Erp, an extracellular protein family specific to mycobacteria

Leila de Mendonça-Lima,^{1†} Mathieu Picardeau,² Catherine Raynaud,¹ Jean Rauzier,¹ Yves-Olivier Goguet de la Salmonière,¹ Lucia Barker,^{1‡} Fabiana Bigi,³ Angel Cataldi,³ Brigitte Gicquel¹ and Jean-Marc Reyrat¹

Author for correspondence: Jean-Marc Reyrat. Tel: +33 1 45688828. Fax: +33 1 45688843.
e-mail: jmreyrat@pasteur.fr

^{1,2} Unité de Génétique Mycobactérienne¹, and Unité de Bactériologie Moléculaire et Médicale², Institut Pasteur, F-75724 Paris Cedex 15, France

³ Instituto de Biotecnologia, CICV-INTA, Moron, Argentina

Erp (exported repeated protein) was originally characterized as a virulence factor in *Mycobacterium tuberculosis* and was thought to be present only in *Mycobacterium leprae* and members of the TB complex. Here it is shown that Erp is a ubiquitous extracellular protein found in all of the mycobacterial species tested. Erp proteins have a modular organization and contain three domains: a highly conserved amino-terminal domain which includes a signal sequence, a central variable region containing repeats based on the motif PGLTS, and a conserved carboxy-terminal domain rich in proline and alanine. The number and fidelity of PGLTS repeats of the central region differ considerably between mycobacterial species. This region is, however, identical in all of the clinical *M. tuberculosis* strains tested. In addition, it is shown here that a *Mycobacterium smegmatis* *erp::aph* mutant displays altered colony morphology which is complemented by all the Erp orthologues tested. The genome sequence flanking the *erp* gene includes cell-wall-related ORFs and displays extensive conservation between saprophytic and pathogenic mycobacteria.

Keywords: exported product, protein family, repeats, *Mycobacterium*

INTRODUCTION

Mycobacterium tuberculosis, the aetiologic agent of tuberculosis, one of the most prevalent causes of death due to bacterial infection worldwide, is a facultative intracellular bacterium that persists within professional phagocytic cells. Erp (exported repeated protein), also known as P36 (Bigi *et al.*, 1995), Pirg or Rv3810 (Cole *et al.*, 1998), has been shown to be an exported product in *M. tuberculosis* using *phoA* technology, has a signal peptide and is therefore probably exported via the Sec pathway (Lim *et al.*, 1995; Berthet *et al.*, 1995). It contains repeated sequences based on a PGLTS motif.

[†] Present address: Dept of Biochemistry and Molecular Biology, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, RJ, Brazil.

[‡] Present address: University of Minnesota, Duluth School of Medicine, Duluth, MN, USA.

The GenBank accession numbers for the sequences reported in this paper are AF213152 (*Mycobacterium smegmatis*), AF213153 (*Mycobacterium marinum*), AF213154 (*Mycobacterium ulcerans*), AF213155 (*Mycobacterium xenopi*) and AF315789 (*Mycobacterium avium*). The accession numbers for the partial sequences of the *erp* gene of *Mycobacterium tuberculosis* clinical isolates are AF165856 (Benin), AF165857 (R.C.A.), AF165858 (Vietnam) and AF165859 (Tahiti).

The gene has been detected in all members of the TB complex (*M. tuberculosis*, *Mycobacterium africanum*, *Mycobacterium bovis*, *M. bovis* BCG and *Mycobacterium microti*), and in *Mycobacterium leprae*, the aetiologic agent of leprosy (Berthet *et al.*, 1995), but it was thought to be restricted to pathogenic species of mycobacteria. Disruption of the gene in *M. tuberculosis* and in *M. bovis* BCG resulted in a marked attenuation of virulence, as judged from survival and multiplication in mouse lungs, spleen, or in bone marrow derived macrophages. However, no differences in growth characteristics were observed in axenic media between wild-type, mutant and complemented H37Rv and BCG strains (Berthet *et al.*, 1998). The Erp protein has also been described as an immunodominant antigen in both lepromatous leprosy and bovine tuberculosis (Cherayil & Young, 1988; Bigi *et al.*, 1995).

