Cloning and Nucleotide Sequence of a Gene (ompS) Encoding the Major Outer Membrane Protein of Legionella pneumophila

PAUL S. HOFFMAN,[1,2]* MURRAY RIPLEY,[1] AND RISINI WEERATNA[1,2]

Department of Microbiology[1] and Division of Infectious Diseases, Department of Medicine,[2]
Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

The major outer membrane protein of Legionella pneumophila is composed of 28- and 31-kDa subunits cross-linked by interchain disulfide bonds. The oligomer is covalently anchored to the underlying peptidoglycan via the 31-kDa subunit. We have cloned the structural gene ompS encoding both proteins. Oligonucleotide probes synthesized from the codons of the N-terminal amino acid sequence of purified 28- and 31-kDa subunits were used to identify cloned sequences. A 2.9-kb HindIII fragment cloned into pBluescript (clone H151) contained the ompS gene. Nucleotide sequence analysis revealed an open reading frame of 891 bp encoding a polypeptide of 297 amino acids. A leader sequence of 21 amino acids was identified, and the mature protein contained 276 amino acids. The deduced amino acid sequence of OmpS matched the experimentally determined amino acid sequence (32 amino acids), with the exception of two cysteine residues. The deduced amino acid sequence was rich in glycine and aromatic amino acids and contained four cysteine residues, two in the amino terminus and two in the carboxy region. Primer extension analysis (total RNA from L. pneumophila) identified the transcription start at 96 to 98 bp upstream of the translation start, but no Escherichia coli-like promoter sequences were evident. While an mRNA transcript from clone H151 was detected, no cross-reactive protein was detected by immunoblotting with either monoclonal or polyclonal antibody. Attempts to subclone the gene in the absence of the putative promoter region (i.e., under the control of the lac promoter) proved unsuccessful, possibly because of overproduction lethality in E. coli. The ompS DNA sequence was highly conserved among the serogroups of L. pneumophila, and related species also exhibited homology in Southern blot analysis at a moderately high stringency. Evidence is presented to suggest that this gene may be environmentally regulated in L. pneumophila.

Legionella pneumophila and related species are a diverse group of environmental microorganisms capable of producing severe lobar pneumonia in humans. These facultative intracellular parasites can invade and colonize a wide range of eukaryotic host cells, including aquatic amoebae (40, 49, 51), alveolar macrophages (28), and vertebrate cell lines (13, 33, 36). The bacteria reside in the phagosomes, in which they abrogate phagolysosomal fusion, a process that does not require de novo protein synthesis (29). These observations suggest that preexisting surface proteins may participate in the pathogenesis process. One surface protein, named Mip (macrophage invasion potentiator), is a 24-kDa protein exhibiting a pI of 9.3 (14). Null mutations in mip result in attenuated virulence for macrophage cell lines and for guinea pigs (10, 11). The most abundant surface protein, referred to as the major outer membrane protein (MOMP), is a porin composed of subunits of 28 and 31 kDa (5, 6, 17, 20, 21). Studies by Payne and Horwitz (37) have demonstrated that the MOMP porin binds the C3b and C3bi factors of the complement system, which mediate phagocytosis via the macrophage integrin receptors CR1 and CR3. Further, complement enhances the binding of MOMP-containing liposomes to macrophages (3). Studies by Quinn et al. (38) also suggest that the MOMP may be directly involved in attachment in a HeLa cell model and also may be a protective immunogen in the guinea pig model. While human and guinea pig convalescent-phase sera contain little or no antibody to the MOMP (42), guinea pigs mount a significant cellular immune response to this antigen (23). Preliminary reports suggest that this antigen may be protective against lethal challenge in the guinea pig model (48).

The MOMP is novel in that one of the subunits is covalently bound to peptidoglycan via a peptide bond, probably to diaminopimelic acid (5). This subunit is covalently cross-linked to 28-kDa subunits via interchain disulfide bonds (5, 27). Progress in assessing the role of this protein in pathogenesis as well as in characterizing the novel structure of this putative porin has been hindered by an inability to clone the structural gene. A number of laboratories, including our laboratory, have been unsuccessful in cloning this gene in expression vectors. The expression of cloned porin genes, particularly from nonenteric pathogens, in Escherichia coli is often inhibitory for growth (2, 19). In some cases, gene fusions in vectors such as lambda gt11 have been used to identify porin sequences in expression libraries (19, 45). Such approaches have been used to clone the 40-kDa MOMP gene (ompl L2) from Chlamydia trachomatis (46).

Our approach to cloning the ompS gene involved the use of oligonucleotide probes generated from amino acid sequence information from the 28- and 31-kDa MOMP subunits. We found that both subunits had a common N-terminal amino acid sequence (27). On the basis of this sequence (GTMGPVWT), a series of oligonucleotides were generated, one of which hybridized to a single site of restricted genomic DNA, as judged by Southern blot analysis. In the present study, we report on the use of this oligonucleotide as a probe for identifying clones containing the putative genes coding for the L. pneumophila MOMP subunits. On the basis of the nucleotide sequence and the deduced amino acid sequence, we have confirmed that an open reading frame (ORF) of 891 bp encodes a polypeptide of 297 amino acids. The gene is

---

* Corresponding author.

poorly expressed under its own promoter in *E. coli* and may be down-regulated during the early stages of invasion.

(A preliminary account of this work was presented at the 1991 American Society for Microbiology general meeting [26].)

MATERIALS AND METHODS

Bacterial strains, protein sequence, and oligonucleotide reagents. A streptomycin-resistant strain (SVir) of *L. pneumophila* Philadelphia 1 (serogroup 1) was used as the source of the 28- and 31-kDa proteins and chromosomal DNA as described elsewhere (24, 27). Other serogroups of *L. pneumophila* and various *Legionella* species used in this study included *L. pneumophila* serogroups 1, 2, 3, 4, 5, and 7; *L. micdadei* (HEBA and Tatlock); *L. jordanis* serogroup 1; and *L. oakridgensis* OR10. Preparation of genomic DNA from these strains has been described previously (24). Purification of the 31- and 28-kDa proteins, peptide maps, and peptide sequencing were as described elsewhere (5, 27). Oligonucleotides used in this study were synthesized at the Molecular Resource Center, Department of Microbiology and Immunology, University of Tennessee, Memphis, and at the Molecular Gene Probe Laboratory, Dalhousie University. Two oligonucleotides (24b and H25 [complement]) used as probes in this study had the following respective sequences: 5'GGTACTATGGGTCCAGTATGGAC3' and 5'GTCCATA CTGGACCCATAGTACC3'. These sequences were based on the amino acid sequence GTMGPVWT.

