JOURNAL OF BACTERIOLOGY, Aug. 1993, p. 4930–4935
0021-9193/93/154930-06$02.00/0

Vol. 175, No. 15

Isolation and Sequencing of the Gene Coding for an Antigenic 34-Kilodalton Protein of *Mycobacterium paratuberculosis*

PHILIPPE GILOT,[1] MYRIAM DE KESEL,[1] LIEVE MACHTELINCKX,[2] MARC COENE,[1] AND CARLO COCITO[1]*

Microbiology and Genetics Unit, Institute of Cell Pathology, University of Louvain Medical School, Brussels 1200,[1] and Innogenetics, Ghent 9052,[2] Belgium

Received 9 March 1993/Accepted 17 May 1993

The gene coding for an antigenic 34-kDa protein of *Mycobacterium paratuberculosis* was isolated and sequenced. The 897-bp open reading frame coded for a novel protein containing specific B epitopes. The occurrence of well-defined hydrophobic and hydrophilic portions suggests the wall location of the protein.

Mycobacterium paratuberculosis, the etiological agent of Johne's disease, contains a major antigen complex called A36 which is highly recognized by sera from cattle infected with *M. paratuberculosis* (6, 9). We have recently shown that a 34-kDa protein of the A36 complex is immunodominant and contains B epitopes specific for all tested mycobacteria, including *Mycobacterium bovis* and *Mycobacterium avium* (6). A 500-bp DNA fragment coding for a portion of the 34-kDa protein has been isolated from a λgt11 genomic library of *M. paratuberculosis* (clone λgt11-a362). This DNA fragment coded for a polypeptide containing B epitopes present in all tested *M. paratuberculosis* strains but not in related bacteria, including many strains of the *M. avium-M. intracellulare-M. scrofulaceum* group (7). In this article, we describe the isolation and the sequence of the complete gene of the 34-kDa protein.

The previously cloned DNA fragment of *M. paratuberculosis* (7) was excised by *Eco*RI digestion from clone λgt11-a362 (18) and inserted into the *Eco*RI site of the sequencing vector pBluescript SK$^+$ (Stratagene, La Jolla, Calif.) (20). The recombinant plasmid was transferred to *Escherichia coli* DH5αF′ by transformation (18), and the insert (boxed in Fig. 1) was sequenced with the universal primers SK and KS (Stratagene) and the T7 sequencing kit used with Deaza G/A T7 sequencing mix (Pharmacia, Uppsala, Sweden).

The gene sequence revealed two open reading frames, possibly in phase with the β-galactosidase gene of λgt11, at each extremity of the inserted mycobacterial DNA. In one direction, a stretch of 299 bp (nucleotides 1340 to 1638 in Fig. 1) corresponding to a 13.6-kDa polypeptide was found, whereas in the opposite direction there was a 179-bp frame (nucleotides 1832 to 1654 in Fig. 1) coding for a 7.2-kDa polypeptide. The original orientation of the inserted mycobacterial DNA fragment in phage λgt11-a362 (unidentifiable after its transfer into the sequencing vector) was established as follows. The mycobacterial DNA insert was cloned in both orientations into the expression plasmid pmTNF.MPH, and both reading frames were translated (Fig. 2). This vector allows the expression of cloned genes as fusion proteins with a polypeptide comprising the first 25 amino acids of mouse tumor necrosis factor and a hexahistidine peptide (4). The resulting recombinant plasmids were introduced into *E. coli* K-12 Δ H1 (ATCC 33767) by transformation (18). Single transformed colonies were grown at 28°C in the presence of tetracycline to an A_{600} of 0.2 and then heat shocked (4 h, 42°C). Cell lysates were tested, in Western blot (immunoblot) experiments, with rabbit antiserum directed against the fusion protein expressed by phage λgt11-a362. Restriction analysis of positive clones (pmTNF.MPHa362) indicated that the correct orientation of the mycobacterial DNA fragment in phage λgt11-a362 involved the transcription of the 299-bp open reading frame, which encoded a 13.6-kDa polypeptide.

