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

ÅSE BENGÅRD ANDERSEN[1]* AND EGON BECH HANSEN[2]†

The Mycobacteria Department, Statens Seruminstitut, Amager Boulevard 80, DK-2300 Copenhagen S,[1] and Gensplejsningsgruppen, The Technical University of Denmark, DK-2800 Lyngby,[2] Denmark

Only a limited number of proteins from *Mycobacterium tuberculosis* have so far been shown to possess species-specific epitopes as defined by monoclonal antibodies. One such protein is protein antigen b (Pab) of molecular weight 38,000, which binds the monoclonal antibodies HYT 28, HAT 2, HBT 12, HGT 3, TB 71, and TB 72. The gene encoding this protein was isolated from a lambda gt11 *M. tuberculosis* DNA library. The nucleotide sequence of the recombinant mycobacterial insert was determined, and an open reading frame of 374 amino acids was identified. The amino acid sequence exhibited 30% homology to a phosphate-binding protein, PstS, from *Escherichia coli*. The *pab* gene was subcloned into pBR322 in conjunction with the *lacZ* gene, and deletions were obtained from the 3' end. The anti-Pab monoclonal antibodies were used to probe crude protein lysates of *E. coli* transformed with the deletion plasmids. The monoclonal antibodies showed two reactivity patterns; one group of antibodies were dependent on the presence of the ultimate 91 amino acids of the protein, whereas another group of antibodies recognized an antigenic domain located on the middle portion of the molecule. None of the antibodies bound to the N-terminal 117-amino-acid peptide.

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 being developed 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* (φ80 d*lacZ*ΔM15)], 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 *Hin*dIII-*Sma*I fragment of lambda AA59 containing the *pab* gene was cloned into the *Hin*dIII and *Eco*RV 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 *Bam*HI 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. *Bam*HI linkers (CCCGGATCCGGG) 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
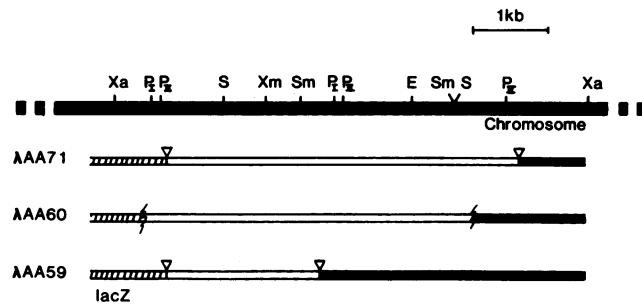
FIG. 1. Physical map of the recombinant lambda phages expressing the Pab protein. ■■, *M. tuberculosis* chromosome; thin bars, the part of the phages carrying the mycobacterial insert; ▨, *lacZ*; ▢, mycobacterial insert; ■■, lambda gt11 arm; ∇, *Eco*RI linker; ⸸, endpoint not precisely mapped, as *Eco*RI linker was missing. Restriction enzyme cleavage sites: E, *Eco*RI; PI, *Pvu*I; PII, *Pvu*II; S, *Sal*I; Sm, *Sma*I; X, *Xho*I. a, Positions determined within a range of 200 base pairs. kb, Kilobase pair.

