Structure and Mapping of Antigenic Domains of Protein Antigen b, a 38,000-Molecular-Weight Protein of *Mycobacterium tuberculosis*

ÅSE BENGÅRD ANDERSEN and E贡N BECH HANSEN

The Mycobacteria Department, Statens Seruminstitut, Amaner Boulevard 80, DK-2300 Copenhagen S, and Genplejsningsgruppen, The Technical University of Denmark, DK-2800 Lyngby, Denmark

Received 8 February 1989/Accepted 9 May 1989

An important long-term goal of mycobacterial research is to identify mycobacterial antigens involved in the induction of protective immunity against tuberculosis. The development of improved diagnostic reagents, such as specific skin test antigens and serological markers, is another important goal. A prerequisite for getting anywhere near such goals is a detailed evaluation of the immunological role of single mycobacterial antigens.

The application of molecular biology techniques to mycobacteria has already proved powerful in dissecting the bacteria. Currently, tools are developing allowing genetic manipulation of mycobacteria (14, 28). Recently, a number of genes have been isolated from mycobacteria, especially from *Mycobacterium tuberculosis*, *Mycobacterium bovis* BCG, and *Mycobacterium leprae* (1. 13, 26, 32, 40, 42). Special interest has been paid to genes encoding proteins belonging to the heat shock proteins, which are highly conserved among procaryotes as well as eucaryotes (5, 17, 27, 39).

We have been interested in proteins which distinguish mycobacteria from other genera to study the interaction of such proteins with the host immune system.

The 38,000-molecular-weight protein antigen b (Pab) of *M. tuberculosis* is by serological methods solely detected in the virulent *M. tuberculosis* and in lesser amounts in the vaccine strain *M. bovis* BCG. Hyperimmune sera from rabbits immunized with other mycobacterial species do not bind affinity-purified Pab. Furthermore, Pab has been able to induce proliferation of immune T lymphocytes from mice, guinea pigs, and humans (16, 35, 38).

In this paper, we present the nucleotide sequence of *pab* and the mapping of two species-specific domains of the molecule defined by monoclonal antibodies (MAbs).

**MATERIALS AND METHODS**

**Bacteria and phages.** The *Escherichia coli* strains used in this study were DH5alpha [F endA1 hsdR17 supE44 thi-1 recA1 gyrA96 relA1 (de80 lacZDAl5)], JM101 (37), and Y1089 (41).

The recombinant lambda gt11 bacteriophages lambda AA59, lambda AA71, and lambda AA60 were isolated in a previous study (1) from an *M. tuberculosis* genomic DNA library constructed in lambda gt11 by R. A. Young (40).

**DNA technology.** Standard procedures were used for the preparation and handling of DNA (21). DNA sequencing was done by the dideoxynucleotide chain termination method using single-stranded DNA of recombinant M13 phages as a template (24). Nucleotide sequence was obtained from both strands.

**Subcloning of *pab* into *pBR322* (6).** A HindIII-Smal fragment of lambda AA59 containing the *pab* gene was cloned into the HindIII and EcoRV sites of pBR322. The resulting plasmid, pAA26, was isolated from *E. coli* DH5alpha in which alpha complementation of β-galactosidase activity could be observed.

**Bal 31 deletions of the *pab* gene.** DNA of the recombinant plasmid pAA26 was cleaved at the unique BamHI site. After ethanol precipitation, the DNA was dissolved in 20 mM Tris hydrochloride buffer (pH 8.0) with 12 mM CaCl₂, 12 mM MgCl₂, 200 mM NaCl, and 1 mM EDTA. The exonuclease Bal 31 (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) was added, and the mixture was incubated at room temperature. Seven samples were taken within 10 min, and the reaction was stopped by adding phenol with 20 mM EDTA. The DNA was treated with T4 DNA polymerase to make the ends blunt. BamHI linkers (CCCGGATCCCGG) were added to the ligation mixture. The ligated DNA was transformed into *E. coli* DH5alpha and plated on Luria-Bertani agar with ampicillin, isopropyl-β-D-thiogalactoside (IPTG), and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal). Blue colonies were picked randomly, and plasmid...
DNA was cleaved with BamHI and EcoRI and analyzed by agarose gel electrophoresis to determine the size of the mycobacterial insert.