We show here that the *erp* gene is present in all mycobacterial species tested and is therefore not a pathogen-specific gene. The *erp* gene was found not only in pathogenic mycobacteria, but also in saprophytic mycobacteria such as *Mycobacterium smegmatis*, in environmental opportunistic pathogenic mycobacteria

such as *Mycobacterium avium*, *Mycobacterium marinum* and *Mycobacterium xenopi*, and in extracellular-toxin-producing mycobacteria such as *Mycobacterium ulcerans*. This allowed the definition of a new protein family that seems specific to the genus *Mycobacterium*.

METHODS

Bacterial strains, plasmids and growth conditions. *Escherichia coli* was routinely grown in liquid or on solid Luria-Bertani (LB) medium (Difco) at 37 °C. Mycobacteria were grown in liquid Middlebrook 7H9 medium (Difco) or Sauton medium (Sauton, 1912) containing 0.05 % (v/v) Tween 80 where indicated, or on Middlebrook 7H10 or 7H11 at 37 °C. When required, antibiotics were included at the following concentrations: ampicillin, 100 µg ml⁻¹; kanamycin, 20 µg ml⁻¹; hygromycin, 200 µg ml⁻¹ (for *E. coli*) or 50 µg ml⁻¹ (mycobacteria).

Cloning of the *M. smegmatis* and *M. xenopi* *erp* genes. Chromosomal DNA was isolated from *M. smegmatis* and *M. xenopi* as described previously (Pelicic *et al.*, 1996). Five micrograms of genomic DNA from either species was digested with *SalI* (*M. smegmatis*) or *PstI*-*XbaI* (*M. xenopi*) and DNA fragments in the 2 kb or 4 kb size range (*M. smegmatis* or *M. xenopi*, respectively) were excised from the gel, purified and ligated to pBluescript II SK vector DNA linearized with the respective restriction enzymes and dephosphorylated, in the case of *SalI*. Colony replicas of the transformants were made on Hybond-N⁺ nylon membranes (Amersham Pharmacia) and screened by hybridization with a ³²P-labelled PCR fragment probe generated by the amplification of genomic DNA with primers *erp*-8 and *erp*-9. Recombinant plasmid DNA from clones yielding strong hybridization signals were recovered by using Qiagen mini-columns. Screening of these partial genomic libraries allowed the identification of recombinant plasmids carrying the *M. smegmatis* 2 kb *SalI* insert (plasmid pML-2A1) or the *M. xenopi* *PstI*-*XbaI* 4.5 kb insert (plasmid pEPX6).

DNA sequence of *erp* genes and flanking regions. Sequences of double-stranded plasmid DNA or PCR-generated fragments were determined by the dideoxy chain-termination

method using a 373-B DNA Analysis System (Applied Biosystems). Oligonucleotide sequences are listed in Table 1.

The DNA sequence of the *M. marinum* and *M. ulcerans* *erp* genes was obtained by sequencing PCR fragments generated from genomic DNA. The DNA sequence of the cloned *M. xenopi* and *M. smegmatis* *erp* genes (pML-2A1) was obtained from sequencing of plasmid DNA with different internal primers, and extended towards the 5' and 3' regions by sequencing of PCR fragments generated from genomic DNA. The region of the *M. avium* genome between contigs 5434 and 5730 (TIGR unfinished genomes; <http://www.tigr.org>), containing the 3' portion of the *erp* gene and not yet available from the genome project, was obtained by PCR with primers *erp*-17 + *erp*-18 (300 bp) and *erp*-18 + *erp*-19 (400 bp).

Sequencing in TB isolates. Five clinical isolates originating from France (hp030496 13), Benin (be290396 18), Vietnam (be290396 29), Tahiti (ta080296 11) and the Central African Republic (rca30696 16) with a distinct spoligotype profile were chosen (Goguet de la Salmonière, 1999), and the central region of the *erp* gene corresponding to amino acids 92–235 was amplified by PCR using primers *erp*Rf and *erp*Rr and sequenced.