DNA was cleaved with *Bam*HI and *Eco*RI 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 2-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*, *phoA*, *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 sequences, 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* DH5alpha transformed with these plasmids were subjected to polyacrylamide gel electrophoresis (Fig. 6A). The deletion plasmids pAA25 and

```
                PB2                                                                                   PB1
TGTTCTTCGACGGCAGGCTGGTGGAGGAAGGGCCCACCGAACAGCTGTTCTCCTCGCCGAAGCATGCGGAAACCGCCCGATACGTCGCCGGACTGTCGGGGGACGTCAAGGACGCCAAGC
      10       20       30       40       50       60       70       80       90      100      110      120

                    S-D        M  K  I  R  L  H  T  L  L  A  V  L  T  A  A  P  L  L  L  A  A  A  G  C  G  S  K  P  P  S
GCGGAAATTGAAGAGCACAGAAAGGTATGGCGTGAAAATTCGTTTGCATACGCTGTTGGCCGTGTTGACCGCTGCGCCGCTGCTGCTAGCAGCGGCGGGCTGTGGCTCGAAACCACCGAG
      130      140      150      160      170      180      190      200      210      220      230      240

     G  S  P  E  T  G  A  G  A  G  T  V  A  T  T  P  A  S  S  P  V  T  L  A  E  T  G  S  T  L  L  Y  P  L  F  N  L  W  G  P
CGGTTCGCCTGAAACGGGCGCCGGCGCCGGTACTGTCGCGACTACCCCCGCGTCGTCGCCGGTGACGTTGGCGGAGACCGGTAGCACGCTGCTCTACCCGCTGTTCAACCTGTGGGGTCC
      250      260      270      280      290      300      310      320      330      340      350      360

  A  F  H  E  R  Y  P  N  V  T  I  T  A  Q  G  T  G  S  G  A  G  I  A  Q  A  A  A  G  T  V  N  I  G  A  S  D  A  Y  L  S
GGCCTTTCACGAGAGGTATCCGAACGTCACGATCACCGCTCAGGGCACCGGTTCTGGTGCCGGGATCGCGCAGGCCGCCGCCGGGACGGTCAACATTGGGGCCTCCGACGCCTATCTGTC
      370      380      390      400      410      420      430      440      450      460      470      480

            pAA23|
  E  G  D  M  A  A  H  K  G  L  M  N  I  A  L  A  I  S  A  Q  Q  V  N  Y  N  L  P  G  V  S  E  H  L  K  L  N  G  K  V  L
GGAAGGTGATATGGCCGCGCACAAGGGGCTGATGAACATCGCGCTAGCCATCTCCGCTCAGCAGGTCAACTACAACCTGCCCGGAGTGAGCGAGCACCTCAAGCTGAACGGAAAAGTCCT
      490      500      510      520      530      540      550      560      570      580      590      600

  A  A  M  Y  Q  G  T  I  K  T  W  D  D  P  Q  I  A  A  L  N  P  G  V  N  L  P  G  T  A  V  V  P  L  H  R  S  D  G  S  G
GGCGGCCATGTACCAGGGCACCATCAAAACCTGGGACGACCCGCAGATCGCTGCGCTCAACCCCGGCGTGAACCTGCCCGGCACCGCGGTAGTTCCGCTGCACCGCTCCGACGGGTCCGG
      610      620      630      640      650      660      670      680      690      700      710      720

                                                                              Sal I
  D  T  F  L  F  T  Q  Y  L  S  K  Q  D  P  E  G  W  G  K  S  P  G  F  G  T  T  V  D  F  P  A  V  P  G  A  L  G  E  N  G
TGACACCTTCTTGTTCACCCAGTACCTGTCCAAGCAAGATCCCGAGGGCTGGGGCAAGTCGCCCGGCTTCGGCACCACCGTCGACTTCCCGGCGGTGCCGGGTGCGCTGGGTGAGAACGG
      730      740      750      760      770      780      790      800      810      820      830      840

  N  G  G  M  V  T  G  C  A  E  T  P  G  C  V  A  Y  I  G  I  S  F  L  D  Q  A  S  Q  R  G  L  G  E  A  Q  L  G  N  S  S
CAACGGCGGCATGGTGACCGGTTGCGCCGAGACACCGGGCTGCGTGGCCTATATCGGCATCAGCTTCCTCGACCAGGCCAGTCAACGGGGACTCGGCGAGGCCCAACTAGGCAATAGCTC
      850      860      870      880      890      900      910      920      930      940      950      960

            pAA17|
  G  N  F  L  L  P  D  A  Q  S  I  Q  A  A  A  A  G  F  A  S  K  T  P  A  N  Q  A  I  S  M  I  D  G  P  A  P  D  G  Y  P
TGGCAATTTCTTGTTGCCCGACGCGCAAAGCATTCAGGCCGCGGCGGCTGGCTTCGCATCGAAAACCCCGGCGAACCAGGCGATTTCGATGATCGACGGGCCCGCCCGGACGGCTACCC
      970      980      990     1000     1010     1020     1030     1040     1050     1060     1070     1080

  I  I  N  Y  E  Y  A  I  V  N  N  R  Q  K  D  A  A  T  A  Q  T  L  Q  A  F  L  H  W  A  I  T  D  G  N  K  A  S  F  L  D
GATCATCAACTACGAGTACGCCATCGTCAACAACCGGCAAAAGGACGCCGCCACCGCGCAGACCTTGCAGGCATTTCTGCACTGGGCGATCACCGACGGCAACAAGGCCTCGTTCCTCGA
     1090     1100     1110     1120     1130     1140     1150     1160     1170     1180     1190     1200

  Q  V  H  F  Q  P  L  P  P  A  V  V  K  L  S  D  A  L  I  A  T  I  S  S  -
CCAGGTTCATTTCCAGCCGCTGCCGCCCGCGGTGGTGAAGTTGTCTGACGCGTTGATCGCGACGATTTCCAGCTAGCCTCGTTGACCACCACGCGACAGCAACCTCCGTCGGGCCATCGG
     1210     1220     1230     1240     1250     1260     1270     1280     1290     1300     1310     1320

                  XmaIII
GCTGCTTTGCGGAGCATGCTGGCCCGTGCCGGTGAAGTCGGCCGCGCTGGCCCGGCCATCCGGTGGTTGGGTGGGATAGGTGCGGTGATCCCGCTGCTTGCGCTGGTCTTGGTGCTGGTG
     1330     1340     1350     1360     1370     1380     1390     1400     1410     1420     1430     1440

GTGCTGGTCATCGAGGCGATGGGTGCGATCAGGCTCAACGGGTTGCATTTCTTCACCGCCACCGAATGGAATCCAGGCAACACCTACGGCGAAACCGTTGTCACCGACGCGTCGCCCATC
     1450     1460     1470     1480     1490     1500     1510     1520     1530     1540     1550     1560

CGGTCGGCGCCTACTACGGGGCGTTGCCGCTGATCGTCGGGACGCTGGCGACCTCGGCAATCGCCCTGATCATCGCGGTGCCGGTCTCTGTAGGAGCGGCGCTGGTGATCGTGGAACGGC
     1570     1580     1590     1600     1610     1620     1630     1640     1650     1660     1670     1680

TGCCGAAACGGTTGGCCGAGGCTGTGGGAATAGTCCTGGAATTGCTCGCCGGAATCCCCAGCGTGGTCGTCGGTTTGTGGGGGGCAATGACGTTCGGGCCGTTCATCGCTCATCACATCG
     1690     1700     1710     1720     1730     1740     1750     1760     1770     1780     1790     1800

                                                      SmaI
CTCCGGTGATCGCTCACAACGCTCCCGATGTGCCGGTGCTGAACTACTTGCGCGGCGACCCGGGCAACGGGGAGGGCATGTTGGTGTCCGGTCTGGTGTTGGCGGTGATGGTCGTTCCCA
     1810     1820     1830     1840     1850     1860     1870     1880     1890     1900     1910     1920

                       SmaI              EcoRI
TTATCGCCACCACCACTCATGACCTGTTCCGGCAGGTGCCGGTGTTGCCCCCGGGAGGGCGCGATCGGGAATTC
     1930     1940     1950     1960     1970     1980     1990
```

