High-Level Heterologous Expression and Secretion in Rapidly Growing Nonpathogenic Mycobacteria of Four Major *Mycobacterium tuberculosis* Extracellular Proteins Considered To Be Leading Vaccine Candidates and Drug Targets

GÜNTER HARTH, BAI-YU LEE, AND MARCUS A. HORWITZ*

Department of Medicine, School of Medicine, University of California, Los Angeles, Los Angeles, California 90095

Mycobacterium tuberculosis, the primary etiologic agent of tuberculosis, is the world's leading cause of death from a single infectious agent, and new vaccines and drugs to combat it are urgently needed. The major extracellular proteins of *M. tuberculosis*, which are released into its phagosome in macrophages, its host cells in humans, are leading candidates for a vaccine and prime targets for new drugs. However, the development of these biologicals has been hampered by the unavailability of large quantities of recombinant extracellular proteins identical to their native counterparts. In this report, we describe the heterologous expression and secretion of four major *M. tuberculosis* extracellular proteins (the 30-, 32, 16-, and 23.5-kDa proteins—the first, second, third, and eighth most abundant, respectively) in rapidly growing, nonpathogenic mycobacterial species. Multiple attempts to obtain secretion of the proteins by using *Escherichia coli*- and *Bacillus subtilis*-based expression systems were unsuccessful, suggesting that high-level expression and secretion of these *Mycobacterium*-specific proteins require a mycobacterial host. All four recombinant proteins were stably expressed from the cloned genes' own promoters at yields that were 5- to 10-fold higher than those observed for the native proteins. The four proteins were purified to apparent homogeneity from culture filtrates by ammonium sulfate precipitation and ion-exchange and molecular sieve chromatography. The recombinant proteins were indistinguishable from their native counterparts by multiple criteria. First, N-terminal amino acid sequence determination demonstrated that processing of the leader peptides was highly accurate. Second, sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis revealed identical migration patterns. Third, mass spectrometry analysis confirmed that differences in mass were ≤5 Da. A homolog of the *M. tuberculosis* 30-kDa protein was identified in *M. smegmatis* by means of DNA analyses and immunoscreening. This is the first time that secretion of recombinant *M. tuberculosis* extracellular proteins in their native form has been achieved. This study opens the door to mass production of correctly processed and secreted extracellular proteins of *M. tuberculosis* in a heterologous host and allows ready evaluation of their biologic and immunologic function.

Tuberculosis (TB) has been declared a global emergency by the World Health Organization, the first disease so designated. The World Health Organization has estimated that 30 million people will die worldwide from TB during the next 10 years. Compounding this tragedy is the worldwide emergence of multidrug-resistant strains of *Mycobacterium tuberculosis*, the primary causative agent of TB. This situation has created an urgent need for a vaccine to prevent TB and for new drugs to combat the disease.

Extracellular proteins of *M. tuberculosis* figure prominently in new vaccine and drug development. Extracellular proteins of *M. tuberculosis* are proteins that are secreted or otherwise released by the bacterium into its extracellular milieu. The major extracellular proteins are released by *M. tuberculosis* organisms growing either extracellularly in broth culture or intracellularly within a phagosome in host mononuclear phagocytes (8, 10, 12, 16).

Purified major extracellular proteins of *M. tuberculosis* are leading vaccine candidates. In previous studies, we have shown that such proteins are capable of inducing protective immunity against aerosol challenge with *M. tuberculosis* in the outbred guinea pig model of pulmonary tuberculosis (10). Guinea pigs are highly susceptible to TB and constitute the most relevant small animal model of TB because they develop a disease that closely mimics TB in humans (10). Purified major extracellular proteins are the only proteins and the only defined molecules yet demonstrated to induce protective immunity. Among the most promising of the major extracellular proteins as vaccine candidates are the first, second, and third most abundant, the 30 (also called antigen 85B)-, 32 (also called antigen 85A)-, and 16-kDa proteins, respectively, and the eighth most abundant, the 23.5-kDa protein (8, 10).

Major extracellular proteins are also prime targets for drug development. Many of these proteins have been found to have enzyme activity, including the 58-kDa protein, glutamine synthetase (7); the 23-kDa protein, superoxide dismutase (25); and the 30- and 32-kDa complex of proteins, mycolyl transferase (2). We have found (7a) that selective inhibition of the extracellular enzymatic activity of one such major extracellular protein, glutamine synthetase, a protein released by pathogenic mycobacterial species but not nonpathogenic ones, inhibits *M. tuberculosis* multiplication both in broth culture and within human macrophages.

Because *M. tuberculosis* is both exceptionally slow growing and biohazardous, the evaluation of major extracellular pro-

* Corresponding author. Mailing address: Department of Medicine, School of Medicine, University of California, Los Angeles, 10833 Le Conte Avenue, Los Angeles, CA 90095. Phone: (310) 206-0074. Fax: (310) 794-7156.

TABLE 1. Bacterial strains and plasmids used for expression studies

Strain or plasmid[a]	Genotype[d]	Growth medium or marker(s)[e]	Source or reference
E. coli			
BL21 (DE3) pLysS	λDE3 lysogen; F⁻ *ompT* (r_B⁻ m_B⁻)	LB broth	Novagen
DH5α	F⁻ *endA1 hsdR17* (r_K⁻ m_K⁺) *supE44 thi-1* λ⁻ *recA1 gyrA96 relA1 deoR* φ80d*lacZ* ΔM15Δ(*lacZYA-argF*)U169	LB broth	GIBCO/BRL
GI698	F⁻ λ⁻ *lacI*q *lacPL8 ampC*::P_trp *c*I	LB broth	Invitrogen
GI724	F⁻ λ⁻ *lacI*q *lacPL8 ampC*::P_trp *c*I	LB broth	Invitrogen
JM109	F′ (*traD36 proAB*⁺ *lacI*q *lacZ*ΔM15) *hsdR17* (r_K⁻ m_K⁺) *endA1recA1supE44 thi-1 gyrA96 relA1* Δ(*lac-proAB*)	LB broth	Invitrogen
B. subtilis BD168	*trpC2*	VY broth	6
M. smegmatis 1-2c[b]		7H9 broth	5
M. vaccae R877R-NCTC11659[b]		7H9 broth	5
Cloning vehicles			
pET20	ColE1 *ori*; pT7	Amp^r	Invitrogen
pET22b(+)	ColE1 *ori*; pT7	Amp^r	Invitrogen
pKK233-2[c]	ColE1 *ori*; p*trc*	Amp^r	1
pRSET-A	ColE1 *ori*; pT7	Amp^r	Invitrogen
pTrxFus	ColE1 *ori*; pT7	Amp^r	Invitrogen
pKH80	*B. subtilis*-*E. coli* shuttle vector; p*eryR*	Amp^r, Ery^r, Tet^r	6
pPL608	pUB110 *ori*; pSPO2	Cam^r, Neo^r	23
pSMT3[b]	*Mycobacterium*-*E. coli* shuttle vector; p*hsp60*	Hyg^r	5

[a] The following were obtained from the American Type Culture Collection: strain BD168 (ATCC 27689), plasmid pKH80 (ATCC 37093), and plasmid pPL608 (ATCC 37108).
[b] Provided by Peadar O'Gaora, Imperial College School of Medicine at St. Mary's, London, United Kingdom.
[c] Provided by Andrew Campbell, Brown University, Providence, R.I.
[d] Genetic markers are abbreviated as follows: *ori*, replication origin; p*eryR*, erythromycin resistance gene promoter; p*hsp60*, 60-kDa heat shock protein promoter; pSPO2, phage SPO2 promoter; pT7, bacteriophage T7 promoter; p*trc*, hybrid *trp-lac* promoter.
[e] Antibiotic resistance markers are abbreviated as follows: Amp, ampicillin; Cam, chloramphenicol; Ery, erythromycin; Hyg, hygromycin; Neo, neomycin; Tet, tetracycline.

teins in a vaccine trial and as drug targets imposes a practical requirement that they be produced in recombinant form in a rapidly growing, nonpathogenic host. However, while expression of these proteins as insoluble intracellular proteins or as secreted fusion proteins has been achieved by us and others (8, 11, 14, 25), high-level expression and secretion of these proteins in native form have proved difficult. Indeed, we shall describe below several unsuccessful attempts to do so by using *Escherichia coli* and *Bacillus subtilis* as host cells. In this paper, we shall demonstrate for the first time the high-level expression and secretion in native form of recombinant major extracellular proteins of *M. tuberculosis* by using *Mycobacterium smegmatis* and *Mycobacterium vaccae*, two rapid growing nonpathogenic mycobacteria, as hosts. We shall also report the sequence of the genes encoding the 16- and the 23.5-kDa *M. tuberculosis* Erdman major extracellular proteins, which have not been previously reported.

