Purification, Characterization, Gene Sequence, and Significance of a Bacterioferritin from *Mycobacterium leprae*

By Maria Cristina V. Pessolani,* Douglas R. Smith,‡
Becky Rivoire,* Julie McCormick,* Stanley A. Hefta,§
Stewart T. Cole,‖ and Patrick J. Brennan*

*From the *Department of Microbiology, Colorado State University, Fort Collins, Colorado 80523; ‡Collaborative Research, Inc., Waltham, Massachusetts 02154; §Beckman Institute of The City of Hope, Duarte, California 91010; and ‖Institut Pasteur, Paris, France*

Summary

The study of tissue-derived *Mycobacterium leprae* provides insights to the immunopathology of leprosy and helps identify broad molecular features necessary for mycobacterial parasitism. A major membrane protein (MMP-II) of in vivo-derived *M. leprae* previously recognized (Hunter, S.W., B. Rivoire, V. Mehra, B.R. Bloom, and P.J. Brennan. 1990. *J. Biol. Chem.* 265:14065) was purified from extracts of the organism and partial amino acid sequence obtained. This information allowed recognition, within one of the cosmids that encompass the entire *M. leprae* genome, of a complete gene, *bfr*, encoding a protein of subunit size 18.2 kD. The amino acid sequence deduced from the major membrane protein II (MMP-II) gene revealed considerable homology to several bacterioferritins. Analysis of the native protein demonstrated the iron content, absorption spectrum, and large native molecular mass (380 kD) of several known bacterioferritins. The ferroxidase-center residues typical of ferritins were conserved in the *M. leprae* product. Oligonucleotides derived from the amino acid sequence of *M. leprae* bacterioferritin enabled amplification of much of the MMP-II gene and the detection of homologous sequences in *Mycobacterium paratuberculosis*, *Mycobacterium avium*, *Mycobacterium tuberculosis*, *Mycobacterium intracellulare*, and *Mycobacterium scrofulaceum*. The role of this iron-rich protein in the virulence of *M. leprae* is discussed.

Leprosy is a chronic disease of the skin and nerves caused by *Mycobacterium leprae*. Chemotherapy is successful but requires protracted multidrug therapy and the continuing evolution of new drug regimens. The world-wide prevalence of leprosy is now estimated at less than 5 million cases compared with 10–12 million in the mid-1980s (1). Much of this decline is attributed to chemotherapy. Nevertheless, leprosy still represents a major public health problem within regions of Latin America, Africa, and Asia, many of them not amenable to the standard drug regimens. Hence, the development of a vaccine that could contribute to the eradication of the disease remains a goal of some leprosy research programs.

The molecular definition of *M. leprae* is fundamental to an understanding of the physiology and metabolism of the bacillus and the immunoregulatory mechanisms underlying the disease and protection against it. The leprosy bacillus grows intracellularly in cells of the reticuloendothelial system and Schwann cells, and it is so well adapted to this habitat that it has evaded axenic cultivation so far (2). The development of the armadillo as a relatively rich source of the bacillus has allowed extensive study of the physiology of the organism

(2) and definition of the chemistry of the highly antigenic glycoconjugates and their roles in pathogenesis (3). Screening of *M. leprae* genomic libraries with antibodies has also permitted the isolation of genes that encode protein antigens, many of which are related to components of the highly conserved family of stress proteins (4).

A landmark in leprosy research that may revolutionize our knowledge of the basis of the obligate parasitism of *M. leprae* is the recently initiated "*M. leprae* genome project" (5). Present achievements from this undertaking include an ordered collection of overlapping clones encompassing the complete chromosome of *M. leprae* in which 72 loci have been mapped (5) and the complete sequence of 27 cosmids representing approximately one third of the genome (6; Smith, D.R., unpublished data). In an attempt to complement this genetic approach, we initiated a program to identify and characterize the major proteins expressed in vivo by the leprosy bacillus. The dominant proteins from the different subcellular fractions of *M. leprae* were identified in terms of relative size and antibody-binding patterns, and some of these were immediately recognized as the previously described stress proteins

(7). However, two major novel proteins were identified and designated major membrane protein I (MMP-I)[1] and major membrane protein II (MMP-II). The present study describes the purification, biochemical characterization and gene sequence of MMP-II, pointing to a novel ~380-kD bacterioferritin (Bfr), an iron-rich protein that probably plays a crucial role in the intracellular existence of *M. leprae*.

Materials and Methods

Purification of MMP-II. *M. leprae* was purified from irradiated armadillo spleens and livers by the Draper protocol as described previously (8). Frozen bacteria (300 mg, dry weight) were resuspended in 3 ml of PBS containing 10 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, and 1 mM benzamidine and disrupted by intermittent probe sonication (W-385 sonicator; Heat Systems-Ultrasonics, Inc., Farmingdale, NY) for four 15-min periods on ice. The suspension was centrifuged at 27,000 g for 30 min to remove unbroken cells and cell walls. The supernatant was further centrifuged at 100,000 g for 2 h to remove membranes and polysomes. Mycobacterial lipoglycans (lipoarabinomannan, lipomannan, and phosphatidylinositol mannosides) were removed from the resulting supernatant by fractionation on DEAE-Sephacel (Pharmacia Biotech, Inc., Piscataway, NJ) equilibrated in 20 mM piperazine buffer, pH 5.7. After washing the column with 160 mM LiClO₄, proteins were eluted with 1 M salt and subjected to gel filtration on a Sephacryl S-200 (Pharmacia Biotech, Inc.) column (1 × 120 cm) equilibrated with PBS. MMP-II in the eluates was monitored by SDS-PAGE. The fractions containing MMP-II were combined, dialyzed against water and concentrated. Further purification of MMP-II was accomplished on a reversed phase HPLC column (model C4; 0.46 × 25 cm; VYDAC, Hesperia, CA). Proteins were eluted from the column at a flow rate of 1 ml/min using a linear gradient from 70% solvent A (trifluoroacetic acid/water, 0.1:99.9), 30% solvent B (trifluoroacetic acid/water/acetonitrile, 0.1:9.9:90) to 30% solvent A/70% solvent B over 40 min. The final highly purified MMP-II was obtained upon elution with a linear gradient from 70 to 100% solvent B in 30 min.