Cloning and nucleotide sequencing of the *ompS* gene. Chromosomal DNA from *L. pneumophila* restricted with *Eco*RI, *Hin*dIII, or a combination of *Hin*dIII and *Pst*I was subjected to electrophoresis in an 0.8% low-melting-temperature agarose gel (SeaPlaque; FMC Bioproducts, Rockland, Maine). DNA fragments in molecular weight ranges previously identified by Southern blotting were excised and ligated into pBluescript (Stratagene), similarly restricted, and purified from low-melting-temperature agarose. The resulting colonies were screened with [³²P]ATP-end-labeled oligonucleotide H25 (reverse primer) as generally described by Sambrook et al. (41). Restriction enzymes, buffers, and procedures were as described by the various manufacturers. Restricted DNAs from clones H151 (2.9-kb *Hin*dIII fragment) and HP246 (900-bp *Hin*dIII-*Pst*I fragment) were subcloned into M13 vectors for sequencing. Sequencing was performed by the dideoxy chain termination method with [³⁵S]dATP (NEN-DuPont, Toronto, Ontario, Canada) and Sequenase (U.S. Biochemical Corp., Cleveland, Ohio) as suggested by the manufacturers. Both DNA strands were sequenced, and the sequences were assembled and analyzed with the Wisconsin Genetics Computer Group sequence analysis programs (Genetics Computer Group, Inc., University of Wisconsin Biotechnology Center, Madison) (12).

Southern and Northern (RNA) blot analyses and primer extension. DNA hybridizations were done essentially as described by Southern (44). Chromosomal DNAs from the various *Legionella* serogroups and species were restricted with *Eco*RI or *Hin*dIII. The DNA probe was radiolabeled by the polymerase chain reaction (PCR). Oligonucleotide primers 24b and R3 (reverse primer hybridizing to an internal region of the ORF) were used to amplify an approximately 500-bp segment of *ompS*. [³²P]dCTP (65 µCi in a 50-µl reaction mixture) was incorporated into the PCR fragments by decreasing the carrier dCTP molar concentration in the deoxynucleotide triphosphate reaction mixture from 200 to 50 µM. The PCR fragments were purified through Sephadex

G-50 as described by Silhavy et al. (43). Approximately 8 × 10⁶ cpm was added to the hybridization mixture following denaturation. Hybridizations and washings were done at a moderately high stringency (15% mismatch in the duplex) and at a high stringency (<5% mismatch in the duplex) as described previously (24).

Total RNA was prepared by a hot sodium dodecyl sulfate-phenol procedure (27). Primer extension was performed by hybridizing probe H25 to approximately 25 µg of total RNA. The oligonucleotide probe was end labeled with [³²P]ATP (125 µCi; NEN-DuPont) and T4 polynucleotide kinase (New England BioLabs, Inc., Beverly, Mass.) as described previously (27). The probes were used either directly or after gel purification on a 20% polyacrylamide–8 M urea gel. The extension reaction was run at 42°C for 1 h in the presence of actinomycin D. Following RNase treatment, phenol extraction, and ethanol precipitation, the cDNA was taken up in loading buffer and 2 to 5 µl was loaded onto the sequencing gel. The extension product was compared with the simultaneously run sequencing reaction products generated from M13 containing a *Hin*dIII-*Pst*I fragment encoding the 5' region of the *ompS* gene and H25 as a primer.

Nucleotide sequence accession number. The nucleotide sequence accession number (GenBank) for *ompS* is M76178.

RESULTS

Cloning strategy. We have assumed, on the basis of various strategies aimed at cloning the MOMP gene into expression vectors, either that the expression of the gene was toxic or that the gene was not expressed in the *E. coli* genetic background. By use of a DNA hybridization approach, the requirement for gene expression could be eliminated and, in the case of toxic gene expression, unexpressed fragments of the desired gene might be cloned. In a previous study, we reported that an oligonucleotide synthesized from the N-terminal amino acid sequence GTMGPVWT hybridized to single sites in genomic DNA restricted with various restriction enzymes (27). Since Southern hybridization analysis provided information on the relative sizes of the desired restriction fragments, we first attempted to clone these sized fragments into pBluescript. We were unsuccessful in obtaining clones of a 1.5-kb *Eco*RI fragment in either pBluescript or M13 phage. The fact that recombinants were obtained (white versus blue colonies or plaques scored with isopropyl-β-D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) suggested no technical difficulty. We then screened clones containing DNA fragments obtained by *Hin*dIII cleavage and by a combination of *Hin*dIII and *Pst*I restriction. Four colonies from a total of 280 white colonies were found strongly positive by Southern colony blot hybridization. Restriction endonuclease analysis of cloned sequences showed that two clones contained a 2.9-kb *Hin*dIII fragment (H151 and H157) and that two contained a 900-bp *Hin*dIII-*Pst*I DNA fragment (HP243 and HP246). Restriction maps for H151 and HP246 are presented in Fig. 1. Interestingly, the 1.5-kb *Eco*RI fragment was contained within the 2.9-kb *Hin*dIII fragment. Attempts to subclone the *Eco*RI fragment proved unsuccessful, suggesting that an ORF, designated *ompS*, may be within the *Eco*RI restriction fragment and that the expression of this gene may produce a toxic product in *E. coli*. While not shown, a screening of the *E. coli* clones for antigen expression with both monoclonal and polyclonal antibodies reactive with both the 28- and the 31-kDa subunits was also negative. The possibility that a transcriptional defect might account for the
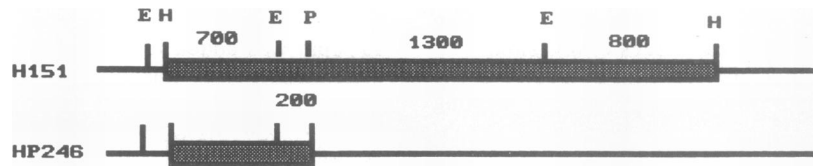
FIG. 1. Restriction endonuclease maps of H151 and HP246 in pBluescript. A 2.9-kb HindIII fragment was cloned into pBluescript. The ORF was localized to the 1.5-kb EcoRI region. Abbreviations: H, HindIII; P, PstI; E, EcoRI. The map distances are given in base pairs. The lac promoter is in the left arm of pBluescript and would permit the transcription of ompS.

lack of expression was examined by Northern blot analysis of clone H151. A low level of transcription was observed for H151 probed with an internal PCR-generated DNA fragment of the ompS gene (Fig. 2). The molecular size of the H151 mRNA was similar to that detected in total RNA preparations from L. pneumophila SVir and from an isogenic protease-deficient strain, PRT8. These results suggest that the ompS promoter region is poorly recognized in E. coli and that perhaps a translational defect might also account for a lack of detectable MOMP.