MATERIALS AND METHODS

Bacterial strains and plasmids. All of the bacterial strains used in this study, their relevant genotypes and growth conditions, and all of the wild-type plasmids used for general cloning purposes are listed in Table 1. All of the recombinant plasmids assayed for expression of the 30-, 32-, 16-, and 23.5-kDa extracellular *M. tuberculosis* Erdman proteins and their relevant genetic features are listed in Table 2.

Cloning procedures and DNA analyses. The isolation of high-molecular-weight genomic DNA of *M. tuberculosis* Erdman by hot-phenol extraction and the procedures for cloning and sequencing of the genes encoding the 30- and 32-kDa proteins have already been described (8). The 16- and 23.5-kDa protein genes were cloned by PCR methodology with primer pairs 1 and 2 (for map positions of the primers on the respective genes, see Fig. 1), which were based on the N-terminal amino acid sequence of the 16-kDa protein (10) and on the published DNA sequence of the 23.5-kDa protein (MPT64) from *M. tuberculosis*

H37Rv (15). The PCR products were then used to identify the corresponding genes from genomic *M. tuberculosis* Erdman DNA by Southern hybridization. Both genes, including extensive flanking regions, were cloned into pUC19 (24). The coding regions for the 30- and 32-kDa proteins are ~1 kbp long and are contained within an ~4.7-kbp *Hin*dIII-*Bam*HI genomic restriction fragment (30-kDa protein) or within an ~4.5 kbp *Eco*RV genomic DNA fragment (32-kDa protein), both of which were completely sequenced by the chain termination method (18). The coding region for the 16-kDa protein is ~0.5 kb long and is contained within an ~5.5 kbp *Bam*HI restriction fragment whose DNA sequence was also completely determined. The coding region for the 23.5-kDa protein is ~0.7 kb long and is contained within an ~9.5-kbp *Bam*HI genomic restriction fragment. This DNA fragment was not completely sequenced; rather, after establishing the DNA sequence of the region encoding the 23.5-kDa protein, we extended the sequence analysis into the 5′ and 3′ flanking regions and used this information to amplify and clone a stretch of ~1.5 kbp, encompassing the 23.5-kDa protein gene, into pUC19. All DNA sequences were analyzed by the GCG sequence analysis software package, version 7, from the University of Wisconsin, Madison.

All recombinant constructs assayed for expression of the 30-, 32-, 16-, and 23.5-kDa proteins were derived from the original, recombinant pUC19 plasmids by direct subcloning of internal restriction fragments or by generating an amplification product flanked by restriction enzyme sites to allow directional ligation of the structural genes encoding either the mature or the full-length protein into the appropriate expression vector sites as listed in Table 2. Constructs in pSMT3 were established in *E. coli* DH5α and then electroporated into *M. smegmatis* 1-2c as previously described (5). For each construct, at least two clones were characterized by restriction enzyme mapping and DNA sequence analysis.

Purification and characterization of recombinant 30-, 32-, 16-, and 23.5-kDa proteins. Expression of recombinant proteins was initially monitored by taking samples of the bacterial cultures at various time points (hourly for up to 8 h after induction for the *E. coli* and *B. subtilis* strains and every 12 h for up to 3 days for the *M. smegmatis* 1-2c cultures), separating culture filtrates and cell pellets by centrifugation and filtration, and analyzing the two fractions on 12.5% denaturing polyacrylamide gels (National Diagnostics). Subsequently, cultures were grown for a period of time at which expression of recombinant proteins was maximal. Extracellular and cell-associated recombinant proteins expressed in *M. smegmatis* 1-2c cultures were purified as previously described (10) and charac-

TABLE 2. Constructs used to assay for expression of recombinant *M. tuberculosis* 30-, 32-, 16-, 23.5-kDa proteins

Protein and plasmid[a]	Host strain	Insert[b]	Expression[c]	Secretion[c]	% Mature protein[d]	Yield (mg/liter) of purified protein
30 kDa						
pET20	*E. coli* BL21	M 30 kDa; 0.85 kb (PCR); *Nde*I/*Eco*RI; →	−	−		
pET22b(+)	*E. coli* BL21	FL 30 kDa; 0.98 kb (PCR); *Nde*I/*Eco*RI; →	+ (fusion)	−	40 (+2 aa)	1
pKK233-2	*E. coli* JM109	M 30 kDa; 0.85 kb (PCR); *Nco*I/*Hin*dIII; →	−	−		
pRSET-A	*E. coli* BL21	M 30 kDa; 0.85 kb (PCR); *Bam*HI/*Eco*RI; →	+ (fusion)	−	10 (+2 aa)	<1
pTrxFus	*E. coli* GI724/698	M 30 kDa; 0.85 kb (PCR); *Kpn*I/*Bam*HI; →	+ (fusion)	−	10 (+2 aa)	<1
pKH80	*B. subtilis* BD168	M 30 kDa; 0.85 kb (PCR); *Bcl*I; →	−	−		
pSMT3*	*M. smegmatis* 1-2c	FL 30 kDa; 4.7 kb (gene); *Hin*dIII/*Bam*HI; ←	+	+	95	2.5
pSMT3*	*M. vaccae* R877R	FL 30 kDa; 4.7 kb (gene); *Hin*dIII/*Bam*HI; ←	+	+	95	2.5
32 kDa						
pET20	*E. coli* BL21	M 32 kDa; 0.9 kb (PCR); *Nde*I/*Eco*RI; →	−	−		
pET22b(+)	*E. coli* BL21	FL 32 kDa; 1.25 kb (PCR); *Nde*I/*Eco*RI; →	+	−	90 (+2 aa)	1
pKK233-2	*E. coli* JM109	FL 32 kDa; 1.25 kb (PCR); *Nco*I/*Hin*dIII; →	−	−		
pPL608	*B. subtilis* BD168	FL 32 kDa; 0.9 kb (PCR); *Sma*I/*Hin*cII; →	−	−		
pSMT3	*M. smegmatis* 1-2c	FL 32 kDa; 4.5 kb (gene); *Eco*RV; →	+ (28°C)	+		
pSMT3	*M. smegmatis* 1-2c	FL 32 kDa; 4.5 kb (gene); *Eco*RV; ←	+ (28°C)	+	95	
pSMT3	*M. smegmatis* 1-2c	FL 32 kDa; 1.5 kb (PCR); *Bam*HI; →	−	−		
pSMT3*	*M. smegmatis* 1-2c	FL 32 kDa; 1.5 kb (PCR); *Bam*HI; ←	+ (28°C)	+	99	2.5
pSMT3	*M. smegmatis* 1-2c	FL 32 kDa; 1.25 kb (PCR); *Bam*HI; →	−	−		
pSMT3	*M. smegmatis* 1-2c	FL 32 kDa; 1.25 kb (PCR); *Bam*HI; ←	+ (fusion; 28°C)	+	90	
pSMT3	*M. vaccae* R877R	FL 32 kDa; 4.5 kb (gene); *Eco*RV; →	+	+ (10%)		
pSMT3	*M. vaccae* R877R	FL 32 kDa; 4.5 kb (gene); *Eco*RV; ←	−	−		
16 kDa						
pSMT3*	*M. smegmatis* 1-2c	FL 16 kDa; 5.5 kb (gene); *Bam*HI; ←	+	+	99	2.5
pSMT3*	*M. vaccae* R877R	FL 16 kDa; 5.5 kb (gene); *Bam*HI; ←	+	+	99	2.5
23.5 kDa						
pSMT3*	*M. smegmatis* 1-2c	FL 23.5 kDa; 1.4 kb (PCR); *Bam*HI; →	+	+	99	7.5

[a] Constructs described in the text are in boldface. For information on plasmids and host strains, see Table 1. Plasmids marked with asterisks were designated pMTB30, pMTB32A, pMTB16, and pMTB23.5 for the 30-, 32-, 16-, and 23.5-kDa proteins, respectively.

