INFECTION AND IMMUNITY, Mar. 1991, p. 1196–1201
0019-9567/91/031196-06$02.00/0

Vol. 59, No. 3

Sequence Diversity of the 60-Kilodalton Protein and of a Putative 15-Kilodalton Protein between the Trachoma and Lymphogranuloma Venereum Biovars of *Chlamydia trachomatis*

LUIS M. DE LA MAZA,* THOMAS J. FIELDER, ELAINE J. CARLSON, BRIAN A. MARKOFF, AND ELLENA M. PETERSON

Department of Pathology, Medical Sciences I, University of California, Irvine, California 92717

Received 3 October 1990/Accepted 6 December 1990

DNA from *Chlamydia trachomatis* serovars L3, C, and E corresponding to the open reading frames of the 60-kDa protein and of a putative 15-kDa protein was sequenced. The open reading frames coding for the 60-kDa protein had 1,641 bp in the three serovars. Compared with the L3 serovar, there were 9 and 11 amino acid changes in the C and E serovars, respectively. The open reading frames corresponding to the putative 15-kDa protein had 450, 456, and 453 bp for the L3, C, and E serovars, respectively. When compared with the L3 serovar, the C and E serovars had 14 and 16 amino acid differences, respectively.

Three Cys-rich proteins with molecular weights of 60,000, 40,000, and 12,000 have been identified in the outer membrane of the elementary bodies (EBs) of *Chlamydia trachomatis* (4, 6, 11). The 40-kDa major outer membrane protein accounts for approximately 60% of the weight of the outer membrane and is acidic, with a pI ranging from 5.3 to 5.5 (4). This protein is a strong immunogen for humans and other mammals, and its molecular and antigenic structure has been extensively characterized (3, 7, 22). The 60-kDa protein appears as a doublet in the lymphogranuloma venereum (LGV) biovar, whereas only a single band has been detected in the trachoma biovar (4). This protein has a neutral isoelectric point (pI 7.3 to 7.7) in the trachoma biovar and contains a net positive charge (pI 8.5 to 9.0) in the LGV group (4). The 60-kDa protein is also strongly immunogenic (24), but all of the antigenic sites so far characterized with monoclonal antibodies have species specificity, suggesting that this protein is well conserved among *C. trachomatis* isolates (14–16). Little is known about the 12-kDa protein, although several proteins with apparent molecular sizes ranging from 9 to 18 kDa have been thought to correspond to this protein (5, 9, 12, 13, 27). In this report, we present the complete DNA sequence of the 60-kDa protein and of a putative 15-kDa protein of the L3, C, and E serovars. Our findings support the evidence for the highly conserved nature of the 60-kDa protein within each biovar. In addition, these results show significant structural variability in the putative 15-kDa protein between the LGV and trachoma biovars and indicate that this putative protein does not correspond to the 12- or 18-kDa protein.

Chlamydia EBs were purified with Renografin 76 (E. R. Squibb and Sons, Inc., Princeton, N.J.) as outlined by Caldwell et al. (6). The DNA from *C. trachomatis* L3 purified EBs was extracted, sonicated, and ligated to *Eco*RI linkers, and a library was constructed in λgt11 cleaved with *Eco*RI (7). The recombinant library was screened in *Escherichia coli* Y1090 (r^- m^+; Promega Biotec, Madison, Wis.) with serum from a rabbit hyperimmunized with *C. trachomatis* L3 EBs (7). One recombinant, λgt11/L3/c9, was further characterized for this study (Fig. 1). The *Pst*I library was constructed by partially cleaving L3 DNA with *Pst*I, ligating it with *Pst*I-cleaved pBluescript SK^- (Stratagene, La Jolla, Calif.), and transforming *E. coli* XL1-Blue (Stratagene) with the ligation product. The *Pst*I library was screened by using λgt11/L3/c9 DNA labeled with ^{32}P. The *Pst*I clone, P4, was prepared for sequencing by overlapping deletions made with the Exomung kit (Stratagene). DNA extracted from purified EBs of the L3, C, and E serovars was used for the polymerase chain reaction (PCR). The oligonucleotides TCGGTCGACATAACAATTTCTACCCG ATGG and TCATTGGGTCTGATCCACCAG were constructed on a DNA synthesizer (model 8600; Biosearch Inc., Burlington, Mass.) and were used to amplify the 5′ and 3′ ends of the chlamydial DNA corresponding to the 60- and 15-kDa open reading frames (ORFs) (2, 9, 25). Other oligonucleotides complementary to internal regions were synthesized to amplify specific sections of the DNA. PCR conditions have been previously described (19, 20). The PCR products were sequenced after cloning into pBluescript SK^-. The DNA was sequenced by the [^{35}S]dATP dideoxy nucleotide chain termination method on both strands from two different clones with synthesized oligonucleotides as primers (21). When there was a discrepancy between the two clones, due to a mutation arising from the PCR, a third clone was sequenced. The DNA sequences of the L3, C, and E serovars were analyzed on a Macintosh computer (Apple Computer Inc., Cupertino, Calif.) with the MacVector software package (International Biotechnologies, Inc., New Haven, Conn.).