Confirmation of this result, and of the corresponding nucleotide sequence, was obtained by sequencing the purified recombinant polypeptide. A sonic extract of the induced recombinant bacteria containing pmTNF.MPHa362 was centrifuged (25,000 × *g*, 15 min, 4°C), and supernatant proteins were precipitated [30% (wt/vol) $(NH_4)_2SO_4$], dissolved (0.5 M NaCl–15 mM potassium phosphate buffer, pH 5.2), and dialyzed overnight at 4°C (against the same buffer). Guanidine chloride was added (final concentration, 4 M in 50 mM phosphate buffer, pH 7.4), and the recombinant polypeptide carrying six adjacent histidines (Fig. 2) was purified from *E. coli* proteins by selective interaction between the polyhistidine peptide and a metal chelate adsorbent (Ni ion affinity chromatography on a chelating fast-flow Sepharose column) (16, 17). After elution (0.5 M NaCl–50 mM potassium phosphate buffer, pH 7.4, containing 4 M guanidine chloride and 150 mM imidazole), the last traces of contaminating *E. coli* proteins were removed by gel filtration (Superdex 75; Pharmacia). The purified recombinant polypeptide (mTNF-a362) was transferred to a nitrocellulose membrane and tested with rabbit antiserum directed against the fusion protein expressed by phage λgt11-a362 by Western blot analyses. This antiserum, which was previously shown to recognize the 34-kDa protein of *M. paratuberculosis* (7), also recognized the mTNF-a362 recombinant polypeptide, thus confirming the correspondence of all these products (Fig. 3).

To allow sequence analysis, the fusion protein was desalted by reverse-phase chromatography (PTC-C18 column; 2.1 by 220 mm; Applied Biosystems, Foster City, Calif.). The fraction that was eluted at 45% acetonitrile contained the fusion protein and was used for peptide mapping. Internal fragmentation was obtained by either enzymatic (0.1 μg of trypsin/μg of protein, 37°C, 18 h) or chemical (3% formic acid, 110°C, 4 h) cleavages, followed by separation of the fragments by reverse-phase chromatography on either PTC-

* Corresponding author.