[b] The description of each insert includes (i) the designation of the structural gene encoded, preceded by the letters FL to indicate that the gene encodes the full-length protein, including the leader peptide, or the letter M to indicate that the gene encodes only the mature protein; (ii) the size of the insert followed, in parentheses, by the word "gene" to indicate that the insert is the original genomic locus or by the abbreviation "PCR" to indicate that the insert is an amplification product of the original genomic locus; (iii) the restriction enzyme site(s) at the vector-insert junction(s); and (iv) the orientation of the insert with regard to the plasmid-encoded promoter of heat shock protein 60.

[c] Expression and secretion of the proteins are indicated by a plus sign, and lack thereof is indicated by a minus sign. Expression of a fusion protein and/or expression only at 28°C is indicated in parentheses in the Expression column. Fusion proteins contain vector-encoded amino acids (between the cloning site and the first ATG) at their N termini. Limitation of secretion to a minority of the expressed protein is indicated in parentheses in the Secretion column.

[d] N termini of mature proteins were determined by amino acid sequence analysis; +2 aa indicates that the mature protein contains two additional N-terminal amino acids.

terized by N-terminal amino acid sequencing (performed at the UCLA Protein Microsequencing Facility) and immunoblotting.

Identification of homologs. The search for homologs to the four *M. tuberculosis* Erdman genes in *M. smegmatis* 1-2c was performed by gene amplification, DNA hybridization, and immunoblotting.

(i) Gene amplification. PCR-based analyses to detect homologs of the *M. tuberculosis* 30-, 32-, 16-, and 23.5-kDa protein genes in *M. smegmatis* used primers 1 and 2 for the 30-, 32-, and 23.5-kDa protein genes and primers 3 and 4 for the 16-kDa protein (Fig. 1) for 40 rounds of amplification at 94, 42, and 70°C with 1 μg of genomic *M. tuberculosis* and *M. smegmatis* DNA and 1 to 2 U of *Taq* DNA polymerase (Fisher, Perkin Elmer). All PCR products were analyzed by agarose gel electrophoresis and visualized by UV transillumination.

(ii) DNA hybridization. Southern DNA hybridizations were performed by using ~5 μg of genomic *M. tuberculosis* and *M. smegmatis* DNA and oligonucleotide probes (primer 3 for the 30- and 32-kDa protein genes, primers 3 and 4 for the 16-kDa protein gene, and primers 1 and 2 for the 23.5-kDa protein gene) radiolabelled to ~5 × 10⁷ to 1 × 10⁸ cpm/μg. Nitrocellulose filters were kept overnight at 50 or 42°C in 5× SSC (1× SSC is 150 mM NaCl plus 15 mM Na citrate) or in Church hybridization buffer (500 mM Na phosphate [pH 7.2], 7% sodium dodecyl sulfate, 1 mM EDTA [3]) and subsequently washed, dried, and autoradiographed on Kodak X-Omat AR5 film for 48 to 72 h at −70°C with a Cronex Lightning-Plus intensifying screen (Du Pont).

(iii) Immunoblotting. Culture filtrates and cell pellets of axenically grown *M. tuberculosis* Erdman and *M. smegmatis* 1-2c were prepared and electrophoresed as described above, transferred to nitrocellulose membranes, and probed with antibodies (diluted 1:5,000) specific for the 30/32-kDa protein complex, the 16-kDa protein, or the 23.5-kDa protein. The antibodies were generated in

rabbits, had reciprocal titers between 2,500 and 10,000, and were shown to recognize native and denatured forms of the four proteins equally well. Following incubation with alkaline phosphatase-conjugated goat anti-rabbit antibodies, proteins were visualized by staining with alkaline phosphatase-specific color development reagents (Bio-Rad).

Reagents and enzymes. All chemicals were purchased from Sigma unless indicated otherwise and were of the highest grade possible. Enzymes were purchased from either New England Biolabs, Fisher/Promega, or Appligene. Oligonucleotides for amplification reactions and DNA sequencing were purchased from either BioSynthesis or GIBCO/BRL, and radiolabelled nucleotides were obtained from Amersham or New England Nuclear.

Nucleotide sequence accession numbers. The nucleotide sequences of the four genes examined in this study are available from GenBank under accession numbers MTB30-8 U38939 (30-kDa protein gene), MTB32-2 U47335 (32-kDa protein gene), U82234 (16-kDa protein gene), and U82235 (23.5-kDa protein gene).