Sequence analysis. The nucleotide sequence of structural gene ompS and flanking sequences and the deduced amino acid sequence are depicted in Fig. 3. The ORF begins approximately 800 bp in from the leftward HindIII site (Fig. 1) and 270 bp into the depicted sequence. A ribosome binding domain (TGGAG) is located 9 bases upstream from the initiation codon. An 891-bp ORF encodes a polypeptide of 297 amino acids, of which the first 21 amino acids are presumably involved in protein export, since the processed protein begins at amino acid 22 (PheAla/GlyThrMet . . .) (slash indicates processing site). With the exception of two cysteine residues, there was good agreement between the deduced and experimentally determined amino acid sequences (32 amino acids), as depicted by the underlined N-terminal region of the sequence in Fig. 3. The N-terminal amino acid sequence of a 19-kDa peptide generated by cyanogen bromide cleavage of the 31-kDa protein was also located in the deduced sequence. However, the differences noted between this sequence and the deduced sequences are

FIG. 2. Northern blot analysis of H151 and L. pneumophila ompS mRNAs. Total RNA was extracted as described in the text from E. coli HB101 (pBluescript) (lane A), E. coli HB101 (clone H151) (lane B), L. pneumophila SVir (lane C), and L. pneumophila PRT8 (protease-deficient avirulent mutant of SVir) (lane D). The mRNA was probed with a PCR-generated ³²P-labeled DNA fragment generated to an internal region of the ompS gene (oligonucleotide primers 24b and R3). The hybridization noted in lane B at a high molecular weight probably represents contaminating plasmid sequences of H151.

most likely due to the poor quality of the amino acid sequence obtained for this peptide. Downstream of the termination codon (TAA) is a palindromic sequence resembling a rho-independent transcription terminator. An obvious E. coli-like promoter sequence was not identified upstream of the translation start. The possibility that the gene was part of an operon was resolved by primer extension analysis. Primer extension revealed that transcription begins within a CCC sequence (bp 172 to 174 in Fig. 3) that is flanked on both sides by an AT-rich sequence and that is 90 to 100 bp upstream of the translation start (Fig. 4).

In a previous study, we predicted that the MOMP contained three cysteine residues on the basis of amino acid composition analysis (5). On the basis of the deduced amino acid sequence, four cysteine residues were identified. Two of these residues are located in the N-terminal region (amino acids 7 and 16 of the processed polypeptide), and two are in the carboxyl region at amino acids 194 and 197. The polypeptide also contains seven methionine residues rather than three, as previously reported on the basis of the amino acid composition analysis. The protein is rich in glycine (11%) and the aromatic amino acids phenylalanine (7.4%), tyrosine (6%), and tryptophan (3.4%). The processed polypeptide exhibits an acidic pI of 4.59, and the amino acid composition, hydrophilicity characteristics, and a lack of long stretches of alpha- or beta-sheet sequences are typical of porin proteins. A search of the protein data base revealed that OmpS was not closely related to other porin proteins, with the exception of a 69-kDa outer membrane protein of Bordetella pertussis which exhibited a similarity of 45% and an identity of 21% (8). The sequence also exhibited 22% identity with a repeat in the CR1 or CR3 integrin, a complement receptor of macrophages (30).

