INFECTION AND IMMUNITY, Sept. 1990, p. 3154–3157
0019-9567/90/093154-04$02.00/0

Vol. 58, No. 9

Nucleotide Sequence of *htpB*, the *Legionella pneumophila* Gene Encoding the 58-Kilodalton (kDa) Common Antigen, Formerly Designated the 60-kDa Common Antigen

JACQUELYN S. SAMPSON,[1]* STEVEN P. O'CONNOR,[1] BRIAN P. HOLLOWAY,[2] BONNIE B. PLIKAYTIS,[1] GEORGE M. CARLONE,[1] AND LEONARD W. MAYER[1]†

Divisions of Bacterial Diseases[1] and Viral Diseases,[2] Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333

Received 6 March 1990/Accepted 26 June 1990

Gene *htpB*, which encodes the 58-kilodalton protein of *Legionella pneumophila*, was cloned in *Escherichia coli* and its complete nucleotide sequence was determined. Analysis of this sequence revealed an open reading frame of 1,644 nucleotides encoding a protein with a predicted molecular mass of 57,952 daltons. Data obtained by amino-terminal sequencing of the purified 58-kilodalton protein agreed, except for one amino acid residue, with the predicted amino acid sequence, identifying this open reading frame as *htpB*. A comparison of the primary structure of this protein to other proteins of similar molecular weights from *E. coli*, *Mycobacterium leprae*, *M. tuberculosis*, and *Coxiella burnetii* revealed significant regions of sequence similarity, which are discussed.

In an earlier study, we identified a genus-wide 58-kilodalton (kDa) protein in *Legionella pneumophila* which reacted with convalescent-phase sera from culture-positive patients with active *L. pneumophila* infections (20). Data reported here suggest that the molecular mass of this protein is 58 kDa, not 60 kDa as previously reported (19). Investigation of this 58-kDa protein in our laboratory and elsewhere has shown the following. (i) It shares cross-reactive epitopes with 38 *Legionella* serogroups and at least 39 other species of gram-negative bacteria (19). (ii) It is a heat shock or stress response protein that exhibits homology to the GroEL protein of *Escherichia coli* (14, 15) and the 65-kDa protein of *Mycobacterium tuberculosis* (24). (iii) It contains *Legionella*-specific epitopes, in addition to epitopes common to other bacteria (19). (iv) The gene that encodes this protein may be part of a heat shock operon similar to *groE* (recently renamed *mop* [2]) in *E. coli* and the *htpB* locus in *Coxiella burnetii* (11). To facilitate identification of the *Legionella*-specific antigenic domains, we cloned and sequenced *htpB*, the *L. pneumophila* 58-kDa protein gene.

Chromosomal DNA was extracted from *L. pneumophila* Philadelphia 1 (27) and purified as previously described (5). The DNA was digested with restriction enzyme *Bam*HI as recommended by the manufacturer (New England BioLabs, Inc., Beverly, Mass.), and the DNA fragments were ligated into *Bam*HI-digested pUC13 (29) with T4 ligase (New England BioLabs). DNA from the ligation was transformed into *E. coli* TB1 (4) by electroporation (Gene Pulser; Bio-Rad Laboratories, Richmond, Calif.), and transformants were selected by plating on L agar containing 200 μg of carbenicillin (Sigma Chemical Co., St. Louis, Mo.) per ml, 0.4 mM isopropyl-β-D-thiogalactopyranoside (Research Organics, Inc., Cleveland, Ohio), and 0.4 μg of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Research Organics). Colonies of transformants were transferred to nitrocellulose and screened for expression of a 58-kDa protein with an anti-58-kDa monoclonal antibody (MAb) pool which did not cross-react with *E. coli* TB1 (J. S. Sampson et al., unpublished data). Of the approximately 3,500 colonies screened, a single positive colony was identified.

Recombinant plasmid pLWM4500 was found to contain a 9.0-kilobase *Bam*HI fragment. Restriction enzyme analysis identified a single *Eco*RI site located 3.2 kilobases into the insert. A subclone consisting of a 3.2-kilobase *Eco*RI-*Eco*RI fragment (the second *Eco*RI site in the polylinker) in pUC13 was shown by screening, as described earlier to encode the *htpB* gene product. One positive subclone with recombinant plasmid pLWM4504 was selected for further study.

