

Characterization of a 34-Kilodalton Protein of *Mycobacterium leprae* That Is Isologous to the Immunodominant 34-Kilodalton Antigen of *Mycobacterium paratuberculosis*

FAUZI S. SILBAQ,¹ SANG-NAE CHO,^{1†} STEWART T. COLE,² AND PATRICK J. BRENNAN^{1*}

Department of Microbiology, Colorado State University, Fort Collins, Colorado 80523,¹ and Institut Pasteur, Paris, France²

Received 19 December 1997/Returned for modification 29 January 1998/Accepted 26 June 1998

During DNA sequence analysis of cosmid L373 from the *Mycobacterium leprae* genome, an open reading frame of 1.4 kb encoding a protein with some homology to the immunodominant 34-kDa protein of *Mycobacterium paratuberculosis*, but lacking significant serological activity, was detected. The DNA sequence predicted a signal peptide with a modified lipoprotein consensus sequence, but the protein proved to be devoid of lipid attachment.

Mycobacterium leprae, the cause of leprosy, is an obligatory intracellular parasite. The prevalence of leprosy has diminished remarkably: currently there are about 1,150,000 cases worldwide compared to about 10 million in 1985 (4, 5). Interest in leprosy research has correspondingly declined (8). One of the remaining challenges is completion of the sequence of the *M. leprae* genome in light of its relatively small size (ca. 2.8 Mb compared to 4.4 Mb for that of *Mycobacterium tuberculosis*), obligate intracellularism, and unique pathogenesis (3, 19, 20). An ordered genomic cosmid library from *M. leprae* has been prepared and used for systematic genomic sequencing analysis (6, 11, 14). We now describe an open reading frame (ORF) of 1,011 bp within one of the original recombinant DNA cosmids (11), the L373 cosmid located at contig 64 near the origin of replication. The deduced protein showed a high level of homology to the *Mycobacterium paratuberculosis* immunodominant 34-kDa cell wall antigen but differs in key respects.

Identification of the gene coding for the 34-kDa isolog. A fragment of 1,429 bp, obtained from the cosmid L373 DNA, was sequenced from both strands. One ORF of 1,011 bp was found starting at position 52 with a typical translation start codon (AUG) and a translation stop codon (UAA) at position 1060. A potential Shine-Dalgarno sequence (GTTGATG) was found five bases upstream from the translation start codon. The proposed ORF (Fig. 1) encoded a 336-amino-acid (aa) polypeptide, and this polypeptide seemed to have a signal peptide, with a lipoprotein consensus sequence, of 24 aa. The proposed mature protein (i.e., devoid of the signal sequence) was 312 aa in length, with a molecular mass of about 31 kDa. A comparison of the DNA sequence encoding the *M. leprae* 34-kDa isolog with that of the corresponding gene of *M. paratuberculosis* (9, 10, 12) showed that the former was longer at the 5' end and that the encoded protein had the comparative properties described in the legend for Fig. 1. The *M. tuberculosis* product has recently appeared in the database (protein

MTCY10D7.20C). Again, there are four transmembrane segments but no signal peptide.

The C terminus: expression, purification, and antigenicity. The 34-kDa protein of *M. paratuberculosis* is an immunodominant antigen in which the specific B-cell epitopes reside within the C-terminal end of the molecule (9, 10). The complete 34-kDa protein-encoding gene of *M. leprae* was amplified from the L373 cosmid DNA by PCR with appropriate sets of primers. The forward primer (5'-GCATGCACCTACTTCCCCGGTAG-3') corresponds to positions 121 to 137 of the ORF and does not code for the signal peptide. The reverse primer (5'-TTAAACCGGCGCTGACC-3') corresponds to the stop codon at positions 1011 to 995 of the ORF. The PCR product was cloned into the pQE-30 expression vector (QIAGEN, Inc., Chatsworth, Calif.). The sequence encoding the C-terminal part of the 34-kDa isolog (14 kDa; corresponding to aa 200 to 312) was cloned with a forward primer (5'-GCATGCGCACC GCGGCTGAATTACGATC-3') starting at position 598 of the ORF and with a reverse primer (see above), and the PCR product was cloned into pQE-30 vector and expressed in *Escherichia coli*. However, only bacteria containing clones that encode the C terminus grew, apparently due to toxicity for *E. coli* of the first hydrophobic 160-aa segment. The recombinant C terminus protein was purified from large-scale expression cultures by the Ni tag affinity method (16), yielding a highly purified product (M_r , ca. 20,000) by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (results not shown). An examination of its seroreactivity in humans showed scant evidence of circulating antibodies in any of the groups of individuals tested with the curious exception of a group consisting of persons who had had contact with multibacillary patients (results not shown). The relative lack of immunoreactivity may be due to the nature of the processing of the recombinant protein in *E. coli*. Nevertheless, in view of the broad relationship of this protein to the one in *M. paratuberculosis*, attempts were made to identify the B-cell epitope responsible for the modest seroreactivity. Three oligopeptides (P-1, P-2, and P-3) (Fig. 1) were synthesized, based on regions of the C terminus which show a high level of diversity at the amino acid sequence level, and were examined for their capacity to block the binding of the 34-kDa antigen to patient antibodies (18). Only peptide P-3, which is located nearest the C terminus, effectively inhibited binding of the 34-kDa antigen to