The multimeric form of MMP-II was obtained by affinity chromatography based on the monoclonal antibody CS-41 (7). CS-41 was purified from ascites fluid by protein G agarose column chromatography (Schleicher & Schuell, Inc., Keene, NH). A 1-ml affinity column was constructed by using the Affinica antibody orientation kit (Schleicher & Schuell, Inc.) as recommended by the manufacturer. The column was equilibrated with PBS, the soluble fraction of the bacteria applied and nonspecifically bound proteins were removed with 1 M NaCl after washing the column with PBS. MMP-II was eluted with 1 M ammonium hydroxide, pH 11.5. Fractions were immediately neutralized with 1 M HCl and dialyzed against 10 mM Tris-HCl, pH 7.5

Amino Acid Sequence of MMP-II. The pure protein was reduced with β-mercaptoethanol and subsequently alkylated with 4-vinylpyridine. The sample was desalted on a Vydac C4 reversed-phase HPLC column (0.46 by 5 cm) under the conditions described above. Fractions were pooled, dried, and treated with each of the following enzymes: endoproteinase Lys-C (from *Lysobacter enzymogenes*) (Boehringer Mannheim Corp., Indianapolis, IN), trypsin (treated with tolylsulfonyl phenylalanyl chloromethyl ketone) (Sigma Chem-

ical Co., St. Louis, MO), and endoproteinase Asp-N (from *Pseudomonas fragi*) (Boehringer Mannheim Corp.). Before use, the trypsin was repurified by reversed phase HPLC (9). Digestions were accomplished on samples of 200–300 pmol as described by Lee and Shively (10). The enzyme/protein ratios were typically 1:50. Peptide mapping was performed using a capillary HPLC system as described by Davis and Lee (11), equipped with a 0.5-mm microcolumn packed with VYDAC C18 resin. The peptides were eluted at a flow rate of 0.02 ml/min using a linear gradient from 100% solvent A (trifluoroacetic acid/water, 0.1:99.9) to 40% solvent A/60% solvent B (trifluoroacetic acid/water/acetonitrile, 0.1:9.9:90) in 60 min. Absorbance at 214 nm was monitored with a SPD-6A UV/VIS monitor equipped with a microanalytical flow cell (Shimadzu Corp., Tokyo, Japan). Peptides were spotted on polyvinylidene difluoride membranes (Millipore Corp., Bedford, MA) and subjected to automated Edman degradation on a gas-phase sequencer built at City of Hope (12) and equipped with a continuous-flow reactor (13). The phenylthiohydantoin amino acid derivatives were indicated by on-line reversed-phase HPLC. Sequences were confirmed by fast atom bombardment-mass spectrometry as described (14).

PCR Amplification of the MMP-II Gene. The reaction mixture contained the DNA template (10 ng) in 100 μl (final volume) of 10 mM Tris-HCl buffer, pH 8.0, containing 50 mM KCl, 2 mM MgCl₂, deoxyribonucleotide triphosphates (250 μM each), oligonucleotide amplimer (1 mM each) and 2.5 U of Taq polymerase (Boehringer Mannheim Corp.). Before addition to the reaction mixture, the template was boiled for 5 min. The reaction was carried out in the DNA Thermal Cycler (Perkin-Elmer Corp., Norwalk, CT) as follows: cycles 1 to 5, 95°C/30 sec, 37°C/1 min, a ramp time of 3 min between 37°C and 70°C, 70°C/1 min; cycles 6 to 34, 95°C/30 sec, 57°C/1 min, 70°C/1 min; cycle 35, 95°C for 30 sec, 57°C for 1 min, and 70°C for 10 min. Degenerate oligonucleotides were derived from the amino acid sequence based on the usage codon table of the first of the *M. leprae* genome cosmid clones to be sequenced (6) and were synthesized on a DNA synthesizer (model 7500; MilliGen/Biosearch, Burlington, MA). The forward primer 5'-ATGCAGGGC(T)GAC(T)CCC(G)GAC(T)GT-3' was derived from amino acid 1 to 7 of the NH₂-terminal sequence. The following alternative backward primers were used: 5'-CACATA-(G)ATA(G)ATA(G)CCG(C)GGCTT-3', derived from peptide T-11; 5'-GTGAAA(G)CCCCAGTTC(T)TCCTGCAT-3', derived from peptide T-16 (T-16a); and 5'-CTCG(C)GTGAAA(G,C)CCCCAG-TTCTCCTGCAT-3', derived from T-16 (T-16b). Oligonucleotides 5'-AG(A)G(A)AAG(A)TAC(T)TGA(G)TTA(G)AT-3' derived from amino acid 22 to 27 of the NH₂-terminal region, were labeled at the 5' end with γ-[³²P]ATP (15) and used as an internal probe to check the specificity of the PCR products on Southern analysis. Hybridization was conducted at 37°C and washed with tetramethylammonium chloride at 44°C. Chemiluminescent detection with Lumi-Phos 530 was achieved with the nonradioactive Genius kit (Boehringer Mannheim Corp.) according to the protocols provided by the manufacturer.

M. leprae Genome Sequencing and Sequence Analysis. The construction of a physical genetic map of the *M. leprae* genome has been described (5). Multiplex sequencing of cosmid B38 was conducted as described by Church and Kieffer-Higgins (16) and will be described in detail elsewhere. The application of multisequence analysis using the program PILEUP from Genetic Computer Grouping (GCG, Wisconsin Package) has been described (17). The hydropathy plot of the predicted amino acid sequence was constructed by the method of Kyte and Doolittle (18).

Southern Hybridization. Genomic DNA from *M. leprae* was purified as described (19). Purification of armadillo genomic DNA

---

[1] *Abbreviations used in this paper:* Bfr, bacterioferritin; MMP-I, major membrane protein I; MMP-II, major membrane protein II; SSC, sodium chloride–sodium citrate.

from uninfected liver and other standard DNA manipulation techniques were performed according to Maniatis et al. (15). 3–5 μg of purified DNA were digested with restriction enzymes and fractionated in 0.8% agarose gel electrophoresis. The 0.33-kb PCR product was labeled with digoxigenin by primer extension with random hexamers (Boehringer Mannheim Corp.). Hybridization was conducted at 42°C for 16 h in buffer containing five times the concentration of standard sodium chloride–sodium citrate (SSC) (1 × SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 0.1% N-lauroylsarcosine, 0.02% SDS, 5% (wt/vol) blocking reagent for nucleic acid hybridization (Boehringer Mannheim Corp.), and 50% formamide (vol/vol). The blots were washed twice for 5 min at room temperature in 2 × SSC with 0.1% SDS. Stringent washes were performed twice for 15 min at 65°C or 40°C in 0.1 × SSC containing 0.1% SDS. Chemiluminescent detection with Lumi-Phos 530 was conducted following the manufacturer recommendations (Boehringer Mannheim Corp.). Membranes were exposed to XAR-2X-RAY film (Eastman Kodak Co., Rochester, NY).