Construction of the *M. smegmatis* *erp* insertional mutant. A *M. smegmatis* *erp* mutant was generated by insertion of a kanamycin (*aph*) resistance cassette into the *erp* coding region. DNA from plasmid pML-2A1, a suicide vector in mycobacteria, was partially digested with *Bam*HI and ligated to a 1264 bp *Bam*HI DNA fragment containing a kanamycin resistance cassette from plasmid pUC4K (Taylor & Rose, 1988). The ligation mixture was used to transform *E. coli* DH5 α and colonies harbouring the desired constructs were selected on LB plates supplemented with kanamycin. Plasmid pML-2A1::*aph* contains the *aph* fragment inserted in the *Bam*HI site at position 503 of the *erp* gene (relative to the start codon). One microgram of pML-2A1::*aph* DNA was introduced by electroporation into *M. smegmatis* mc²155, essentially as described by Pelicic *et al.* (1996). The electroporated bacteria were plated on 7H10 plates supplemented with kanamycin. Colonies were screened by PCR with oligonucleotide primers flanking the insertion site (primers *erp*-11 and *erp*-R) in order to identify double-recombination events.

Construction of complemented strains. Constructs for complementation of the *M. smegmatis* *erp*::*aph* insertional mutant were made in the integrative vector pNIP40b (Berthet *et al.*, 1998). Vector DNA was linearized with *XbaI* and dephosphorylated with 0.1 U shrimp alkaline phosphatase (USB) prior to ligation. For the *M. smegmatis* mc²155 or the *M. leprae* complementation constructs (pML-30 and pML-40, respectively), the following strategy was used: a 426 bp *M. tuberculosis* MT103 PCR fragment generated with primers *erp*-C1 and *erp*-C3, containing the *erp* regulatory region and signal peptide up to a unique *Eco*RI site (at position 129 in relation to the start codon), was ligated to either (1) an 805 bp PCR fragment generated with primers *erp*-C4 and *erp*-C6, containing the *M. smegmatis* *erp* coding sequence, starting at position 120 in relation to the start codon and ending at the stop codon, or (2) a 586 bp PCR fragment generated with primers *erp*-C5 and *erp*-C7, containing the *M. leprae* *erp* coding sequence, starting at position 132 in relation to the start codon and ending at the stop codon. The ligation mixtures were used to transform *E. coli* DH5 α and transformants were selected on LB plates supplemented with hygromycin and kanamycin. Plasmid DNA was used to transform the *M. smegmatis* mc²155 *erp*::*aph* mutant strain

Table 1. Oligonucleotide primers used in this study

Primer	Nucleotide sequence (5'–3')
<i>erp</i> -8	GTGCCGAACCGACGCCGACG
<i>erp</i> -9	GGCACC GGCGGCAGGTTGATCCCCG
<i>erp</i> -11	CCTGATTGCCGCAACTCAGC
<i>erp</i> -17	CGATCACGCAGGCATGCAC
<i>erp</i> -18	CCGCGATGGCGGTGAGCA
<i>erp</i> -19	CGTACCCGATCTGGGCGAC
<i>erp</i> -C1	GCTCTAGACGAGCGGTCATCGGTTGCATAGGG
<i>erp</i> -C3	CGGAATTCATGGTGCTCGGGCCGCTC
<i>erp</i> -C4	CGGAATTCACCCAGGCCGCGCTGGTCACC
<i>erp</i> -C5	CGGAATTCAAACAAGCAGCATCGATAGCC
<i>erp</i> -C6	GCTCTAGATCAGGCAGGCGGCGGCACGGGTGC
<i>erp</i> -C7	GCTCTAGACTACGTGACAGGAATCAGTGATAT
<i>erp</i> -Rf	CCTGGCCTGACTAGTCCGGGATTG
<i>erp</i> -Rr	ACTGGCGCCCAACAGGTTGGCCAC

by electroporation. Transformants were selected on 7H10 plates supplemented with hygromycin and kanamycin.

