Identification, Sequencing, and Expression of *Mycobacterium leprae* Superoxide Dismutase, a Major Antigen

HARRY S. THANGARAJ, F. IAN LAMB,† ELAINE O. DAVIS, PETER J. JENNER, LOICE H. JERYKUMAR, AND M. JOSEPH COLSTON*

Laboratory for Leprosy and Mycobacterial Research, National Institute for Medical Research, The Ridgeway, London NW7 1AA, United Kingdom

Received 27 December 1989/Accepted 30 March 1990

The gene encoding a major 28-kilodalton antigen of *Mycobacterium leprae* has now been sequenced and identified as the enzyme superoxide dismutase (SOD) on the basis of the high degree of homology with known SOD sequences. The deduced amino acid sequence shows 67% homology with a human manganese-utilizing SOD and 55% homology with the *Escherichia coli* manganese-utilizing enzyme. The gene is not expressed from its own promoter in *E. coli* but is expressed from its own promoter in *Mycobacterium smegmatis*. The amino acid sequences of epitopes recognized by monoclonal antibodies against the 28-kilodalton antigen have been determined.

Leprosy is a chronic mycobacterial disease caused by the obligate intracellular parasite *Mycobacterium leprae*. *M. leprae* resides and multiplies inside the macrophages of the skin and Schwann cells of peripheral nerves. Repeated attempts to cultivate the organism in vitro have been unsuccessful, although disseminated experimental infection can be achieved in the nine-banded armadillo (11). The availability of large amounts of *M. leprae* from infected armadillos has enabled the production of monoclonal antibodies against *M. leprae* antigens (5, 7, 12, 31) and the production of genomic libraries of *M. leprae* DNA (6, 33). The isolation of the genes encoding a number of major antigens by screening with monoclonal antibodies (33) has allowed their detailed characterization.

In this study, we have concentrated on one of these antigens, an *M. leprae* protein with an apparent molecular size of 28 kilodaltons (kDa) (31). On the basis of monoclonal antibody cross-reaction studies, this protein is homologous to a 23-kDa protein found in *Mycobacterium tuberculosis*, which also appears to be a major antigen (D. B. Young, personal communication). The entire nucleotide sequence of the *M. leprae* gene encoding the 28-kDa antigen has been determined, and the identity of this antigen as a superoxide dismutase (SOD) has been established by its close homology with SODs from other species. The antigen has been expressed in *Escherichia coli* under control of the tac promoter; it has also been expressed in *Mycobacterium smegmatis*, in this case by using transcriptional and translational control signals present in the cloned DNA. Finally, we have identified the epitopes recognized by monoclonal antibodies SA1.B11H and SA1.D2D, which were originally used to identify this protein (31).

**MATERIALS AND METHODS**

Cloning of the *M. leprae* 28-kDa antigen gene. Recombinant DNA manipulations were performed by standard procedures, essentially as described by Maniatis et al. (16). The 2.6-kilobase (kb) insert of the lambda gt11 clone Y3164 (kindly provided by R. A. Young, Whitehead Institute, Boston, Mass.), which encodes epitopes of the *M. leprae* 28-kDa antigen, was subcloned into the *EcoRI* site of pUC8 to produce clone pIL153, and the 1.9-kb *EcoRI-EcoRI* fragment of Y3164 was also subcloned into pUC8 (28) to produce clone pIL152 (Fig. 1). A cosmid library consisting of *M. leprae* DNA (6) partially digested with *PstI* (kindly provided by J. Clark-Curtiss, Washington University, St. Louis, Mo.) was screened by colony hybridization by using the insert of pIL152 as probe. DNA from resulting positive colonies was restricted with *BamHI* and *PstI* and Southern blotted with the same probe. Positively hybridizing fragments were cloned; a 3.5-kb *PstI* fragment was cloned into pUC8 to generate plasmid pH7, and a 4.2-kb *BamHI* fragment was cloned into pUC18 (28) to generate clone pH720 (Fig. 1). Cosmid DNA manipulation was as described elsewhere (6, 9).

**DNA sequencing.** Random fragments of pH7 were prepared by making nested deletions (3) and by preparation of 200- to 400-base-pair (bp) DNA fragments by sonication at three times 30 W for 5 s each time. These were subcloned in pUC vectors and were sequenced according to protocols provided with the Sequenase kit (U.S. Biochemicals Corp., Cleveland, Ohio). DNA sequences were assembled and analyzed by using the Analyseq program (24), to search for protein coding regions.