Physical and Chemical Assays. Visible- and UV-absorption spectra of the purified protein were obtained with a Varian Cary 118-C double-beam spectrophotometer (Varian Instrument Business, San Fernando, CA) in 1-cm pathlength microcuvettes. Absorption maxima of the reduced product were obtained after mixing in a few crystals of sodium dithionite. The iron content of the purified protein was determined by inductively coupled plasma–optical emission spectroscopy on an atomcomp (model 975; Thermo Jarrell Ash, Franklin, MA). The iron content was related to protein content as estimated by amino acid analysis in a Beckmen High Performance Amino Acid Analyzer System (Model 6300; Beckman Instruments Inc., Fullerton, CA) after the sample had been hydrolyzed with 6 M HCl for 24 h at 110°C. Molecular mass estimation was performed by gel filtration on Superose 6 fast performance liquid chromatography (FPLC) and Sephacryl S-300 columns (Pharmacia Biotech, Inc.). Columns were equilibrated with 10 mM Tris-HCl pH 7.5 containing 0.15 M NaCl and calibrated using four standard proteins: aldolase (150 kD), catalase (232 kD), ferritin (440 kD), and thyroglobulin (669 kD) (Pharmacia LKB). SDS-PAGE was per-
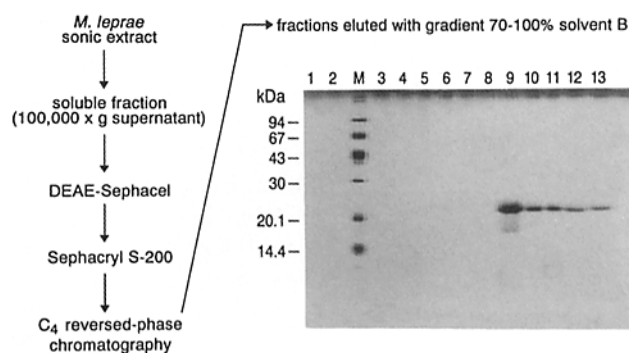
Figure 1. Purification of MMP-II. Silver-stained SDS-PAGE of purified protein after elution from C4 reversed-phase column. Molecular mass standards are indicated on the left side.

formed on uniform 15% gels or 4–20% gradient gels according to Laemmli (20) and proteins were stained with silver nitrate or Coomassie blue.

Results

Amino Acid Sequence of MMP-II. MMP-II was initially identified as a major band of 22 kD by SDS-PAGE analysis of the membrane fraction of leprosy bacillus (7). However, after disrupting the cells with prolonged sonic oscillation, considerable quantities of MMP-II were present in the soluble fraction of the bacteria. The protein was purified from this source by conventional chromatography (Fig. 1). Most of the lipoarabinomannan, lipomannan, and phosphatidylinositol mannosides were successfully removed by passing the soluble fraction through DEAE-Sephacryl. Further fractionation of the soluble proteins was achieved by Sephacryl S-200 gel filtration chromatography from which MMP-II was eluted close

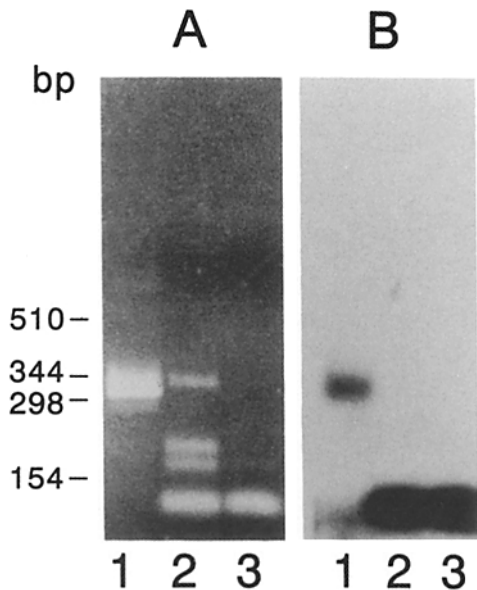**Table 1.** *Partial Amino Acid Sequence of MMP-II of M. leprae*

Digest/peptide	Mass predicted	Mass observed	Sequence
LYS-C:			
K-8	3575.9	3475.4	MQGDPDVLRLLNEQLTS(E)LTA(I NQYFLHSK
TRYPSIN:			
T-11	1089.4	1089.5	LKPG I I MCR
T-16	1611.8	1611.6	MQENWGFTELAER
T-18	1415.7	1415.6	I LLLDGLPNYQR
T-23	2463.8	2463.5	LLNEQLTSELTA I NQYFLHSK
ASP-N:			
A-12	963.0	962.7	DGLPNYQR
NH₂-terminal:			
	NA	NA	MQGDPDVLRLLNEQLTSE

Digestion conditions with the three different endoproteinases, Lys-C, trypsin, and Asp-N, are described in the text. References to the procedures used for amino acid sequencing and mass determination are provided in the text. NA, not applicable.

to the void volume. Rechromatography of the MMP-II–containing fractions on a C4 reversed-phase HPLC column provided highly purified protein (Fig. 1). From 300 mg (dry weight) of *M. leprae*, 14 mg of soluble protein were recovered of which 50 μg were pure MMP-II.

Amino acid sequencing of MMP-II involved digestion with three different endoproteinases: trypsin, Lys-C, and Asp-N. Table 1 shows the list of peptides whose sequences were confirmed by mass analysis. In one instance, K8, the sequence data were incomplete; however, the correct assignment was achieved by overlapping the sequence with T23 and from the mass data. The NH₂-terminal sequence as obtained on this occasion was different in position 2 from that reported previously (7) and is now in accord with the gene sequence (see below).

Amplification of the MMP-II Gene by PCR. Oligonucleotides with low degeneracy were derived from the amino acid sequence and three alternative reactions were run. The PCR products were analyzed by electrophoresis on agarose gel (Fig. 2 A), yielding multiple products in reactions 1 and 2. A Southern blot of the same gel hybridized with the ³²P-labeled internal probe showed a single band in each lane (Fig. 2 B). The measured sizes in base pairs of the hybridizing fragments were 330 bp for reaction 1, which accounts for 67% of the complete gene, and 100 bp for reactions 2 and 3. The product of PCR reaction 1 was employed for purposes of cloning the 330 bp fragment in the pCR II™ vector (Invitrogen, San Diego, CA). The fragment was purified by agarose gel electrophoresis, labeled with digoxigenin, and employed as a probe in subsequent experiments.

Nucleotide Sequence of the MMP-II Gene. As part of the *M. leprae* genome project being carried out at the Pasteur Institute and Collaborative Research, Inc. (6), a sequence was identified with homology to several Bfr genes. The predicted product of this gene turned out to be identical to that of MMP-II and therefore was presumed to originate from the same gene; the sequence of the *M. leprae* Bfr gene and surrounding regions are given in Fig. 3. The Bfr gene is located within cosmid B38 at positions 15,000–16,549 (the sequence data are available from Genbank under accession number

Figure 2. Amplification of the MMP-II gene by PCR. (*A*) Agarose gel electrophoresis of the PCR products; and (*B*) Southern blot with an internal probe labeled with ³²P. The forward primer was derived from the first seven amino acids of the protein. The alternative backward primers T-11, T-16a, and T16b were used in reactions 1, 2, and 3, respectively.