The strategy for cloning the DNA coding for the 60- and 15-kDa ORFs and the genetic map of this region are shown in Fig. 1. The DNA sequences and the computer-derived amino acid sequences of the 60-kDa proteins of *C. trachomatis* serovars L3, C, and E are shown in Fig. 2A. The DNA sequence of the L3 serovar is identical to that of the L1 and L2 serovars previously published by Clarke et al. (9) and Allen and Stephens (2), respectively. The PCR product sequenced had a 35-bp segment of DNA in front of the first Met of the 60-kDa ORF. The ORFs of all three serovars had a total of 1,641 bp, starting with a Met codon and terminating with a single stop codon. The calculated molecular weights of the 547-amino-acid 60-kDa proteins of the L3, C, and E serovars were 58,777, 58,675, and 58,703, respectively (Ta-

* Corresponding author.

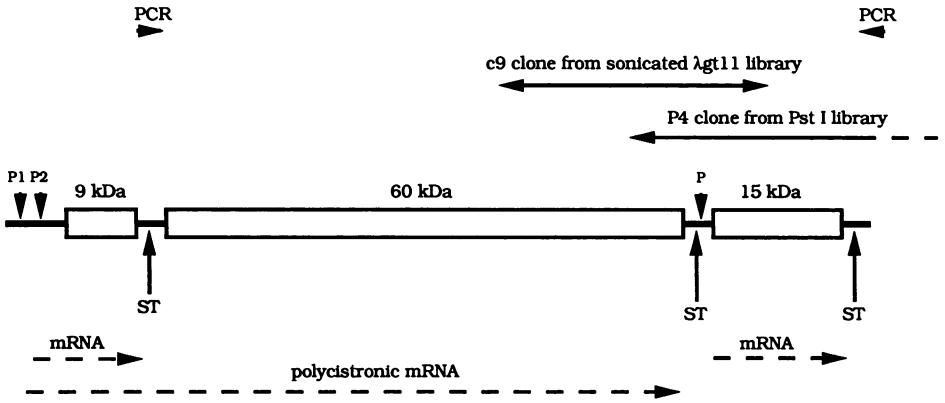

FIG. 1. Diagram of the cloning strategy for the 60- and 15-kDa ORFs of *C. trachomatis* serovars L3, C, and E. The λgt11/L3/c9 original clone was detected with rabbit immune serum to L3 from an L3 DNA library cloned in λgt11. The P4 clone was isolated from a *Pst*I library by using the λgt11/L3/c9 DNA as a probe. The rest of the L3 DNA and the complete DNA fragments of the C and E serovars were obtained after amplification by the PCR with primers located upstream from the 60-kDa ORF and at the end of the ORF of the 15-kDa protein (arrows). Abbreviations: P, promoter; ST, stem-loop.