**Expression in *E. coli***. An *AsuII-EcoRI* fragment from pIL152 (Fig. 1) was subcloned in both orientations in the *BamHI* site of the *tac* promoter vector pDR540 (21), producing clones pH740 and pH741. These clones, in *E. coli* JM103 (21), and clones in pIL152, pIL153 and pH77, in *E. coli* NM522 (18), along with nonrecombinant vectors as controls were grown in L broth to an optical density at 560 nm of 0.5 prior to further culture in the presence of 10 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 2 h. Cultures of 50 ml were harvested by centrifugation, suspended in 3 ml of 0.1 M Tris hydrochloride (pH 8.0)—1 mM phenylmethylsulfonyl fluoride and sonicated at 80 W for 3 × 1 min. After removal of insoluble debris by centrifugation, samples of 10 μl were mixed with equal volumes of 100 mM Tris hydrochloride (pH 8.4)—250 mM sucrose—1 mM EDTA—2% sodium dodecyl

*Corresponding author.
†Present address: Department of Medical Microbiology, St. Bartholomews Hospital Medical College, London, EC1A 7BE, United Kingdom.
sulfate–5 mM dithiothreitol, heated to 95°C for 10 min, and analyzed on 12% polyacrylamide gels (14). Proteins were transferred to nitrocellulose by using a semidy blotter according to the protocol of the manufacturer, and were probed with monoclonal antibodies SA1.B11H and SA1.D2D (kindly provided by D. B. Young, Medical Research Council Unit for Tuberculosis and Related Diseases, Hammersmith Hospital, London, United Kingdom) at 1:1,000 and 1:2,000, respectively. Immune complexes were detected with a horseradish peroxidase-conjugated rabbit anti-mouse antibody, and blots were developed with 4-chloro-1-naphthol.

Expression in M. smegmatis. The 4.2-kb BamHI insert of pH720 was recloned into a unique BamHI site of the Mycobacterium fortuitum-E. coli shuttle vector PYUB12 (23), and the resulting clone was designated pH724. This plasmid was electroporated into M. smegmatis mc²155 (a gift from W. Jacobs, Albert Einstein College of Medicine, Yeshiva University, New York, N.Y.) (S. B. Snapper, R. E. Melton, T. Kiesser, and W. R. Jacobs, Jr., submitted for publication) essentially as described by Snapper et al. (23), but using 10% glycerol as the electroporation buffer. Kanamycin-resistant colonies containing pH742 were grown for 24 h in Lecom medium containing Tween 80 (10 g of peptone, 10 g of LabLemco powder, 5 g of NaCl, and 0.5 g of Tween 80 per liter; pH 7.4), and cultures were harvested and analyzed by miniplasmid preparation (23) and Western blotting (immunoblotting) as described above.

Identification of epitopes recognized by monoclonal antibodies. A library of 39 overlapping polypeptides (15 to 19 residues) was prepared by a simultaneous multiple-peptide solid-phase method (8) employing a polymide resin (1) and 9-fluorenyl methoxycarbonyl chemistry. The purity of peptides was confirmed by amino acid analysis. Two micrograms of each peptide in trifluoroacetic acid were dried onto microdilution plates (Nunc Maxisorb, Gibco Ltd., Uxbridge, Middlesex, United Kingdom) in a vacuum desiccator, and antibody binding was assayed by enzyme-linked immunosorbent assay. After incubation with 100 μl of 0.25 M carbonate buffer (pH 9.6) at 37°C for 1 h, nonspecific binding was blocked by incubation with 200 μl of 1% dried milk (Marvel) in 0.05% Tween 20 in phosphate-buffered saline (PBSTM) for 3 h at 37°C, and peptides were incubated with 100 μl of SA1.B11H or SA1.D2D (1:1,000 to 1:2,000) in PBSTM at 4°C overnight. After extensive washing with 0.05% Tween 20 in saline, 100 μl of alkaline phosphatase-conjugated goat anti-mouse serum (1:1,000 in PBSTM) was added and incubated for 2 h at 37°C. Wells were further washed with 0.05% Tween 20 in saline, and were developed with p-nitrophenol phosphate. Absorbances were determined using an automatic reader at 404 nm, and results were averaged over two experiments.