Plasmid DNA was prepared by standard techniques (13, 16), and restriction sites for *Bgl*II, *Hin*dIII, *Nsi*I, and *Pst*I were mapped in the insert. Also, expression of a 58-kDa protein by this clone was examined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and immunoblot analysis as described previously (19). Total cellular protein bound to nitrocellulose were reacted with either rabbit polyclonal anti-58-kDa serum (Fig. 1A) or the MAb pool described above (Fig. 1B). In panel A, the anti-58-kDa protein serum cross-reacted with the GroEL protein in the negative controls (TB1 containing a random *L. pneumophila* DNA fragment cloned into pUC13; lanes 1 and 3) and a 58-kDa protein purified from *L. pneumophila* (lane 4). The recombinant strain (lane 2) expressed a protein which binds the antibody and comigrated with both the purified 58-kDa protein and GroEL. Similar results were observed for an identical blot exposed to the MAb pool (panel B), except that these antibodies did not cross-react with GroEL.

To determine the nucleotide sequence of *htpB*, restriction fragments from pLWM4504 were subcloned into bacteriophage M13 and their sequences were determined by the method of Sanger et al. (21) with T7 polymerase (Sequenase; U.S. Biochemical Corp., Cleveland, Ohio). Reaction products were electrophoresed on 6 and 8% polyacrylamide gels and autoradiographed. Because their sizes precluded sequencing entirely across most of the templates, 11 oligonucleotides were synthesized (3, 8, 25) and used as primers in place of the M13 universal primer. The sequences of both strands of DNA were determined. The sequence data ob-

* Corresponding author.
† Present address: Division of Vector-Borne Infectious Diseases, Center for Infectious Diseases, Centers for Disease Control, Ft. Collins, CO 80522.

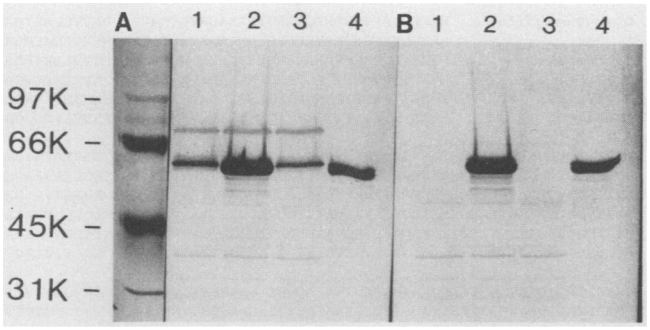

FIG. 1. Immunoblot analysis of the 58-kDa protein expressed by recombinant *E. coli* TB1(pLWM4504). Panel A was tested with rabbit polyclonal anti-58-kDa serum, and panel B was tested with the MAb pool. Lanes: 1 and 3, *E. coli* TB1 control (pUC13 with a random *L. pneumophila* fragment) whole-cell lysate; 2, *E. coli* (pLWM4504); 4, purified 58-kDa protein. Cells were adjusted to the same optical density before solubilization and sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Molecular weights (in thousands) are indicated on the left.

tained were analyzed with the DNASTAR software package (DNASTAR Inc., Madison, Wis.) and with the University of Wisconsin Genetics Computer Group sequence analysis software package (7). Amino-terminal sequence determination of the purified 60-kDa protein was done as previously described (9, 28).

A total of 2,016 nucleotides were sequenced. Analysis of the sequence revealed one open reading frame of 1,644 nucleotides (Fig. 2). This open reading frame was confirmed by correlating the derived amino acid sequence with the N-terminal amino acid sequence of the purified *L. pneumophila* 58-kDa protein. With the exception of one amino acid residue, the derived amino acid sequence from nucleotides +4 to +96 (Fig. 2) correlated with the 31 N-terminal amino acids determined by automated sequencing of the purified protein. There were two methionine start codons (ATG) which could be used to initiate protein synthesis at nucleotides −6 and +1 (Fig. 2). N-terminal amino acid sequencing of the protein identified alanine as the first amino acid in the purified protein (Fig. 2). Therefore, we believe that the sequence begins at the second methionine (ATG), which is absent in the purified protein because of posttranslational excision, in keeping with the observation of Hirel et al. (10). They studied methionine excision and observed that its occurrence is inversely related to the size of the side chain on the second amino acid of the protein. The location of the putative ribosomal binding site or Shine-Dalgarno sequence (22) is an additional proof that the second ATG begins the sequence. The sequence AGGAGA (with five of six nucleotides matching consensus) is eight nucleotides upstream from the second ATG, an interval more in agreement with that reported for ribosomal binding site locations.