* Corresponding author. Mailing address: Department of Microbiology, Colorado State University, Fort Collins, CO 80523. Phone: (970) 491-6700. Fax: (970) 491-1815. E-mail: pbrennan@cvmbs.colostate.edu.

† Permanent address: Department of Microbiology, Yonsei University College of Medicine, Seoul, Republic of Korea.

M. leprae 34kDa isolog

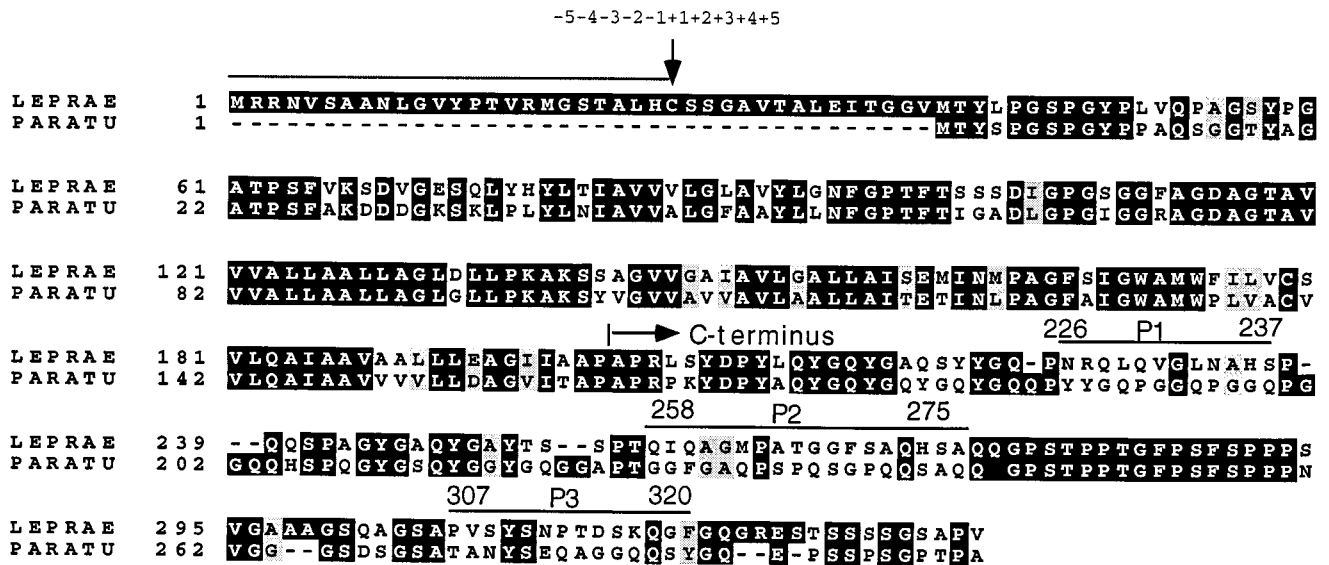


FIG. 1. Comparison of the deduced amino acid sequences of the *M. leprae* (LEPRAE) and *M. paratuberculosis* (PARATU) 34-kDa proteins. The global comparison shows 62% identity between proteins of *M. leprae* and *M. paratuberculosis*. The comparison also shows a higher level of homology (73% identity) at the N terminus (170 aa), which comprises the four predicted transmembrane regions, than at the C terminus (46% identity; the last 142 aa) (1, 2). The configuration of the proposed N terminus is based on the presence of positively charged amino acids in the N terminus and hydrophobic amino acids along the signal sequence, in addition to a pseudolipoprotein consensus sequence. The signal sequence is indicated by the upper line with a small arrow for the cleavage site of the signal peptidase. The oligopeptides that were synthesized and used for the serum blocking experiments are marked P-1, P-2, and P-3. Dashes show identity between the two sequences. The arrow above *M. leprae* aa 201 indicates the first amino acid of the C-terminal protein expressed in *E. coli*. Gray shading represents conserved substitution; black shading represents identical amino acids; no shading represents different sequences.