RESULTS

Determination of the DNA sequences encoding the 30-, 32-, 16-, and 23.5-kDa proteins and their genetic loci. We have previously reported the cloning and sequencing of the genes encoding the 30- and 32-kDa proteins, including extensive flanking DNA sequences (8). The DNA fragments containing the genes for the 16- and 23.5-kDa proteins and extensive flanking regions were isolated from high-molecular-weight
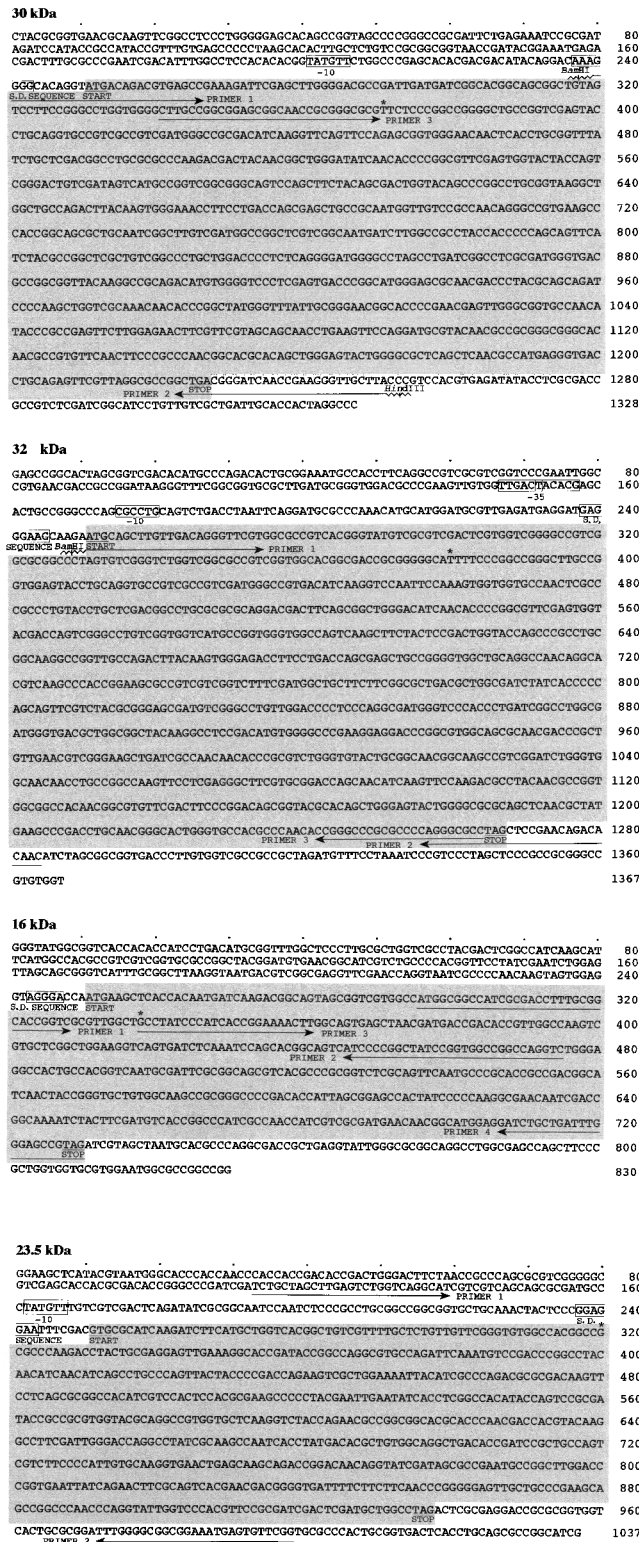
30 kDa

```
CTACGCGGTGAACGCAAGTCTGGCCTCCCTGGGGGAGCACAGCCGGTAGCCCCGGGCCGCGATTCTGAGAAATCCGCGAT   80
AGATCCATACCGCCATACCGTTTGTGAGCCCCTTAAGCACACTTGCTCTGTCCGCGGCGGTAACCGATACGGAAATGAGA  160
CGACTTTGCGCCCGAATCGACATTTGGCCTCCACACACGGTATGTTCTGGCCCGAGCACACGACGACATACAGGACAAAG  240
                                         -10
GGGCACAGGTATGACAGACGTGAGCCGAAAGATTCGAGCTTGGGGACGCCGATTGATGATCGGCACGGCAGCGGCTGTAG  320
S.D. SEQUENCE START                          PRIMER 1
TCCTTCCGGGCCTGGTGGGGCTTGCCGGCGGAGCGGCAACCGCGGGCGCGTTCTCCCGGCCGGGGGCTGCCGGTCGAGTAC  400
                                         PRIMER 3
CTGCAGGTGCCGTCGCCGTCGATGGGCCGCGACATCAAGGTTCAGTTCCAGAGCGGTGGGAACAACTCACCTGCGGTTTA  480
TCTGCTCGACGGCCTGCGCGCCCAAGACGACTACAACGGCTGGGATATCAACCACCCCGGCGTTCGACGTGGTACTACCAGT  560
CGGGACTGTCGATAGTCATGCCGGTCGGCGGGCAGTCCAGCTTCTACAGCGACTGGTACAGCCCGGCCTGCCGGTAAGGCT  640
GGCTGCCAGACTTACAAGTGGGAAACCTTCCTGACCAGCGAGCTGCCGCAATGGTTGTCCGCCAACAGGGCCGTGAAGCC  720
CACCGGCAGCGCTGCAATCGGCTTGTCGATGGCCGGCTCGTCGGCAATGATCTTGGCCGCCTACCACCCCCAGCAGTTCA  800
TCTACGCCGGCTCGCTGTCGGCCCTGCTGGACCCCCTCAGGGGATGGGGCCTAGCCTGATCGGCCTCGCGATGGGTGAC  880
GCCGGCGGTTACAAGGCCGCAGACATGTGGGGTCCCTCGAGTGACCCGGCATGGGAGCGCAACGACCCTACGCAGCAGAT  960
CCCCAAGCTGGTCGCAAACAACACCCGGCTATGGGTTTATTGCGGGAACGGCACCCCGAACGAGTTGGGCGGTGCCAACA 1040
TACCCGCCGAGTTCTTGGAGAACTTCGTTCGTAGCAGCAACCTGAAGTTCCAGGATGCGTACAACGCCCGCGGGCGGGCAC 1120
AACGCCGTGTTCAACTTCCCCAACGGCACGCACGCAGCTGGGAGTACTGGGGCGCTCAGCTCAACGCCATGAGGGTGAC 1200
CTGCAGAGTTCGTTAGGCGCCGGCTGACGGGATCAACCGAAGGGTTGCTTACCCGGTCCACGTGAGATATACCTCGCGACC 1280
        PRIMER 2       STOP                      H.I.H.I.L.1
GCCGTCTCGATCGGCATCCTGTTGTCGCTGATTGCACCACTAGGCCC                                  1328
```

32 kDa

```
GAGCCGGCACTAGCGGTCGACACATGCCCAGACACTGCGGAAATGCCACCTTCAGGCCGTCGCGTCGGTCCCGAATTGGC   80
CGTGAACGACCGCCGGATAAGGGTTTCGGCGGTGCGCTTGATGCGGGTGGACGCCCGAAGTTGTGGTTGACTACACGAGC  160
                                                       -35
ACTGCCGGGCCCAGCGCCTGCAGTCTGACCTAATTCAGGATGCGCCCAAACATGCATGGATGCGTTGAGATGAGGATGAG  240
                -10                                                S.D.
GGAAGCAAGAATGCAGCTTGTTGACAGGGTTCGTGGCGCCGTCACGGGTATGTCGCGTCGACTCGTGGTCGGGGCCGTCG  320
SEQUENCE BamHI START            PRIMER 1
GCGCGGCCCTAGTGTCGGGTCTGGTCGGCGCCGTCGGTGGCACGGCGACCGCGGGGGCATTTTCCCGGCCGGGCTTGCCG  400
GTGGAGTACCTGCAGGTGCCGTCGCCGTCGATGGGCCGTGACATCAAGGTCCAATTCCAAAGTGGTGGTGCCAACTCGCC  480
CGCCCTGTACCTGCCTGCAGCCTGCGCGCGCAGGACGACTTCAGCGGCTGGGACATCAACCACCCCGGCGTTCGAGTGGT  560
ACGACCAGTCGGGCCTGTCGGTGGTCATGCCGGTGGGTGGCCAGTCAAGCTTCTACTCCGACTGGTACCAGCCCGCCTGC  640
GGCAAGGCCGGTTGCCAGACTTACAAGTGGGAGACCTTCCTGACCAGCGAGCTGCCGGGGTGGCTGCAGGCCAACAGGCA  720
CGTCAAGCCCACCGGAAGCGCCGTCGTCGGTCTTTCGATGGCTGCTTCTTCGGCGCTGACGCTGGCGATCTATCACCCCG  800
AGCAGTTCGTCTACGCGGGAGCGATGTCGGGCCTGTTGGACCCCTCCCAGGCGATGGGTCCCACCCTGATCGGCCTGGCG  880
ATGGGTGACGCTGGCGGCTACAAGGCCTCCGACATGTGGGGCCCGAAGGAGGACCCGGCTGGCAGCGCAACGACCCGCT  960
GTTGAACGTCGGGAAGCTGATCGCCAACAACACCCGCTGTGGGTGTACTGCGGCAACGGCAAGCCGTCGGATCTGGGTG 1040
GCAACAACCTGCCGGCCAAGTTCCTCGAGGGCTTCGTGCGGACCAGCAACATCAAGTTCCAAGACGCCTACAACGCCGGT 1120
GGCGGCCACAACGGCGTGTTCGACTTCCCCGACAGCGGTACGCACAGCTGGGAGTACTGGGGCGCGCAGCTCAACGCTAT 1200
GAAGCCCGACCTGCAACGGGCACTGGGTGCCACGCCCAACACCGGGCCCGCGCCCCAGGGCGCCTAGCTCCGAACAGACA 1280
                         PRIMER 3        PRIMER 2       STOP
CAACATCTAGCGGCGGTGACCCTTGTGGTCGCCGCCGTAGATGTTCCTAAATCCCGTCCCTAGCTCCCGCCGCGGGCC 1360
GTGTGGT                                                                          1367
```