FIG. 2. Nucleotide sequence of the *pab* gene and the deduced amino acid sequence of the gene product. The potential ribosome-binding site is underlined and indicated with S-D. The two sequences showing homology to the putative *phoE* Pribnow boxes (which are CAGGTT and GACTCT) are underlined and marked PB2 and PB1. Endpoints of deletions carried by the plasmids pAA17 and pAA23 are indicated with vertical bars. The *Eco*RI site at position 1988 is not a mycobacterial sequence but an *Eco*RI linker used in the construction of the lambda gt11 gene library. The *Eco*RI linker in lambda AA59 is located at position 41.

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 pat-

terns. 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).

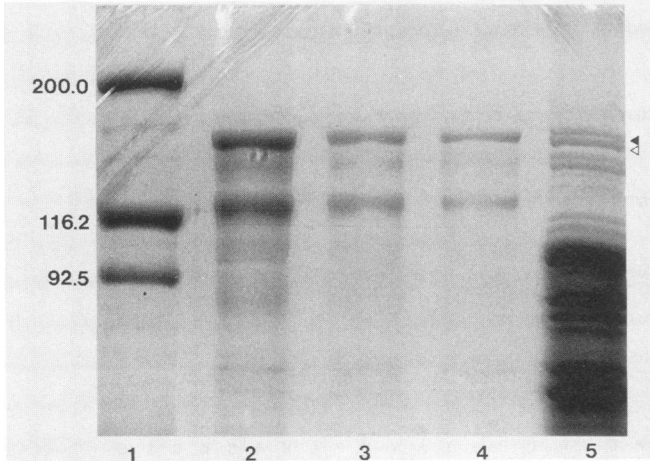Various theoretical methods claiming to predict B-cell

FIG. 3. Coomassie-stained polyacrylamide gel. Lanes: 1, molecular weight markers (values given on the left are in thousands); 2, 3, and 4, dilute samples of crude protein extracts of *E. coli* DH5alpha transformed with pAA26 separated on 7.5% polyacrylamide gel; 5, *E. coli* DH5alpha transformed with pBR322. ◁, RpoB; ◄, RpoC (molecular weights, 150,615 and 155,159, respectively).