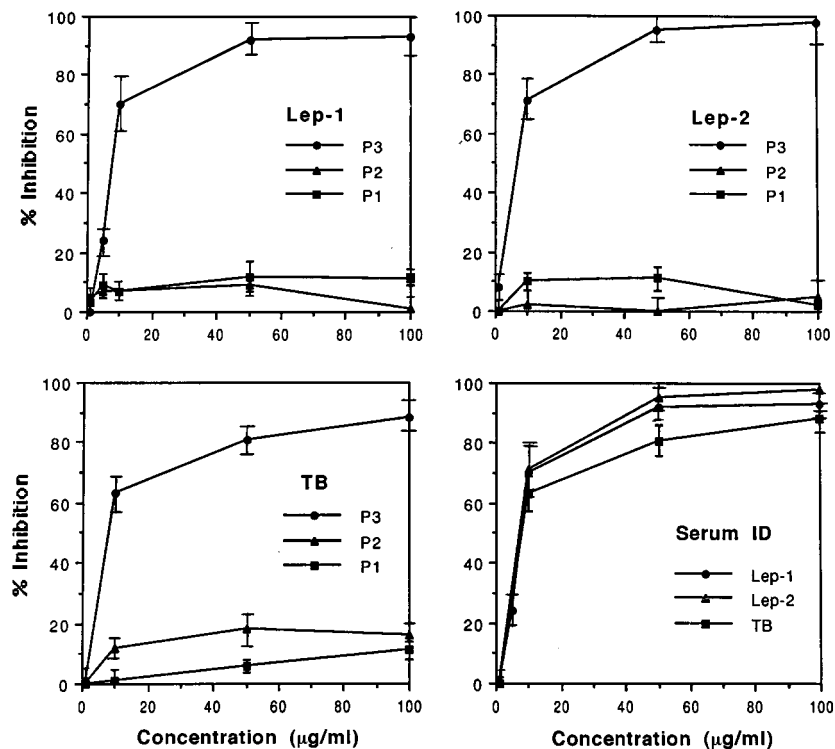


FIG. 2. Comparison of oligopeptides (P1 to P3) in terms of the inhibition of the seroreactivity of the recombinant 34-kDa antigen against sera from leprosy and tuberculosis patients. Lep-1, leprosy patient 1; Lep-2, leprosy patient 2; TB, tuberculosis patient 1; Serum ID, inhibition by the P-3 peptide.

TABLE 1. Comparison of lipoprotein consensus elements of the 38- and 19-kDa proteins of *M. tuberculosis* and the 34-kDa isolog of *M. leprae* with the established lipoprotein consensus element of *E. coli*^a

Protein	Sequence adjacent to Cys	Suitability for acylation
38-kDa <i>M. tuberculosis</i> antigen	Ala Ala Gly Cys Gly Ser	Productive
19-kDa <i>M. tuberculosis</i> antigen	Leu Ser Gly Cys Ser Ser	Productive
34-kDa <i>M. leprae</i> isolog	Ala Leu His Cys Ser Ser	Abortive
<i>E. coli</i> lipoprotein consensus	Leu Ala Gly Cys Ser Ser	Productive

^a Some of this information was adapted from the study of D. B. Young and T. R. Garbe (21).

antibodies (Fig. 2), indicating that the major B-cell epitope of the 34-kDa isolog of *M. leprae* also resides at the C terminus.