Preparation of protein samples for SDS-PAGE and immunoblotting. Protein samples were prepared from bacteria grown in 25 ml Sauton medium containing 0.05% Tween 80, under agitation. Briefly, the culture supernatants were obtained after centrifugation at 4000 g for 10 min, and filtered through a 0.2 µm Millex-GV (Millipore) filter to remove any remaining cells, and the proteins were precipitated with 17% (v/v) TCA and analysed by SDS-PAGE and Western blotting.

SDS-PAGE and immunoblotting. Proteins were resolved by SDS-PAGE using 12% polyacrylamide gels (Laemmli, 1970) and transferred to Hybond-C membranes (Amersham Pharmacia). A rabbit Erp antiserum (Berthet *et al.*, 1998) was used as first antibody at a 1:5000 dilution, and bound antibodies were revealed using a horseradish-peroxidase-coupled donkey anti-rabbit antibody (Amersham Pharmacia) at 1:10000 dilution. Detection was performed with the ECL system (Amersham Pharmacia).

Computer methods. The amino acid sequence alignments used to produce the identity between orthologues were generated by ALIGN from the Wisconsin Package, version 9.1-Unix with a PAM250 matrix, whereas multiple alignments were generated using PILEUP (Pearson & Lipman, 1988). Domain analysis was performed using ProDom (Corpet *et al.*, 2000). Boxshade (<http://www.ch.embnet.org/software/box.html>) was used to highlight similarity between proteins. Public databases were searched using either BLASTP or PSI-BLAST algorithms (Altschul *et al.*, 1997).

RESULTS AND DISCUSSION

Definition of the Erp family

The Erp protein was thought to be present only in pathogenic mycobacteria. However, a 36 kDa secreted protein that reacted with an anti-Erp rabbit serum was

detected in *M. smegmatis* and in other non-TB mycobacteria (Bigi *et al.*, 1999; and our unpublished results). This led us to investigate whether Erp homologues were present in other pathogenic and non-pathogenic mycobacterial species. Using a combination of degenerate primers for amplification, PCR fragments encompassing the *erp* gene were directly sequenced using a primer walking strategy for *M. avium*, *M. marinum* and *M. ulcerans*. A short specific PCR fragment was used as a probe to screen a partial genomic library of *M. smegmatis* and *M. xenopi* and the selected recombinant plasmids were sequenced. The DNA sequences were assembled and the deduced amino acid sequences were shown to contain repeated sequences based on the PGLTS motif, as already described for the *M. tuberculosis* and *M. leprae* proteins. Fig. 1 shows a schematic representation of a multiple alignment of all the deduced amino acid sequences of the Erp homologues.

Three different protein domains can be identified within Erp. The amino-terminal domain (amino acids 1–80), which includes the signal sequence, is highly conserved between species, showing more than 70% sequence identity. The central domain, which contains the repeated region, is variable in two aspects. First, the absolute number of repeats is variable, from 4 in *M. leprae*, to 24 in *M. xenopi*. Second, the sequence of the PGLTS repeat is variable, so that in some species, such as *M. xenopi* or *M. smegmatis*, half the repeats contain two mismatches (see Fig. 1). The amino acid sequence of the carboxy-terminal domain shows more than 50% identity between species, and the extreme carboxy-terminus of the protein is composed of a stretch rich in proline and alanine of variable length. None of these

Fig. 1. Schematic representation of the Erp orthologues. The conserved amino- and carboxy-terminal domains are shown as solid black boxes (■). The repeats are indicated by hatched or white boxes: ▨, perfect PGLTS repeats; □, repeats containing one mismatch; □, repeats containing two mismatches. All protein features and sequences as well as the protein alignment are included as supplementary data with the on-line version of this paper (available at <http://mic.sgmjournals.org>).