Analysis of the nucleotide sequence for promoters did not identify any regions with sequences similar to consensus −10 and −35 sequences. One explanation for this lack of possible promoter sequences is suggested by studies showing that *htpB* may exist as part of a bicistronic operon similar to *groE* and a related *C. burnetii* heat shock operon. Hoffman et al. (11) cloned a gene encoding an *L. pneumophila* 58-kDa protein and found that a 15-kDa heat shock protein was expressed concurrently with the 58-kDa protein. If the *L. pneumophila* gene is arranged in an operon similar to its other bacterial homologs, promoter sequences would be expected to be quite a distance upstream and would be

```
▼-210
CGTGGTGAGATCATTGCTGTCGGCGCTGGCAAAGTATTGGAAAACGGTGATGTTCGTGCTTTA
▼-147
GCAGTTAAAGTAGGTGATGTGGTATTGTTTGGCAAGTACTCTGGTACTGAAGTTAAAGTTGAT
▼-84
GGAAAAGAATTAGTTGTGATGCGCGAAGACGACATCATGGGTGTAATCGAGAAGTAATCTAAT
▼-21          ▼1
TGTTTAAAGGAGACAATGATAATGGCTAAAGAATTACGTTTTGGTGATGACGCTCGCCTACAA
      S-D          M  A  K  E  L  R  F  G  D  D  A  R  L  ▪
▼43
ATGCTTGCTGGTGTTAATGCATTAGCAGATGCGGTTCAAGTTACTATGGGTCCACGTGGTCGT
M  L  A  G  V  N  A  L  A  D  A  V  Q  V  T  M  G  P  R  G  R
▼106
AATGTTGTATTGGAAAAATCTTATGGCGCTCCTACTGTAACTAAAGACGGTGTGTCTGTTGCC
N  V  V  L  E  K  S  Y  G  A  P  T  V  T  K  D  G  V  S  V  A
▼169
AAAGAAATTGAGTTTGAGCATCGTTTCATGAACATGGGCGCTCAAATGGTTAAAGAAGTGGCT
K  E  I  E  F  E  H  R  F  M  N  M  G  A  Q  M  V  K  E  V  A
▼232
TCTAAAACTTCTGATACTGCTGGTGATGGTACTACTACTGCAACAGTATTGGCTCGTTCTATT
S  K  T  S  D  T  A  G  D  G  T  T  T  A  T  V  L  A  R  S  I
▼295
CTTGTTGAAGGTCACAAAGCAGTTGCTGCTGGTATGAATCCAATGGATCTCAAACGCGGTATT
L  V  E  G  H  K  A  V  A  A  G  M  N  P  M  D  L  K  R  G  I
▼358
GATAAAGCAGTATTAGCAGTTACCAAAAAATTACAAGCTATGTCTAAGCCATGCAAAGACAGC
D  K  A  V  L  A  V  T  K  K  L  Q  A  M  S  K  P  C  K  D  S
▼421
AAAGCTATTGCTCAAGTTGGAACTATTTCTGCTAATTCCGATGAAGCGATTGGTGCTATCATT
K  A  I  A  Q  V  G  T  I  S  A  N  S  D  E  A  I  G  A  I  I
▼484
GCTGAAGCAATGGAAAAAGTTGGTAAAGAGGGTGTTATTACCGTTGAAGATGGTAATGGATTG
A  E  A  M  E  K  V  G  K  E  G  V  I  T  V  E  D  G  N  G  L
▼547
GAAAATGAGCTTTCTGTTGTTGAAGGTATGCAATTGATCGCGGTACATTCTCCATACTTTATC
E  N  E  L  S  V  V  E  G  M  Q  L  I  A  V  H  S  P  Y  F  I
▼610
AACAACCAGCAAAACATGAGCTGTGAACTTGAGCATCCATTCATTTTATTGGTTGACAAAAAA
N  N  Q  Q  N  M  S  C  E  L  E  H  P  F  I  L  L  V  D  K  K