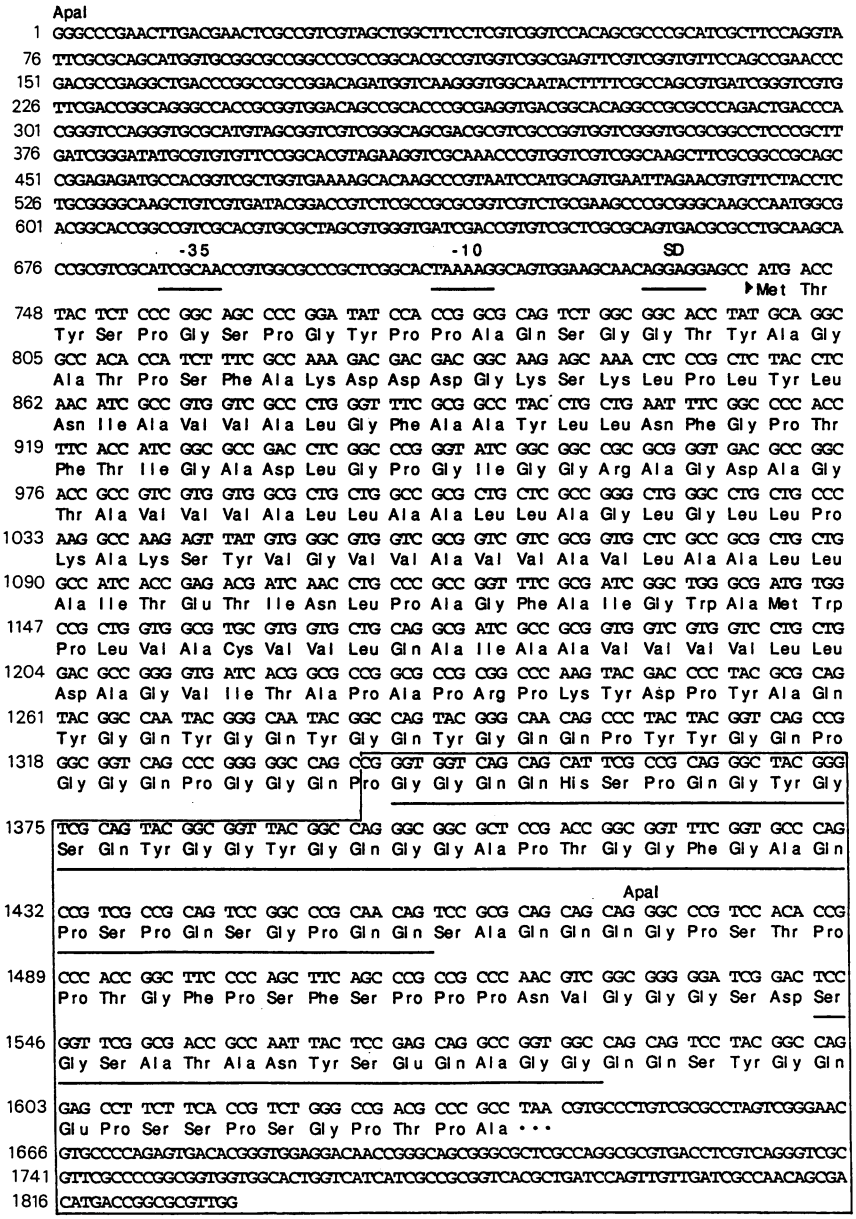

FIG. 1. Nucleotide sequence of the region containing the gene encoding the 34-kDa protein of *M. paratuberculosis*. The sequence of the gene and of its translation product and the potential promoter region (−10 and −35 sequences) and Shine-Dalgarno sequence (SD) are shown. The translation initiation codon, ATG, is indicated by an arrow, and the stop codon is indicated by dots. The amino acids confirmed by internal sequencing of the recombinant polypeptide are underlined. The sequence corresponding to the DNA fragment inserted in clone λgt11-a362 is boxed.

C18 (see above) or RP300-C4 (Aquapor; 2.1 by 100 mm; Applied Biosystems) columns. Peptides were sequenced by the Edman degradation method (8) (model 427A protein sequenator with an on-line 120A amino acid analyzer; Applied Biosystems). The determined peptide sequence (underlined in Fig. 1) corresponded to the 299-bp reading frame of the mycobacterial DNA fragment inserted in the λgt11-a362 clone.

Since the 299-bp reading frame was shown to correspond to the carboxyl-terminal portion of the 34-kDa protein, as indicated by the stop codon at position 1636, the remaining part of the gene was cloned as follows. The mycobacterial DNA insert in the pBluescript SK$^+$ plasmid (see above) was hydrolyzed by endonuclease *Apa*I (18); two fragments including nucleotides 1340 to 1477 and 1478 to 1832 were obtained (Fig. 1). The former segment, corresponding exclusively to the coding region of the 34-kDa protein, was isolated by electroelution from a 2% agarose gel, labeled by random priming with [α-^{32}P]dCTP (18), and used as a probe to detect the remaining part of the gene. For this purpose, this probe was denatured and hybridized with the *Apa*I restriction fragments of the *M. paratuberculosis* genome, which were previously fractionated by agarose gel electrophoresis, denatured, and transferred to a nylon filter (18). The labeled DNA probe was hybridized with blotted DNA (48 h at 49°C in 50% deionized formamide, 0.9 M NaCl, 90