Distribution of ompS in legionellae. Previous studies in our laboratory have demonstrated that several Legionella species express disulfide-cross-linked and peptidoglycan-bound outer membrane proteins (5). Furthermore, on the basis of immunoblot studies with polyclonal anti-Legionella MOMP serum, we reported that the MOMPs of the various species may share common epitopes (6). To address the possibility that the genes encoding these proteins from the various species might be genetically related, we probed chromosomal digests from selected Legionella species and serogroups with an internal PCR-radiolabeled fragment of the ompS gene. Figure 5A depicts the results of Southern hybridization analysis performed at a moderate stringency (15% mismatch in the duplex). At this stringency, all Legionella species examined exhibited related sequences. The weakest signal was noted for L. jordanis, which does not express a 28-kDa MOMP (5). The strongest signals were observed for the serogroups of L. pneumophila. However, multiple bands were seen with the serogroups, suggesting that there might be either related genes encoding outer membrane proteins or possibly cryptic genes. At higher

```
   1  CGAATAAATTTTGGCGCAAATTAATAGTCATGCACTCCCCCTCATTACTACAAAATTAAAGGAAAAAGCAGACTACTTGC  80

  81  TGGCGTATAAATCAATAACAGCAGATACTGATCAAATCTTACATTTAATGTTTAAATCAATGAGTTAAATAACTTTAATT  160

 161  TAATAAAATTACCCTTATTATTTGATGAAGAATCAAATAGATTCTGTTAGAATTCGCTAATCCTAGATGTTTAAAGGAAT  240
      ------->mRNA

 241  AACAATAATAAATCAGTGGAGAACGGGATATGTTTAGTTTGAAAAAAACAACTGCGGCTGTATTTGCTCTGGGAAGCAGT  320
                                MetPheSerLeuLysLysThrThrAlaAlaValPheAlaLeuGlySerSer

 321  GCTTTGTTTGCAGGTACGATGGGTCCAGTATGTACGCCAGGCAATGTGACTGTTCCTTGCGAAAGAACCGCATGGGATAT  400
      AlaLeuPheAlaGlyThrMetGlyProValCysThrProGlyAsnValThrValProCysGluArgThrAlaTrpAspIl
                                  W                                H    G    A

 401  TGGTATTACCGCACTTTATCTGCAGCCAACTTATGATGCTGATTGGGGCTATAATGGTTTTACTGATGTTGGTGGCTGGA  480
      eGlyIleThrAlaLeuTyrLeuGlnProThrTyrAspAlaAspTrpGlyTyrAsnGlyPheThrAspValGlyGlyTrpA

 481  GAAATTGGCATGATGTTGATTTAGAGTGGGATTGGGGCTTTAAATTAGAAGGTTCTTATCACTTCAACACAGGCAATGAC  560
      rgAsnTrpHisAspValAspLeuGluTrpAspTrpGlyPheLysLeuGluGlySerTyrHisPheAsnThrGlyAsnAsp

 561  ATCAACGTCAATTGGTACCATTTGGATGCCAACACCGATCATTGGGCGTTTGTTGATCAATTGCACGCCTACGATATTAC  640
      IleAsnValAsnTrpTyrHisLeuAspAlaAsnThrAspHisTrpAlaPheValAspGlnLeuHisAlaTyrAspIleTh

 641  CTGGGATGCAGTGAATGCTGAACTGGGGCAATTTGTTGATTTTAGCGCGAACAAAAAAATGCGCTTCCATGGCGGCGTTC  720
      rTrpAspAlaValAsnAlaGluLeuGlyGlnPheValAspPheSerAlaAsnLysLysMetArgPheHisGlyGlyValG

 721  AGTATGCCTACATCAAAACAGACGTTAATCGTCACTTAAACGGCTTCTTTCTAGATAACTTCAATTCCAAATTCAATGGA  800
      lnTyrAlaTyrIleLysThrAspValAsnArgHisLeuAsnGlyPhePheLeuAspAsnPheAsnSerLysPheAsnGly
                                     Y   F              A   F   N

 801  TTTGGTCCCCGTACCGGCCTTGATATGAATTATGTATTCGGTAATGGGTTTGGTATTTATGCAAAATCAGCTGTAGCAAT  880
      PheGlyProArgThrGlyLeuAspMetAsnTyrValPheGlyAsnGlyPheGlyIleTyrAlaLysSerAlaValAlaIl

 881  CCTGGTCGGTACAAGCAAATTCGTGGATAACTGTACTGTCTGTGGTTTTTCTTATGGCTCCAAAAATGCAATAGTGCCAG  960
      eLeuValGlyThrSerLysPheValAspAsnCysThrValCysGlyPheSerTyrGlySerLysAsnAlaIleValProG

 961  AAGTTGAAATGAAGTTAGGGAGCAGATTATACCTATGCAATGGCTCAAGGTGATGTGACTCTGGATGTGGGTTACATGTGG  1040
      luValGluMetLysLeuGlyAlaAspTyrThrTyrAlaMetAlaGlnGlyAspValThrLeuAspValGlyTyrMetTrp

1041  TTCAATTACTTCAATGCCTTACATAATACTGCTGCCGTGAATGTTGGCTTAGGCACTTCTCTGGAAACCGACTTCTCTGC  1120
      PheAsnTyrPheAsnAlaLeuHisAsnThrAlaAlaValAsnValGlyLeuGlyThrSerLeuGluThrAspPheSerAl

1121  ATCTGGCCCCTACATAGGCTTAAAATATGTTGGCAATGTCTAAATAACCCCAAATTATGCTGCTAATCTAGTCATGATCA  1200
      aSerGlyProTyrIleGlyLeuLysTyrValGlyAsnValEnd

1201  AGCAGCAGTATGCAAAAAACCCGTATTGACATAGGCTGCTAATACGGGTTTTTTTTATTTCTCTCTGTTTAATTAAGTTA  1280
      --------------->                    <---------------

1281  GTTATCCTATCTAAGCTAACTTGTAATTATATGATAATTTCCCAATGAAATTGAGATAACTTTATGCGAAACAATAATTG  1360

1361  GTCCGTGTTGGGCAACATCACCAATGAAAAATCACATCCGTGATCTGCTGCGAAAACTATAGTAAACGCAATAATCTGCT  1440

1441  GGTTCTGCCTGATGCATAATTCTTTTCCATTCCAAAGATCGTGGGTTATAATGAAACCATCTATCACAAACCGTTTATAA  1520

1521  AAATAAACATGCCTGATGTTCTCGGCAAACCTGATGCAACCCAGATTCCAGATGACGTCTGGCTTGACATTCTTATCAAT  1600

1601  TATCTTCAACCACAAGATATTGTAAACCTGTCTGAAATCAACAATCGCTTGAGCAGACTTTTCAAAACTCAACACACTCA  1680

1681  ATCTGAGCAATTGAACCAAAAAAAAAA
```
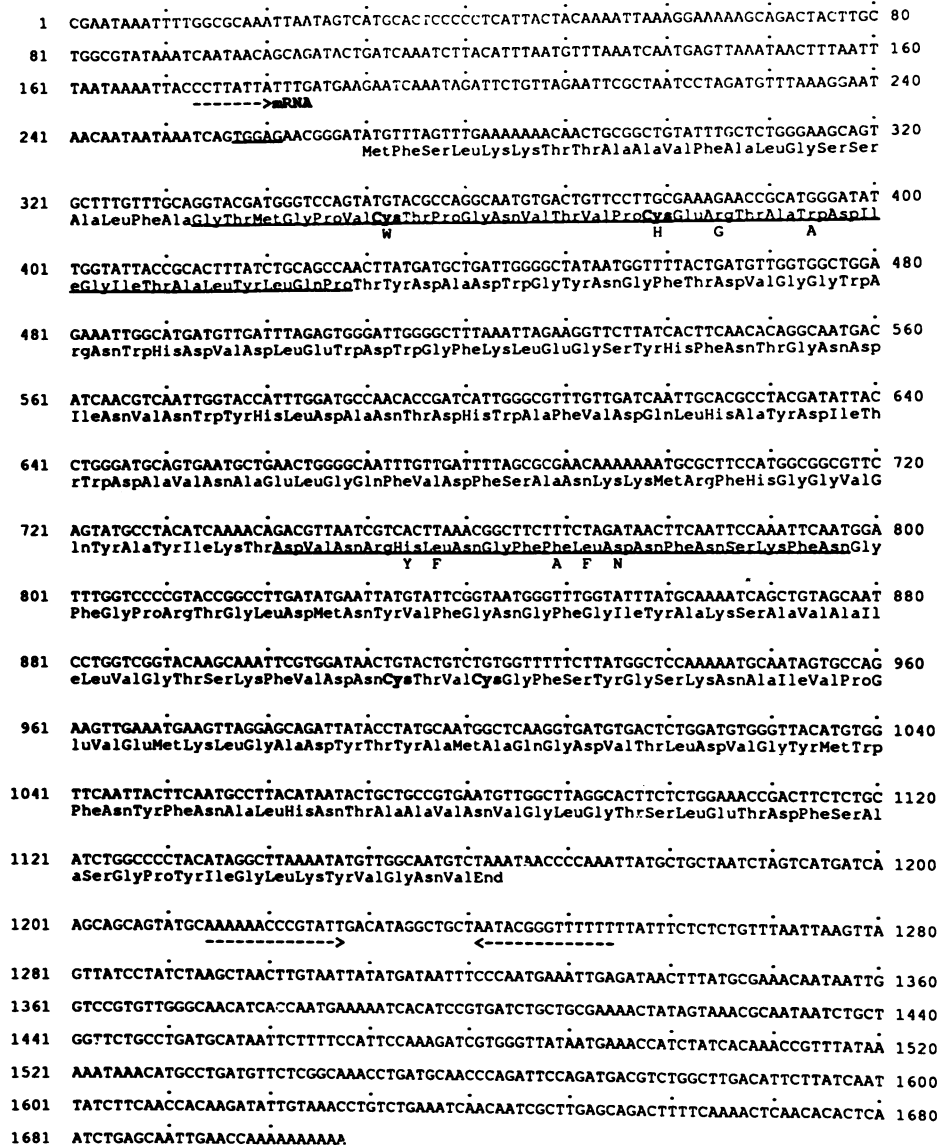
FIG. 3. Nucleotide sequence and derived amino acid sequence of *ompS*. The transcription start at bp 173 is indicated by the arrow, and the termination sequence is indicated by the inverted arrows in the region from positions 1201 to 1280 of the sequence. The putative ribosome binding site upstream of the translation start is underlined, as are the regions of the deduced amino acid sequence determined by sequencing of the 31- and 28-kDa proteins (27). Differences from peptide sequencing are indicated by the single-letter format. The positions of the 4 cysteine residues are depicted in boldface type.