16 kDa

```
GGGTATGGCGGTCACCACACCATCCTGACATGCGGTTTGGCTCCCTTGCGCTGGTCGCCTACGACTCGGCCATCAAGCAT   80
TCATGGCCACGCCGTCGTCGGTGCGCCGGCTACGGATGTGAACGGCATCGTCTGCCCCACGGTTCCTATCGAATCTGGAG  160
TTAGCAGCGGTCATTTGCGGCTTAAGGTAATGACGTCGGCGAGGTTCGAACCAGGTAATCGCCCCAACAAGTAGTGGAG  240
GTAGGGACCAATGAAGCTCACCACAATGATCAAGACGGCAGTAGCGGTCGTGGCCATCGGCAAGCCCTTTGCGG         320
S.D. SEQUENCE START
CACCGGTGCGGTTGGCTGCCTATCCCATCACCGGAAAACTTGGCAGTGAGCTAACGATGACCGACACCGTTGGCCAAGTC  400
                                         PRIMER 2
GTGCTCGGCTGGGAAGGTCAGTGATCTCAAATCCAGCACGGCAGTCATCCCCGGCTATCCGGTGGCCGGCCAGGTCTGGA  480
                                         PRIMER 3
GGCCACTGCCACGGTCAATGCGATTCGCGGCCAGCGTCACGCCCGCGGTCTCGCAGTTCAATGCCCGCCACCGCCGACGGCA  560
TCAACTACCGGGTGCTGTGGCAAGCCGCGGGCCCCGACACCATTAGCGGAGCCACTATCCCCAAGGCGAACAATCGACC  640
GGCAAAATCTACTTCGATGTCACCGGCCCATCGCCAACCATCGTCGCGATGAACAACGGCATGGAGGATCTGCTGATTTG  720
                                                        PRIMER 4
GAGCCGTAGATCGTAGCTAATGCACGCCCAGGCGACCGCTGAGGTATTGGGCGCGGCAGGCCTTGGCGAGCCAGCTTCCC  800
        STOP
GCTGGTGGTGCGCGTGGAATGGCGCCGGCCCGG                                                830
```

23.5 kDa

```
GGAAGCTCATACGTAATGGGCACCCACCAACCCACCACCGACACCGACTGGGACTTCTAACCGCCCAGCGCGTCGGGGGC   80
GTCGAGCACCACGCGACACCGGCCCGATCGATCTGCTAGCCTTGAGTCTGGTCAGGCATCGTCGTCAGCAGCGCGATGCC  160
                                                        PRIMER 1
CTATGTTTGTCGTCGACTCAGATATCGCGGCAATCCAATCTCCCGCCTGCGGCCGGCGGTGCCGTGCAAACTACTCCCGGAG  240
 -10                                                             S.D.
GAAGTTCGACGTGCGCATCAAGATCTTCATGCTGGTCACGGCTGTCGTTTTGCTCTGTTGTTCGGGTGTGGCCACGGGGAC  320
SEQUENCE START
CGCCCAAGACCTACTGCGAGGAGTTGAAAGGCACCGATACCGGCCAGGCGTGCCAGATTCAAATGTCCGACCCGGCCTAC  400
AACATCAACATCAGCCTGCCCAGTTACTACCCCGACCAGAAGTCGCTGGAAAATTACATCGCCCAGACGCGCGACAAGTT  480
CCTCAGCGCGGCCACATCGTCCACTCCACGCGAAGCCCCCTACGAATTGAATATCACCTCGGCCACATACCAGTCCGCGA  560
TACCGCCGCGTGGTACGCAGGCCGTGGTGCTCAAGGTTACCAGAACGCCGGCGGCACGCACCCAACGACCACGTACAAG  640
GCCTTCGATTGGGACCAGGCCTATCGCAAGCCAATCACCTATGACACGCTGTGGCAGGCTGACACCGATCCGCTGCCAGT  720
CGTCTTCCCCATTGTGCAAGGTGAACTGAGCAAGCAGACCGGACAACAGGTATCGATAGCGCCGAATGCCGGCTTGGACC  800
CGGTGAATTATCAGAACTTCGCAGTCACGAACGACGGGGTGATTTTCTTCTTCAACCCGGGGAGTTGCTGCCCGAAGCA  880
GCCGGCCCAACCCAGGTATTGGTCCCACGTTCCGCGATCGACTCGATGCTGGCCTAGACTCGCGAGGACCGCGCGGTGGT  960
CACTGCGCGGATTTGGGGGCGGCGAAATGAGTGTTCGGTGCGCCCACTGCGGTGACTCACCTGCAGCGCCGGCATCG    1037
            PRIMER 2
```

FIG. 1. DNA sequences of the 30-, 32-, 16-, and 23.5-kDa protein genes. The sequences show the 5′ to 3′ sense DNA strands of the four structural genes (shaded box). The initiator methionine codons of the genes are underlined and marked START. The first codon of each mature protein is indicated by an asterisk. The termination codons are underlined and marked STOP. The structural genes are flanked by 250 nucleotides at the 5′ end and 100 nucleotides at the 3′ end (unshaded). Sequences are further annotated by boxes matching Shine-Dalgarno sequences (S.D. SEQUENCE) and boxes indicating tentatively assigned −10 and −35 promoter elements. Furthermore, the map positions of

genomic *M. tuberculosis* Erdman DNA by PCR methodology and Southern hybridization. Figure 1 shows the DNA sequences of the coding strands in 5′ to 3′ direction for the four protein genes plus the 250 bp upstream and the 100 bp downstream of the structural genes.

The genetic loci for three of the four genes are known. The 30-kDa protein (also designated FbpB) gene maps to cosmid Y207, the 32-kDa protein (also designated FbpA) gene maps to cosmids Y430 and Y302, and the 23.5-kDa protein (also designated MPT64) gene maps near the 3′ end of cosmid Y39. No map position has been described for the 16-kDa protein gene, although a homology search, using the entire DNA sequence of the 5.5-kb *Bam*HI DNA fragment containing the 16-kDa protein gene, shows limited sequence identity with cosmid tbc2 (17).

Analysis of DNA sequences flanking the structural genes of the 30-, 32-, 16-, and 23.5-kDa proteins. (i) G+C content. As with most protein-coding genes of *M. tuberculosis*, the overall G+C content of the genes encoding the 30-, 32-, 16-, and 23.5-kDa proteins, including their leader peptides, is ∼65% and the codon usage is very similar to that described for mycobacteriophage L5 (9).

(ii) Leader peptides. All four proteins are preceded by leader peptides rich in hydrophobic amino acids; the hydrophobicity ranges from 65 to 75%. The lengths of the leader peptides vary greatly—40 amino acids for the 30-kDa protein, 43 amino acids for the 32-kDa protein, most likely 29 amino acids for the 16-kDa protein, and 23 amino acids for the 23.5-kDa protein. The start site of the leader peptide for the 16-kDa protein is uncertain, since the DNA sequence upstream of the mature protein contains three in-frame methionine codons (Fig. 1). However, only the methionine codon at position −29 with regard to the alanine residue in the +1 position of the mature protein is preceded by a strong Shine-Dalgarno sequence, GGAGG, 14 bp upstream of the ATG codon (20).