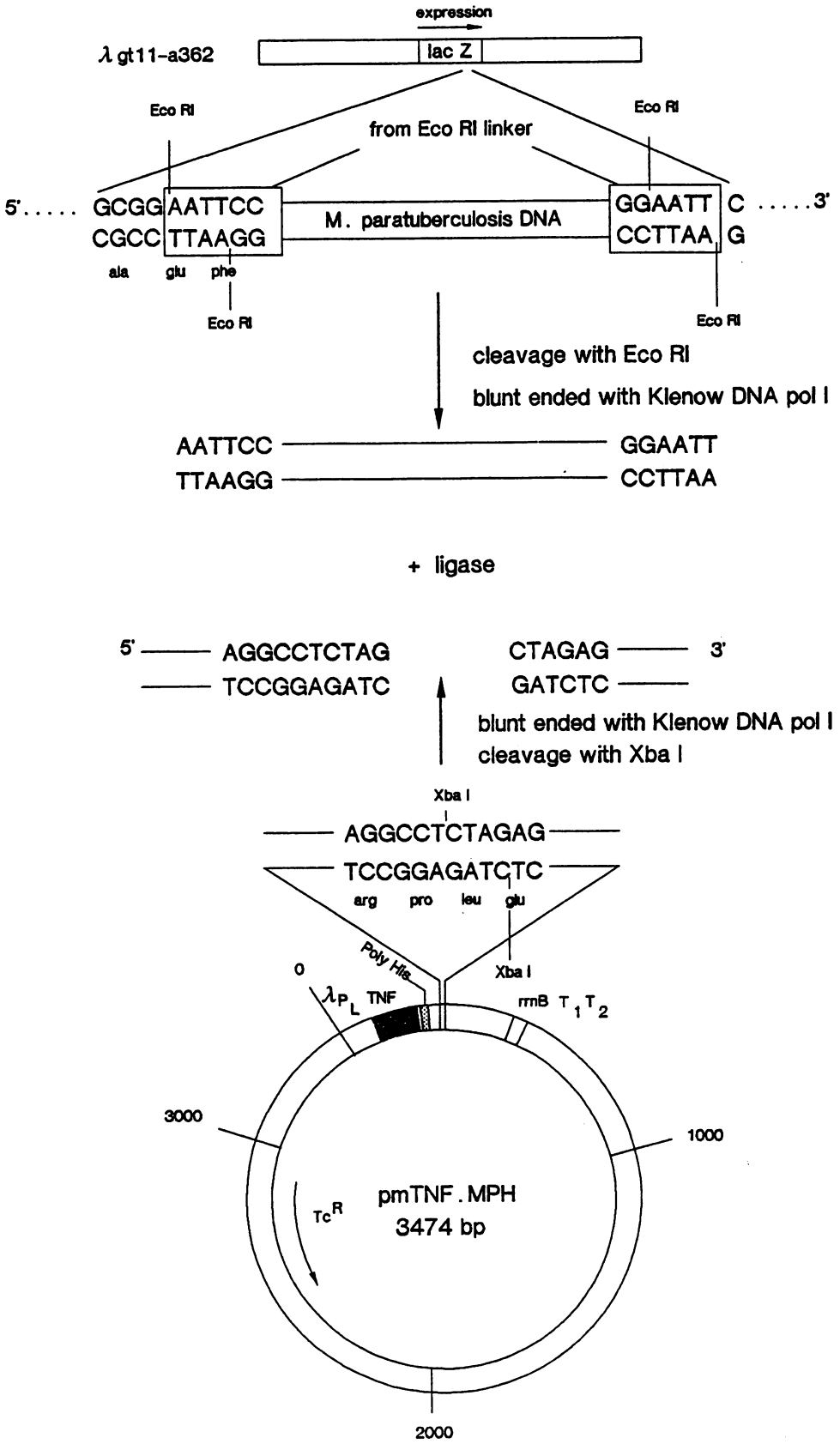

FIG. 2. Cloning and expression of the mycobacterial insert a362. The strategy adopted for cloning of the a362 insert into expression vector pmTNF.MPH is outlined.

