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

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), Prg or Rv3810 (Cole et al., 1998), has been shown to be an exported product in M. tuberculosis using pboA 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.

The gene has been detected in all members of the TB complex (M. tuberculosis, Mycobacterium africanaum, 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.

1,2 Unité de Génétique Mycobactérienne1, and Unité de Bactériologie Moléculaire et Médicale2, Institut Pasteur, F-75724 Paris Cedex 15, France
3 Instituto de Biotecnología, CICV-INTA, Moron, Argentina

† 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.
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 either species as described previously (Pelcic 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 dyeoxy 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 (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 erpF and erpR and sequenced.

Construction of the M. smegmatis erp insertion 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 BamHI and ligated to a 1264 bp BamHI DNA fragment containing a kanamycin resistance cassette from plasmid pUC4K (Taylor & Rose, 1988). The ligation mixture was used to transform E. coli DH5α 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 BamHI 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 Pelcic 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 insertion 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 C-1 and C-3, containing the erp regulatory region and signal peptide up to a unique EcoRI site (at position 129 in relation to the start codon), was ligated to either (1) an 805 bp PCR fragment generated with primers C-4 and C-6, 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 C-5 and C-7, 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 DH5α 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

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide sequence (5′–3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>erp-8</td>
<td>GTCGCCAGACCCGACGCCGAG</td>
</tr>
<tr>
<td>erp-9</td>
<td>GCCACCACCCGCCGAGTGTCCGCCGAG</td>
</tr>
<tr>
<td>erp-11</td>
<td>CCTGATTGTCGCCAATCTCAG</td>
</tr>
<tr>
<td>erp-17</td>
<td>CGTACCCGAGGCAGTCCAC</td>
</tr>
<tr>
<td>erp-18</td>
<td>CGCCGATGCGGCTGAGCA</td>
</tr>
<tr>
<td>erp-19</td>
<td>CGTACCCGATCTGCGCCGAC</td>
</tr>
<tr>
<td>erp-C1</td>
<td>GCTCTAGACGACGGCTATCAGGTTCATAGGG</td>
</tr>
<tr>
<td>erp-C3</td>
<td>CGAACATATGGTGCGCCGCGCTCT</td>
</tr>
<tr>
<td>erp-C4</td>
<td>CGAACATACCCAGCGCGGCCGCTGTCACC</td>
</tr>
<tr>
<td>erp-C5</td>
<td>CGAACATACCCAGCGCGCCGCGCTGTCACC</td>
</tr>
<tr>
<td>erp-C6</td>
<td>GCTCTAGACGACGGCTATCAGGTTCATAGGG</td>
</tr>
<tr>
<td>erp-C7</td>
<td>GCTCTAGACGACGGCTATCAGGTTCATAGGG</td>
</tr>
<tr>
<td>erp-Rf</td>
<td>CTGGCGCTACTGAGTCCGCGATG</td>
</tr>
<tr>
<td>erp-Rr</td>
<td>ACTGGCACCACAGGCCGAGTGC</td>
</tr>
</tbody>
</table>
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 **pilup** (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 **psiblast** 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 carboxyl-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
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 product. Nevertheless, the extensive 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...
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**


Received 22 February 2001; revised 30 March 2001; accepted 5 April 2001.