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, 32, 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

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

```
Pab    1   MKIRLHTLLAVLTAAPLLLAAAGCGSKPPSGSPETGAGAGTV ATTPASSPVTLAETGST
                                                 *  *  **  *    **    *
PhoS   1                                          MKVMRTTVATVVAATLSMSAFSVFAEASLT


      60   LLYPLFNLWGPA FHERYPNVT   ITAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMA
            *      *       *   *      ** **  *  *  * **  ***** **
      31   GAGATFPAPVYAKWADTYQKETGNKVNYQGIGSSGGVKQIIANTVDFGASDAPLS DEKL


     116   AHKGLMNIALAISAQQVNYNLPG VSEHLKLNGKVLAAMYQGTIKTWDDPQIAALNPGVN
            *  **          *         * **  *  * * **  *  * * ** ***  ** ****
      90   AQEGLFQFPTVIGGVVLAVNIPGLKSGELVLDGKTLGDIYLGKIKKWDDEAIAKLNPGLK


     175   LPGTAVVPLHRSDGSGDTFLFTQYLSKQDPEGWGKSPGFGTTVDFPAVPGALGENGNGGM
            **          * ****  * ** ** *   * *     * * **    *  **  ** *
     150   LPSQNIAVVRRADGSGTSFVFTSYLAK VNEEWKNNVGTGSTV   KWPIGLGGKGNDGI


     235   VTGCAETPGCVAYIGISFLDQASQRGLGEAQLGNSSGNFLLPDAQSIQAAAAG FASKTP
            **       **      * *  *     *      *       * ** *  ***
     206   AAFVQRLPG    AIGYVEYAYAKQNNLAYTKLISADGKPVSPTEENFANAAKGADWSKTF


     294   ANQAISMIDGPAPDGYPIINYEYAIVNNRQKDAATAQTLQAFLHWAITDGNK ASFLDQV
            * *      *   *   **           **        *  **   * * *   **
     263   A QDLTNQKG   EDAWPITSTTFILIHKDQKKPEQGTEVLKFFDWAYKTGAKQANDLD


     353   HFQPLPPAVVKLSDALIATISS
            **  **     *    *
     318   YASLPDSVVEQVRAAWKTNIKDSSGKPLY
```

FIG. 4. Alignment of the amino acid sequence of Pab with that of PstS (also designated PhoS). *, Identical amino acids.