The start site of the leader peptide for the 23.5-kDa protein is also uncertain (Fig. 1). The GTG methionine codon 23 amino acids upstream of the first residue of the mature protein, alanine, is preceded 14 bp upstream by a stronger Shine-Dalgarno sequence (GGAGGAA) than the in-frame ATG methionine codon 17 amino acids upstream of the start of the mature protein, which is preceded 9 bp upstream by AAGA. Interestingly, the leader peptide of the 23.5-kDa protein contains two tandemly arranged cysteine residues, which are infrequently found in leader peptides. The site of leader peptide cleavage was determined by sequencing the N termini of several preparations of each of the four proteins. In each case, both the −1 and −3 positions are occupied by alanine residues, consistent with the von Heijne rule for prokaryotic leader peptides (21).

(iii) Promoter elements. Apart from the Shine-Dalgarno motif present within 20 bp upstream of each of the four structural genes, the 5′ flanking regions are devoid of extended domains of DNA sequence similarity. Moreover, standard *E. coli*-type promoter elements are not present within 250 bp upstream of the structural genes (13). Based on mRNA and primer extension analyses, we previously described the motif TATGTT −47 bp upstream of the 30-kDa protein gene as a potential candidate for the −10 promoter element (8). Other

the primers used for PCR amplifications and Southern blot analyses are indicated by arrows. Wavy lines at the 5′ ends of primers indicate 10 nucleotides consisting of 4 randomly selected nucleotides followed by 6 nucleotides specifying the recognition site of the restriction enzyme appearing as a superscript.

investigators have described promoter elements upstream of the 32-kDa protein gene and shown that, despite their apparent similarity at the DNA and protein levels, the genes encoding the 30- and 32-kDa proteins do not have any regulatory elements in common in their 5′ flanking regions (11). The transcriptional start sites for the 16- and 23.5-kDa protein genes have not been mapped. The DNA sequences upstream of the 16-kDa protein gene show very little similarity to the upstream sequences of the other three genes. Interestingly, the aforementioned −10 promoter element of the 30-kDa protein gene is also found ~87 bp upstream of the 23.5-kDa protein gene but there is no apparent similarity with regard to a potential −35 promoter element.

(iv) Transcription units. The genes encoding the 30- and 32-kDa proteins apparently represent single transcription units (8, 11). Transcript analyses of the 16- and 23.5-kDa protein genes have not been performed. The DNA sequences of the 3′ flanking regions of the four genes differ greatly, and none of the typical transcription termination elements described for *E. coli* mRNA, such as extensive hairpin loops and runs of five or six uracil residues (4), or other consensus elements have been identified.

Description of constructs used to express recombinant 30-, 32-, 16-, and 23.5-kDa proteins. We first focused on optimizing the expression of the two most abundant extracellular proteins, the 30- and 32-kDa proteins. As seen in Tables 1 and 2, while expression and leader peptide processing of these two proteins are possible in an *E. coli*-based expression system, the yield of soluble, correctly processed recombinant protein is far greater in the two mycobacterial systems used—*M. smegmatis* 1-2c and *M. vaccae* R877R. The fact that the mycobacterial systems allowed the use of DNA inserts in the mycobacterium-*E. coli* shuttle vector pSMT3 that were much larger than the actual genes encoding these two *M. tuberculosis* proteins suggests that the mycobacterial transcription apparatus recognized and utilized the proper promoter elements upstream of the structural genes and that the mycobacterial secretion machinery very efficiently released the processed proteins into the extracellular milieu. Since the two mycobacterial systems proved highly successful for expression of the 30- and 32-kDa proteins, we also subcloned the genes for the 16- and 23.5-kDa proteins in this vector. The resulting recombinant plasmids were designated pMTB30, pMTB32A, pMTB16, and pMTB23.5, respectively (Table 2).

Several important issues remain unresolved. First, we do not know why constructs containing the 32-kDa protein gene express recombinant protein only when inserted in one orientation, although the DNA insert can be cloned and stably maintained in both orientations. Second, we do not know why expression of the 32-kDa protein is optimal at 28°C and dramatically decreases at 37°C, a temperature at which the other recombinant proteins are produced in abundance. Finally, we do not know why the *M. tuberculosis* 23.5-kDa DNA insert can be cloned and stably maintained in recombinant pSMT3 constructs in only one orientation with respect to the plasmid-encoded promoter for the mycobacterial 60-kDa heat shock protein.

Analysis of expression of recombinant proteins in *M. smegmatis* 1-2c. As discussed above, the mycobacterial expression systems proved far superior to the *E. coli*-based systems with regard to yield, accuracy, and efficiency of leader peptide processing and extracellular release of the four recombinant *M. tuberculosis* proteins. Since we observed no significant differences between *M. smegmatis* 1-2c and *M. vaccae* R877R in the expression levels of the four recombinant proteins, we selected

one of these expression systems, *M. smegmatis* 1-2c, for studies aimed at optimizing expression of the recombinant proteins.

We initially investigated expression of the recombinant proteins during different phases of growth of the bacteria, from the early log phase (~2.5×10^6 cells per ml) to the stationary phase (~2.5×10^9 to 5×10^9 cells per ml), occurring after 3 days of growth at 37°C (except for cultures expressing the 32-kDa protein, which were grown at 28°C). The observed yields of each of the four expressed and released recombinant proteins were strictly dependent on the total cell mass and increased in parallel with the bacterial growth curve. Bacteria expressing a recombinant 30-, 16-, or 23.5-kDa protein could even be held at the stationary phase for 1 to 2 days with no detectable loss or degradation of the released recombinant protein. Growth of bacteria expressing the recombinant 32-kDa protein was slower at the lower growth temperature of 28°C; however, the yield of expressed and released 32-kDa protein was comparable to that of the 30-kDa protein, provided the cultures were harvested at the late logarithmic phase.

The constitutive expression of all four recombinant proteins did not lead to intracellular accumulation of these proteins. The secretion machinery was apparently able to keep pace with protein expression and very efficiently released large quantities of the recombinant proteins into the culture medium. For each of the four proteins, the amount of expressed protein that remained in the intracellular milieu never exceeded 10%.

Purification of extracellular recombinant 30-, 32-, 16-, and 23.5-kDa proteins. The recombinant proteins were isolated from *M. smegmatis* 1-2c culture filtrates by ammonium sulfate precipitation and purified by ion-exchange and molecular sieve chromatography as described for the native proteins (10). All purified proteins remained soluble in aqueous buffers and did not show even trace contamination with other proteins by sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis. Each preparation was essentially free of endotoxin as determined by the *Limulus* amoebocyte lysate assay (<0.025 endotoxin U/mg or <5 pg of endotoxin/mg).

Purified recombinant proteins were analyzed by four techniques, the results of which are presented in Fig. 2. (i) N-terminal amino acid sequence determination demonstrated that >95 to 99% of the released recombinant proteins had an N-terminal amino acid sequence identical to that of the native molecule. Where different processing of the released recombinant proteins was observed, it involved removal of the leader peptide plus the first amino acid or second two amino acids of the mature protein. (ii) Sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis showed that each recombinant protein migrated at the same rate as its native protein counterpart. (iii) Mass spectrometry analysis of purified recombinant and native proteins revealed that both the recombinant and native proteins in a pair had the same mass (±5 Da) and were virtually indistinguishable. (iv) Immunoblot analyses showed that the recombinant proteins reacted as strongly as the native proteins with antibodies raised in rabbits against the relevant purified native *M. tuberculosis* protein.