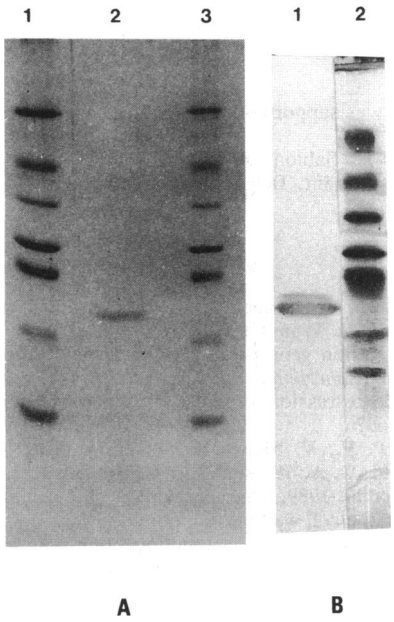

FIG. 3. SDS-PAGE and Western blot analysis ot the recombinant polypeptide. The expression product (1 μg) of clone pmTN F.MPHa362 was electrophoresed on a polyacrylamide gel (SDS–14% PAGE) and either stained with Coomassie blue (panel A, lane 2) or transferred to a nitrocellulose membrane which was incubated with rabbit antiserum directed against the fusion protein expressed by phage λgt11-a362. Membrane-bound immunoglobulins were revealed by peroxidase-labeled protein A (panel B, lane 1), as previously described (9). Molecular mass standards (66.0, 45.0, 36.0, 29.0, 24.0, 20.1, and 14.2 kDa) were stained with either Coomassie blue (panel A, lanes 1 and 3) or india ink (panel B, lane 2).

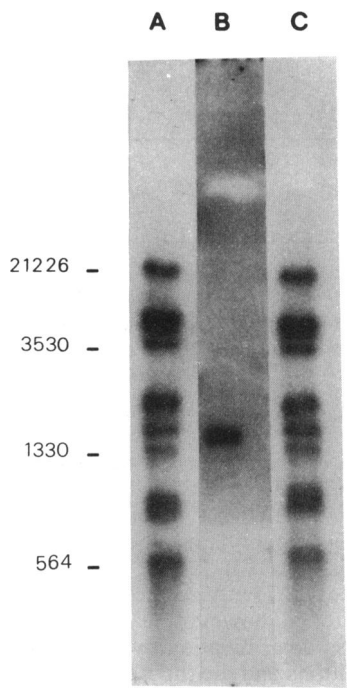

FIG. 4. Identification of the *M. paratuberculosis* DNA fragment containing the beginning of the 34-kDa-protein gene. *Apa*I-digested DNA from *M. paratuberculosis* ATCC 19698 was electrophoresed on a 0.6% agarose gel and hybridized after Southern blotting with a ^{32}P-labeled probe corresponding to nucleotides 1340 to 1477 of the 34-kDa-protein gene (lane B). Molecular mass markers of the indicated sizes (λ DNA cut with *Hin*dIII and *Eco*RI), ^{32}P labeled with the Klenow fragment of DNA polymerase I, were electrophoresed and blotted (lanes A and C). Autoradiography of the membranes was performed at −70°C.

mM sodium citrate, 0.5% sodium dodecyl sulfate [SDS], 0.1% bovine serum albumin, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% denatured salmon sperm DNA). After hybridization, the membranes were repeatedly washed at 60°C (with 0.3 M NaCl–30 mM sodium citrate solution, with the same solution plus 0.1% SDS, and finally with 15 mM NaCl–1.5 mM sodium citrate solution) and autoradiographed. As shown in Fig. 4, a single DNA fragment of 1,500 bp was detected by probe hybridization. According to its molecular mass, this DNA fragment should contain the gene portion corresponding to the NH_2-terminal part of the 34-kDa protein and the sequence preceding it. The *Apa*I-cleaved mycobacterial genome fragments located in the 1,500-bp region, after electroelution from the agarose gel, were cloned into the *Apa*I site of the pBluescript SK$^+$ sequencing vector. Recombinant plasmids were transferred into *E. coli* DH5αF′, and the resulting DNA library was screened (colony hybridization) with the aforementioned probe (18). The mycobacterial DNA insert of a clone thus isolated was sequenced (nucleotides 1 to 1477 in Fig. 1). The orientation of the cloned segment, with respect to the mycobacterial DNA inserted in clone λgt11-a362, was identified owing to the 137-bp sequence (nucleotides 1340 to 1477 in Fig. 1) overlapping the two DNA segments.

The complete sequence of the 34-kDa protein gene and of the adjacent DNA segments is displayed in Fig. 1. It contains an open reading frame of 897 bp, which starts with codon ATG at position 742 and ends with the TAA triplet at position 1638. A Shine-Dalgarno sequence, AGGAGG, fully homologous to the consensus sequence of *E. coli* (19), is located 4 bases upstream of the start codon. A transcriptional promoter is expected to be present in the region preceding the ATG start codon. Promoter-like sequences, similar to those described for *E. coli*, are the hexanucleotide TAAAAG (homology of 4 of 6 bases with the consensus sequence, TATAAT) (10) and the box TCGCAA (50% homologous to the consensus sequence, TTGACA) (10). The length of the 19-nucleotide-long spacer between these two promoter regions is in the 15- to 20-base range of *E. coli* transcriptional promoters (10). The occurrence of promoter sequences different from those found in *E. coli* is still possible. It is known, in fact, that many promoters of cloned mycobacterial genes do not function in *E. coli* (5).