A

B
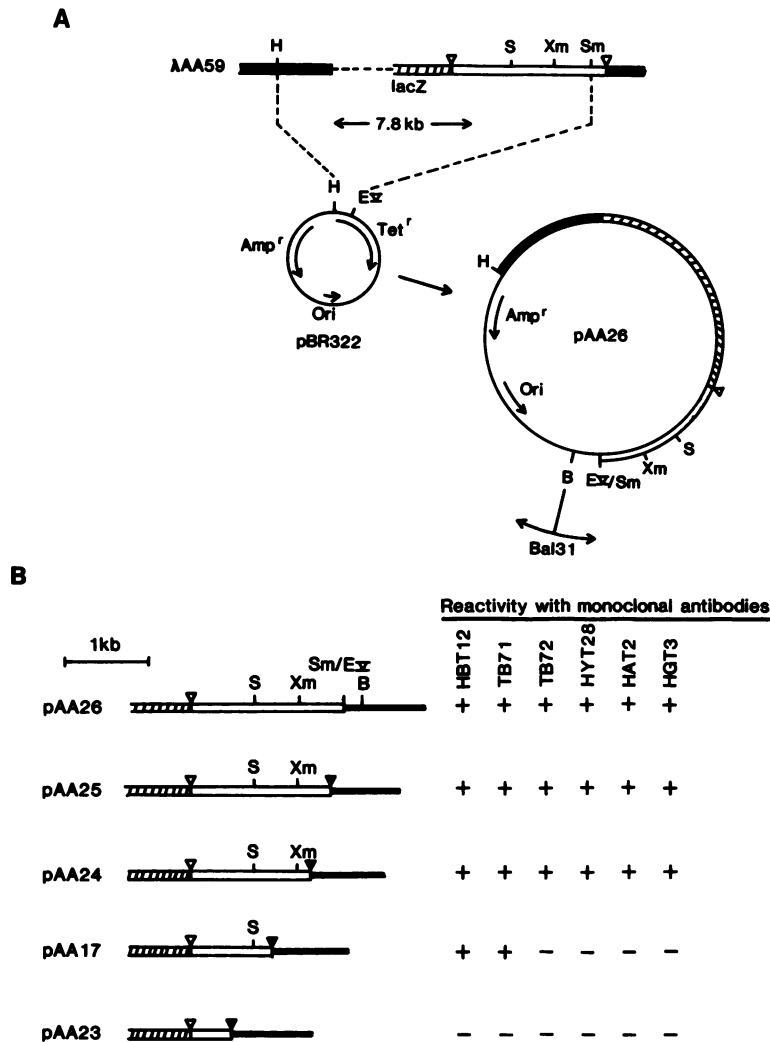
FIG. 5. Subcloning of *lacZ* and *pab* into pBR322 (A) and map of deletion plasmids obtained by *Bal* 31 exonuclease digestion (B). Restriction enzyme cleavage sites: EV, *Eco*RV; B, *Bam*HI; H, *Hin*dIII; S, *Sal*I; Sm, *Sma*I; Xm, *Xma*III. kb, Kilobase pairs.

in the first two positions." The same phenomenon was observed in the *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 *E. coli*. PstS is a phosphate-binding protein of the phosphate-specific transport (Pst) system of *E. coli* which is activated during phosphate starvation. *pstS* is the first gene in the Pst operon, which consists of five genes: *pstS*, *pstC*, *pstA*, *pstB*, and *phoU* (31). With phosphate excess, the Pst operon down regulates the genes of the phosphate regulon, including *pstS*. Whether the function of Pab is to bind and transport phosphate in mycobacteria remains to be tested. In *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 *M. tuberculosis* has not been determined as yet.

The *pab* gene was subcloned into a plasmid in conjunction

with the *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 *lacZ* start codon as well as from the *pab* start codon.

It has previously been observed that the phage lambda AA60 produced the recombinant gene product in *E. coli* at 30°C independently of IPTG addition (1). This might be explained by the observed homology to the *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.