Identification of homologs of the 30-, 32-, 16-, and 23.5-kDa protein genes and gene products in *M. smegmatis* 1-2c. A survey of the literature revealed only one report of an *M. smegmatis* protein related to one of the four *M. tuberculosis* proteins described in this study (19). This was an ~25,000-kDa mycolic acid exchange enzyme, most likely representing the homolog of the 30-kDa protein of *M. tuberculosis*, which is also associated with mycolic acid transferase activity (2). We extended these studies by probing for the potential presence of the aforementioned four genes or their encoded proteins in *M. smegmatis* by
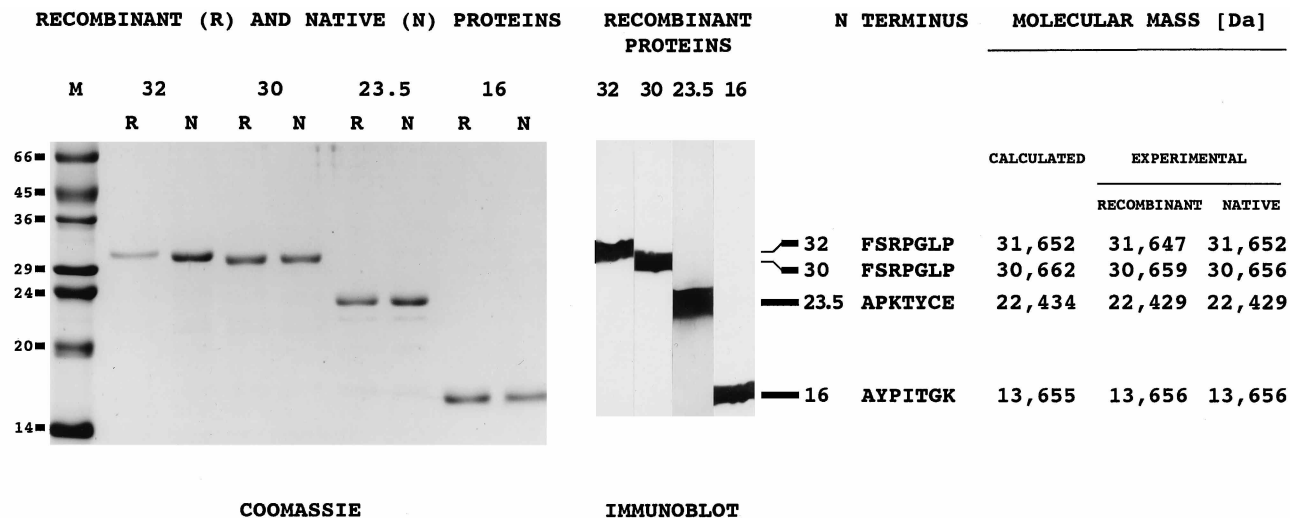
RECOMBINANT (R) AND NATIVE (N) PROTEINS

RECOMBINANT PROTEINS

N TERMINUS

MOLECULAR MASS [Da]

			CALCULATED	EXPERIMENTAL	
				RECOMBINANT	NATIVE
32	FSRPGLP		31,652	31,647	31,652
30	FSRPGLP		30,662	30,659	30,656
23.5	APKTYCE		22,434	22,429	22,429
16	AYPITGK		13,655	13,656	13,656

COOMASSIE

IMMUNOBLOT

FIG. 2. Purified recombinant and native 32-, 30-, 23.5-, and 16-kDa proteins. On the left are Coomassie blue R-stained, purified recombinant and native proteins, including molecular mass standards (lane M) in kilodaltons. In the center are recombinant proteins stained with antibodies specifically raised against each individual purified native protein. On the right is a characterization of each of the proteins by its N terminus, calculated molecular mass (based on the DNA sequence of the structural gene), and experimentally determined molecular mass (based on mass spectrometry analysis of the native and recombinant proteins). Mass spectrometry analyses were performed by Diane Wong and Bradford Gibson at the University of California, San Francisco.

three different approaches, the results of which are presented in Fig. 3.

(i) Gene amplification. UV visualization of amplification products by using denatured genomic *M. tuberculosis* Erdman and *M. smegmatis* 1-2c DNAs and oligonucleotide primers 1 and 2 for the 30-, 32-, and 23.5-kDa protein genes and primers 3 and 4 for the 16-kDa protein gene, revealed strong and specific bands for each of the four genes in *M. tuberculosis* DNA. Each band was of the expected size, ~1 kbp for each of the 30- and 32-kDa protein genes, ~0.5 kbp for the 16-kDa protein gene, and ~0.7 kbp for the 23.5-kDa protein gene.

Primers specific for the 30-kDa protein gene yielded an amplification product in *M. smegmatis* DNA that was slightly less than 1 kbp. However, no amplification products were detected with the primer pairs specific for the 32-, 16-, and 23.5-kDa proteins.

(ii) DNA hybridization. All four gene-specific oligonucleotides (specified in Materials and Methods) detected unique *M. tuberculosis* restriction fragments. However, again, only the probe specific for the 30-kDa protein gene hybridized to a unique *M. smegmatis* 1-2c restriction fragment (data not shown).

DNA AMPLIFICATION PRODUCTS

M. TUBERCULOSIS

M. SMEGMATIS

IMMUNOBLOT

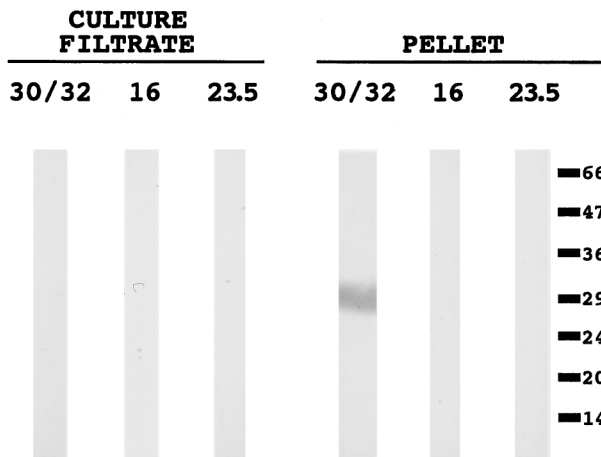### M. SMEGMATIS

CULTURE FILTRATE

PELLET
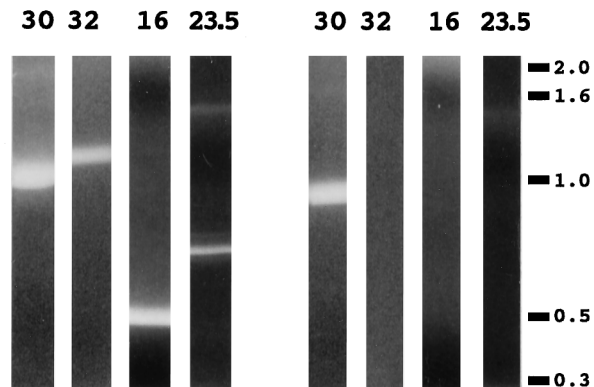
FIG. 3. Search for homologs of the *M. tuberculosis* 30-, 32-, 16-, and 23.5-kDa proteins in *M. smegmatis*. On the left are DNA amplification products of the 30-, 32-, 16-, and 23.5-kDa protein genes in *M. tuberculosis* Erdman and *M. smegmatis* 1-2c. The amplification primers used are described in Materials and Methods. Marker sizes are indicated on the right in kilobase pairs. On the right are immunoblots of extracellular (CULTURE FILTRATE) and cell-associated (PELLET) fractions of *M. smegmatis* probed with antibodies to the four proteins. Molecular mass standards are indicated on the right in kilodaltons.

(iii) Immunoblotting. Antibodies specific to the 30/32-kDa protein complex, the 16-kDa protein, or the 23.5-kDa protein reacted exclusively with their cognate antigens in both the extracellular and cell-associated fractions of *M. tuberculosis*; reaction with the extracellular fraction was substantially stronger (data not shown). In contrast, only an antibody specific for the 30/32-kDa protein complex reacted with its cognate antigen in *M. smegmatis*, and then only with the cell-associated fraction. The antibodies did not recognize any extracellular *M. smegmatis* proteins.