The overall base composition of the 34-kDa protein gene, 70% guanine plus cytosine, is in agreement with the high G+C content of *M. paratuberculosis* chromosomal DNA (67 G+C) (12). The percents G+C occurrence in the first, second, and third positions of the codons were 70.6, 56.5, and 86.2, respectively. The codon usage for the 34-kDa protein was found to be similar to those described for the few known genes of the *M. tuberculosis* complex, with a strong preference for codons with a G or a C in the third position (5). Glycine was the most abundant amino acid of the 34-kDa protein (17.8% of total amino acids), followed by alanine and proline (13.4 and 11.7%, respectively). The exceptionally high frequency of a G or a C in the first position of the codons could be explained by the relative abundance of glycine, alanine, and proline, which are all specified by codons beginning with either a G or a C.

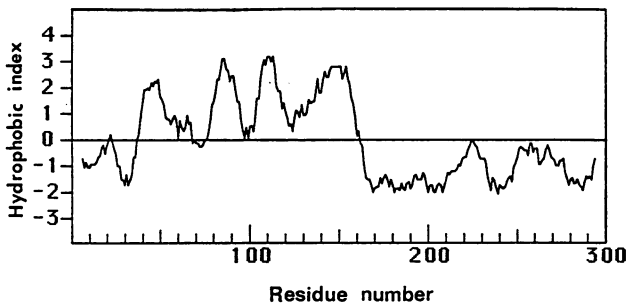

FIG. 5. Kyte-Doolittle hydrophobicity profile of the 34-kDa protein of *M. paratuberculosis*.

The protein corresponding to the open reading frame of the cloned gene herewith reported was initially recognized in a *M. paratuberculosis* lysate fractionated by electrophoresis on a denaturing polyacrylamide gel (SDS-polyacrylamide gel electrophoresis [PAGE]). Under these experimental conditions, the size of the identified protein was estimated to be 34 kDa (6). There is, however, a discordance between the molecular mass determination by electrophoresis under denaturing conditions and the theoretical molecular mass. As a matter of fact, the sum of the molecular masses of the amino acids encoded by the nucleotide sequence of the open reading frame is 29,565 Da. Such a discrepancy might be due to the high proline content of the protein (11.7%), which reduces the mobility of the protein in the SDS-PAGE fractionation, thus leading to an overestimation of its size, a phenomenon previously observed with other proteins (1, 3, 21). In this connection, it is worth noting that the relative proline content of the carboxyl-terminal moiety of the 34-kDa protein is even higher than that of the whole protein. In fact, the recombinant polypeptide mTNF-a362 contains 14.3% proline. Its molecular mass was estimated to be 22,500 Da by its electrophoretic mobility (by SDS-PAGE) (Fig. 3) but only 14,976 Da according to the sum of the amino acid molecular masses.

A hydrophilicity profile of the 34-kDa protein was drawn with the "DNA Strider" computer program (14) by the method of Kyte and Doolittle (13). The profile shown in Fig. 5 identifies two distinct regions: a strongly hydrophobic portion corresponding to amino acids 40 to 160 and a highly hydrophilic part encompassing amino acids 161 to 298. Such a peculiar feature is compatible with a peripheral location of the 34-kDa protein within the mycobacterial cell, the hydrophobic NH_2 moiety being buried within the envelope and the hydrophilic COOH part being exposed at the cell surface. This view is supported by the observation that B epitopes localized in the carboxyl-terminal part of the protein (amino acids 200 to 298, expressed by clone a362) were found by immune electron microscopy to be present at the cell surface (7). Some other antigenic proteins associated with cell wall polymers have also been reported by others (2, 7, 11, 15).

In conclusion, this article describes the isolation and sequencing of a new mycobacterial gene and the identification of a protein endowed with interesting biochemical and immunological properties. A search in the EMBL data bank (release 31) indicated the uniqueness of the 34-kDa protein, with no resemblance to other known proteins.