RESULTS

Sequence analysis. Sequencing of the insert of plasmid pH720 revealed an open reading frame of 621 bp encoding a protein with a calculated molecular size of 23 kDa; the coding sequence has been published elsewhere (25). Figure 2 shows the coding sequence plus 469 bp of upstream and 447 bp of downstream sequences. Comparison of this sequence with the National Biomedical Research Foundation and Swiss protein sequence data bases showed that the open reading frame encodes a protein homologous to several iron and manganese-utilizing SODs. The primary structure suggests that it is probably a manganese-utilizing enzyme rather than an iron-utilizing enzyme, on the basis of the analysis of Parker and Blake (20). The presence of a manganese-utilizing SOD in M. leprae has been previously described (13, 30). Comparison of the deduced amino acid sequence with the human liver manganese SOD (4) and with the N-terminal sequences of four mycobacterial SODs (19) is shown in Fig. 3. Clearly, the N-terminal sequence of the M. leprae enzyme resembles the consensus mycobacterial sequence. Mycobacterial sequences more closely resemble eucaryotic mitochondrial manganese SODs than other bacterial SODs (19). The enzyme with the greatest overall homology with M. leprae SOD is the human liver manganese SOD, which shows 67% similarity; this homology is maintained throughout the whole coding sequence (Fig. 3).

Immediately upstream of the initiator valine is a likely ribosome-binding sequence, showing homology with the consensus Shine-Dalgarno sequence of E. coli. We have sequenced 469 bp of DNA upstream of the start of the coding sequence and have been unable to identify with certainty promoter sequences, on the basis of comparison with known E. coli promoters; the regions showing greatest homology with the E. coli consensus sequences for the −10 and −35 regions are marked in Fig. 2.

Expression of the M. leprae SOD. Clones in pUC vectors expressed a 28-kDa polypeptide immunoreactive with mono-
Identification of epitopes recognized by monoclonal antibodies. Monoclonal antibody SA1.B11H bound strongly to peptides 8 and 9 but not to any other peptides in the enzyme-linked immunosorbent assay (Table 1). These peptides share the core sequence NDALAKLDEA, which is thus the maximum epitope for this antibody. SA1.D2D bound only to peptide 2, which shares the core sequence WDYAA with the adjacent peptides, to which the antibody did not bind. The epitope thus consists of this core sequence, plus at least one residue at either end, giving a minimum epitope of DWD YAAAL and a maximum epitope of LP DLDW YAAAL EP HLSG. Since these two monoclonal antibodies were originally used to identify the 28-kDa antigen and since they clearly recognize peptides within the SOD sequence, this confirms the identity of the 28-kDa antigen as SOD.

**DISCUSSION**

A great deal of interest has recently focused on the nature of the major antigens of the mycobacterium *M. leprae*. Several of these have now been identified and characterized at the molecular level, and a number of them have been found to be homologs of heat shock proteins (32), whose general function is to enable organisms to cope with environmental stress. Our identification of the major *M. leprae* 28-kDa antigen as SOD indicates the theme of immunodominant antigens as cytoprotective agents, given that mycobacterial mycobacteria occupy an environmental niche in which they are subjected to severe oxidative stress.

Our initial aim was to determine the primary structure of the major 28-kDa antigen of *M. leprae* as a starting point for investigations of its role in the immune response to *M. leprae*. This antigen was one of a small number of *M. leprae* proteins found to be recognized by monoclonal antibodies derived from mice immunized with *M. leprae* (31) and subsequently described as “major” or “dominant” mycobacterial antigens (29, 33). The protein was detectable in infected tissue of 55% of leprosy patients, including some who had minimal bacillary infection (10). Computer analysis of the sequence demonstrated a high degree of homology with several SODs and a consideration of the analysis of Parker and Blake (20) suggests that the *M. leprae* enzyme is probably a manganese-containing SOD (Mn-SOD). Copper-zinc-utilizing (Cu-Zn) SODs are the predominant species of the enzyme in eucaryotes, whereas Mn-SODs are present in procaryotes and eucaryotic mitochondria. Iron-utilizing enzymes are present in procaryotes and a few eucaryotes. While there is significant structural homology between iron and manganese enzymes, there is none between these two forms and Cu-Zn SODs (27). The identity of the 28-kDa protein as SOD has been confirmed by identifying the epitopes recognized by monoclonal antibodies SA1.B11H and SA1.D2D as being short linear peptides contained within the SOD sequence (Table 1). The primary structure of the *M. leprae* enzyme retains the major characteristics of the other manganese- or iron-utilizing SODs, including conservation of residues involved in the enzyme reaction and in metal ion binding (2). It is of interest that the enzyme with the greatest overall similarity with the *M. leprae* protein is the human enzyme. There are three possible explanations for this. First, the *M. leprae* and human enzymes may have evolved under similar selective pressures arising from the obligatory intracellular nature of the *M. leprae* habitat. Second, alternatively, horizontal gene transfer may have occurred such as has been suggested for Cu-Zn SOD-encoding genes of leognathid fish and their long-standing
symbiont Photobacterium leiognathi (17). Third, the M. leprae SOD shows greater similarity to eucaryotic Mn-SOD sequences than to other bacterial sequences (data not included); since eucaryotic Mn-SODs are of mitochondrial origin, this suggests that mycobacteria may be premitochondrial eucaryotic endosymbionts or are closely related to such mitochondrial progenitors.