The N terminus: evidence for a signal sequence, expression of the complete and truncated 34-kDa protein, and absence of acylation. The crucial difference between the *M. paratuberculosis* and *M. leprae* products is evidence that the *M. leprae* product has a signal element, namely, a 24-aa segment containing positively charged amino acids (Arg, Lys, and His) at the N terminus, and a total content of about 50% hydrophobic amino acids (Gly, Val, Leu, and Ala) (Fig. 1). In addition, there were aspects of the well-known lipoprotein consensus sequence (7, 13, 21), particularly the presence of a Cys residue, at the predicted signal peptidase site (-1 to -5 residues from Cys), and thus there was the possibility of Cys acylation. However, this must be regarded as a modified lipoprotein consensus element, since the normally conserved Gly residue before Cys is replaced by His (Table 1). To examine the relevance of the signal sequence of the *M. leprae* 34-kDa isolog, a forward primer, 5'-ACCGCCGCAACGTAAGCGCTG-3', and a reverse primer, 5'-GAATTCGGTTTATTCCGGCTGACC-3', were used to generate the whole gene, with the signal peptide corresponding to bases at positions 1 to 22 of the ORF. A second forward primer, 5'-ACTGCAGCAGTGGCGCCGTG A-3', was used to generate the coding sequence for the 34-kDa antigen lacking the signal peptide, starting from the codon for the Cys residue and corresponding to bases 71 to 91 of the ORF. The two forms of the coding sequence for the 34-kDa antigen were cloned into pMV261 vector. As a control, *Mycobacterium smegmatis* was electrotransformed with the pMV261 vector. The polyclonal mouse serum against the affinity-purified C-terminal polypeptide (BALB/c mice repeatedly immunized with 40 µg of the protein in Freund's incomplete adjuvant) was applied to the products from the three *M. smegmatis* clones (Fig. 3). Clearly, there was expression of specific proteins by both clones containing the two versions of the 34-kDa protein gene. Despite the difference in the sizes of the two inserts, the protein products were similar in size and heterogeneity; the clone without the signal peptide expressed more of its protein. Thus, this approach did not provide definitive evidence of the nature of the product found in *M. leprae*, i.e., whether the product was cleaved or uncleaved. However, Western blots (17) of subcellular fractions of *M. leprae* (15) provided unequivocal proof of the presence of two native versions of the 34-kDa protein (Fig. 3), presumably with and without the signal sequence. These products were not readily distinguishable in the transformed *M. smegmatis*, due to their heterogeneity. The majority of the native protein was found in the cytosolic fraction despite the implications that it is associated with the membrane or cell wall. Two forms of the *M. paratuberculosis* 34-kDa protein were previously observed in *M. paratuberculosis*, with molecular masses of 31 and 34 kDa (9, 10, 12). Since no signal peptide was observed, it was assumed that the former was a degradation product of the native 34-kDa protein.

Thus, the genetic evidence suggested that both the whole 34-kDa protein and the 31-kDa product devoid of the signal sequence were expressed in *M. smegmatis* and that in *M. leprae* two distinct products were present, presumably also the entire polypeptide with and without the leader sequence. Moreover, in *M. leprae*, the products were much more discrete than the recombinant versions. To address the question of whether the N-terminal Cys of the truncated product was the site of acylation, both the recombinant (carrying pMV and expressing the 34-kDa protein with and without the proposed signal sequence) and the control (electrotransformed with pMV261) versions of *M. smegmatis* were grown in Luria-Bertani medium with 25 µg of kanamycin per ml until early logarithmic phase and [¹⁴C]palmitic acid (50 mCi/mmol; 50 µCi), [³H]glycerol (10 Ci/mmol; 100 µCi), or [¹⁴C]acetate (50 mCi/mmol; 100 µCi) (NEN Life Science Products, Boston, Mass.) was added for 8 h at 37°C with vigorous agitation. Cells were harvested, and subcellular fractions were isolated and subjected to PAGE and autoradiography. No difference in the labelling profiles of the protein populations was observed, and there was no evidence of labelling of the proteins in the 34-kDa range. Thus, we conclude that the Ala Leu His Cys Ser Ser sequence in the *M. leprae* product is an abortive lipoprotein consensus element, clearly defective in terms of acylation and as an effective signal peptidase cleavage site. The implications of the absence of the signal sequence and the glyceryl diacyl unit in *M. leprae*, features shared by all gram-positive, gram-

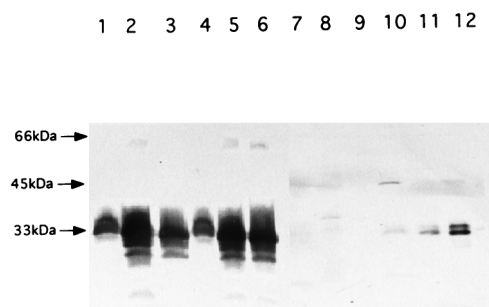


FIG. 3. Expression and localization of the *M. leprae* 34-kDa protein with and without the signal peptide. The gene was amplified by PCR, cloned into pMV261, and transformed into *M. smegmatis* mc²155. Cells were grown and fractionated into cell wall, membrane, and cytosol fractions as described in the text. The different sera were analyzed by Western blotting of SDS-15% PAGE gel using sera from a mouse immunized with the C-terminal portion of the 34-kDa protein expressed in *E. coli*. Lanes 1, 2, and 3 are the cell wall, membrane, and cytosolic fractions, respectively, of *M. smegmatis* cells expressing the 34-kDa protein with the signal peptide. Lanes 4, 5, and 6 indicate the cell wall, membrane, and cytosolic fractions, respectively, of *M. smegmatis* cells expressing the 34-kDa protein without the signal. Lanes 7, 8, and 9 are the cell wall, membrane, and cytosolic fractions, respectively, of *M. smegmatis* containing vector pMV261 as the control. Lanes 10, 11, and 12 show *M. leprae* cell wall, membrane, and cytosolic fractions, respectively.

negative, and cultivable *Mycobacterium* spp. (7, 21), are matters for further investigation.