▼673
GTTTCCAGTATTCGTGAAATGTTGTCCGTATTGGAAGGTGTTGCCAAATCTGGTCGTCCTTTA
V  S  S  I  R  E  M  L  S  V  L  E  G  V  A  K  S  G  R  P  L
▼736
TTGATCATTGCAGAAGATGTTGAAGGCGAAGCTTTAGCTACTCTGGTAGTCAACAACATGCGC
L  I  I  A  E  D  V  E  G  E  A  L  A  T  L  V  V  N  N  M  R
▼799
GGTATTGTAAAAGTATGTGCTGTCAAAGCGCCTGGTTTTGGTGATCGCCGCAAAGCGATGTTG
G  I  V  K  V  C  A  V  K  A  P  G  F  G  D  R  R  K  A  M  L
▼862
CAAGACATTGCTATTTTGACTAAGGGTCAAGTTATTTCTGAAGAAATTGGCAAGAGCTTGGAA
Q  D  I  A  I  L  T  K  G  Q  V  I  S  E  E  I  G  K  S  L  E
▼925
GGTGCTACTCTGGAAGATCTTGGTAGTGCTAAGCGAATCGTTGTTACCAAAGAAAACACTACT
G  A  T  L  E  D  L  G  S  A  K  R  I  V  V  T  K  E  N  T  T
▼988
ATCATTGATGGTGAAGGAAAGGCAACTGAAATTAATGCTCGTATTACTCAAATTCGTGCACAA
I  I  D  G  E  G  K  A  T  E  I  N  A  R  I  T  Q  I  R  A  Q
▼1,051
ATGGAAGAAACCACTTCTGATTACGATAGAGAAAAATTACAAGAGCGCGTTGCTAAACTAGCT
M  E  E  T  T  S  D  Y  D  R  E  K  L  Q  E  R  V  A  K  L  A
▼1,114
GGTGGTGTTGCTGTTATCAAAGTTGGCGCTGCTACAGAAGTTGAAATGAAAGAGAAGAAAGCA
G  G  V  A  V  I  K  V  G  A  A  T  E  V  E  M  K  E  K  K  A
▼1,177
CGTGTTGAAGATGCTCTTCATGCTACTCGCGCTGCAGTAGAAGAAGGTATCGTTGCCGGTGGT
R  V  E  D  A  L  H  A  T  R  A  A  V  E  E  G  I  V  A  G  G
▼1,240
GGTGTTGCCTTGATTCGTGCTCAGAAAGCTCTTGATTCATTGAAAGGCGATAATGACGATCAA
G  V  A  L  I  R  A  Q  K  A  L  D  S  L  K  G  D  N  D  D  Q
▼1,303
AATATGGGTATCAATATTTTACGTCGCGCTATTGAATCTCCAATGCGTCAAATTGTTACTAAC
N  M  G  I  N  I  L  R  R  A  I  E  S  P  M  R  Q  I  V  T  N
▼1,366
GCAGGATATGAAGCTTCTGTTGTAGTAAACAAGGTAGCTGAGCACAAAGACAACTACGGTTTC
A  G  Y  E  A  S  V  V  V  N  K  V  A  E  H  K  D  N  Y  G  F
▼1,429
AACGCTGCAACTGGTGAATACGGTGATATGGTTGAGATGGGTATTCTTGATCCAACCAAAGTA
N  A  A  T  G  E  Y  G  D  M  V  E  M  G  I  L  D  P  T  K  V
▼1,492
ACCCGTATGGCTCTGCAAAATGCAGCTTCTGTAGCCAGTTTGATGTTGACTACTGAATGTATG
T  R  M  A  L  Q  N  A  A  S  V  A  S  L  M  L  T  T  E  C  M
▼1,555
GTTGCTGATCTGCCTAAGAAAGAAGAAGGTGTTGGTGCCGGCGATATGGGCGGCATGGGCGGA
V  A  D  L  P  K  K  E  E  G  V  G  A  G  D  M  G  G  M  G  G
▼1,618                1,644▼
ATGGGTGGCATGGGCGGAATGATGTAATTTTCTCCCGGTTATCCAATAAAAAACCCGCTGTAA
M  G  G  M  G  G  M  M
▼1,680
AGGCGTTTTTTATTAATGTTTATTTTGTTCTTTGACTGTTTTGAACCTATCCTTAATAGCGTG
▼1,743                                                        1,806▼
AATAAATAACAAGGAAAGAATAATGAGTAAAGTAATTAAAAAGCTGGAAGTCGCTCTTGCTGA
```