ble 1). According to the findings of Allen and Stephens (2), the 60-kDa protein of the LGV serovars is posttranslationally processed at two cleavage sites that account for the doublet observed by gel electrophoresis. The signal peptides are 22 and 40 amino acids in length, thus yielding mature proteins 525 and 507 amino acids in length, respectively (Table 1). In contrast, the trachoma serovars have a single signal peptide of 40 amino acids. Two amino acid substitutions at positions 33 (Phe to Ser) and 34 (Ile to Leu) occurred in the signal peptide of the trachoma serovars in comparison with the LGV serovars. When the molecular structures of the processed 60-kDa proteins of the biovar trachoma serovars C and E were compared with that of the L3 serovar, the following changes were found (Table 1; Fig. 2A). The C serovar had 10 bp changes and the E serovar had 13 bp changes relative to the LGV serovars. These resulted in a total of 7 amino acid changes in the C serovar and 9 amino acid changes in the E serovar. Seven of these changes were the same for the two trachoma serovars, and two were unique to the E serovar. All seven changes common to the C and E serovars were also present in the B serovar (1, 25). The B serovar had one additional unique substitution relative to the L3 serovar (Ile instead of Val at position 233) according to Watson et al. (25) (isolate Jali 20/OT) but not according to Allen et al. (1) (isolate B/Tw-5/OT). Three of the seven common changes were from a Lys in the LGV serovars to a Glu in the trachoma serovars and resulted in a significant change in the pIs of these proteins. The calculated pIs were 8.93, 6.84, and 6.84 for the L3, C, and E serovars, respectively (Table 1). There was a total of 24 Cys residues in the 60-kDa protein, mainly located in the center of the protein in two groups of 8 residues at positions 187 to 225 (Cys-1) and between residues 411 and 455 (Cys-2). Computer analysis of the complete ORF indicated that the signal peptide was hydrophobic, whereas the region corresponding to the first 50 residues in the N terminus was highly hydrophilic, likely surface exposed, and potentially moderately antigenic. Both Cys-rich regions were mostly hydrophobic.

A second ORF, 450 bp in length, in the L3 serovar was found to code for a putative sulfur-rich protein with four Cys residues and four Met residues, starting with a Met and ending with two stop codons (Fig. 2B). The calculated molecular weight was 15,835 and the pI was 5.08 for the L3 putative protein. Similar ORFs were also found in the C and E serovars. Comparison of the three DNA sequences indicated that there were 14 bp changes in the C serovar and 17 bp changes in the E serovar relative to the L3 serovar. In addition, the E serovar contained one extra codon and the C serovar had 6 bp more than the L3. When compared with the L3 serovar, these changes resulted in 16 different amino acids for the E serovar and 14 different amino acids for the C serovar. Of these changes, 12 were common to the C and E serovars and 3 were unique to the E serovar. Among the common substitutions two Met residues of L3 were replaced by Ile residues in the C and E serovars. The changes between the trachoma and the LGV biovars, although not particularly clustered, tended to be located in the center of the molecule, with the N and the C termini being well conserved. These amino acid changes resulted in molecular weights of 16,131 and 16,018 for the C and E serovars, respectively. However, both ORFs had the same estimated pI of 5.11.

The intergenic region between the 60- and 15-kDa ORFs of the L3, C, and E serovars had a total of 177 bp (Fig. 2B). At 25 bp downstream from the stop codon of the 60-kDa protein there was a Rho-independent transcriptional terminator with a stem of 10 bp, a 1-bp mismatch and a 4-nucleotide loop (ΔG = −22.6 kcal [ca. −94.6 kJ]/mol) (23). The stem was followed by seven T's in the L3 and E serovars and by eight T's in the C serovar sequence. Twelve base pairs downstream from the T's there was a run of eight A's interrupted by a T that could potentially form a second loop with the cluster of T's (ΔG = 0 kcal/mol). In the L3 serovar a Rho-independent terminator was located 72 bp downstream from the second stop codon of the 15-kDa protein with a perfectly matched 10-bp stem and a 4-nucleotide loop (ΔG = −24.0 kcal [ca. −116 kJ]/mol), followed by a run of 6 T's interrupted by 1 A.