Figure 3. Nucleotide sequence of the *M. leprae* Bfr gene. The 1.5 kb sequence is derived from a 37 kb sequence of cosmid B38 (Smith D. R., unpublished results; the sequences data are available from Genbank under accession number L01095). The amino acid sequence is shown below the coding region. Some of the peptides sequenced are indicated. The nucleotides below the asterisk symbols are analogous to an iron box sequence. The nucleotides below the plus symbols correspond to the ribosome binding site.

L01095). A multisequence analysis using the program PILEUP (GCG Wisconsin Package) revealed strong sequence conservation with several other Bfrs. The alignments between the sequences from *Azotobacter vinelandii* (21; GB/M83692), *Escherichia coli* (22; GB/M27176), *M. paratuberculosis* (23; Pir/A44893), *Nitrobacter winogradskyi* (24; SP/p13570), and the cyanobacterium *Synechocystes* P.C.C. 6803 (25; SP/P24602) (Fig. 4) shows appreciable homology to the translated *M. leprae* Bfr gene. Of these, the *M. leprae* sequence is most closely related to the partial sequence from *M. paratuberculosis*. This alignment reveals a number of invariant residues near the NH₂ terminus of the proteins and three regions of strong sequence conservation at residues 21–30, 61–70, and 130–138. An alignment between *M. leprae* Bfr and H-like ferritin also indicates that the established and putative ferroxidase-center residues of ferritins are absolutely conserved or conservatively substituted in MMP-II. These residues are Glu 18, Tyr 25 Asp 50, Glu 51, Hist 54, Glu 94, and Glu 127, and, as previously shown by Andrews et al. (26) and Grossman et al. (21), are conserved in all Bfrs characterized so far (Fig. 4). In addition, prediction of MMP-II secondary structure employing the Peptide Structure Program (GCG Wisconsin Package) indicated a high (∼80%) α-helix content, consistent with the presence of the ferritin-like four-helix bundle characteristic of other Bfrs (21, 26). The hydropathy plot (not shown) of the deduced protein sequence conducted as described (18)
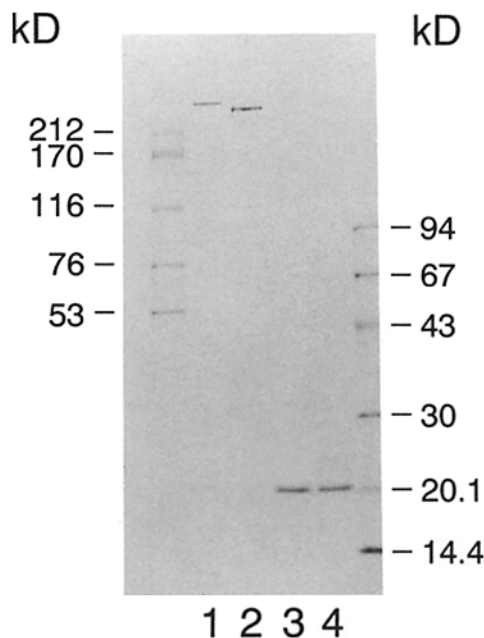
showed the occurrence of only a few short stretches of hydrophobic amino acids consistent with a soluble/peripheral membrane protein.

At the nucleic acid level, the sequence contains a ribosome binding site with significant homology to the 3′ end of the *M. leprae* 16S rRNA (27) preceding a methionine start codon at nucleotide 630 in Fig. 3. A sequence with significant homology to an iron box (28), but without any dyad symmetry, occurs in an AT-rich region surrounding nucleotide 540 which may correspond to the transcriptional initiation site of the gene. Several regions of dyad symmetry were noted in the sequence, including one centered at nucleotide 177 upstream of the gene and at nucleotide 1077 near the translation termination site.

Biochemical Characterization of M. leprae Bfr. Crystallographic studies (29) and electron microscopic imaging (23) demonstrate that Bfrs share similar quaternary structure to ferritins. Ferritins are multimeric proteins that contain 24 identical subunits forming a spherical shell enclosing a central cavity that contains up to ∼4,500 Fe(III) atoms of non-haem iron (30). However, Bfrs differ from ferritins in that they are haem proteins containing protoporphyrin IX as prosthetic group (31–33). To establish whether the *M. leprae* Bfr also shared some of these characteristics, chemical analyses were conducted on the native protein. Small quantities of the pure protein were obtained in the native form by affinity chromatography based on the monoclonal antibody GS-41.
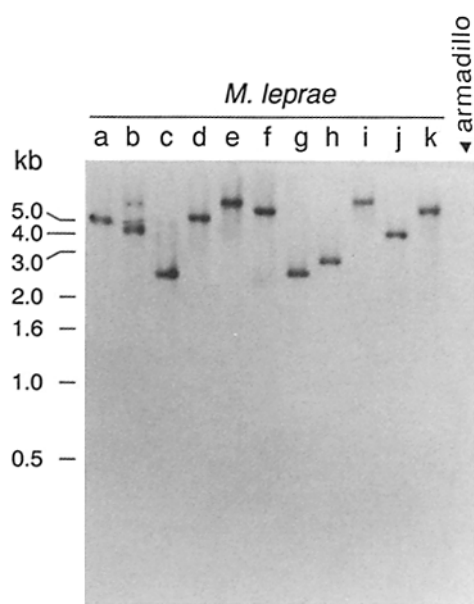
```
               1                                                  50
(A.vibfr)   MKGDKIVIQH LNKILGNELI AINQYFLHAR MYEDWGLEKL GKHEYHESID
(E.cobfr)   MKGDTKVINY LNKLLGNELV AINQYFLHAR MFKNWGLKRL NDVEYHESID
(M.lepbfr)  MQGDPDVLRL LNEQLTSELT AINQYFLHSK MQENWGFTEL AERTRVESFD
(M.paratb)  MQGDPEVLRL LNEQLTSELT AINQYFLHSK MQDNWGFTEL AE........
(N.wibfr)   MKGDPKVIDY LNKALRHELT AINQYWLHYR LLDNWGIKDL AKKWRAESIE
(Cyanobfr)  MKGKPAVLAQ LHKLLRGELA ARDQYFIHSR MYQDWGLEKL YSRIDHEMQD
Consensus   MKGDP-V--- LNK-L--ELT AINQYFLHSR MY-NWGL-EL AE---HESID
                            +          +                        +

               51                                                 100
(A.vibfr)   EMKHADKLIK RILFLEGLPN LQELGKLLIG EHTKEMLECD LKLEQAGLPD
(E.cobfr)   EMKHADRYIE RILFLEGLPN LQDLGKLNIG EDVEEMLRSD LALELDGAKN
(M.lepbfr)  EMRHAEAITD RILLLDGLPN YQRIGSLRVG QTLREQFEAD LAIEYEVMSR
(M.paratb)  .......... .......... .......... .......... ..........
(N.wibfr)   .......... .......... .......... .......... ..........
(Cyanobfr)  ETHA...... .......... .......... .......... ..........
Consensus   EM-HAD---- RILFLEGLPN -Q--G-L-IG ----E-L--D L--E------
            + +

               101                                                150
(A.vibfr)   LKAAIAYCES VGDYASRELL EDILESEEDH IDWLETQLDL IDKIGLENYL
(E.cobfr)   LREAIGYADS VHDYVSRDMM IEIIRDEEGH IDWLETELDL IQKMGLQNYL
(M.lepbfr)  LKPGIIMCRE KQDSTSAVLL EKIVADEEEH IDYLETQLAL MGQLGEELYS
(M.paratb)  .......... .......... .......... .......... ..........
(N.wibfr)   .......... .......... .......... .......... ..........
(Cyanobfr)  .......... .......... .......... .......... ..........
Consensus   L---I----- --D--S--LL --I---EE-H IDWLET-L-L ----G---Y-
                                                +

               151
(A.vibfr)   QSQMDE...
(E.cobfr)   QAQIREEG.
(M.lepbfr)  AQCVSRPPS
(M.paratb)  .........
(N.wibfr)   .........
(Cyanobfr)  .........
Consensus   ---------
```