	<i>glf</i>	<i>erp</i>	<i>csp</i>
<i>M. tuberculosis</i>	+	+	+
<i>M. bovis</i> BCG	+	+	+
<i>M. leprae</i>	+	+	+
<i>M. avium</i>	+	+	+
<i>M. xenopi</i>	+	+	+
<i>M. smegmatis</i>	+	+	+
<i>C. diphtheriae</i>	+	-	+

Fig. 2. Conservation of the *erp* genomic environment. +, Presence of the gene; -, absence of the gene; +*, presence of a pseudogene; +†, partial sequence.

three domains is significantly similar to any protein sequence in public databases.

Differences between species in the sequence of the central region of the *erp* gene led us to look for sequence variation between characterized isolates of *M. tuberculosis*. Five clinical isolates from France, Benin, Vietnam, Tahiti and the Central African Republic with different spoligotype profiles were studied (Goguet de la Salmonière, 1999). The central region of the *erp* gene was amplified by PCR and sequenced. No differences in DNA sequence were observed, despite the different spoligotype profiles and geographical origins, demonstrating that the repeated region was not subject to allelic variation.

During the completion of this work, the genomic sequences of *M. smegmatis* (<http://www.tigr.org/tdb/mdb/mdbinprogress.html>) and *Mycobacterium paratuberculosis* (<http://www.cbc.umn.edu/cgi-bin/blasts/AGAC.restrict/blastn.cgi>) were released, showing the *erp* gene to be present in both species, thereby confirming and extending our data. It is thus very likely that *erp* is common to all members of the genus *Mycobacterium*.

Conservation of the *erp* genomic environment

The completed genome sequence of *M. tuberculosis* shows that *erp* lies between *glf* (Rv3809) and *csp* (Rv3811), two genes probably involved in cell-wall elaboration (Cole *et al.*, 1998; Weston *et al.*, 1998). Using *in silico* methods, when the sequences were available (*M. tuberculosis*, *M. bovis*, *M. leprae* or *M. avium*), or through PCR amplification with *glf*- and *csp*-specific primers for *M. smegmatis*, *M. xenopi* and *M. ulcerans*, we showed that the genetic context of the *erp* locus is conserved in all mycobacteria tested (Fig. 2). However, in *M. leprae*, numerous point mutations in the *csp* gene probably lead to the absence of an active product, suggesting that *csp* may not be required for a strictly intracellular lifestyle. Nevertheless, the extensive

Fig. 3. Immunoblot of denaturing SDS-PAGE of Erp in supernatant. WT, *M. smegmatis* wild-type; *erp*, *M. smegmatis* *erp::aph*; comp, *M. smegmatis* *erp::aph* (pIPX70). Purified recombinant Erp (30 ng) is used as an internal control.

genomic conservation observed suggests a strong selective pressure that has maintained this locus unchanged from saprophytic to pathogenic mycobacteria.

In *Corynebacterium diphtheriae* (<http://www.sanger.ac.uk>), a close relative of the mycobacteria belonging to the order *Actinomycetales*, genomic analysis shows the presence of both *glf* and *csp* (Fig. 2). Contrary to mycobacteria, *erp* is not present between these two genes, being replaced in this species by the homologues of Rv0836c and Rv0837c, the function of which are unknown. However, as the genome of *C. diphtheriae* has not yet been completely sequenced, we cannot rule out the possibility that an orthologue of *erp* is present somewhere else in the genome.

Phenotypic characterization of the *M. smegmatis* *erp::aph* mutant

To demonstrate that the 36 kDa protein of *M. smegmatis* reacting with the anti-Erp serum was indeed due to the Erp homologue, an *erp::aph* derivative with a kanamycin cassette inserted within the *erp* gene was constructed by allelic exchange. This *erp* mutant was then transformed with pIPX70, an integrative vector expressing TB *erp* under its own promoter, which had already been used to complement the *erp* mutant of *M. tuberculosis* (Berthet *et al.*, 1998). Fifty micrograms of total protein contained in the culture supernatants of *M. smegmatis* mc²155, the *M. smegmatis* *erp::aph* mutant and the isogenic complemented strain was probed with the anti-Erp antiserum. No cross-reacting band was detected in the *M. smegmatis* *erp* mutant strain (Fig. 3), demonstrating that this cross-reaction is indeed due to the product of the *erp* gene of *M. smegmatis*.