FIG. 2. Nucleotide sequence of the *htpB* gene. The predicted amino acid sequence of the 58-kDa protein is shown below the nucleotide sequence. Amino acids that are underlined are those which agree with residues obtained from automated amino-terminal sequencing. The glutamine indicated by shading was ambiguous on N-terminal analysis and appeared to be C. The putative Shine-Dalgarno sequence is underlined. A dot marks the position of the termination codon.

separated from the 58-kDa protein gene by the sequences encoding the smaller protein. Unfortunately, we cannot confirm this, since our data do not extend to this region. The deduced amino acid sequence was 547 amino acids long with a computed molecular mass of 57,952 daltons. This compares favorably with the previously reported molecular masses of 58 and 60 kDa (determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis [19, 20]) and 63 kDa (chromatographically determined [18]). Hydrophobicity analysis as described by Hopp and Woods (12) revealed no significant areas of hydrophobicity within the sequence. The computed isoelectric point was 5.27, which differs slightly from the value 5.8 reported by Pau et al. (18).

Since our ultimate purpose was identification of epitopes both specific and common to *L. pneumophila*, it was of interest to compare the primary structure of this protein with its immunologically related counterparts in *E. coli*, *M. tuberculosis*, *M. leprae*, and *C. burnetii*. Figure 3 shows the complete alignment of all five amino acid sequences, with the consensus sequence shown below. Analysis of this alignment reveals the following. (i) A strong sequence similarity exists between *E. coli* GroEL and the *L. pneumophila* 58-kDa protein (73.8%) and between the *Legionella* and *Coxiella* proteins (75.3%). (ii) Lesser sequence similarity exists between the *L. pneumophila* 58-kDa protein and the *M. tuberculosis* 65-kDa protein (57.4%) or the *M. leprae* 65-kDa protein (56.8%). Long stretches of sequence similarity among the five proteins are located throughout the protein sequence. This confirms reports by others of the highly conserved nature of this gene, even among bacteria which are phylogenetically distant (23, 26).

Several investigators have defined epitopes for the 65-kDa protein of *M. leprae* by using (MAbs) (1, 17). Some of these same *M. leprae* MAbs have been shown to cross-react with *M. tuberculosis* and other bacterial species, specifically, *E. coli* and *L. pneumophila* (23). By using the published epitope map (23) and defined epitopes of these cross-reactive *M. leprae* MAbs (1), we have located their probable position within the *L. pneumophila* 58-kDa protein. MAb Y1.2, which reacts with all of the above-mentioned bacteria, maps to an area of moderate homology in the N terminus, where 9 (53%) of 17 amino acids match. The epitope sequence for MAb IIH9, which reacts similarly to MAb Y1.2, contains 12 (75%) of 16 amino acids which match. MAb IIIC8, which does not react with *M. tuberculosis* but does cross-react with *E. coli* and *L. pneumophila*, binds to a carboxy-terminal epitope where 5 (46%) of 11 amino acids match. This suggests that conservation of epitopes, as well as of the primary sequence, is a characteristic of this protein and is one basis for cross-reactivity among the bacteria whose sequences appear in Fig. 3. Moreover, these analyses provide a probable molecular explanation for the existence of some of the *Legionella* 58-kDa common epitopes which have been serologically demonstrated (19). These areas of moderate-to-high homology are in contrast to the site for MAb IIC8 binding, where only 3 (25%) of 12 amino acids match. The facts that this MAb reacts only with *Mycobacterium* spp. and its epitope lies in a more divergent area of the gene suggest that genus-specific epitopes lie in regions of high heterogeneity. Peptide mapping will be a definitive way to confirm the locations of these epitopes and others of importance.