Figure 4. Multisequence alignment of known Bfr sequences produced with the Genetics Computer Group PILEUP. The amino acids above the plus symbols correspond to the established and putative ferroxidase-center residues. Program parameters were: plurality: 3.00; threshold: 1.00; average weight 0.54; average mismatch −0.40.

Figure 5. Coomassie blue–stained SDS-PAGE of the multimeric and monomeric forms of *M. leprae* Bfr. Lane *1*, horse spleen ferritin (44 kD); lane *2*, affinity-purified *M. leprae* Bfr plus sample buffer (not heated); lane *3*, affinity-purified *M. leprae* Bfr plus sample buffer (boiled for 10 min.); lane *4*, *M. leprae* Bfr purified by reversed phase HPLC plus sample buffer (not heated). Molecular mass markers were loaded on the right and left lanes of the gel.
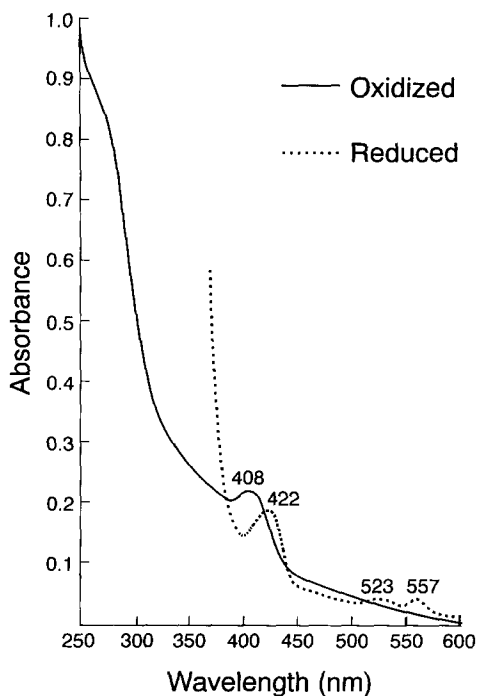
Figure 6. Absorption spectrum of *M. leprae* Bfr in 10 mM Tris-HCl buffer pH 7.5. The protein was reduced by adding few crystals of Na₂S₂O₄ to the sample in cuvette.

Figure 7. Estimation of the copy number of the Bfr gene in the *M. leprae* chromosome. Southern blot analysis of *M. leprae* DNA digested with BamHI (lane *a*), EagI (lane *b*), EcoRI (lane *c*); EcoRV (lane *d*), HindIII (lane *e*), KpnI (lane *f*), PstI (lane *g*), SspI (lane *h*), SmaI (lane *i*), SpeI (lane *j*) and Xho (lane *k*), and hybridized with the digoxigenin-labeled 0.33-kb PCR product. Armadillo DNA digested with EcoRI was used as a control. The numbers on the left show the sizes of the 1-kb ladder (Life Technologies, Inc., Gaithersburg, MD).

Fig. 5 shows the presence of multimeric and monomeric forms of the Bfr on SDS-PAGE in a 4–20% gradient gel. The affinity purified protein, when mixed with sample buffer and immediately loaded on the gel, behaved like a high molecular mass molecule (Fig. 5, lane *2*), close in size to ferritin (440 kD) (Fig. 5, lane *1*). By heating the sample before loading on the gel, the multimer dissociated into subunits of approximate molecular mass 20 kD (Fig. 5, lane *3*). Fig. 5, lane *4* shows the subunit form as obtained after reversed-phase chromatography (Fig. 1). The molecular mass of the multi-

meric form of Bfr was estimated by gel filtration to be 380 kD. Also, the absorption spectrum of Bfr from *M. leprae* showed a prominent Soret band at 408 nm (Fig. 6), and reduction with sodium dithionite shifted the Soret band to 422 nm. The α and β bands are also detectable with respective maxima at 557 and 523 nm. *M. leprae* Bfr is iron-rich, since, on the basis of a molecular mass of 380 kD, it contained of the order of 1,000–4,000 atoms of iron per molecule of protein as estimated on two different preparations of the pure protein.
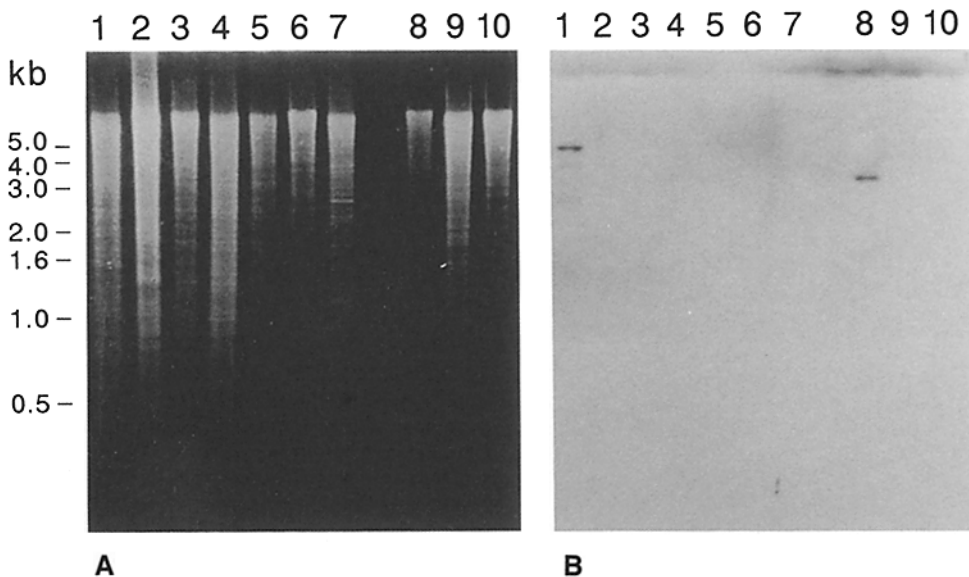
Figure 8. Search for the Bfr gene in other mycobacteria. Agarose gel electrophoresis (*A*) and Southern blot hybridization patterns (*B*) of DNA isolated from *M. leprae* (lane *1*), *M. tuberculosis* H37Rv (lane *2*), *M. avium* serovar 2 (lane *3*), serovar 4 (lane *4*), serovar 8 (lane *5*), *M. smegmatis* (lane *6*), *M. xenopi* (lane *7*), *M. paratuberculosis* (lane 8), *M. intracellulare* (lane *9*), and *M. scrofulaceum* (lane *10*) cleaved with EcoRI. The DNAs were hybridized to the digoxigenin-labeled 0.33-kb PCR product using high stringent conditions (washes in 0.1 SSC at 68°C). The numbers on the left show the sizes of the 1-kb ladder.