The results from these three approaches indicate that while there is a counterpart of the 30-kDa protein in *M. smegmatis*, the 32-, 16-, and 23.5-kDa proteins are specific for *M. tuberculosis.*

DISCUSSION

This study demonstrates the high-level expression and secretion of recombinant *M. tuberculosis* major extracellular proteins in a form indistinguishable from that of the native proteins. The host species used, *M. smegmatis* and *M. vaccae*, are nonpathogenic, grow 10 times faster than *M. tuberculosis*, and produce 5 to 10 times as much extracellular protein per liter as *M. tuberculosis*, effectively a 50- to 100-fold increase in yield per unit of time.

Repeated attempts to obtain high-level expression and secretion in *E. coli* were unsuccessful, prompting us first to evaluate expression systems in *B. subtilis*, which, like *M. tuberculosis*, is gram positive, and then in nonpathogenic mycobacteria. Previously, heat shock proteins of *M. tuberculosis* have been produced in high yield in *E. coli* intracellularly (14). However, these proteins are highly conserved and are present in both prokaryotes and eukaryotes. The extracellular proteins of *M. tuberculosis* studied here are evidently not highly conserved in bacteria; a homolog for only one of them was present in *M. smegmatis*. Our experience suggests the possibility that high-level expression and especially secretion of relatively nonconserved *Mycobacterium*-specific proteins require a mycobacterial host. In this regard, it is interesting that the overall DNA homology between *M. tuberculosis* and *M. smegmatis* is only ~15% (22).

The large amount of recombinant proteins expressed in *M. smegmatis* and released as soluble molecules into the extracellular milieu, the accuracy of leader peptide processing leading to recombinant proteins indistinguishable from their native counterparts, and the constitutive expression of recombinant proteins all argue for the feasibility of unraveling some of the components and mechanisms involving mycobacterial protein secretion in *M. smegmatis*. Since *M. smegmatis* is both rapidly growing and nonpathogenic, it may yield answers faster than slowly growing, pathogenic *M. tuberculosis.*

Although both *M. tuberculosis* and *M. smegmatis* express and secrete the four major extracellular proteins constitutively, they do so at different rates. By far the most abundant extracellular protein in *M. tuberculosis* cultures is the 30-kDa protein, followed, in descending order of abundance, by the 32-, 16-, and 23.5-kDa proteins—the approximate ratio of production is 3:2:1.5:1. However, in *M. smegmatis*, the 23.5-kDa protein is expressed at a rate at least twofold higher than that of the other three proteins. The approximate ratio of production is 1.5:1.5:1:3 for the 30-, 32-, 16-, and 23.5-kDa proteins, respectively. The basis for the observed differences is unknown.

In conclusion, our study makes possible the high-level production of recombinant *M. tuberculosis* major extracellular proteins by using nonpathogenic species of mycobacteria. The use of rapidly growing, nonpathogenic host species to produce these proteins avoids the exceptional biohazard of *M. tuberculosis* and effectively opens the door to mass production of these proteins for use in vaccines and as targets for new drug development.

REFERENCES

1. **Amann, E., and J. Brosius.** 1985. "ATG vectors" for regulated high-level expression of cloned genes in *Escherichia coli*. Gene **40:**183–190.
2. **Belisle, J. T., T. Sievert, K. Takayama, and G. S. Besra.** 1995. Identification of a mycolyltransferase from *Mycobacterium tuberculosis*: extracellular release and characterization of its enzymatic activity, p. 212–216. *In* Program of the 30th Joint Conference on Tuberculosis and Leprosy 1995. U.S.-Japan Cooperative Medical Science Program, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Ft. Collins, Colo.
3. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:**1991–1995.
4. **Das, A.** 1993. Control of transcription termination by RNA-binding proteins. Annu. Rev. Biochem. **62:**893–930.
5. **Garbe, T. R., J. Barathi, S. Barnini, Y. Zhang, C. Abou-Zeid, D. Tang, R. Mukherjee, and D. B. Young.** 1994. Transformation of mycobacterial species using hygromycin resistance as selectable marker. Microbiology **140:**133–138.
6. **Hardy, K., S. Stahl, and H. Küpper.** 1981. Production in *B. subtilis* of hepatitis B core antigen and a major antigen of foot and mouth disease virus. Nature **293:**481–483.
7. **Harth, G., D. L. Clemens, and M. A. Horwitz.** 1994. Glutamine synthetase of *Mycobacterium tuberculosis*: extracellular release and characterization of its enzymatic activity. Proc. Natl. Acad. Sci. USA **91:**9342–9346.
7a. **Harth, G., and M. A. Horwitz.** Unpublished data.
8. **Harth, G., B.-Y. Lee, J. Wang, D. L. Clemens, and M. A. Horwitz.** 1996. Novel insights into the genetics, biochemistry, and immunocytochemistry of the 30-kilodalton major extracellular protein of *Mycobacterium tuberculosis*. Infect. Immun. **64:**3038–3047.
9. **Hatfull, G. F., and G. J. Sarkis.** 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: phage system for mycobacterial genetics. Mol. Microbiol. **7:**395–405.
10. **Horwitz, M. A., B.-W. E. Lee, B. J. Dillon, and G. Harth.** 1995. Protective immunity against tuberculosis induced by vaccination with major extracellular proteins of *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA **92:** 1530–1534.
11. **Kremer, L., A. Baulard, J. Estaquier, J. Content, A. Capron, and C. Locht.** 1995. Analysis of the *Mycobacterium tuberculosis* 85A antigen promoter region. J. Bacteriol. **177:**642–653.
12. **Lee, B.-Y., and M. A. Horwitz.** 1995. Identification of macrophage and stress-induced proteins of *Mycobacterium tuberculosis*. J. Clin. Invest. **96:**245–249.
13. **McClure, W. R.** 1985. Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. **54:**171–204.
14. **Mehlert, A., and D. B. Young.** 1989. Biochemical and antigenic characterization of the *Mycobacterium tuberculosis* 71 kD antigen, a member of the 70 kD heat-shock protein family. Mol. Microbiol. **3:**125–130.
15. **Oettinger, T., and Å. B. Andersen.** 1994. Cloning and B-cell-epitope mapping of MPT64 from *Mycobacterium tuberculosis* H37Rv. Infect. Immun. **62:**2058–2064.
16. **Pal, P. G., and M. A. Horwitz.** 1992. Immunization with extracellular protein of *Mycobacterium tuberculosis* induces cell-mediated immune response and substantial protective immunity in a guinea pig model of pulmonary tuberculosis. Infect. Immun. **60:**4781–4792.
17. **Philipp, W. J., S. Poulet, K. Eiglmeier, L. Pascopella, V. Balasubramanian, B. Heym, S. Bergh, B. R. Bloom, W. R. Jacobs, Jr., and S. T. Cole.** 1996. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. Proc. Natl. Acad. Sci. USA **93:**3132–3137.
18. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
19. **Sathyamoorthy, N., and K. Takayama.** 1987. Purification and characterization of a novel mycolic acid exchange enzyme from *Mycobacterium smegmatis*. J. Biol. Chem. **262:**13417–13423.

20. **Shine, J., and L. Dalgarno.** 1974. The N-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
21. **Von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14:**4683–4690.
22. **Wayne, L. G.** 1984. Mycobacterial speciation, p. 25–65. *In* G. P. Kubica and L. G. Wayne (ed.), The mycobacteria. Marcel Dekker, Inc., New York, N.Y.
23. **Williams, D. M., E. J. Duvall, and P. S. Lovett.** 1981. Cloning restriction fragments that promote expression of a gene in *Bacillus subtilis*. J. Bacteriol. **146:**1162–1165.
24. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
25. **Zhang, Y., R. Lathigra, T. Garbe, D. Catty, and D. Young.** 1991. Genetic analysis of superoxide dismutase, the 23 kilodalton antigen of *Mycobacterium tuberculosis*. Mol. Microbiol. **5:**381–391.