Nucleotide sequence accession number. The nucleotide sequence of the 34-kDa protein of *M. paratuberculosis* has been submitted to the EMBL data library under accession number X68102.

This work was supported by a grant from the Belgian association IRSIA.

We thank E. Sablon, J. Raymackers, and W. Van Arnhem (Innogenetics, Ghent, Belgium) for help in peptide and DNA sequencing.

REFERENCES

1. **Aldovini, A., C. Debouck, M. B. Feinberg, M. Rosenberg, S. K. Arya, and F. Wong-Staal.** 1986. Synthesis of the complete trans-activation gene product of human T-lymphotropic virus type III in *Escherichia coli*: demonstration of immunogenicity *in vivo* and expression *in vitro*. Proc. Natl. Acad. Sci. USA **83**:6672–6676.
2. **Barnes, P. F., V. Mehra, G. R. Hirschfield, S. J. Fong, C. Abou-Zeid, G. A. W. Rook, S. W. Hunter, P. J. Brennan, and R. L. Modlin.** 1989. Characterization of T-cell antigens associated with the cell wall protein-peptidoglycan complex of *Mycobacterium tuberculosis*. J. Immunol. **143**:2656–2662.
3. **Cherayil, B. J., and R. A. Young.** 1988. A 28-kDa protein from *Mycobacterium leprae* is a target of the human antibody response in lepromatous leprosy. J. Immunol. **141**:4370–4375.
4. **Cocito, C., M. Coene, M. De Kesel, and P. Gilot.** October 1992. International patent WO92/16628.
5. **Dale, J. W., and A. Patki.** 1990. Mycobacterial gene expression and regulation, p. 172–198. *In* J. McFadden (ed.), Molecular biology of the mycobacteria. Surrey University Press, London.
6. **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.
7. **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.
8. **Edman, P., and G. Begg.** 1967. A protein sequenator. Eur. J. Biochem. **1**:80–91.
9. **Gilot, P., M. De Kesel, M. Coene, and C. Cocito.** 1992. Induction of cellular immune reactions by A36, an antigen complex of *Mycobacterium paratuberculosis*. Comparison of A36 and johnin components. Scand. J. Immunol. **36**:811–821.
10. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
11. **Hirschfield, G. R., M. McNeil, and P. J. Brennan.** 1990. Peptidoglycan-associated polypeptides of *Mycobacterium tuberculosis*. J. Bacteriol. **172**:1005–1013.
12. **Imaeda, T., G. Broslawski, and S. Imaeda.** 1988. Genomic relatedness among mycobacterial species by nonisotopic blot hybridization. Int. J. Syst. Bacteriol. **38**:151–156.
13. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
14. **Marck, C.** 1988. "DNA strider": a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16**:1829–1836.
15. **McNeil, M. R., and P. J. Brennan.** 1991. Structure, function and biogenesis of the cell envelope of mycobacteria in relation to bacterial physiology, pathogenesis and drug resistance: some thoughts and possibilities arising from recent structural information. Res. Microbiol. **142**:451–463.
16. **Porath, J., J. Carlsson, I. Olsson, and G. Belfrage.** 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. Nature (London) **258**:598–599.
17. **Porath, J., and B. Olin.** 1983. Immobilized metal ion affinity adsorption and immobilized metal ion affinity chromatography of biomaterials. Serum protein affinities for gel-immobilized iron

and nickel ions. Biochemistry **22**:1621–1630.
18. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
20. **Short, J. M., J. M. Fernandez, J. A. Sorge, and W. D. Huse.** 1988. λ ZAP: a bacteriophage λ expression vector with in vivo excision properties. Nucleic Acids Res. **16**:7583–7600.
21. **Thole, J. E. R., L. F. E. M. Stabel, M. E. G. Suykerbuyk, M. Y. L. De Wit, P. R. Klatser, A. H. J. Kolk, and R. A. Hartskeerl.** 1990. A major immunogenic 36,000-molecular-weight antigen from *Mycobacterium leprae* contains an immunoreactive region of proline-rich repeats. Infect. Immun. **58**:80–87.