Nucleotide sequence accession number. The sequence obtained during this study has been assigned GenBank accession no. U82111.

This work was supported by funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, contract AI-55262, the Heiser Foundation for Research on Leprosy and Tuberculosis, and the Association Française Raoul Follereau.

REFERENCES

1. Altschul, S. F. 1993. A protein alignment scoring system sensitive to all evolutionary distances. *J. Mol. Evol.* **36**:290–300.
2. Altschul, S. F., M. S. Boguski, W. Gish, and J. C. Wootton. 1994. Issues in searching molecular sequence databases. *Nat. Genet.* **6**:119–129.
3. Anonymous. 1996. Setting priorities. *Int. J. Lepr.* **64**:591–592.
4. Anonymous. 1997. Global case detection in leprosy. *Weekly Epidemiol. Rec.* **72**:165–172.
5. Anonymous. 1997. Progress towards leprosy elimination. *Weekly Epidemiol. Rec.* **72**:173–180.
6. Bergh, S., and S. T. Cole. 1994. MycDB: an integrated mycobacterial database. *Mol. Microbiol.* **12**:517–534.
7. Braun, V., and H. C. Wu. 1994. Lipoproteins, structure, function, biosynthesis and model for protein export, p. 319–341. In J.-M. Ghuyssen and R. Hakenbeck (ed.), *New comprehensive biochemistry*. Elsevier Science, Amsterdam, The Netherlands.
8. Brennan, P. J. 1995. Prospects for contributions from basic biological research to leprosy control. *Int. J. Lepr.* **63**:285–286.
9. De Kesel, M., P. Gilot, M. Coene, and C. Cocito. 1992. Composition and immunological properties of the protein fraction of A36, a major antigen complex of *Mycobacterium paratuberculosis*. *Scand. J. Immunol.* **36**:201–212.
10. De Kesel, M., P. Gilot, M.-C. Misonne, M. Coene, and C. Cocito. 1993. Cloning and expression of portions of the 34-kilodalton-protein gene of *Mycobacterium paratuberculosis*: its application to serological analysis of Johne's disease. *J. Clin. Microbiol.* **31**:947–954.
11. 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**:197–206.
12. Gilot, P., M. De Kesel, L. Machtelinckx, M. Coene, and C. Cocito. 1993. Isolation and sequencing of the gene coding for an antigenic 34-kilodalton protein of *Mycobacterium paratuberculosis*. *J. Bacteriol.* **175**:4930–4935.
13. Heijne, G. V. 1985. Signal sequences, the limits of variation. *J. Mol. Biol.* **184**:99–105.
14. Honore, N. T., S. Bergn, S. Chanteau, F. Doucetpopulaire, K. Eiglmeier, T. Garnier, C. Georges, P. Iaunois, T. Limpai boon, S. Newton, K. Niang, P. Del Portillo, G. R. Ramesh, P. Reddi, P. R. Ridet, N. Sittisombut, S. W. Hunter, and S. T. Cole. 1993. Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rif-Str regions. *Mol. Microbiol.* **7**:207–214.
15. 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**:14065–14068.
16. Takacs, B. T., and M.-F. Gordon. 1981. Preparation of clinical grade proteins produced by recombinant DNA technologies. *J. Immunol. Methods* **143**:231–240.
17. Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76**:4350–4354.
18. Voller, A., D. E. Bidwell, and A. Bartlett. 1979. The enzyme-linked immunosorbent assay (ELISA), p. 1–40. Dynatech Laboratories, Alexandria, Va.
19. Young, D. B. 1996. Genome sequencing and its potential applications. *Int. J. Lepr.* **64**:968.
20. Young, D. B., and S. T. Cole. 1993. Leprosy, tuberculosis and the new genetics. *J. Bacteriol.* **175**:1–6.
21. Young, D. B., and T. R. Garbe. 1991. Lipoprotein antigens of *Mycobacterium tuberculosis*. *Res. Microbiol.* **142**:55–65.

Editor: E. I. Tuomanen