The endpoints of the deletions were determined by DNA sequencing by the method of Maxam and Gilbert (22).

Crude protein extracts of E. coli. Single colonies of E. coli carrying the recombinant pBR plasmids were inoculated into Luria-Bertani broth and grown at 37°C to an optical density at 460 nm of 0.5 when induced by the addition of IPTG to a final concentration of 5 mM. Growth was continued for 2 h. The bacteria were harvested and suspended in sample buffer (1/10 the original volume) (62 mM Tris hydrochloride [pH 6.8], 2% sodium dodecyl sulfate, 0.7 M mercaptoethanol) and boiled for 2 min. Viscosity of released chromosomal DNA was reduced by sonication.

Polyacrylamide gel electrophoresis and immunoblotting. Samples of crude E. coli protein extracts (usually 10 µl) were separated on sodium dodecyl sulfate-7.5% polyacrylamide gels before being stained with Coomassie brilliant blue or transferred onto nitrocellulose sheets. The nitrocellulose-bound samples were probed with MAbS as described earlier (2). The detecting antibody was horseradish peroxidase-coupled rabbit anti-mouse immunoglobulin (P260; Dakopatts, Glostrup, Denmark) preabsorbed to E. coli. The MAbS HAT 2, HYT 28, and HBT 12 were produced at the Statens Seruminstitut as described in references 2, 19, and 25. HGT 3 was kindly provided by G. Kadival and S. D. Chaparas (15), and TB 71 and TB 72 were provided by J. Ivanyi (9).

RESULTS

Nucleic acid sequence determination of the pab gene. The pab gene was isolated in a previous study on three recombinant lambda gt11 phages (1). The phages were found because they produced a protein that was recognized by the M. tuberculosis-complex-specific MAbS HAT 2, HYT 28, and HBT 12. Two of the phages, lambda AA59 and lambda AA71, expressed the recombinant gene product as a fusion protein to β-galactosidase as well as a molecule of molecular weight 36,000. The third phage, lambda AA60, only produced the 36,000-molecular-weight protein. Physical maps of the phages are shown in Fig. 1.

The sequence of 1,993 base pairs from the mycobacterial inserts of lambda AA59 and lambda AA60 was determined. The sequence is shown in Fig. 2. The direction of transcription and the reading frame of the pab gene were identified as the reading frame fused to the lacZ reading frame in lambda AA59. An open reading frame of 374 amino acids was found extending from a putative start codon, a GTG triplet, at position 152 to 154 to a TAG stop codon at position 1274 to 1276. The molecular weight of this protein is 38,200. On the basis of the sequence data, the fusion protein produced by the phage lambda AA59 consists of the entire Pab protein linked to β-galactosidase via 35 nonsense amino acids. The molecular weight of the fusion protein was calculated to be 156,660. This is consistent with the observed mobility of the fusion protein in polyacrylamide gels compared with the mobility of the RNA polymerase subunits RpoB and RpoC from E. coli, which are of molecular weights 150,615 and 155,159, respectively (the Protein Sequence Data Base of the Protein Identification Resource, National Biomedical Research Foundation, Washington, D.C.) (Fig. 3).

The start codon was followed by a sequence of four adenines. A potential ribosome-binding sequence was observed five bases upstream from the start codon.

The base composition was 64.5% guanines and cytosines. Within the individual codons, however, the bases were not evenly distributed. Each of the first two bases of each codon was a guanine or a cytosine in only 55.7% of the cases, whereas the third base was either a guanine or a cytosine in 82.3% of the codons.

A 30% homology between Pab and PstS (also designated PhoS) of E. coli was found by comparing the amino acid sequence to the sequences compiled in the Protein Sequence Data Base (Fig. 4). The pstS gene belongs to the phosphate regulon which consists of at least four genes (pstS, pslA, phoE, and phoB) scattered around the chromosome. They are all induced during phosphate starvation by the gene product of phoB. The regulatory regions of these genes exhibit common features differing somewhat from those of the typical E. coli consensus promoters (20, 30, 33). No significant homologies to the so-called pho boxes, which are 18-nucleotide-long sequences located upstream from the Pribnow boxes, were found within the region sequenced. However, two sequences resembling two proposed phoE Pribnow sequences were observed (Fig. 2). In E. coli, the pstS gene is followed by the gene pstC. Another open reading frame was observed from a GTG triplet at position 1383 to 1385 extending beyond the fragment sequenced. The deduced protein sequence of this open reading frame did not show significant homology to the pstC gene product.