Copy Number and Conservation of the Bfr Gene. To determine whether there are multiple homologous copies of the Bfr gene in the *M. leprae* chromosome, genomic DNA from *M. leprae* was digested with 11 different enzymes followed by hybridization with the 330-bp PCR-amplified fragment of the gene labeled with digoxigenin. A single band was detected in each case (Fig. 7). This result, together with the information that no tandem copies of the Bfr gene occur in the cosmid clone, demonstrated that only a single copy of the Bfr gene exists in *M. leprae*. To determine whether other mycobacteria possess the Bfr gene or homologous sequences, chromosomal DNA from nine different species were digested with EcoRI, fragments separated on a gel and hybridized with the 330 bp PCR-amplified product of the Bfr gene. In high stringency (washes in 0.1 × SSC at 68°C), hybridizing bands were detected only in *M. leprae* and *M. paratuberculosis* (Fig. 8), in agreement with the high homology of the amino acid sequence between *M. leprae* Bfr and antigen D. At stringent conditions that allowed 20% of mismatches between probe and target DNA (washes in 0.1 × SSC at 40°C), a hybridizing band was observed in the other mycobacteria tested, with the exception of *M. smegmatis* and *M. xenopi*.

Discussion

Iron is essential to the growth of all living cells; it is required for the transport and storage of oxygen, and as a catalyst in electron transport processes (34). In an aerobic atmosphere, however, iron is extremely insoluble and available in limited amounts. Organisms have therefore developed mechanisms for solubilizing iron and, in many cases, for storing it within particular molecules. Animals have solved the problem of iron acquisition by using ferritin as a ubiquitous iron storage compound and transferrin and lactoferrin for movement of iron (35). Little or no free iron then exists within animal bodies and pathogenic bacteria must acquire iron from such sources. Iron metabolism is therefore considered to play a central role in virulence, and a mechanism to scavenge iron is a requisite for microbial pathogenicity (36). In this context, the observation that one of the most abundant proteins present in in vivo grown *M. leprae* is a Bfr, a protein involved in ferrometabolism, must be of considerable significance.

This present development was due to our technical ability to isolate, even from scarce tissue-derived *M. leprae*, sufficient amounts of this major protein to obtain partial amino acid sequence, to estimate the size and iron content, and, through absorption spectroscopy, establish other relationships to known Bfrs. Oligonucleotides based on the amino acid sequences and PCR allowed amplification of a fragment accounting for ∼70% of the gene. Recognition of the entire gene encoding MMP-II was possible through the new *M. leprae* genome project. At present, the *M. leprae* chromosome is represented by four "contigs" (or blocks of contiguous DNA sequences) or overlapping cosmids, and, by summing their sizes, one can estimate the chromosome to have a size of ∼2.8 megabases. The sequence of one cosmid has been described in detail (6), and the sequences of 26 more (18 sequences in Genbank), or about one third of the *M. leprae* chromosome, have

been determined (Smith, D., unpublished data). Over 300 *M. leprae* genes have been identified, and, in many cases, precise functions can be attributed on the basis of extensive sequence homology. However, the present work, in that it defines a protein that corresponds to a gene identified in this fashion, extends the implications of the research beyond the level of comparisons with known gene sequences. Clearly, the derived amino acid sequence from the MMP-II gene showed significant homology with Bfrs already described in the literature (Fig. 4), and it also shared 100% identity with the first 42 amino acids of antigen D, a Bfr isolated from *M. paratuberculosis* (23, 37). Bfrs are haem proteins, and the optical spectrum of reduced MMP-II showed α, β, and Soret bands at 557, 523, and 422 nm, respectively, demonstrating the presence of protoporphyrin IX as a prosthetic group. Recently, a bis-methionine axial ligation of haem has been demonstrated in Bfr from *Pseudomonas aeroginosa* (38), and, based on this information, alternative models involving two conserved methionine residues have been postulated as providing the haem ligands in other Bfrs (21, 26). With the inclusion of the *M. leprae* Bfr, it can be seen that only the methionine at position one remains conserved among the sequences of all Bfrs (Fig. 4). These results open up new questions about the haem binding site in Bfrs. Fig. 4 also shows that in *M. leprae* Bfr, like other Bfrs, the ferroxidase-center residues are conserved, which in ferritins are involved in iron oxidation. This evidence supports recent indications that Bfrs are distantly related to ferritins, constituting a single family of iron proteins (21, 26).

The grouping of all ferritins in a single family and their phylogenetic distribution in bacteria, fungi, plants, and animals implicate ferritin as an obligate protein for aerobic metabolism. Accordingly, we investigated the conservation of the Bfr gene in several mycobacterial species. The presence of sequences homologous to the *M. leprae* Bfr gene was clearly evident in *M. avium*, *M. tuberculosis*, *M. paratuberculosis*, *M. intracellulare*, and *M. scrofulaceum*, and the absence of hybridizing bands in *M. smegmatis* and *M. xenopi* at the stringency used may merely indicate a higher divergence between the sequences, since these species are taxonomically more distant from *M. leprae* (39). Among these mycobacteria, the *M. paratuberculosis* Bfr shared the highest homology with *M. leprae* Bfr (Fig. 8). Interestingly, in terms of iron acquisition, both *M. leprae* and *M. paratuberculosis* behave similarly. During the early stages of growth after isolation from the host, *M. paratuberculosis* demonstrates an obligate need for exogenous mycobactins, exochelins or an excess of citric ferric (40). Thus, the similarity between the Bfrs of *M. leprae* and *M. paratuberculosis* raises the possibility that both acquire iron in vivo through a similar mechanism in which Bfr plays a central role, thereby imposing high conservation on the primary sequences.