We, as others, have found that the *L. pneumophila* 58-kDa protein gene is highly conserved. Cross-reactivity patterns indicate that in addition to the four discussed here, there are similar proteins in other organisms (19), which most likely

```
L. pneumophila      M-AKELRFGDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYGA
E. coli             MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGA
C. burnetii         MAAKVLKFSHEVLHAMSRGVEVLANAVKVTLGPKGRNVVLDKSFGA
M. tuberculosis     M-AKTIAYDEEARRGLERGLNALADAVKVTLGPKGRNVVLEKKWGA
M. leprae           M-AKTIAYDEEARRGLERGLNSLADAVKVTLGPKGRNVVLEKKWGA
Consensus           M AK : :.::.   :  G:: LADAV:VT:GP:GRNVVL:K..GA

PTVTKDGVSVAKEIEFEHRFMNMGAQMVKEVASKTSDTAGDGTTTATVLARSILVEGHKAVAAGM
PTITKDGVSVAREIELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGM
PTITKDGVSVAKEIELEDKFENMGAQMVKEVASRTSDDAGDGTTTATVLAQAILVEGIKAVIAGM
PTITNDGVSIAKEIELEDPYEKIGAELVKEVAKKTDDVAGDGTTTATVLAQALVREGLRNVAAGA
PTITNDGVSIAKEIELEDPYEKIGAELVKEVAKKTDDVAGDGTTTATVLAQALVKEGLRNVAAGA
PT:T:DGVS:A:EIELE: : ::GA::VKEVA.::.D AGDGTTTATVLAQ::: EG :.V AG

NPMDLKRGIDKAVLAVTKKLQAMSKPCKDSKAIAQVGTISANSDEAIGAIIAEAMEKVGKEGVIT
NPMDLKRGIDKAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVIT
NPMDLKRGIDKAVTAAVAELKKISKPCKDQKAIAQVGTISANSDKSIGDIIAEAMEKVGKEGVIT
NPLGLKRGIEKAVEKVTETLLKGAKEVETKEQIAATAAISA-GDQSIGDLIAEAMDKVGNEGVIT
NPLGLKRGIEKAVDKVTETLLKDAKEVETKEQIAATAAISA-GDQSIGDLIAEAMDKVGNEGVIT
NP::LKRGI:KAV  .. .L   :   .. ..IA..::ISA :D.::G :IAEAM:KVG:EGVIT

VEDGNGLENELSVVEGMQLI-AVHSPYFINNQQNMSCELEHPFILLVDKKVSSIREMLSVLEGVA
VEDGTGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAVA
VEDGSGLENALEVVEGMQFDRGYLSPYFINNQQNMSAELENPFILLVDKKISNIRELIPLLENVA
VEESNTFGLQLELTEGMRFDKGYISGYFVTDPERQEAVLEDPYILLVSSKVSTVKDLLPLLEKVI
VEESNTFGLQLELTEGMRFDKGYISGYFVTDAERQEAVLEEPYILLVSSKVSTVKDLLPLLEKVI
VE:::.:: .L.:.EGM::  :  S YF:...:  .  LE P:ILL...K:S.::::::::LE V

KSGRPLLIIAEDVEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTKGQVISEEIG
KAGKPLLIIAEDVEGEALATAVVNTIRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTVISEEIG
KSGRPLLVIAEDIEGEALATLVVNNIRGVVKVAAVKAPGFGDRRKAMLQDIAVLTGGKVISEEVG
GAGKPLLIIAEDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAMLQDMAILTGGQVISEEVG
QAGKSLLIIAEDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAMLQDMAILTGAQVISEEVG
 :G::LL:IAED:EGEAL:T VVN.:RG. K  AVKAPGFGDRRKAMLQD:A.LT : VISEE:G

KSLEGATLEDLGSAKRIVVTKENTTIIDGEGKATEINARITQIRAQMEETTSDYDREKLQERVAK
MELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDYDREKLQERVAK
LSLEAASLDDLGSAKRVVVTKDDTTIIDGSGDAGDIKNRVEQIRKEIENSSSDYDKEKLQERLAK
LTLENADLSLLGKARKVVVTKDETTIVEGAGDTDAIAGRVAQIRQEIENSDSDYDREKLQERLAK
LTLENTDLSLLGKARKVVMTKDETTIVEGAGDTDAIAGRVAQIRTEIENSDSDYDREKLQERLAK
 .LE :.L. LG AKR:V:.K:.TTI::G G....I .R::QIR ::E::.SDYDREKLQER:AK

LAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALIRAQKALDSLKGDNDDQ
LAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALIRVASKLADLRGQNEDQ
LAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVALIRVLKSLDSVEVENEDQ
LAGGVAVIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLLQAAPTLDELKLEG-DE
LAGGVAVIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLLQAAPALDKLKLTG-DE
LAGGVAVIK.GAATEVE:KE:K R:EDA::.::AAVEEG:V:GGGV:L::.   L..:   . D:

NMGINILRRAIESPMRQIVTNAGYEASVVVNKVAEHK-DNYGFNAATGEYGDMVEMGILDPTKVT
NVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKGGD-GNYGYNAATEEYGNMIDMGILDPTKVT
RVGVEIARRAMAYPLSQIVKNTGVQAAVVADKVLNHKDVNYGYNAATGEYGDMIEMGILDPTKVT
ATGANIVKVALEAPLKQIAFNSGLEPGVVAEKVRNLP-AGHGLNAQTGVYEDLLAAGVADPVKVT
ATGANIVKVALEAPLKQIAFNSGMEPGVVAEKVRNLS-VGHGLNAATGEYEDLLKAGVADPVKVT
. G ::   A:E:P::QI. N G E::VV.:.V .    ..G.NA.T. Y.:::  G: DP.KVT

RMALQNAASVASLMLTTECMVADLP-KKEEGV-GAGD-MGGMGGMGGMGGMM
RSALQYAASVAGLMITTECMVTDLP-KNDAADLGAA---GGMGGMGGMGGMM
RTALQNAASIAGLMITTECMVTEAPKKKEESMPGGGD-MGGMGGMGGMGGMM
RSALQNAASIAGLFLTTEAVVADKPEKEKASVPGGGD-MGGMDF
RSALQNAASIAGLFLTTEAVVADKPEKTAAPASDPTGGMGGMDF
R ALQ AAS:A:L.:TTE :V:D P K..:   :     GGM:
```