Reactivity pattern of anti-Pab MAbS. Because of our interest in the immunological potential of the molecule, we wanted to map the domains to which the monoclonal anti-Pab antibodies bound. We therefore constructed truncated versions of the fusion protein. By manipulating the gene encoding the fusion protein rather than the nonfused protein, as in lambda AA60, we were able to monitor that the truncated molecules were actually being produced. Because of the high molecular weight of β-galactosidase, fusion products to this protein could easily be identified on Coomassie-stained polyacrylamide gels.

The pab gene was subcloned from lambda AA59 into pBR322 in conjunction with the lacZ gene, as shown in Fig. 5A, and Bal 31 exonuclease deletions were obtained from the 3' end of the gene. Four plasmids were recovered with differently sized mycobacterial inserts (Fig. 5B).

Crude protein extracts of E. coli DH5α transformed with these plasmids were subjected to polyacrylamide gel electrophoresis (Fig. 6A). The deletion plasmids pAA25 and...
pAA24 produced fusion proteins of the same size as the starting plasmid pAA26. However, the fusion proteins produced by pAA17 and pAA23 seemed to migrate faster than the full-length molecule.

The reactivity of a panel of anti-Pab MAbs towards these truncated fusion proteins was analyzed in immunoblotting experiments. The results are summarized in Fig. 5B, and two examples are shown in Fig. 6B and C. The MAbs were divided into two groups according to their reactivity patterns. One group reacted only with the full-length fusion protein, whereas the other group recognized the protein even though 91 of the C-terminal amino acids were missing. None of the Mabs bound to the 117-amino-acid N-terminal peptide produced by pAA23. The observed proteolytic degradation of the fusion proteins was not reduced by transforming the plasmids into an *E. coli* strain lacking the *lon* proteases (data not shown).

Various theoretical methods claiming to predict B-cell
epitopes are now available. The amino acid sequence was analyzed by a program, Protean I (Proteus Biotechnology Limited), that predicts B-cell epitopes on the basis of the deduced secondary structure of the molecule (10). A hydrophilicity profile of the protein was deduced by the program of Hopp and Woods (12). In Fig. 7, the highest-ranked epitopes, as predicted by the Protean I program, are shown together with the hydrophilicity profile.

**DISCUSSION**

A stretch of 1,993 base pairs of *M. tuberculosis* DNA containing the *pab* gene has been sequenced. The structural gene was considered to extend 1,122 base pairs from a GTG start codon. The deduced molecular weight of this gene product, 38,200, is in accordance with the molecular weight of the native protein as judged by its mobility in polyacrylamide gels.

GTG is used as a start codon in only one of the mycobacterial genes sequenced so far (36). However, only a very limited number of mycobacterial gene sequences have been established as yet. The GTG codon is followed by a sequence of four adenine residues which is one of the preferred sequences to follow an initiator codon (29). Five base pairs upstream from the start codon, a putative ribosome-binding site was observed, strengthening the likelihood that this GTG codon is a start codon. The overall GC content of the *pab* gene, 64.5%, is in accordance with results obtained from hybridization studies of sheared chromosomal DNA (4) and with sequence data from other mycobacterial genes (7, 26, 36).

Shinnick previously observed that the GC content of the gene encoding the 65-kilodalton protein was not evenly distributed throughout the gene: the GC content was 55% in the first two positions of each codon versus 87% in the third position (26). This was interpreted as "a strategy allowing the organism to have a high GC content without limiting the access to amino acids whose codons contain A or T residues..."
in the first two positions. The same phenomenon was observed in the \textit{pab} gene: 55.7\% G+C residues in the first two positions versus 82.3\% in the third position.