As previously described for *M. bovis* BCG in liquid culture (Berthet *et al.*, 1998), the *M. smegmatis* *erp* mutant strain grew as well as the wild-type strain in both rich and minimal media, suggesting that the *erp* mutation did not impair the basic metabolism of the bacterium (data not shown). Despite the absence of growth defects in liquid culture, the *M. smegmatis* *erp::aph* mutant displayed a characteristic altered

Fig. 4. Photomicrographs of colonies of *M. smegmatis*. (A) *M. smegmatis* wild-type; (B) *M. smegmatis* *erp::aph*; (C) *M. smegmatis* *erp::aph* (pIPX70); and (D) *M. smegmatis* *erp::aph* (pML-40).

colony morphology on solid medium (Fig. 4). Whereas wild-type colonies were flat and spread out, fully in contact with the agar surface (Fig. 4A), the colonies formed by the *erp* mutant were raised, growing upwards, touching the surface over a much smaller area (Fig. 4B). The wild-type colony phenotype was restored by re-integration of one copy of the *M. smegmatis* (data not shown), *M. tuberculosis* or *M. leprae* *erp* genes (Fig. 4C, D, respectively). This straightforward phenotypic characterization indicated that the altered colony morphology is due to the absence of the Erp protein itself, and not to a polar effect on upstream or downstream genes. The capacity of different Erp orthologues to complement the morphological abnormality observed in the mutant suggests that the molecules of this family share similar functions with regard to colony morphology and might, directly or indirectly, contribute to cell-wall structure. In this regard, it is worth noting that the central part of the Erp protein sequence differs greatly between species, and these differences in the number and in the sequence of the PGLTS repeat seem to have no effect on morphology, suggesting that flanking domains may have a more important structural role in maintaining surface integrity. Indeed, all of the Erp orthologues tested were able to complement this morphological abnormality in the mutant.

The morphological defect observed for the *erp* mutant of *M. smegmatis* led us to re-examine this phenomenon in *M. bovis* BCG in which it had only been observed after animal infection and organ recovery (Berthet *et al.*, 1998). Serial dilutions of *M. bovis* BCG and the *erp* mutant strain were plated so as to obtain about 10 colonies per Petri dish. Under such conditions, the phenotype of the *M. bovis* BCG *erp* mutant was similar to that of the *M. smegmatis* *erp* mutant strain (data not shown). This phenotype was not observed if a larger inoculum was used, which may explain why it has been primarily observed after organ recovery, where bacteria are plated in order to obtain isolated colonies.

The requirement for Erp in *M. tuberculosis* for a successful murine infection is clear (Berthet *et al.*, 1998). However, this ubiquity in the mycobacterial genus shows that it is not a pathogen-specific gene. It is, however, possible to speculate that the Erp protein, which has an extracellular localization, may have an important role in cell-wall structure. Several studies have shown that genes encoding cell-wall-related products are involved in the virulence of the tubercle bacillus (Camacho *et al.*, 1999; Cox *et al.*, 1999; Armitige *et al.*, 2000; Glickman *et al.*, 2000). One future challenge is the identification of the molecular function of the Erp protein and its role in *M. tuberculosis* virulence.

ACKNOWLEDGEMENTS

L. M.-L. is funded by the Oswaldo Cruz Foundation (Fiocruz) and Pasteur Institute. M. P. thanks the Fondation de France for financial support. D. Kahn is acknowledged for help in protein domain analysis. We thank P. Small for the kind gift of *M. ulcerans* genomic DNA, and C. Le Dantec for *M. avium* genomic DNA. V. Pelicic and W. Degrave are gratefully acknowledged for critical reading of the manuscript. J.-M. R. is chargé de recherche at Inserm.