While it has been found that mycobacterial genes are not generally expressed in E. coli, some of the highly conserved heat shock genes are (15, 22, 26), which suggests that not only are the heat shock functions themselves highly conserved, but their regulatory mechanisms are also conserved. Unlike these heat shock genes, the SOD gene of M. leprae does not appear to be expressed from its own regulatory signals in E. coli since expression was under the control of the lacZ or tac promoter (Fig. 3). On the other hand, high-level expression was obtained in M. smegmatis, probably from the M. leprae transcriptional and translational signals. In addition, we were unable to identify with certainty an upstream region resembling an E. coli promoter. Thus, it appears that while the protein coding sequence for SOD is highly conserved, presumably because of functional constraints, regulatory mechanisms have evolved differently. This is probably a reflection of the importance of SOD to a bacterium living in an oxidatively stressful environment.

The development of any bacterial disease involves a complex interaction between the pathogen and its host. This is particularly true of an obligate intracellular pathogen such as M. leprae. In this study, we have identified another mycobacterial protein which is a major antigen in mice, which, while not being part of the global heat shock response, shares a number of similarities with the heat shock proteins; it is highly conserved and is a component of the

FIG. 3. N-terminal sequences of four mycobacterial SODs (19) and the complete sequences of the M. leprae SOD and human liver Mn-SOD. Boxed residues show residues conserved between mycobacterial SODs and residues conserved between M. leprae and human SODs. (), Undetermined residues in the mycobacterial sequences; —, deletions; *, positions of stop codons with respect to translated sequences.

FIG. 4. Expression of M. leprae SOD in E. coli and M. smegmatis as determined by immunoblotting with the M. leprae-specific monoclonal antibody SA1.11H. For details of constructs, see legend to Fig. 1. Lanes pH740 and pH741 show expression of E. coli containing vector pDR40 with the gene cloned in the correct orientation and lack of expression in the reverse orientation, respectively. Lanes pIL152 show expression in E. coli containing the gene cloned into pUC8, in IPTG-induced (+) and noninduced (++) cultures. Lane pH742 shows the expression of the gene from a BamHI fragment cloned into pYUB12 in M. smegmatis. Lane pYUB12 shows M. smegmatis containing vector alone, and lane pH20 shows lack of expression of the same BamHI fragment in E. coli. MSL, Sonicate of M. leprae; MWM, prestained molecular-size markers (Amersham International).
bacterial defense mechanism against intracellular attack by phagocytes. While the coding sequence for this gene is highly conserved, the sequences regulating its expression do not appear to be so. In addition to its role as a mycobacterial antigen, we are currently investigating the importance of the enzyme and its regulation to intracellular survival of mycobacteria.

ACKNOWLEDGMENTS

We thank R. A. Young (Whitehead Institute, Boston, Mass.), J. Clark-Curtiss (Washington University, St. Louis, Mo.), and W. Jacobs (Albert Einstein College of Medicine, New York) for their gifts of cloned genes, DNA libraries, vectors, and mycobacterial strains. We also thank D. B. Young (Hammersmith Hospital, London, United Kingdom) and S. R. Young (London School for Tropical Medicine and Hygiene, London, United Kingdom) for the supply of monoclonal antibodies.

Harry Thangaraj is supported by The Leprosy Mission, London, and Loice Jeyakumar receives a Fellowship from the Commonwealth Scholarship Commission, United Kingdom.

LITERATURE CITED


<table>
<thead>
<tr>
<th>Peptide no.</th>
<th>Amino acid sequence*</th>
<th>OD^</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epitope of SA1.D2D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>VAETYLPDLWDYAA</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>LPDLWDYAALEPHISG</td>
<td>0.81</td>
</tr>
<tr>
<td>3</td>
<td>WDALEPHISGEINEI</td>
<td>0.00</td>
</tr>
<tr>
<td>Epitope of SA1.B11H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>KHHAAYVKVGNDALK</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>YVKGVNDALKDEA</td>
<td>1.39</td>
</tr>
<tr>
<td>9</td>
<td>NDLALKDEATAKDDHSA</td>
<td>1.20</td>
</tr>
<tr>
<td>10</td>
<td>AKLDEARKDDHSAIFLNEK</td>
<td>0.00</td>
</tr>
</tbody>
</table>

* Amino acid sequences of peptides from the synthetic peptide library given as their single letter code. Only those peptides which were positive plus the flanking peptides have been included.

^ The optical density (OD) is given as the mean of quadruplicate determinations with the blank optical density subtracted. The optical densities of the positive peptides and their flanking peptides are given.