated 5' region for *mas* mRNA. An RNA-folding computer program predicted double-stranded hairpin structures at the 5' end of *mas* mRNA (Fig. 6c). Northern blots showed smears of *mas* mRNA (but correct size transcripts for other mRNAs on the same blot) and the primer extension showed discrete bands in all samples, indicating that the predicted structure likely exists and stabilizes the 5' region of the mRNA. It is known that the half-life of an mRNA species can be determined by features near its 5' end since endonucleolytic cleavage at the 5' end plays a regulatory role in mRNA degradation (for review see Higgins, 1991). It may be possible that the 5' structure contributes to the unusual mRNA stability/turnover of *mas*, as seen on Northern blots (Fig. 5b). An extended –10 promoter structure was observed in the putative –10 region of the *mas* upstream sequence. In a systematic study of mycobacterial promoters, extended –10 promoters were identified structurally and their function was analysed by site-specific mutagenesis (Bashyam & Tyagi, 1998). They showed that sequence changes in the TGN motif reduced the transcription levels of the reporter gene controlled by the promoter. When the TGN structure was introduced into non-TGN promoters the transcription level of the reporter genes was increased. It is possible that because of the lack of a typical –35 sequence a TGN structure is employed to strengthen the *mas* promoter activity. More significantly, the extended –10 motif of *mas* could be used to enhance transcription initiation when the bacteria are exposed to environmental stresses (such as those inside macrophages), as it has been reported that in *E. coli* the extended –10 promoters facilitate the formation of RNA polymerase and –10 region complex under cold stress (Burns *et al.*, 1996).

***rplE*.** Multiple bands were observed in the Northern blot analysis using the *rplE* gene probe. An earlier study of the expression of a ribosomal protein operon, *spc*, in *E. coli* (Mattheakis & Nomura, 1988) also showed several discrete RNA bands ranging from 8.4 kb to a few hundred bp in size as detected by an *rplX* probe, the gene adjacent to *rplE*. The authors speculated that the 8.4 kb band resembled the full-length *spc/α* operon co-transcription product and the rest of the bands were processed products. The size of the *rplE*-(H) high molecular mass band observed in our Northern blot analysis (Fig. 5h) could only represent co-transcription

of the first four ORFs predicted as 1.4 kb in the *spc* operon (*rplN*, *rplX*, *rplE* and *rpsN*). The lower molecular mass band may possibly derive from products of processing, premature termination or even a transcript of the *rplE* gene alone. Inspection of the genomic organization of the *spc* operon in *E. coli*, *B. subtilis* and *M. tuberculosis* shows that the spacing sequence between the first four closely spaced ORFs (*rplN*, X, E and *rpsN*) and the immediate downstream ORF (*rpsH*) is 33, 31 and 163 bp, respectively, with the remaining ORFs all closely located. This increased spacing on the mycobacterial genome could indicate that the first four ORFs are independently transcribed. Ribosomal protein genes are generally considered as examples of housekeeping genes. It was a surprise to identify one of those genes as differentially expressed in mycobacteria. To date we have not found similar reports in a prokaryotic system. However, in eukaryotes it has been demonstrated that genes encoding ribosomal proteins, including L5 (*rplE*), are differentially regulated. Using *in situ* hybridization, it has been shown (Scholnick *et al.*, 1997) that in *Xenopus* the expression level of ribosomal protein genes for S22, L1 and L5 is differentially regulated in a developmental specific manner. More interestingly, genes for ribosomal protein L4 and L5 were found to be overexpressed in a doxorubicin-resistant human colon cancer cell line and the study indicated that the up-regulation of these genes is the response of the cancer cells to the drug pressure (Bertram *et al.*, 1998). Similarly, one could assume that the up-regulation of *rplE* (L5), and possibly the other three adjacent genes, in *M. bovis* BCG while inside macrophages represents an adaptive gene expression response involving a mechanism of translational control by these regulatory ribosome-associated proteins.

***GroEL-2*.** It is known that elevated expression of mycobacterial heat-shock proteins is induced not only by temperature changes but also by other environmental conditions, including intra-macrophage environments (Alavi & Affronti, 1994; Lee & Horowitz, 1995; Monahan *et al.*, 2001). Our study shows increased transcription of *groEL-2* by *M. bovis* BCG inside macrophages, thus confirming the protein expression data we have reported using the same macrophage infection model (Monahan *et al.*, 2001), and further verifies the subtractive hybridization enrichment strategy used here. Although the role of heat-shock proteins in pathogenesis is not fully understood, it is widely accepted that heat-shock proteins are crucial for the intracellular survival of pathogens.