stringencies for genomic DNA restricted with *Hin*dIII (5% mismatch in the duplex), the hybridizations noted with the other *Legionella* species disappeared (Fig. 5B), along with the high signal intensity of the multiple hybridizations noted for the serogroups. Southern blot analysis of *L. pneumophila* serogroup 1 DNA restricted with *Hin*dIII and *Eco*RI identified a single fragment of 1.5 kb, confirming that *ompS* exists in a single copy (data not shown).

DISCUSSION

We have cloned and sequenced the structural gene encoding the 28- and 31-kDa subunits of the *L. pneumophila* MOMP oligomer. The MOMP is the most abundant protein synthesized by *L. pneumophila*, and recent studies have implicated this protein in pathogenesis (3, 37, 38) and immu-

nity (23, 48). DNA sequence analysis revealed an 891-bp ORF encoding a polypeptide of 297 amino acids. The polypeptide contained a 21-amino-acid signal sequence, and the mature protein contained 276 amino acids. The deduced amino acid sequence of the MOMP exhibited four cysteine residues and in general was rich in glycine and aromatic amino acids. With the exception of cysteine residues at positions 7 and 16 of the processed polypeptide, there was good correlation between the deduced amino acid sequence and that obtained by direct sequencing of the purified MOMP subunits (27). In this study, we confirmed the results of earlier work regarding the number of cysteine residues in the monomers. Two cysteine residues were found in the amino terminus while two were found in the carboxyl region of the molecule. A search of nucleic acid and protein data
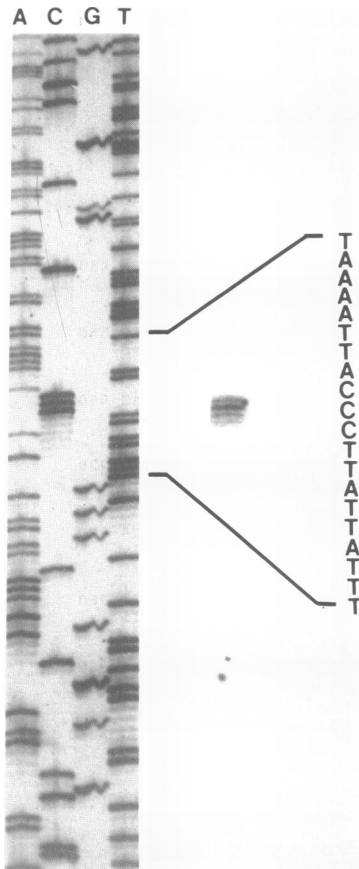
FIG. 4. Primer extension analysis of the transcription start of *ompS*. Total RNA was isolated from cells grown in buffered yeast extract medium for 24 h. RNA (25 μg) was annealed to oligonucleotide H25, which had been end labeled with [^{32}P]ATP. The primer was extended with 50 U of avian myeloblastosis virus reverse transcriptase, and the cDNA was suspended in sample buffer and resolved on a polyacrylamide gel beside a sequencing ladder generated from the same oligonucleotide primer with a *Hind*III-*Pst*I DNA fragment in M13 as the template. The transcription start is located in a CCC sequence flanked by AT-rich sequences.