Several differences have been found at the molecular level between the *Chlamydia* trachoma and the LGV biovars. For example, we determined by restriction endonuclease analysis DNA cleavage patterns that could differentiate between the two biovars (17, 18). Batteiger et al. (4) also showed significant pI differences between the 60-kDa proteins of the two biovars. In spite of these changes we have shown here that the 60-kDa protein is well conserved in *C. trachomatis*.

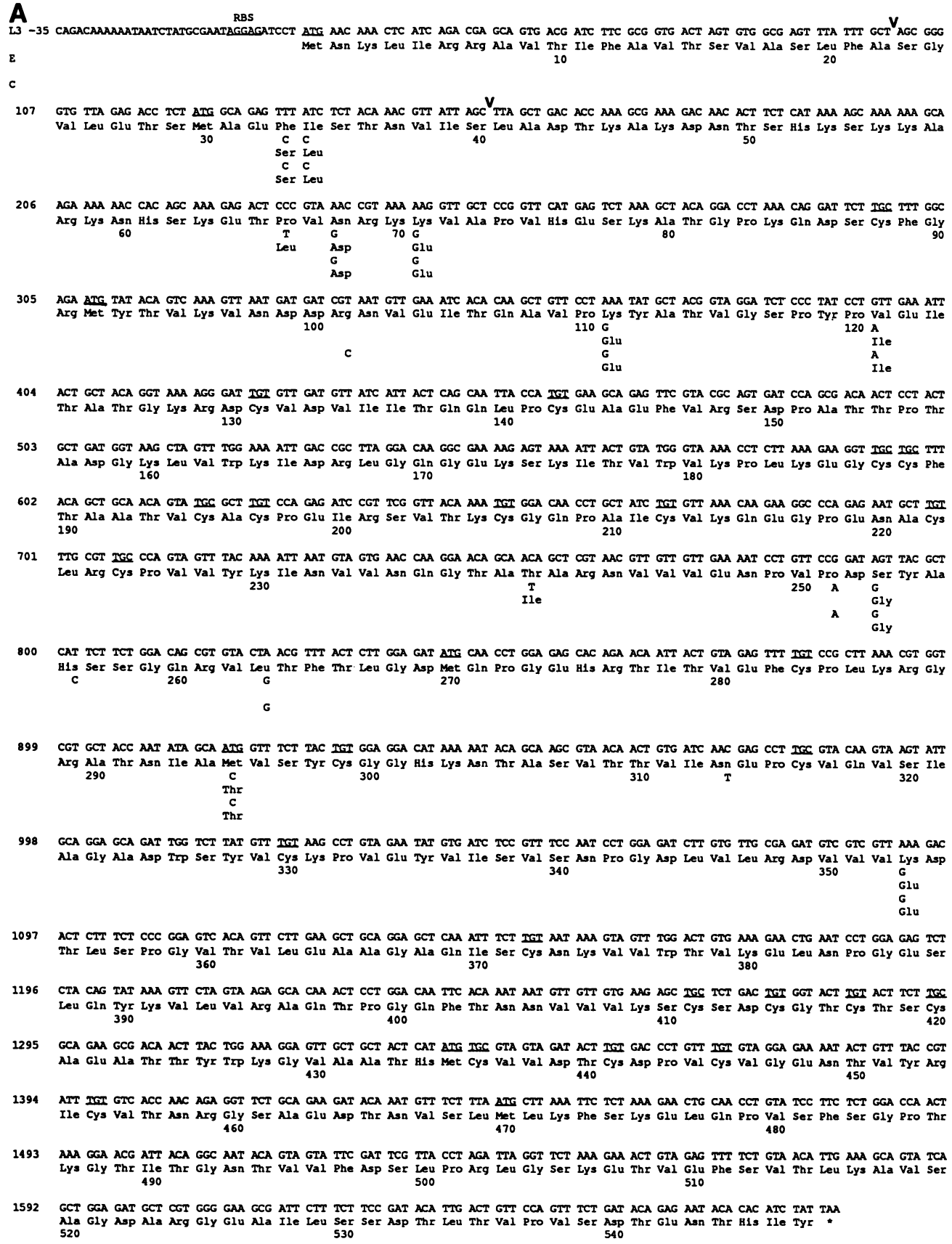

FIG. 2. DNA sequences and computer-translated amino acid sequences of the *C. trachomatis* L3, C, and E serovars. (A) The 60-kDa proteins, including the 5′-flanking sequences. The cleavage sites for the two signal peptides are marked with arrowheads. (B) The putative 15-kDa proteins, including the 5′- and 3′-flanking sequences. Potential stem-loop structures are marked by arrows, and the potential transcription start point is indicated by an arrowhead. The putative ribosome binding sites, the −10 and −35 regions, and the Cys and Met residues are underlined.