Bal 31 exonuclease deletions were obtained from the 3′ end of the *pab* gene of pAA26, and two plasmids, pAA17 and pAA23, appeared to encode truncated fusion proteins. The anti-Pab MAbs presently available were probed against these fusion proteins. The binding of the MAbs HYT 28,
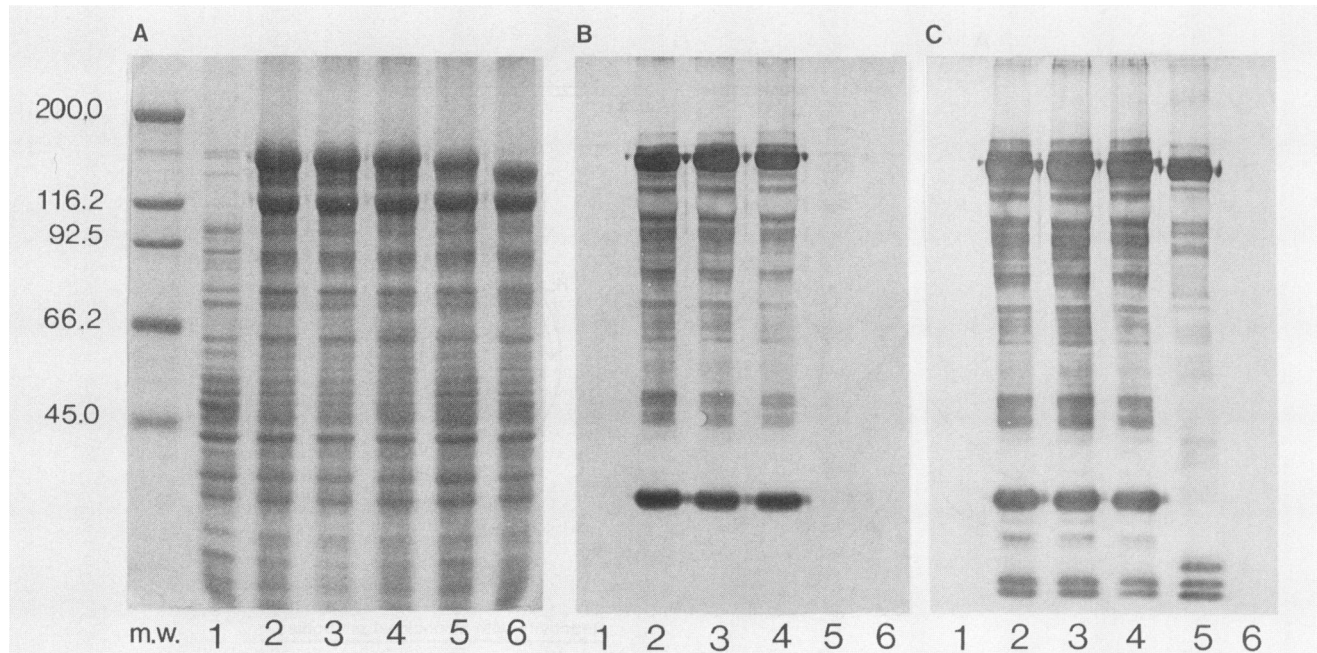
FIG. 6. (A) Coomassie-stained polyacrylamide gel. Crude protein extracts of *E. coli* DH5alpha transformed with the deletion plasmids (see text) separated on 7.5% polyacrylamide gel. Lanes: m.w., molecular weight markers (values given on the left are in thousands); 1, pBR322; 2, pAA26; 3, pAA25; 4, pAA24; 5, pAA17; 6, pAA23. (B and C) Immunoblot analyses. The proteins from two gels identical to the one shown in panel A were transferred onto nitrocellulose and incubated with HYT 28 (panel B) and HBT 12 (panel C).

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 eptitope(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*.
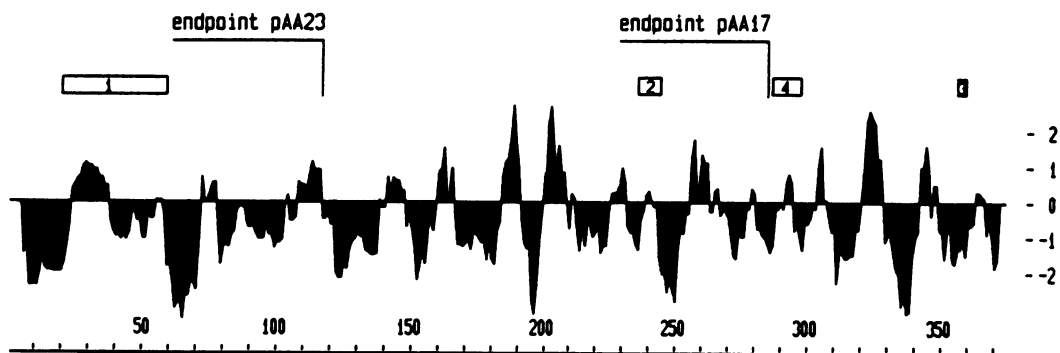
FIG. 7. Hydrophilicity profile of the Pab protein. The averaged hydrophilicity values are plotted versus the position of each residue along the amino acid sequence (12). Open boxes show the four best-graded B-cell epitopes predicted by the Protean I program (10). The individual ranking is indicated by numbers; box 1 indicates the best-graded epitope.