The function of Bfrs is still unclear. Based on their structural similarity to ferritins, it has been proposed that Bfrs are involved in the safe storage of iron in iron-overloading conditions. However, Bfrs differ from their eukaryote analogs in two major respects which may be indicative of unique functions for Bfrs. First, they contain haem residues with proper-

ties similar to those of b-type cytochromes. Indeed, *E. coli* and *A. vinelandii* Bfrs were first isolated as cytochrome b_1 and $b_{557.5}$, respectively, and a role for these in electron transport systems has been proposed (24, 31, 41). The interaction between haem components and the ability of Bfrs to store nonhaem iron is, however, not understood. Second, in contrast to plants and animals in which ferritin synthesis is induced by iron (30), the Bfrs subunits retain a constant concentration regardless of the iron status of the cell (25). Also, in iron-deficient cyanobacterium *Synechocystes*, Bfrs seem to play a role as intermediates in iron acquisition, acting as a temporary depository (25). However, in overloaded cells, Bfr iron represents a very low proportion of the total iron content, and low molecular mass molecules carry the bulk of the cellular iron content and fulfill the iron storage function. These results are supported by studies performed on *E. coli* in which Bfr iron was shown to represent no more than 1% of the total cellular iron, the bulk of iron in the iron-rich cells being in the form of aggregates (42, 43).

The mechanism adopted by *M. leprae* to acquire iron inside macrophages is unknown. Macrophages are one of the major iron storage cells of the human body, and most iron is found as ferritin in the cytosol and as hemosiderin inside lysosomal membranes (35). Two major mechanisms by which bacteria remove iron from host compounds have been described. Microbes such as the pathogenic *Neisseria* spp., possess iron acquisition systems using surface receptors for host Fe-binding proteins (28). However, in most mycobacteria, iron acquisition mechanisms involve two different types of siderophores, the extracellular exochelins and the cell-bound mycobactins (44). Also, a set of iron-regulated envelope proteins has been implicated as exochelins receptors (45). In

M. leprae, the presence of exochelins and mycobactins has not yet been demonstrated, but an exochelin-mediated iron uptake was demonstrated by using exochelin from either *M. neoaurum* or an armadillo derived mycobacterium, ADM 8563. Also, iron-regulated envelope proteins have been detected in *M. leprae*, indicating that the leprosy bacillus is growing under iron deficiency in the host (2). Might *M. leprae* Bfr be playing the same function in ferrometabolism as that observed in cyanobacteria? The abundance of Bfr found in armadillo-derived *M. leprae*, as well as the high iron content of the protein, supports the idea that Bfr might be functioning as a temporary depository for iron in iron deprivation. Moreover, the detection of a potential iron box around the promoter region of the *M. leprae* Bfr gene could explain the overexpression of Bfr in an organism grown under conditions of iron deficiency. The iron box associated with Bfr gene constitutes the second example of an iron box–like sequence in *M. leprae*, the first one was observed close to a gene encoding a putative secreted/cell wall associated protein of 28 kD (46). However, the relevance of these sequences in *M. leprae* is presently unknown.

The relevance of *M. leprae* Bfr as an antigen for purposes of diagnosis or vaccine development has not yet been explored. In the context of Johne's disease, *M. paratuberculosis* Bfr is an immunodominant B cell antigen, and is the key component of a diagnostic test (37). The cloning and overexpression of the *M. leprae* Bfr gene will allow an investigation of the immunological properties of this protein in leprosy. However, a clearer definition of the role of Bfr in bacterial ferrometabolism and virulence will be better achieved through isolation and characterization of Bfr mutants from cultivable pathogenic mycobacteria.

References

1. Noordeen, S.K., L.L. Bravo, and T.K. Sundaresan. 1992. Estimated number of leprosy cases in the world. *Bull. WHO.* 70:7.
2. Wheeler, P.R. 1990. Recent research into the physiology of *Mycobacterium leprae*. *Adv. Microb. Physiol.* 31:71.
3. Brennan, P.J. 1989. Structures of mycobacteria: recent developments in defining cell wall carbohydrates and proteins. *Rev. Infect. Dis.* 11(Suppl. 2):S420.
4. Young, D.B., S.H.E. Kaufmann, P.W.M. Hermans, and J.E.R. Thole. 1992. Mycobacterial protein antigens: a compilation. *Mol. Microbiol.* 6:133.

5. Eiglmeier, K., N. Honore, S.A. Woods, B. Caudron, and S.T. Cole. 1993. Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. *Mol. Microbiol.* 7:197.
6. Honore, N., S. Bergh, S. Chanteau, F. Doucet-Populaire, K. Eiglmeier, T. Garnier, C. Georges, P. Launois, T. Limpaiboon, S. Newton, et al. 1993. Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rif-Str regions. *Mol. Microbiol.* 7:207.
7. Hunter, S.W., B. Rivoire, V. Mehra, B.R. Bloom, and P.J.

Brennan. 1990. The major native proteins of the leprosy bacillus. *J. Biol. Chem.* 265:14065.

8. Hunter, S.W., T. Fujiwara, and P.J. Brennan. 1982. Structure and antigenicity of the major specific glycolipid antigen of *Mycobacterium leprae. J. Biol. Chem.* 257:15072.

9. Titani, K., T. Saagawa, K. Resing, and K.A. Walsh. 1982. A simple and rapid purification of commercial trypsin and chymotrypsin by reverse-phase high-performance liquid chromatography. *Anal. Biochem.* 123:408.

10. Lee, T.D., and J.E. Shively. 1990. Enzymatic and chemical digestion of proteins for mass spectrometry. *Methods Enzymol.* 193:361.

11. Davis, M.T., and T.D. Lee. 1992. Analysis of peptide mixtures by capillary. High performance liquid chromatography: a practical guide to small-scale separations. *Protein Science.* 1:935.

12. Hawke, D.H., D.C. Harris, and J.E. Shively. 1985. Microsequence analysis of peptides and proteins. V. Design and performance of a novel gas-liquid-solid phase instrument. *Anal. Biochem.* 147:315.

13. Shively, J.E., P. Miller, and M. Ronk. 1987. Microsequence analysis of peptides and proteins. VI. A continuous flow reactor for a sample concentration and sequence analysis. *Anal. Biochem.* 163:517.

14. Lee, B.-Y., S.A. Hefta, and P.J. Brennan. 1992. Characterization of the major membrane protein of virulent *Mycobacterium tuberculosis. Infect. Immun.* 60:2066.