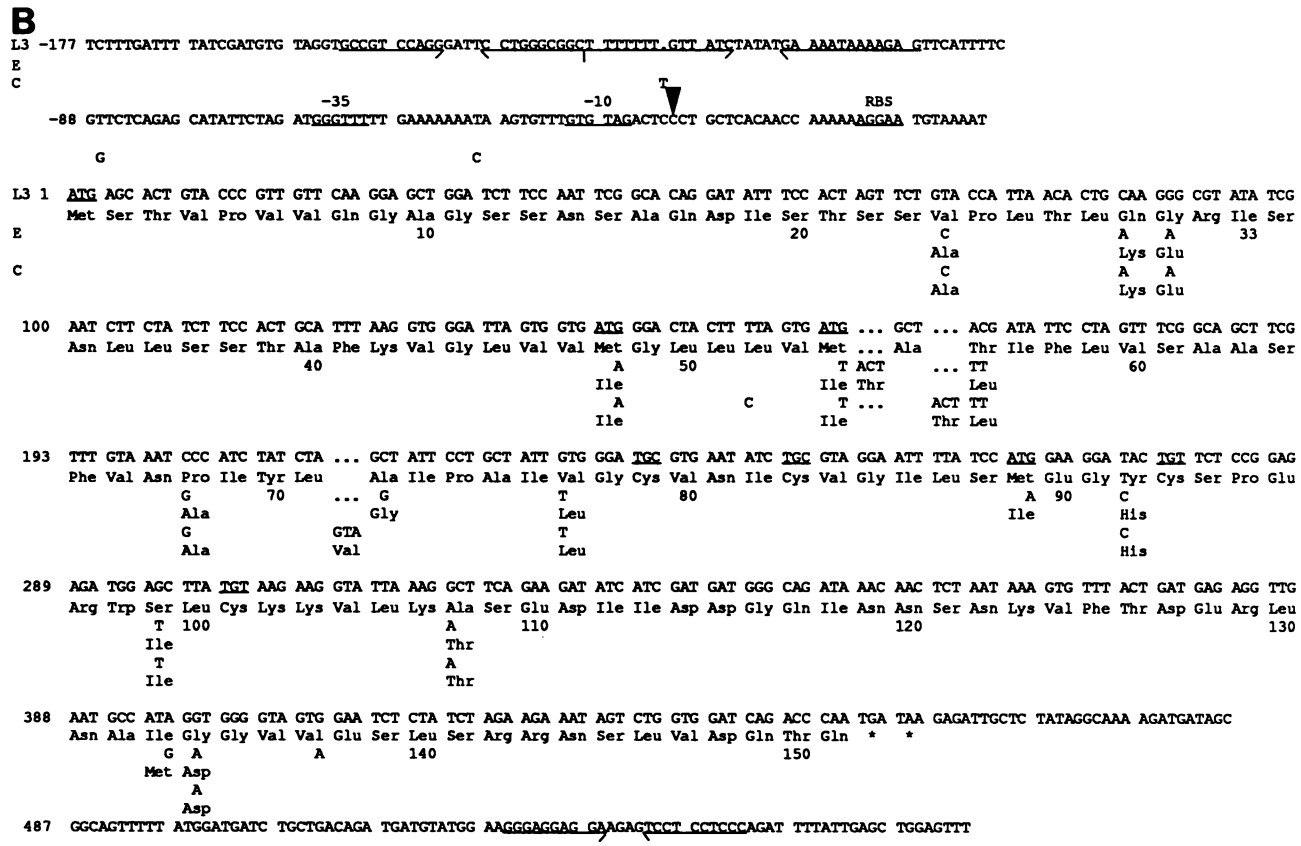

FIG. 2—*Continued*

The DNA sequences of the three LGV serovars so far characterized are identical, and the numbers of residue changes in the trachoma biovar represent less than 1% of the total protein, although they result in a significant difference in the pI of this protein between isolates of the two biovars (4). Interestingly, within the trachoma biovar the three serovars so far sequenced, B, C, and E, shared most of the changes. At this point, we do not know whether the unique changes in the serovars are specific for that particular serovar or are the result of random mutations occurring in individual isolates. DNA sequences of multiple isolates will have to be analyzed to answer this question. The changes appear to be greater among biovars than among serovars, and the differences among the trachoma serovars in the B and C complex do not appear to be consistent. In the Cys-2-rich region of the 60-kDa protein there is a neutral hydropathic area with a very strong possibility of a β-turn that includes the motif Cys-Gly-Thr-Cys (amino acids 414 through 417). A homologous sequence, Cys-Gly-Pro-Cys, is found in the active site of *E. coli* thioredoxin, an enzyme with disulfide isomerase activity; it also serves as an electron donor for ribonucleotide reductase (10). The likelihood that an enzymatic mechanism is responsible for the crosslinking of the outer membrane proteins of *C. trachomatis* has been proposed (15).

In 1987, Newhall (14) described three low-molecular-

TABLE 1. Amino acid composition of the 60-kDa protein

Serovar	Signal polypeptide (no. of amino acids)	No. of amino acid residues																				Calculated mol wt	Estimated pI	No. of amino acids
		Nonpolar								Polar							Acidic		Basic					
		Ala	Val	Leu	Ile	Pro	Met	Phe	Trp	Gly	Ser	Thr	Cys	Tyr	Asn	Gln	Asp	Glu	Lys	Arg	His			
L3	+40	42	72	27	23	29	7	13	5	34	42	55	24	12	25	16	24	29	38	22	8	58,777	9.00	547
	−22	38	69	25	21	29	6	11	5	34	40	53	24	12	24	16	24	29	37	20	8	56,387	8.70	525
	−40	37	67	24	19	29	5	10	5	33	36	51	24	12	23	16	24	27	37	20	8	54,519	8.93	507
C	+40	42	71	28	23	29	6	12	5	35	42	56	24	12	24	16	25	32	35	22	8	58,675	7.38	547
	−40	37	66	24	20	29	4	10	5	34	35	52	24	12	22	16	25	30	34	20	8	54,477	6.84	507
E	+40	42	71	29	24	28	6	12	5	35	42	55	24	12	24	16	25	32	35	22	8	58,703	7.38	547
	−40	37	66	25	21	28	4	10	5	34	35	51	24	12	22	16	25	30	34	20	8	54,505	6.84	507