FIG. 3. Comparison of amino acid sequences of the 58-kDa protein of *L. pneumophila*, the GroEL protein of *E. coli*, the 62-kDa protein of *Coxiella burnetii*, and the 65-kDa proteins of *M. tuberculosis* and *M. leprae*. The consensus sequence is shown below. Colons represent highly conserved amino acids, and periods represent moderately conserved amino acids (6).

means that homology exists between their respective sequences and that of the protein from *L. pneumophila*. The function of this protein in *L. pneumophila* and other species of bacteria is not completely understood; however, this sequence conservation may indicate a similarity in biological function.

Nucleotide sequence accession number. The data in Fig. 2 have been submitted to GenBank under accession no. M35149.

We are grateful to W. Lanier Thacker for help in preparing *Legionella* DNA, Carol Alosio for MAbs, Kristen Birkness for technical advice, Thomas Shinnick and Sally Berish for helpful discussions and suggestions, Harold Russell and Carolyn Black for

help in manuscript preparation, and Sharon Tart for clerical assistance.

LITERATURE CITED

1. **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.
2. **Bachmann, B. J.** 1987. Linkage map of *Escherichia coli* K-12, edition 7, p. 807–876. *In* F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
3. **Becker, C. R., J. W. Efcavitch, C. R. Heiner, and N. F. Kaiser.** 1985. Use of a reverse phase column for the HPLC purification of synthetic oligonucleotides. J. Chromatogr. **326**:293–299.
4. **Bethesda Research Laboratories.** 1984. *E. coli* TB1—host for pUC plasmids. Focus **6**:7.
5. **Brenner, D. J., A. G. Steigerwalt, G. W. Gorman, H. W. Wilkinson, W. F. Bibb, M. Hackel, R. L. Tyndall, J. Campbell, J. C. Feeley, W. L. Thacker, P. Skaliy, W. T. Martin, B. J. Brake, B. S. Fields, H. V. McEachern, and L. K. Corcoran.** 1985. Ten new species of Legionella. Int. J. Syst. Bacteriol. **35**:50–59.
6. **Dayhoff, M. O., R. V. Eck, and C. M. Park.** 1972. A model of evolutionary change in proteins, p. 89–99. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5. National Biomedical Research Foundation, Silver Spring, Md.
7. **Devereau, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
8. **Heron, E. J.** 1984. System for automated DNA synthesis. Am. Biotechnol. Lab. **2**:52–54.
9. **Hewick, R. M., M. W. Hunkapiller, L. E. Hood, and W. J. Dreyer.** 1981. A gas-liquid phase peptide and protein sequenator. J. Biol. Chem. **256**:7990–7997.
10. **Hirel, P.-H., J.-M Schmitter, P. Dessen, G. Fayat, and S. Blanquet.** 1989. Extent of N-terminal methionine excision from *E. coli* proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. USA **86**:8247–8251.
11. **Hoffman, P. S., C. A. Butler, and F. Quinn.** 1989. Cloning and temperature-dependent expression in *Escherichia coli* of a *Legionella pneumophila* gene coding for a genus-common 60-kilodalton antigen. Infect. Immun. **57**:1731–1739.
12. **Hopp, T. P., and K. P. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
13. **Ish-Horowicz, D., and J. F. Burke.** 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. **9**:2989–2998.