LITERATURE CITED

1. **Andersen, Å. B., A. Worsaae, and S. D. Chaparas.** 1988. Isolation and characterization of recombinant lambda gt11 bacteriophages expressing eight different mycobacterial antigens of potential immunological relevance. Infect. Immun. **56:**1344–1351.
2. **Andersen, Å. B., Z.-L. Yuan, K. Hasløv, B. Vergmann, and J. Bennedsen.** 1986. Interspecies reactivity of five monoclonal antibodies to *Mycobacterium tuberculosis* as examined by immunoblotting and enzyme-linked immunosorbent assay. J. Clin. Microbiol. **23:**446–451.
3. **Anderson, D. C., M. E. Barry, and T. M. Buchanan.** 1988. Exact definition of species-specific and cross-reactive epitopes of the 65-kilodalton protein of *Mycobacterium leprae* using synthetic peptides. J. Immunol. **141:**607–613.
4. **Baess, I., and B. Mansa.** 1978. Determination of genome size and base ratio on deoxyribonucleic acid from mycobacteria. Acta Pathol. Microbiol. Scand. Sect. B **86:**309–312.
5. **Baird, P. N., L. M. C. Hall, and A. R. M. Coates.** 1988. A major antigen from *Mycobacterium tuberculosis* which is homologous to the heat shock proteins groES from *E. coli* and the htpA gene product of *Coxiella burneti*. Nucleic Acids Res. **16:**9047.
6. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, and H. W. Boyer.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2:**95–113.
7. **Booth, R. J., D. P. Harris, J. M. Love, and J. D. Watson.** 1988. Antigenic proteins of *Mycobacterium leprae*. Complete sequence of the gene for the 18-kDa protein. J. Immunol. **140:** 597–601.
8. **Bothamley, G. H., J. S. Beck, G. M. T. Schreuder, J. D'Amaro, R. R. P. de Vries, T. Kardjito, and J. Ivanyi.** 1989. Association of tuberculosis and *Mycobacterium tuberculosis*-specific antibody levels with HLA. J. Infect. Dis. **159:**549–555.
9. **Coates, A. R. M., J. Hewitt, B. W. Allen, J. Ivanyi, and D. A. Mitchison.** 1981. Antigenic diversity of *Mycobacterium tuberculosis* and *Mycobacterium bovis* detected by means of monoclonal antibodies. Lancet **ii:**167–169.
10. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120:**97–120.
11. **Hopp, T. P.** 1986. Protein surface analysis. Methods for identifying antigenic determinants and other interaction sites. J. Immunol. Methods **88:**1–18.
12. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78:**3824–3828.
13. **Husson, R. N., and R. A. Young.** 1987. Genes for the major protein antigens of *Mycobacterium tuberculosis*: the etiologic agents of tuberculosis and leprosy share an immunodominant antigen. Proc. Natl. Acad. Sci. USA **84:**1679–1683.
14. **Jacobs, W. R., Jr., M. Tuckman, and B. R. Bloom.** 1987. Introduction of foreign DNA into mycobacteria using a shuttle phasmid. Nature (London) **327:**532–534.
15. **Kadival, G. V., and S. D. Chaparas.** 1987. Production, characterization, and species specificity of five monoclonal antibodies