weight Cys-rich proteins in *C. trachomatis*. Two of them, the 10- and 12.5-kDa proteins, were heavily labeled with [^{35}S]Cys, whereas the 15-kDa protein appeared to be only moderately labeled. The 12-kDa protein was found by Batteiger et al. (4) to have a larger molecular mass (12,500 versus 12,000 Da) and a lower pI (5.4 versus 6.9) in the LGV biovar when compared with those of the trachoma biovar. DNA sequence analysis by Clarke et al. (9) demonstrated the presence of a 450-bp ORF potentially coding for a 15,818-Da protein downstream from the 60-kDa protein in the L1 (440) strain (Fig. 1). We have now found the same structural organization for the *C. trachomatis* L3, C, and E serovars. Clarke et al. (9) thought that this ORF corresponded to the 12-kDa Cys-rich outer membrane protein and to the 14- to 15-kDa protein described by Zhang et al. (27). However, recently, Lambden et al. (13) described a 9-kDa protein from the L1 serovar with 88 amino acids and 13 Cys residues located upstream from the 60-kDa protein (Fig. 1) and suggested that this protein may correspond to the 12-kDa Cys-rich protein. Furthermore, sequencing of the 9-kDa ORF of the B serovar demonstrated the presence of two amino acid changes that resulted in a higher predicted pI value (7.61) than the equivalent for the L1 serovar (pI 6.02) (8). Allen et al. (1) have further substantiated, by using monoclonal and polyclonal antibodies and in vitro expression in *E. coli*, that the 9-kDa protein corresponds to the 12-kDa outer membrane protein. The differences between the calculated and the observed mobilities in gels of this protein were ascribed by Allen et al. to posttranslational modification occurring in vivo. Finally, based on the reactivity of monoclonal antibodies, Newhall (14) suggested that the 15-kDa Cys-rich protein corresponded to the 18-kDa eukaryotic binding protein (26). DNA sequence analysis of the 18-kDa protein by Kaul et al. (12) showed a 162-amino-acid ORF with two Cys residues and no significant homology to the 15-kDa protein described here. In summary, the putative 15-kDa protein described by Clarke et al. (9) and by us here probably corresponds to the 15-kDa Cys-rich protein found by Newhall (14) but not to the 12- and 18-kDa proteins characterized by others.

EMBL accession numbers are X54388, X54389, and X54390 for *C. trachomatis* TW-3, Bour, and 404, respectively.

This study was supported in part by Public Health Service grant A1-26807 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