A search for homologous proteins in the Protein Sequence Data Base revealed 30\% homology between the Pab protein and PstS of \textit{E. coli}. PstS is a phosphate-binding protein of the phosphate-specific transport (Pst) system of \textit{E. coli} which is activated during phosphate starvation. \textit{pstS} is the first gene in the Pst operon, which consists of five genes: \textit{pstS}, \textit{pstC}, \textit{pstA}, \textit{pstB}, and \textit{phoU} (31). With phosphate excess, the Pst operon down regulates the genes of the phosphate regulon, including \textit{pstS}. Whether the function of Pab is to bind and transport phosphate in mycobacteria remains to be tested. In \textit{E. coli}, PstS is localized in the periplasm, and a precursor form with a signal peptide of 25 amino acids has been identified (30). The N terminus of Pab resembles a signal sequence (23, 34). Two charged amino acid residues, a lysine and an arginine, are followed by a chain of 22 nonpolar amino acids followed by a lysine residue. The ultrastructural location of Pab in \textit{M. tuberculosis} has not been determined as yet.

The \textit{pab} gene was subcloned into a plasmid in conjunction with the \textit{lacZ} gene. The resultant plasmid, pAA26, produced a fusion protein as well as a nonfused gene product, as was seen with the phages lambda AA59 and lambda AA71. This could be explained by posttranslational cleavage or by initiation of translation from the \textit{lacZ} start codon as well as from the \textit{pab} start codon.

It has previously been observed that the phage lambda AA60 produced the recombinant gene product in \textit{E. coli} at 30\(^\circ\)C independently of IPTG addition (1). This might be explained by the observed homology to the \textit{phoE} Pribnow sequences (33). We do not know which of these sequences is the most likely to be used, as we do not know the starting point of the transcription. The promoter active in lambda AA60 might, however, be located even further upstream from the sequence given in Fig. 2, as lambda AA60 carries more than 300 base pairs upstream from the GTG start codon.

\textit{Bal} 31 exonuclease deletions were obtained from the 3' end of the \textit{pab} gene of pAA26, and two plasmids, pAA17 and pAA23, appeared to encode truncated fusion proteins. The anti-Pab MAb presently available were probed against these fusion proteins. The binding of the MAb HYT 28,

![Diagram](image-url)
HAT 2, HGT 3, and TB 72 was dependent on the presence of the ultimate 91 amino acids of the protein. We cannot rule out that the epitope(s) visualized by these MAbs is dependent on the correct folding of the protein which might be destroyed when a part of the molecule is removed. However, the Protean I computer program predicted two B-cell epitopes of the linear type within the ultimate fragment of the molecule.

None of the MAbs reacted with the 117 N-terminal amino acids produced by pAA23 which, according to the Protean I program, contains the best-graded B-cell epitope. The plasmid pAA17 produced a fragment that was recognized by HBT 12 as well as TB 71. Within this fragment, the Protean I program points out only one B-cell epitope.

Recently, Anderson and co-workers used the algorithm by Hopp and Woods (11, 12) and Levitt (18) to obtain a hydrophilicity profile of the 65-kilodalton protein from M. leprae (3). Peptides corresponding to the most hydrophilic portions of the molecule were synthesized, and 10 out of 14 B-cell epitopes could be mapped by this method.

As seen in Fig. 7, the sequences given the best grades by the Protean I program do not overlap the hydrophilic peaks. We intend to investigate which—if any—of the computer predictions hold true by having peptides synthesized and study their binding to the MAbs.

Recently, Bothamley et al. (8) have shown a strong correlation between antibody titers of patient sera towards the TB 71 and TB 72 defined epitopes and sputum-positive tuberculosis, especially in HLA-DR2 individuals. Production of peptides covering the antigenic domains of the Pab molecule may prove useful in the study of pathogenic events induced by M. tuberculosis.
ACKNOWLEDGMENTS

We cordially thank Juraj Ivanyi for encouraging this study and for supplying the MAb TB 71 and TB 72. We thank Raju Lathigra and Ann Rees for assistance with computer analyses; Jerry Zoon, Britta Dahl, and Otto Dahl for synthesizing oligonucleotide primers for the sequencing reactions; Gurusur Kadival and Sotiros D. Chaparas for supplying the MAb HGT 3; Katrine Bildstedt and Iben Nielsen for excellent technical assistance; and Grethe Jensen for typing the manuscript.

This work was initiated when A.B.A. was a Fogarty fellow in the laboratory of S. D. Chaparas, Food and Drug Administration, Bethesda, Md. This work has received financial support from the World Health Organization Programme for Vaccine Development, from the Carlsberg Foundation, and from the Research Center for Medical Biotechnology in Denmark.

LITERATURE CITED

35. Worsaae, A., L. Ljungqvist, K. Hasløv, I. Heron, and J. Benned-