***furB*.** The ferric uptake regulatory protein, encoded by the gene *fur*, is not only a regulator for genes involved in iron acquisition but also a global regulator for genes involved in other metabolic pathways, oxidative- and acid-stress responses, chemotaxis, bioluminescence and production of toxins and other virulence factors (for review see Escolar *et al.*, 1999). Knowledge about *fur* derives mostly from studies in *E. coli* and other Gram-negative bacteria and little is known about the gene in mycobacteria. Wong *et al.* (1999) observed, using two-

dimensional gels, that in mycobacteria Fur is increased under high-iron conditions *in vitro* and decreased in low-iron conditions, consistent with a negative repressor role for Fur. There are two homologues of *fur* in the *M. tuberculosis* genome: *furA* and *furB* (Cole *et al.*, 1998); we have matched the N-terminal Fur peptide sequence described by Wong *et al.* (1999) with the deduced amino acid sequence from *furA*. In our study *furB* was up-regulated in macrophage-phagocytosed mycobacteria, suggesting *furB* is a possible low-iron-induced, positive regulator for intracellularly expressed genes.

In this study the up-regulation of gene expression was mostly around threefold, according to our quantitative phosphorimage analysis of Northern blots. Possible heterogeneity in terms of the stage of phagocytosis within the infected THP-1 cell population could obscure true levels of mycobacterial differential gene expression and may account for the low levels of up-regulation observed. A study (Wilson *et al.*, 1999) of isoniazid-induced changes in gene expression in *M. tuberculosis* using microarray hybridization showed that most of the expression changes reported were between 2·1- and 3·5-fold. The study by Alland *et al.* (1998) also demonstrated a fourfold increase of *ask/asd* expression in isoniazid-induced *M. tuberculosis*. Together, these and our own study indicate that two- to threefold changes at the transcription level represent significant changes in mRNA that result in substantial phenotypic alteration in response to environmental change. Correlation between mRNA and protein levels has not been systematically investigated for mycobacteria.

In summary, this technique can be used to identify differentially expressed genes in prokaryotes, such as mycobacteria recovered from infected cultured cells or animal models, from which the availability of RNA is limited and where rRNA predominates in the cDNA populations. Comparison of the non-pathogenic vaccine strain *M. bovis* BCG with the virulent *M. tuberculosis* may well highlight key determinants of pathogenesis in tuberculosis. We have identified a number of *M. bovis* BCG genes expressed during macrophage interaction using this selection procedure and confirmed these by Northern blotting. We showed that the expression of *mas*, a major cell wall synthesis gene, as well as a closely located and functionally related gene, *fadD28*, is increased, which may reflect alterations of the mycobacterial cell wall necessary for intracellular survival. Mycobacterial cell wall components are important drug targets; therefore an understanding of the changes in gene expression leading to changes in the mycobacterial cell wall during interaction with the host may facilitate the design of new drugs.

ACKNOWLEDGEMENTS

This work was supported by a grant from GlaxoWellcome Research and Development. P. D. B. acknowledges the support of the MRC (grant no. G9328099) and GlaxoWellcome Action TB, and I. M. M. the support of EU BIOMED 2 (BMH4-CT97-2671). We thank Hasan Al-Ghusein at St George's Hospital Medical School for the gift of *M. tuberculosis* genomic DNA.