15. Sambrook, J., T. Mantis, and E.F. Fritsch. 1989. Molecular cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 2nd ed. 63.

16. Church, G.M., and S. Kieffer-Higgins. 1988. Multiplex DNA sequencing. *Science (Wash. DC).* 240:185.

17. Smith, D.R., and J.M. Calvo. 1980. Nucleotide sequence of the *E. coli* gene coding for dihydrafolate reductase. *Nucleic Acids Res.* 8:2255.

18. Kyte, J., and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105.

19. Jacobs, W.R., Jr., G.V. Kalpana, J.D. Cirillo, L. Pascopella, S.B. Snapper, R.A. Undani, W. Jones, R.G. Barletta, and B.R. Bloom. 1991. Genetic systems for mycobacteria. *Methods Enzymol.* 204:537.

20. Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.).* 227:680.

21. Grossman, M.J., S.M. Hinton, V. Minak-Bernero, C. Slaughter, and E.I. Stiefel. 1992. Unification of the ferritin family of proteins. *Proc. Natl. Acad. Sci. USA.* 89:2419.

22. Andrews, S.C., P.M. Harrison, and J.R. Guest. 1989. Cloning, sequencing, and mapping of the bacterioferritin gene (bfr) of *Escherichia coli* K-12. *J. Bacteriol.* 171:3940.

23. Brooks, B.W., N.M. Young, D.C. Watson, R.H. Robertson, E. Sugden, K.H. Nielsen, and S.A. Becker. 1991. *Mycobacterium paratuberculosis* antigen D: characterization and evidence that it is a bacterioferritin. *J. Clin. Microbiol.* 29:1652.

24. Kurokawa T., Y. Fukumori, and T. Yamanaka. 1989. *Nitrobacter winogradskyi* cytochrome b-559: a nonhaem iron-containing cytochrome related to bacterioferritin. *Biochim. Biophys. Acta.* 976:135.

25. Laulhere, J.P., A.M. Laboure, O. Van Wuytswinkel, J. Gagnon, and J.F. Briat. 1992. Purification, characterization and function of bacterioferritin from the cyanobacterium *Synechocystis* P. C. C. 6803. *Biochem. J.* 281:785.

26. Andrews, S.C., J.M.A. Smith, S.J. Yewdall, J.R. Guest, and P.M. Harrison. 1991. Bacterioferritins and ferritins are distantly related in evolution. Conservation of ferroxidase-centre residues. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 293:164.

27. Liesack, W., C. Pitulle, S. Sela, and E. Stackerbrandt. 1990. Nucleotide sequence of the 16S rRNA from *Mycobacterium leprae. Nucleic Acids Res.* 18:5558.

28. Litwin, C.M., and S.B. Calderwood. 1993. Role of iron in regulation of virulence genes. *Clin. Microbiol. Rev.* 6:137.

29. Smith, J.M.A., G.C. Ford, P.M. Harrison, J. Yariv, and A. Kalb. 1989. Molecular size and symmetry of the bacterioferritin of *Escherichia coli.* X-ray crystallography characterization of crystal forms. *J. Mol. Biol.* 205:465.

30. Theil, E.C. 1987. Ferritin: structure, gene regulation and cellular function in animals, plants and microorganisms. *Annu. Rev. Biochem.* 56:289.

31. Stiefel, E.I., and G.D. Watt. 1979. Azotobacter cytochrome $b^{557.5}$ is a bacterioferritin. *Nature (Lond.).* 279:81.

32. Moore, G.R.S., S. Mann, and J.V. Bannister. 1986. Isolation and properties of the complex non heme-iron-containing cytochrome b557 (bacterioferritin) from *Pseudomonas aeruginosa. J. Inorg. Biochem.* 28:329.

33. Yariv, J., A.J. Kalb, R. Sperling, E.R. Bauminger, S.G. Cohen, and S. Ofer. 1982. The composition and the structure of bacterioferritin of *Escherichia coli. Biochem. J.* 197:171.

34. Neilands, J.B. 1972. Evolution of biochemical iron binding centers. *Struct. Bond.* 11:145.

35. Finch, C.A., and H. Huebers. 1982. Perspectives in iron metabolism. *N. Engl. J. Med.* 306:1520.

36. Griffiths, E. 1985. Candidate virulence markers. *In* The Virulence of *Escherichia coli.* SGM Series. M. Surman, editor. Academic Press, London. 193–226.

37. Sugden, E.A., B.W. Brooks, N.M. Young, and D.C. Watson. 1991. Chromatographic purification and characterization of antigens A and D from *Mycobacterium paratuberculosis* and their use in enzyme-linked immunosorbent assays for the diagnosis of paratuberculosis in sheep. *J. Clin. Microbiol.* 29:1659.

38. Cheesman, M.R., A.J. Thomson, C. Greenwood, G.R. Moore, and F. Kadir. 1990. Bis-methionine axial ligation of haem in bacterioferritin from *Pseudomonas aeruginosa. Nature (Lond.).* 346:771.

39. Smida, J., J. Kazda, and E. Stackebrant. 1988. Molecular genetic evidence for the relationship of *Mycobacterium leprae* to slow growing pathogenic mycobacteria. *Int. J. Lepr.* 56:449.

40. Merkal, R.S., and B.J. Curran. 1974. Growth and metabolic characteristics of *Mycobacterium paratuberculosis. Appl. Microbiol.* 28:276.

41. Smith, J.M.A., A.V. Quirk, R.W.H. Plank, F.M. Diffin, G.C. Ford, and P.M. Harrison. 1988. The identity of *Escherichia coli* bacterioferritin and cytochrome b_1. *Biochem. J.* 255:737.

42. Bauminger, E.R., S.G. Cohen, D.P.E. Dickson, A. Levy, S. Ofer, and J. Yariv. 1980. Mössbauer spectroscopy of *Escherichia coli* and its iron-storage proteins. *Biochim. Biophys. Acta* 623:237.

43. Matzanke, B.F., G.I. Muller, E. Bill, and A.X. Trautwein. 1989. Iron metabolism of *Escherichia coli* studied by Mössbauer spectroscopy and biochemical methods. *Eur. J. Biochem.* 183:371.

44. Ratledge, C. 1982. Nutrition, growth and metabolism. *In* The Biology of the Mycobacteria. C. Ratledge and J. Stanford, editors. Academic Press, London. 186–212.

45. Sritharan M, R.M. Hall, A.M. Messenger, and C. Ratledge. 1987. Iron transport in *Mycobacterium smegmatis*: occurrence of iron-regulated envelope proteins as potential receptors for iron uptake. *J. Gen. Microbiol.* 133:2107.

46. Dale, J.W., and A. Palki. 1990. Mycobacterial gene expression and regulation. *In* Molecular Biology of the Mycobacteria. J. McFadden, editor. Surrey University Press, London. 173–198.