to *Mycobacterium tuberculosis*. J. Clin. Microbiol. **25:**76–80.
16. **Kadival, G. V., S. D. Chaparas, and D. Hussong.** 1987. Characterization of serologic and cell-mediated reactivity of a 38-kDa antigen isolated from *Mycobacterium tuberculosis*. J. Immunol. **139:**2447–2451.
17. **Lathigra, R. B., D. B. Young, D. Sweetser, and R. A. Young.** 1988. A gene from *Mycobacterium tuberculosis* which is homologous to the DnaJ heat shock protein of *E. coli*. Nucleic Acids Res. **16:**1636.
18. **Levitt, M.** 1976. A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. **104:**59–107.
19. **Ljungqvist, L., A. Worsaae, and I. Heron.** 1988. Antibody responses against *Mycobacterium tuberculosis* in 11 strains of inbred mice: novel monoclonal antibody specificities generated by fusions, using spleens from BALB.B10 and CBA/J mice. Infect. Immun. **56:**1994–1998.
20. **Magota, K., N. Otsuji, T. Miki, T. Horiuchi, S. Tsunasawa, J. Kondo, F. Sakiyama, M. Amemura, T. Morita, H. Shinagawa, and A. Nakata.** 1984. Nucleotide sequence of the *phoS* gene, the structural gene for the phosphate-binding protein of *Escherichia coli*. J. Bacteriol. **157:**909–917.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
23. **Michaelis, S., and J. Beckwith.** 1982. Mechanism of incorporation of cell envelope proteins in *Escherichia coli*. Annu. Rev. Microbiol. **36:**435–465.
24. **Sanger, F., A. R. Coulson, G. F. Hong, D. F. Hill, and G. B. Peterson.** 1982. Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. **162:**729–772.
25. **Schou, C., Z.-L. Yuan, A. B. Andersen, and J. Bennedsen.** 1985. Production and partial characterization of monoclonal hybridoma antibodies to *Mycobacterium tuberculosis*. Acta Pathol. Microbiol. Immunol. Scand. Sect. C **93:**265–272.
26. **Shinnick, T. M.** 1987. The 65-kilodalton antigen of *Mycobacterium tuberculosis*. J. Bacteriol. **169:**1080–1088.
27. **Shinnick, T. M., M. H. Vodkin, and J. C. Williams.** 1988. *Mycobacterium tuberculosis* 65-kilodalton antigen is a heat shock protein which corresponds to common antigen and to the *Escherichia coli* GroEL protein. Infect. Immun. **56:**446–451.
28. **Snapper, S. B., L. Lugosi, A. Jekkel, R. E. Melton, T. Kieser, B. R. Bloom, and W. R. Jacobs, Jr.** 1988. Lysogeny and transformation in mycobacteria: stable expression of foreign genes. Proc. Natl. Acad. Sci. USA **85:**6987–6991.
29. **Stormo, G. D., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10:**2971–3011.
30. **Surin, B. P., D. A. Jans, A. L. Fimmel, D. C. Shaw, G. B. Cox, and H. Rosenberg.** 1984. Structural gene for the phosphate-repressible phosphate-binding protein of *Escherichia coli* has its own promoter: complete nucleotide sequence of the *phoS* gene. J. Bacteriol. **157:**772–778.
31. **Surin, B. P., H. Rosenberg, and G. B. Cox.** 1985. Phosphate-specific transport system of *Escherichia coli*: nucleotide sequence and gene-polypeptide relationships. J. Bacteriol. **161:** 189–198.
32. **Thole, J. E. R., W. J. Keulen, A. H. J. Kolk, D. G. Groothuis, L. G. Berwald, R. H. Tiesjema, and J. D. A. van Embden.** 1987. Characterization, sequence determination, and immunogenicity of a 64-kilodalton protein of *Mycobacterium bovis* BCG expressed in *Escherichia coli* K-12. Infect. Immun. **55:**1466–1475.
33. **Tommassen, J.** 1987. Expression of outer membrane PhoE protein in *Escherichia coli* K-12, p. 26–30. *In* A. Torriani-Gorini, F. G. Rothman, S. Silver, A. Wright, and E. Yagil (ed.), Phosphate metabolism and cellular regulation in microorganisms. American Society for Microbiology, Washington, D.C.
34. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12:**5145–5164.
35. **Worsaae, A., L. Ljungqvist, K. Hasløv, I. Heron, and J. Benned-**

sen. 1987. Allergenic and blastogenic reactivity of three antigens from *Mycobacterium tuberculosis* in sensitized guinea pigs. Infect. Immun. **55:**2922–2927.

36. **Yamaguchi, R., K. Matsuo, A. Yamazaki, C. Abe, S. Nagai, K. Terasaka, and T. Yamada.** 1989. Cloning and characterization of the gene for immunogenic protein MPB64 of *Mycobacterium bovis* BCG. Infect. Immun. **57:**283–288.

37. **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.

38. **Young, D., L. Kent, A. Rees, J. Lamb, and J. Ivanyi.** 1986. Immunological activity of a 38-kilodalton protein purified from *Mycobacterium tuberculosis*. Infect. Immun. **54:**177–183.

39. **Young, D., R. Lathigra, R. Hendrix, D. Sweetser, and R. A. Young.** 1988. Stress proteins are immune targets in leprosy and tuberculosis. Proc. Natl. Acad. Sci. USA **85:**4267–4270.

40. **Young, R. A., B. R. Bloom, C. M. Grosskinsky, J. Ivanyi, D. Thomas, and R. W. Davis.** 1985. Dissection of *Mycobacterium tuberculosis* antigens using recombinant DNA. Proc. Natl. Acad. Sci. USA **82:**2583–2587.

41. **Young, R. A., and R. W. Davis.** 1983. Yeast RNA polymerase II genes: isolation with antibody probes. Science **222:**778–782.

42. **Young, R. A., V. Mehra, D. Sweetser, T. Buchanan, J. Clark-Curtiss, R. W. Davis, and B. R. Bloom.** 1985. Genes for the major protein antigens of the leprosy parasite *Mycobacterium leprae*. Nature (London) **316:**450–452.