REFERENCES

- Altschul, S. F., Madden, T. L., Shaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of database search programs. *Nucleic Acids Res* **17**, 3389–3402.
- Armitige, L. Y., Chinnaswamy, J., Wanger, A. R. & Norris, S. J. (2000). Disruption of the genes encoding antigen 85A and antigen 85B of *Mycobacterium tuberculosis* H37Rv: effect on growth in culture and macrophages. *Infect Immun* **68**, 767–778.
- Berthet, F.-X., Rauzier, J., Lim, E. M., Philipp, B., Gicquel, B. & Portnoi, D. (1995). Characterization of the *Mycobacterium tuberculosis* *erp* gene encoding a potential cell surface protein with repetitive structures. *Microbiology* **141**, 2123–2130.
- Berthet, F.-X., Lagranderie, M., Gounon, P. & 9 other authors (1998). Attenuation of virulence by disruption of *Mycobacterium tuberculosis* *erp* gene. *Science* **282**, 759–762.
- Bigi, F., Alito, A., Fisanotti, J. C., Romano, M. I. & Cataldi, A. (1995). Characterization of a novel *Mycobacterium bovis* secreted antigen containing PGLTS repeats. *Infect Immun* **63**, 2581–2586.
- Bigi, F., Taboga, O., Romano, M. I., Alito, A., Fisanotti, J. C. & Cataldi, A. (1999). Expression of the *Mycobacterium bovis* P36 gene in *M. smegmatis* and the baculovirus/insect cell system. *Braz J Med Biol Res* **32**, 29–37.
- Camacho, L. R., Ensergueix, D., Perez, E., Gicquel, B. & Guilhot, B. (1999). Identification of a virulence gene cluster of *M. tuberculosis* by signature-tagged transposon mutagenesis. *Mol Microbiol* **34**, 257–267.
- Cherayil, B. G. & Young, R. A. (1988). A 28 kDa protein from *Mycobacterium leprae* is a target of the human antibody response in lepromatous leprosy. *J Immunol* **141**, 4370–4375.
- Cole, S. T., Brosch, R., Parkhill, J. & 38 other authors (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.
- Corpet, F., Servant, F., Gouzy, J. & Kahn, D. (2000). ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res* **28**, 267–269.

- Cox, J. S., Chen, B., McNeil, M. & Jacobs, W. R. (1999).** Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature* **402**, 79–83.
- Glickman, M. S., Cox, J. S. & Jacobs, W. R. (2000).** A novel mycolic acid cyclopropane synthetase is required for coding, persistence and virulence of *Mycobacterium tuberculosis*. *Mol Cell* **5**, 717–727.
- Goguet de la Salmonière, Y. O. (1999).** *Protéines exportées de Mycobacterium tuberculosis pour l'étude des interactions avec l'hôte et méthodologie pour l'épidémiologie moléculaire*. PhD thesis, University of Paris.
- Laemmli, U.K. (1970).** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
- Lim, E. M., Rauzier, J., Timm, J., Torrea, G., Murray, A., Gicquel, B. & Portnoï, D. (1995).** Identification of *Mycobacterium tuberculosis* DNA sequences encoding exported proteins, using *phoA* gene fusions. *J Bacteriol* **177**, 59–65.
- Pearson, W. R. & Lipman, D. J. (1988).** Improved tools for biological sequence analysis. *Proc Natl Acad Sci USA* **85**, 2444–2448.
- Pellicic, V., Reytrat, J.-M. & Gicquel, B. (1996).** Generation of unmarked directed mutations in mycobacteria, using sucrose counter-selectable suicide vectors. *Mol Microbiol* **20**, 919–925.
- Sauton, B. (1912).** Sur la nutrition minérale du bacille tuberculeux. *C R Acad Sci Ser III Sci Vie* **135**, 860–863.
- Taylor, L. A. & Rose, R. E. (1988).** A correction in the nucleotide sequence of the Tn903 kanamycin resistance determinant in pUC4K. *Nucleic Acids Res* **16**, 358.
- Weston, A., Stern, R. J., Lee, R. E. & 7 other authors (1998).** Biosynthetic origin of mycobacterial cell wall galactofuranosyl residues. *Tuber Lung Dis* **78**, 123–131.
-
- Received 22 February 2001; revised 30 March 2001; accepted 5 April 2001.