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,† STEWART T. COLE,2 AND PATRICK J. BRENNAHM

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 (6). The complete 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

<table>
<thead>
<tr>
<th>LEPRAE</th>
<th>PARATU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MRRNVSAANLGYVDVTMGCSSAATETGTGVTY</td>
<td>----PSGPGYPLVPAPAGSPGC</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>ATRPLVTPKCLGVSGQFLKLGYPLFGPSSGPPAGDAATV</td>
<td>----PSGPGYPLVPAPAGSPGC</td>
</tr>
<tr>
<td>61</td>
<td>62</td>
</tr>
<tr>
<td>VFALLALLAGLPLPKEALSAGVVGAVLALLAENEGPAGFPGWHFGPLVA</td>
<td>226 P1 237</td>
</tr>
<tr>
<td>121</td>
<td>82</td>
</tr>
<tr>
<td>VJALLALLAGLPLPKEALSAGVVGAVLALLAENEGPAGFPGWHFGPLVA</td>
<td>258 P2 275</td>
</tr>
<tr>
<td>181</td>
<td>142</td>
</tr>
<tr>
<td>VLAIAAVAVVENAEASAPFLHSDPYLCYGQGASYQGFCGWQYPFGQGPPS</td>
<td>285 P3 307</td>
</tr>
<tr>
<td>239</td>
<td>202</td>
</tr>
<tr>
<td>GQHHRPGQSGSSGYSQCYGACG</td>
<td></td>
</tr>
<tr>
<td>295</td>
<td>262</td>
</tr>
<tr>
<td>VCGAAGSPQGSGATANQFQGCGESSESKSQGSPGTPA</td>
<td></td>
</tr>
</tbody>
</table>

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.
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′-ACCGCCGAACGTGAGCCGTG-3′, and a reverse primer, 5′-GAATTCGTTTATTCCGGCTGACC-3′, were used to generate the whole gene, with the signal peptide corresponding to bases 1 to 22 of the ORF. A second forward primer, 5′-ACTGCAGCAGTGGCGCCGTG-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 antigen were cloned into pMV261 vector. As a control, Mycobacterium smegmatis mc2155. Cells were grown and fractionated into cell wall, membrane, and cytosol fractions as described in the text. The different fractions 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 cytosol 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. U82111.3

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**


Editor: E. I. Tuomanen