14. **Lema, M. W., A. Brown, C. A. Butler, and P. S. Hoffman.** 1988. Heat-shock response in *Legionella pneumophila* Can. J. Microbiol. **34**:1148–1153.
15. **Lema, M. W., A. Brown, and G. C. C. Chen.** 1985. Altered rate of synthesis of specific peptides in the legionellae in response to growth temperature. Curr. Microbiol. **12**:347–352.
16. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 93–94. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Mehra, V., D. Sweetser, and R. A. Young.** 1986. Efficient mapping of protein antigenic determinants. Proc. Natl. Acad. Sci. USA **83**:7013–7017.
18. **Pau, C. P., B. B. Plikaytis, G. M. Carlone, and I. M. Warner.** Purification, partial characterization, and seroreactivity of a genus-wide 60-kilodalton *Legionella* protein antigen. J. Clin. Microbiol. **26**:67–71.
19. **Plikaytis, B. B., G. M. Carlone, C.-P. Pau, and H. W. Wilkinson.** 1987. Purified 60-kilodalton *Legionella* protein antigen with *Legionella*-specific and nonspecific epitopes. J. Clin. Microbiol. **25**:2080–2084.
20. **Sampson, J. S., B. B. Plikaytis, and H. W. Wilkinson.** 1986. Immunologic response of patients with legionellosis against major protein-containing antigens of *Legionella pneumophila* serogroup 1 as shown by immunoblot analysis. J. Clin. Microbiol. **23**:92–99.
21. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
22. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *E. coli* 16s ribosomal RNA: complementarity to nonsense triplets and ribosomal binding sites. Proc. Natl. Acad. Sci. USA **71**: 1342–1346.
23. **Shinnick, T. M., D. Sweetser, J. E. R. Thole, J. D. A. van Embden, and R. A. Young.** 1987. The etiologic agents of leprosy and tuberculosis share an immunoreactive protein antigen with the vaccine strain *Mycobacterium bovis* BCG. Infect. Immun. **55**:1932–1935.
24. **Shinnick, T. M., M. H. Vodkin, and J. C. Williams.** 1988. The *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.
25. **Sinha, N. D., J. Baernat, J. McManus, and H. Koester.** 1984. Polymer support oligonucleotide synthesis XVII: use of β-cyanoethyl *N,N*-dialkylamino-*N*-morpholino phosphoramidite of deoxynucleotides for synthesis of DNA fragments simplifying deprotection and isolation of the final product. Nucleic Acids Res. **12**:4539–4540.
26. **Vodkin, M. H., and J. C. Williams.** 1988. A heat shock operon in *Coxiella burnetii* produces a major antigen homologous to a protein in both mycobacteria and *Escherichia coli*. J. Bacteriol. **170**:1227–1234.
27. **Wilkinson, H. W.** 1987. Hospital-laboratory diagnosis of *Legionella* infections, p. 3. Second printing, January 1988 (revised). Centers for Disease Control, Atlanta.
28. **Wilson, K. J., D. R. Dupont, P. M. Yuan, M. W. Hunkapillar, and T. D. Schlabach.** 1987. Practical and theoretical aspects of microbore HPLC of proteins, peptides, and amino acid derivatives, p. 21. *In* J. L'Ilalien (ed.), Proteins: structure and function. Selected Proceedings from the First Symposium of American Protein Chemists. Plenum Publishing Corp., New York.
29. **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.