FIG. 5. Southern blot analysis of *Eco*RI (A)- and *Hind*III (B)-restricted DNA from selected *Legionella* species and serogroups of *L. pneumophila*. Genomic DNA was prepared from *L. pneumophila* serogroup 7 (lane 1), *L. micdadei* HEBA (lane 2), *L. pneumophila* serogroup 1 SVir (lane 3), *L. jordanis* serogroup 1 (lane 4), *L. pneumophila* serogroup 4 (lane 5), *L. pneumophila* (unknown serogroup) (lane 6), *L. pneumophila* serogroup 1 Avir (lane 7), *L. pneumophila* serogroup 2 Togus-1 (lane 8), *L. pneumophila* serogroup 5 (lane 9), *L. micdadei* Tatlock (lane 10), *L. oakridgensis* OR10 (lane 11), and *L. pneumophila* serogroup 3 (lane 12). In panel A, the hybridization was done at a decreased stringency (15% mismatch in the DNA duplex) to detect relatedness among the different species. The stringency conditions for the hybridization depicted in panel B was at a 5% mismatch in the DNA duplex. The DNA probe used in these experiments was generated by PCR with oligonucleotides 24b and R3 as primers, and the amplified PCR fragment was radiolabeled as described in the text.

bases revealed that no other proteins showed amino acid sequence homology with OmpS. There was no homology with the *Chlamydia* sp. 40-kDa MOMP, which contains nine cysteine residues and participates in inter- and intramolecular disulfide bonding (46). Interestingly, some homology was noted with the 69-kDa outer membrane protein of *B. pertussis* and with the macrophage integrin protein CR1, a receptor for complement (3, 37). The homology with CR1 appeared to be in a variable-repeat region (34). While the observed similarities might be merely coincidence, it should be noted here that while C3 components of complement bind to the *L. pneumophila* MOMP, the specific binding site has not been identified.

Disulfide-cross-linked and peptidoglycan-bound outer membrane proteins are a characteristic feature of many members of the genus *Legionella* (5). A commercial monoclonal antibody diagnostic test for legionellosis recognizes an epitope on the MOMPs shared by all serogroups of *L. pneumophila* (18). Moreover, a study by Butler et al. (6) suggested that the MOMPs of other *Legionella* species might also contain genus-common epitopes. Southern blot analysis
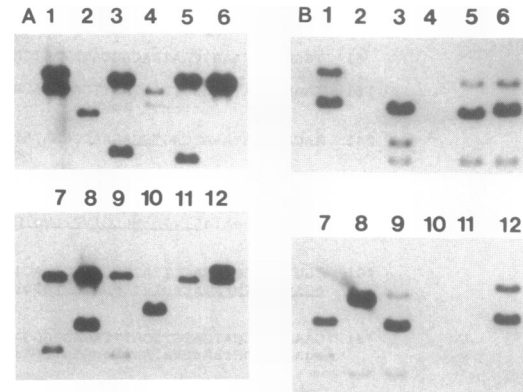
confirmed that all of the serogroups examined in this study contained highly conserved sequences, although some restriction polymorphism was evident. Under moderate-stringency conditions, DNA-DNA hybridizations were also noted with several other *Legionella* species. A number of other genes which contain genus-common sequences have been described for the legionellae; these include *mip* (14), *htpAB* (25, 38), and a gene encoding a 19-kDa peptidoglycan-associated protein (15, 22, 32). In contrast, the gene encoding the cytotoxic metalloprotease is common only to the species *L. pneumophila* (39).

Since the MOMP is the most abundant protein synthesized by *L. pneumophila*, information regarding the expression of the gene is of particular interest. Although primer extension analysis identified the transcription start at 97 bp upstream of the translation start, no *E. coli*-like promoter sequences were seen in regions 10 (TAATAAAAT) and 35 (TCAAT GAG) bp upstream. These promoter sequences differ from those noted for other cloned *L. pneumophila* genes expressed in *E. coli*, including *mip* (14), *htpAB* (25), the protease gene (4), and *recA* (52). An unusual tandem promoter sequence has been reported for the *omp1L2* gene of *C. trachomatis* (47). The *omp1L2* gene is developmentally regulated, and the MOMP is the most abundant protein synthesized by *C. trachomatis*. In contrast, Northern blot analysis of *L. pneumophila* RNA revealed a single transcript of approximately 1 kb (27), confirming a single promoter and a monocystronic gene structure. Interestingly, secondary structure predictions for the mRNA showed a substantial ability of the 5' region to form loops, which may function to stabilize the message in *L. pneumophila* but possibly affect translation in *E. coli* (9). The *ompS* gene, like the chlamydial *omp1L2* gene, may be environmentally regulated. We have

observed that little radiolabel is incorporated into the 28-kDa MOMP subunit of virulent *L. pneumophila* cells during the early stages of invasion of HeLa cell or L-cell monolayers (1 to 3 h postinfection) (16). In contrast, radiolabel is incorporated into stress proteins, a phenomenon that can also be reproduced in tissue culture medium in the absence of host cells (25). When mRNA levels were monitored, the *ompS* message was decreased relative to the *htpAB* message. The observation that avirulent isogenic strains of *L. pneumophila* do not regulate MOMP levels or show decreased mRNA levels when cells are placed in tissue culture medium implies that these cells no longer sense changes in the environment or perhaps no longer produce the necessary regulatory factors. Environmentally regulated transcription has been well characterized for *ompF-ompC* porin genes of *E. coli* (34) and for the *vir* (*bvg*) genes encoding many of the membrane-associated proteins of *B. pertussis* (1). We are presently constructing gene fusions to begin addressing the regulatory aspects of *ompS* gene expression as well as putative regulatory differences between virulent and avirulent isogenic strains.

The *L. pneumophila* MOMP (putative porin) (17) is unique from the porins of other bacteria in that it is covalently bound to peptidoglycan and the subunits are cross-linked via interchain disulfide bonds (5, 27). On the basis of the deduced amino acid sequence, we found two cysteine residues in the N-terminal amino acid sequence (positions 7 and 16 of the processed polypeptide) and two at positions 194 and 197 in the carboxyl region. The first two cysteine residues are found in an amino acid sequence of high β-sheet-forming potential, suggesting that at least one of the cysteine residues may be within the outer membrane lipid bilayer, while the other might be either external or near the outer surface. These cysteine residues are separated by an essentially hydrophobic stretch of eight amino acids, and the cysteine residue at position 16 is followed by the charged amino acids glutamate and arginine. Cysteine residue 194 is preceded by the charged amino acids aspartate and asparagine and separated from cysteine residue 197 by two hydrophobic amino acids. While all cysteine residues could potentially participate in interchain disulfide bonding among subunits, the minimum number of bonds that each subunit could participate in and still maintain the trimeric form would be two. It is conceivable that subtle changes in the conformation of the putative porin might result from different combinations of disulfide bonding among the cysteine residues. Such combinations might also be envisioned to change the pore size or other physical characteristics of the molecule. Variable pore sizes noted for *Pseudomonas aeruginosa* OprF have been attributed to different patterns of disulfide bond cross-linkages (35). A prominent difference between virulent and avirulent isogenic strains of *L. pneumophila* is in the tolerance of sodium chloride (7). While the porin molecule has been shown to exhibit an anion preference (17), changes in conformation could affect the magnitude of this selectivity. It is perhaps noteworthy that the amino-terminal cysteine residues are in a region substantially rich in proline. Recently, the Mip protein has been shown to be highly homologous with a class of proteins known as peptidylprolyl *cis-trans* isomerases (50). These enzymes are capable of changing proline residues from one isomer to the other (31). Such changes might affect the secondary and perhaps, in the case of interchain disulfide bonds, the quaternary structures of the porin. The role of chaperone proteins in porin assembly has only recently been addressed (25), and continued study of these mechanisms