1. **Allen, J. E., M. C. Cerrone, P. R. Beatty, and R. S. Stephens.** 1990. Cysteine-rich outer membrane proteins of *Chlamydia trachomatis* display compensatory sequence changes between biovariants. Mol. Microbiol. **4**:1543–1550.
2. **Allen, J. E., and R. S. Stephens.** 1989. Identification by sequence analysis of two-site posttranslational processing of the cysteine-rich outer membrane protein 2 of *Chlamydia trachomatis* serovar L2. J. Bacteriol. **171**:285–291.
3. **Baehr, W., Y.-X. Zhang, T. Joseph, H. Su, F. E. Nano, K. D. E. Everett, and H. D. Caldwell.** 1988. Mapping antigenic domains expressed by *Chlamydia trachomatis* major outer membrane protein genes. Proc. Natl. Acad. Sci. USA **85**:4000–4004.
4. **Batteiger, B. E., W. J. Newhall, and R. B. Jones.** 1985. Differences in outer membrane proteins of the lymphogranuloma venereum and trachoma biovars of *Chlamydia trachomatis*. Infect. Immun. **50**:488–494.
5. **Bavoil, P., L. Palmer, S. Falkow, and D. Gump.** 1986. Characterization and cloning of cysteine-rich proteins from *Chlamydia trachomatis*, p. 97–100. *In* D. Oriel, G. Ridgway, J. Schachter, D. Taylor-Robinson, and M. E. Ward (ed.), Chlamydial infections. Cambridge University Press, Cambridge.
6. **Caldwell, H. D., J. Kromhout, and J. Schachter.** 1981. Purification and partial characterization of the major outer membrane protein of *Chlamydia trachomatis*. Infect. Immun. **31**:1161–1176.
7. **Carlson, E. J., E. M. Peterson, and L. M. de la Maza.** 1989. Cloning and characterization of a *Chlamydia trachomatis* L3 DNA fragment that codes for an antigenic region of the major outer membrane protein and specifically hybridizes to the C- and C-related complex serovars. Infect. Immun. **57**:487–494.
8. **Clarke, I. N., S. J. Hayes, and P. R. Lambden.** 1990. Comparative sequence analysis of the 9kDa cysteine rich protein from the lymphogranuloma venereum and trachoma biovars of *Chlamydia trachomatis*, p. 125–128. *In* W. R. Bowie, H. D. Caldwell, R. P. Jones, P.-A. Mardh, G. L. Ridgway, J. Schachter, W. E. Stamm, and M. E. Ward (ed.), Chlamydial infections. Cambridge University Press, Cambridge.
9. **Clarke, I. N., M. E. Ward, and P. R. Lambden.** 1988. Molecular cloning and sequence analysis of a developmentally regulated cysteine-rich outer membrane protein from *Chlamydia trachomatis*. Gene **71**:307–314.
10. **Edman, J. C., L. Ellis, R. W. Blacher, R. A. Roth, and W. J. Rutter.** 1985. Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin. Nature (London) **317**:267–270.
11. **Hatch, T. P., I. Allan, and J. H. Pearce.** 1984. Structural and polypeptide differences between envelopes of infective and reproductive life cycle forms of *Chlamydia* spp. J. Bacteriol. **157**:13–20.
12. **Kaul, R., K. L. Roy, and W. M. Wenman.** 1987. Cloning, expression and primary structure of a *Chlamydia trachomatis* binding protein. J. Bacteriol. **169**:5152–5156.