REFERENCES

- Alavi, M. R. & Affronti, L. F. (1994). Induction of mycobacterial proteins during phagocytosis and heat shock: a time interval analysis. *J Leukoc Biol* **55**, 633–641.
- Alland, D., Kramnik, I., Weisbrod, T. R., Cerny, R., Miller, L. P., Jacobs, W. R., Jr & Bloom, B. R. (1998). Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): The effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* **95**, 13227–13232.
- Bashyam, M. D. & Tyagi, A. K. (1998). Identification and analysis of “extended –10” promoters from mycobacteria. *J Bacteriol* **180**, 2568–2573.
- Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. & Small, P. M. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**, 1520–1523.
- Bercovier, H., Kafri, O. & Sela, S. (1986). Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. *Biochem Biophys Res Commun* **136**, 1136–1141.
- Bertram, J., Palfner, K., Hiddemann, W. & Kneba, M. (1998). Overexpression of ribosomal proteins L4 and L5 and the putative alternative elongation factor PTI-1 in the doxorubicin resistant human colon cancer cell line LoVoDxR. *Eur J Cancer* **34**, 732–736.
- Burns, H. D., Belyaeva, T. A., Busby, S. J. & Minchin, S. D. (1996). Temperature-dependence of open-complex formation at two *Escherichia coli* promoters with extended –10 sequences. *Biochem J* **317**, 305–311.
- Butcher, P. D., Mangan, J. A. & Monahan, I. M. (1998). Intracellular gene expression. Analysis of RNA from mycobacteria in macrophages using RT-PCR. *Methods Mol Biol* **101**, 285–306.
- Camacho, L. R., Ensergueix, D., Perez, E., Gicquel, B. & Guilhot, C. (1999). Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol Microbiol* **34**, 257–267.
- Clark-Curtiss, J. E. (1998). Identification of virulence determinants in pathogenic mycobacteria. *Curr Top Microbiol Immunol* **225**, 57–79.
- 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.
- Cox, J. S., Chen, B., McNeil, M. & Jacobs, W. R., Jr (1999). Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature* **402**, 79–83.
- Cunningham, A. F. & Spreadbury, C. L. (1998). Mycobacterial stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton alpha-crystallin homolog. *J Bacteriol* **180**, 801–808.
- Dolin, P. J., Raviglione, M. C. & Kochi, A. (1994). Global tuberculosis incidence and mortality during 1990–2000. *Bull WHO* **72**, 213–220.
- Escolar, L., Perez-Martin, J. & de Lorenzo, V. (1999). Opening the iron Box: transcriptional metalloregulation by the Fur protein. *J Bacteriol* **181**, 6223–6229.
- Fitzmaurice, A. M. & Kolattukudy, P. E. (1998). An acyl-CoA synthase (*acoas*) gene adjacent to the mycocerosic acid synthase (*mas*) locus is necessary for mycocerosyl lipid synthesis in *Mycobacterium tuberculosis* var. *bovis* BCG. *J Biol Chem* **273**, 8033–8039.