might lead to new insight into assembly processes external to the cytoplasmic membrane. It will be important in future studies to address the possibility that conformational changes, perhaps mediated by the mechanisms described here, might be partially responsible for the phenomenon of avirulence acquired through the selection for high salt tolerance under laboratory conditions.

REFERENCES

1. Arico, B., J. F. Miller, C. Roy, S. Stibitz, D. Monack, S. Falkow, R. Gross, and R. Rappuoli. 1989. Sequences required for expression of *Bordetella pertussis* virulence factors share homology with prokaryotic signal transduction proteins. Proc. Natl. Acad. Sci. USA 86:6671–6675.
2. Barlow, A. K., J. E. Heckles, and I. N. Clarke. 1987. Molecular cloning and expression of *Neisseria meningitidis* class 1 outer membrane protein in *Escherichia coli* K-12. Infect. Immun. 55:2734–2740.
3. Bellinger-Kawahara, C., and M. A. Horwitz. 1990. Complement component C3 fixes selectively to the major outer membrane protein (MOMP) of *Legionella pneumophila* and mediates phagocytosis of liposome-MOMP complexes by human monocytes. J. Exp. Med. 172:1201–1210.
4. Black, W. J., F. D. Quinn, and L. S. Tompkins. 1990. *Legionella pneumophila* zinc metalloprotease is structurally and functionally homologous to *Pseudomonas aeruginosa* elastase. J. Bacteriol. 172:2608–2613.
5. Butler, C. A., and P. S. Hoffman. 1990. Characterization of a major 31-kilodalton peptidoglycan-bound protein of *Legionella pneumophila*. J. Bacteriol. 172:2401–2407.
6. Butler, C. A., E. D. Street, T. P. Hatch, and P. S. Hoffman. 1985. Disulfide-bonded outer membrane proteins in the genus *Legionella*. Infect. Immun. 48:14–18.
7. Catrenich, C. E., and W. Johnson. 1989. Characterization of the selective inhibition of growth of virulent *Legionella pneumophila* by supplemented Mueller-Hinton medium. Infect. Immun. 57:1862–1864.
8. Charles, I. G., G. Dougan, D. Pickard, S. Chatfield, M. Smith, P. Novotny, P. Morrissey, and N. F. Fairweather. 1989. Molecular cloning and characterization of protective outer membrane protein P.69 from *Bordetella pertussis*. Proc. Natl. Acad. Sci. USA 86:3554–3558.
9. Chen, L.-H., S. A. Emory, A. L. Bricker, P. Bouvet, and J. G. Belasco. 1991. Structure and function of a bacterial mRNA stabilizer: analysis of the 5′ untranslated region of *ompA* mRNA. J. Bacteriol. 173:4578–4586.
10. Cianciotto, N. P., B. I. Eisenstein, C. H. Mody, and N. C. Engleberg. 1990. A mutation in the *mip* gene results in an attenuation of *Legionella pneumophila* virulence. J. Infect. Dis. 162:121–126.
11. Cianciotto, N. P., B. I. Eisenstein, C. H. Mody, G. B. Toews, and N. C. Engleberg. 1989. A *Legionella pneumophila* gene encoding a species-specific surface protein potentiates initiation of intracellular infection. Infect. Immun. 57:1255–1262.
12. Devereux, J., P. Haberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the Vax. Nucleic Acids Res. 12:387–389.
13. Dreyfus, L. A. 1987. Virulence associated ingestion of *Legionella pneumophila* by HeLa cells. Microb. Pathog. 3:45–52.
14. Engleberg, N. C., C. Carter, D. R. Weber, N. P. Cianciotto, and B. I. Eisenstein. 1989. DNA sequence of *mip*, a *Legionella pneumophila* gene associated with macrophage infectivity. Infect. Immun. 57:1263–1270.
15. Engleberg, N. C., E. Pearlman, D. Dixon, and B. I. Eisenstein.

1986. Antibodies isolated by using cloned surface antigens recognize antigenically related components of *Legionella pneumophila* and other *Legionella* species. J. Immunol. **136**:1415–1417.

16. Fernandez, R. C., S. H. S. Lee, R. Weeratna, M. Ripley, and P. S. Hoffman. Unpublished data.

17. Gabay, J. E., M. Blake, W. D. Niles, and M. A. Horwitz. 1985. Purification of *Legionella pneumophila* major outer membrane protein and demonstration that it is a porin. J. Bacteriol. **162**:85–91.

18. Gosting, L. H., K. Cabrian, J. C. Sturge, and L. C. Goldstein. 1984. Identification of a species-specific antigen in *Legionella pneumophila* by a monoclonal antibody. J. Clin. Microbiol. **20**:1031–1035.

19. Gotschlich, E. C., M. E. Seiff, M. S. Blake, and M. Koomey. 1987. Porin protein of *Neisseria gonorrhoeae*: cloning and gene structure. Proc. Natl. Acad. Sci. USA **84**:8135–8139.

20. Hindahl, M. S., and B. H. Iglewski. 1984. Isolation and characterization of the *Legionella pneumophila* outer membrane. J. Bacteriol. **159**:107–113.

21. Hindahl, M. S., and B. H. Iglewski. 1986. Outer membrane proteins from *Legionella pneumophila* serogroups and other *Legionella* species. Infect. Immun. **51**:94–101.

22. Hindahl, M. S., and B. H. Iglewski. 1987. Cloning and expression of a common *Legionella* outer membrane antigen in *Escherichia coli*. Microb. Pathog. **2**:91–99.