13. **Lambden, P. R., J. S. Everson, M. E. Ward, and I. N. Clarke.** 1990. Sulfur-rich proteins of *Chlamydia trachomatis*: developmentally regulated transcription of polycistronic mRNA from tandem promoters. Gene **87**:105–112.
14. **Newhall, W. J.** 1987. Biosynthesis and disulfide cross-linking of outer membrane components during the growth cycle of *Chlamydia trachomatis*. Infect. Immun. **55**:162–168.
15. **Newhall, W. J.** 1988. Macromolecular and antigenic composition of chlamydiae, p. 47–70. *In* A. L. Barron (ed.), Microbiology of chlamydiae. CRC Press, Inc., Boca Raton, Fla.
16. **Newhall, W. J., and M. B. Basinski.** 1986. Purification and structural characterization of chlamydial outer membrane proteins, p. 93–96. *In* D. Oriel, G. Ridgway, J. Schachter, D. Taylor-Robinson, and M. E. Ward (ed.), Chlamydial infections. Cambridge University Press, Cambridge.
17. **Peterson, E. M., and L. M. de la Maza.** 1983. Characterization of chlamydia DNA by restriction endonuclease cleavage. Infect. Immun. **41**:604–608.
18. **Peterson, E. M., and L. M. de la Maza.** 1988. Restriction endonuclease analysis of DNA from *Chlamydia trachomatis* biovars. J. Clin. Microbiol. **26**:625–629.
19. **Peterson, E. M., B. A. Markoff, J. Schachter, and L. M. de la Maza.** 1990. The 7.5-kb plasmid present in *Chlamydia trachomatis* is not essential for the growth of the microorganism. Plasmid **23**:144–148.
20. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, E. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.
21. **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.
22. **Stephens, R. S., E. A. Wagar, and G. K. Schoolnik.** High-resolution mapping of serovar-specific and common antigenic determinants of the major outer membrane protein of *Chlamydia trachomatis*. J. Exp. Med. **167**:817–831.
23. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C.**

Uhlenbeck, D. M. Crothers, and J. Gralla. 1973. Improved estimation of secondary structure of ribonucleic acids. Nature (London) **246**:40–41.

24. **Wagar, E. A., J. Schachter, P. Bavoil, and R. S. Stephens.** 1990. Differential human serologic response to two 60,000 molecular weight *Chlamydia trachomatis* antigens. J. Infect. Dis. **162**:922–927.
25. **Watson, M. W., P. R. Lambden, M. E. Ward, and I. N. Clarke.** 1989. *Chlamydia trachomatis* 60 kDa cysteine rich outer membrane protein: sequence homology between trachoma and LGV biovars. FEMS Microbiol. Lett. **65**:293–298.
26. **Wenman, W. M., and R. U. Meuser.** 1986. *Chlamydia trachomatis* elementary bodies possess proteins which bind to eucaryotic cell membranes. J. Bacteriol. **165**:602–607.
27. **Zhang, Y.-X., N. G. Watkins, S. Stewart, and H. D. Caldwell.** 1987. The low-molecular-mass, cysteine-rich outer membrane protein of *Chlamydia trachomatis* possesses both biovar- and species-specific epitopes. Infect. Immun. **55**:2570–2573.