- 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 USA* **96**, 11554–11559.
- Henkin, T. M., Moon, S. H., Mattheakis, L. C. & Nomura, M. (1989).** Cloning and analysis of the *spc* ribosomal protein operon of *Bacillus subtilis*: comparison with the *spc* operon of *Escherichia coli*. *Nucleic Acids Res* **25**, 7469–7486.
- Higgins, C. F. (1991).** Stability and degradation of mRNA. *Curr Opin Cell Biol* **3**, 1013–1018.
- Hubank, M. & Schatz, D. G. (1994).** Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acid Res* **22**, 5640–5648.
- Kinger, A. K. & Tyagi, J. S. (1993).** Identification and cloning of genes differentially expressed in the virulent strain of *Mycobacterium tuberculosis*. *Gene* **131**, 113–117.
- Kolattukudy, P. E., Fernandes, N. D., Azad, A. K., Fitzmaurice, A. M. & Sirakova, T. D. (1997).** Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol Microbiol* **24**, 263–270.
- Lee, B.-Y. & Horowitz, M. A. (1995).** Identification of macrophage and stress-induced proteins of *Mycobacterium tuberculosis*. *J Clin Invest* **96**, 245–249.
- Lindahl, L. & Zengel, J. M. (1986).** Ribosomal genes in *Escherichia coli*. *Annu Rev Genet* **20**, 297–326.
- Mangan, J. A. & Butcher, P. D. (1998).** Analysis of mycobacterial differential gene expression by RAP-PCR. *Methods Mol Biol* **101**, 307–322.
- Mangan, J. A., Sole, K. M., Mitchison, D. A. & Butcher, P. D. (1997).** An effective method of RNA extraction from bacteria refractory to disruption, including mycobacteria. *Nucleic Acid Res* **25**, 675–676.
- Mattheakis, L. C. & Nomura, M. (1988).** Feedback regulation of the *spc* operon in *Escherichia coli*: translational coupling and mRNA processing. *J Bacteriol* **170**, 4484–4492.
- Mekalanos, J. J. (1992).** Environmental signals controlling expression of virulence determinants in bacteria. *J Bacteriol* **174**, 1–7.
- Minnikin, D. E. (1982).** Lipids: Complex lipids, their chemistry, biosynthesis and roles. In *The Biology of Mycobacteria*, pp. 95–184. Edited by C. Ratledge & J. Stanford. San Diego: Academic Press.
- Monahan, I. M., Betts, J., Banerjee, D. K. & Butcher, P. D. (2001).** Differential expression of mycobacterial proteins following phagocytosis by macrophages. *Microbiology* **147**, 459–471.
- Patel, B. K., Banerjee, D. K. & Butcher, P. D. (1991).** Characterization of the heat shock response in *Mycobacterium bovis* BCG. *J Bacteriol* **173**, 7982–7987.
- Plum, G. & Clark-Curtiss, J. E. (1994).** Induction of *Mycobacterium avium* gene expression following phagocytosis by human macrophages. *Infect Immun* **62**, 476–483.
- Ponnambalam, S., Webster, C., Bingham, A. & Busby, S. (1986).** Transcription initiation at the *Escherichia coli* galactose operon promoters in the absence of the normal –35 region sequences. *J Biol Chem* **261**, 16043–16048.
- Rainwater, D. L. & Kolattukudy, P. E. (1985).** Fatty acid biosynthesis in *Mycobacterium tuberculosis* var. *bovis* *Bacillus Calmette-Guerin*. Purification and characterization of a novel fatty acid synthase, mycocerosic acid synthase, which elongates n-fatty acyl-CoA with methylmalonyl-CoA. *J Biol Chem* **260**, 616–623.
- Raviglione, M. C., Snider, D. E., Jr & Kochi, A. (1995).** Global epidemiology of tuberculosis: Morbidity and mortality of a worldwide epidemic. *JAMA* **273**, 220–226.
- Rivera-Marrero, C. A., Burroughs, M. A., Masse, R. A., Vannberg, F. O., Leimbach, D. L., Roman, J. & Murtagh, J. J., Jr (1998).** Identification of genes differentially expressed in *Mycobacterium tuberculosis* by differential display PCR. *Microb Pathog* **25**, 307–316.
- Robinson, K. A., Robb, F. T. & Schreier, H. J. (1994).** Isolation of maltose-regulated genes from the hyperthermophilic archaeum, *Pyrococcus furiosus*, by subtractive hybridization. *Gene* **148**, 137–141.
- Rook, G. A. W. & Bloom, B. R. (1994).** Mechanisms of pathogenesis in tuberculosis. In *Tuberculosis: Pathogenesis, Protection and Control*, pp. 485–501. Edited by B. R. Bloom. Washington, DC: American Society for Microbiology.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989).** *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977).** DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.
- Scholnick, J., Sinor, C., Oakes, J., Outten, W. & Saha, M. (1997).** Differential expression of *Xenopus* ribosomal protein gene *XlrpS1c*. *Biochim Biophys Acta* **1354**, 72–82.
- Sturgill-Koszycki, S., Haddix, P. L. & Russell, D. G. (1997).** The interaction between *Mycobacterium* and the macrophage analyzed by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* **18**, 2558–2565.
- Utt, E. A., Brousal, J. P., Kikuta-Oshima, L. C. & Quinn, F. D. (1995).** The identification of bacterial gene expression differences using mRNA-based isothermal subtractive hybridization. *Can J Microbiol* **41**, 152–156.
- Wilson, M., DeRisi, J., Kristensen, H. H., Imboden, P., Rane, S., Brown, P. O. & Schoolnik, G. K. (1999).** Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc Natl Acad Sci USA* **96**, 12833–12838.
- Wong, D. K., Lee, B. Y., Horwitz, M. A. & Gibson, B. W. (1999).** Identification of Fur, aconitase, and other proteins expressed by *Mycobacterium tuberculosis* under conditions of low and high concentrations of iron by combined two-dimensional gel electrophoresis and mass spectrometry. *Infect Immun* **67**, 327–336.
- Wu, G., Su, S. & Bird, R. C. (1994).** Optimization of subtractive hybridization in construction of subtractive cDNA libraries. *Genet Anal Tech Appl* **11**, 29–33.
- Yuan, Y., Zhu, Y., Crane, D. D. & Barry, C. E., III (1998).** The effect of oxygenated mycolic acid composition on cell wall function and macrophage growth in *Mycobacterium tuberculosis*. *Mol Microbiol* **29**, 1449–1458.
- Zuker, M. (1989).** On finding all suboptimal foldings of an RNA molecule. *Science* **244**, 48–52.

Received 22 February 2001; revised 7 April 2001; accepted 19 April 2001.