23. Hoffman, P. S. Unpublished data.

24. Hoffman, P. S., C. A. Butler, and F. D. 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.

25. Hoffman, P. S., L. Houston, and C. A. Butler. 1990. *Legionella pneumophila* htpAB heat shock operon: nucleotide sequence and expression of the 60-kilodalton antigen in *L. pneumophila*-infected HeLa cells. Infect. Immun. **58**:3380–3387.

26. Hoffman, P. S., and M. Ripley. 1991. Cloning and characterization of the MOMP gene of *Legionella pneumophila*, abstr. D-174, p. 107. Abstr. 91st Gen. Meet. Am Soc. Microbiol. 1991.

27. Hoffman, P. S., J. H. Seyer, and C. A. Butler. 1991. Molecular characterization of the 28- and 31-kilodalton subunits of the *Legionella pneumophila* major outer membrane protein. J. Bacteriol. **174**:908–913.

28. Horwitz, M. A., and S. C. Silverstein. 1980. The Legionnaires' disease bacterium (*Legionella pneumophila*) multiples intracellularly in human monocytes. J. Clin. Invest. **66**:441–450.

29. Horwitz, M. A., and S. C. Silverstein. 1983. Intracellular multiplication of Legionnaires' disease bacteria (*Legionella pneumophila*) in human monocytes is reversibly inhibited by erythromycin and rifampin. J. Clin. Invest. **71**:15–26.

30. Klickstein, L. B., W. W. Wong, J. A. Smith, J. H. Weis, J. G. Wilson, and D. T. Fearon. 1987. Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domains that are composed of short consensus repeats characteristic of C3/C4 binding proteins. J. Exp. Med. **165**:1095–1112.

31. Liu, J., and C. T. Walsh. 1990. Peptidyl-prolyl *cis-trans*-isomerase from *Escherichia coli*: a periplasmic homolog of cyclophilin that is not inhibited by cyclosporin A. Proc. Natl. Acad. Sci. USA **87**:4028–4032.

32. Ludwig, B., A. Schmid, R. Marre, and J. Hacker. 1991. Cloning, genetic analysis, and nucleotide sequence of a determinant coding for a 19-kilodalton peptidoglycan-associated protein (Ppl) of *Legionella pneumophila*. Infect. Immun. **59**:2515–2521.

33. Marra, A., M. A. Horwitz, and H. A. Shuman. 1990. The HL-60 model for the interaction of human macrophages with the Legionnaires' disease bacterium. J. Immunol. **144**:2738–2744.

34. Mizuno, T., and S. Mizushima. 1990. Signal transduction and gene regulation through the phosphorylation of two regulatory components: the molecular basis for the osmotic regulation of

the porin genes. Mol. Microbiol. **4**:1077–1082.

35. Moore, R. A., W. A. Woodruff, and R. E. W. Hancock. 1987. Antibiotic uptake pathways across the outer membrane of *Pseudomonas aeruginosa*. Antibiot. Chemother. **39**:172–181.

36. Oldham, L. J., and F. G. Rodgers. 1985. Adhesion, penetration and intracellular replication of *Legionella pneumophila*: an in vitro model of pathogenesis. J. Gen. Microbiol. **131**:697–706.

37. Payne, N. R., and M. A. Horwitz. 1987. Phagocytosis of *Legionella pneumophila* is mediated by human monocyte complement receptors. J. Exp. Med. **166**:1377–1389.

38. Quinn, F. D., C. A. Butler, and P. S. Hoffman. 1987. Characterization and cloning of the disulfide-cross-linked major outer membrane protein of *Legionella pneumophila* H116. J. Cell. Biochem. **1987**:116.

39. Quinn, F. D., and L. S. Tompkins. 1989. Analysis of a cloned sequence of *Legionella pneumophila* encoding a 38 kD metalloprotease possessing hemolytic and cytotoxic activities. Mol. Microbiol. **3**:797–805.

40. Rowbotham, L. J. 1980. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoeba. J. Clin. Pathol. **33**:1179–1183.

41. 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.

42. 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.

43. Silhavy, T. J., M. L. Berman, and L. W. Enquist. 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

44. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

45. Stephens, R. S., C.-C. Kuo, G. Newport, and N. Agabian. 1985. Molecular cloning and expression of *Chlamydia trachomatis* major outer membrane protein antigens in *Escherichia coli*. Infect. Immun. **47**:713–718.

46. Stephens, R. S., G. Mullenbach, R. Sanchez-Pescador, and N. Agabian. 1986. Sequence analysis of the major outer membrane protein gene from *Chlamydia trachomatis* serovar L2. J. Bacteriol. **168**:1277–1282.

47. Stephens, R. S., E. A. Wagar, and U. Edman. 1988. Developmental regulation of tandem promoters for the major outer membrane protein gene of *Chlamydia trachomatis*. J. Bacteriol. **170**:744–750.

48. Tartakovsky, I. S., V. F. Konyukxov, V. V. Petrosov, and S. V. Prosorovsky. 1990. Protective effect of various antigens in experimental legionellosis, abstr. 33. European Working Group on *Legionella* Infections.

49. Thacker, W. L., H. W. Wilkinson, B. B. Plikaytis, A. G. Steigerwalt, W. R. Mayberry, C. W. Moss, and D. J. Brenner. 1985. Second serogroup of *Legionella feeleii* strains isolated from humans. J. Clin. Microbiol. **22**:1–4.

50. Tropschug, M., E. Wachter, S. Mayer, E. R. Schonbrunner, and F. X. Schmid. 1990. Isolation and sequence of an FK506-binding protein from *N. crassa* which catalyzes protein folding. Nature (London) **346**:674–677.

51. Wadowsky, R. M., L. J. Butler, M. K. Cook, S. M. Verma, M. A. Paul, B. S. Fields, G. Keleti, J. L. Sykora, and R. B. Yee. 1988. Growth-supporting activity for *Legionella pneumophila* in tap water cultures and implication of hartmannellid amoebae as growth factors. Appl. Environ. Microbiol. **54**:2677–2682.

52. Zhao, X., and L. A. Dreyfus. 1990. Expression and nucleotide sequence analysis of the *Legionella pneumophila* recA gene. FEMS Microbiol. Lett. **70**:227–232.