INFECTION AND IMMUNITY, May 1989, p. 1621–1625
0019-9567/89/051621-05$02.00/0

Vol. 57, No. 5

Cloning and Sequence Analysis of the Major Outer Membrane Protein Genes of Two *Chlamydia psittaci* Strains

Y.-X. ZHANG,[1]†* S. G. MORRISON,[1] H. D. CALDWELL,[1] AND W. BAEHR[2]

Laboratory of Microbial Structure and Function, National Institute of Allergy and Infectious Diseases, Hamilton, Montana 59840,[1] and Cullen Eye Institute, Baylor College of Medicine, Houston, Texas 77030[2]

Received 30 November 1988/Accepted 18 January 1989

We cloned and sequenced the gene encoding the major outer membrane protein (MOMP) of two *Chlamydia psittaci* strains, guinea pig inclusion conjunctivitis (GPIC) strain 1, and meningopneumonitis (Mn) strain Cal-10. Intraspecies alignment of the two *C. psittaci* MOMP genes revealed 80.6% similarity, and interspecies comparison of *C. trachomatis* and *C. psittaci* MOMP genes yielded about 68% similarity. As found previously for *C. trachomatis* MOMP sequences, stretches of predominantly conserved sequences of GPIC and Mn MOMPs were interrupted by four variable domains whose locations were identical to those of *C. trachomatis* MOMPs. Seven of eight cysteine residues were found at precisely the same positions in GPIC, Mn, and *C. trachomatis* MOMPs, emphasizing their importance in structure and function of the protein. Collectively, these results indicate that *C. psittaci* and *C. trachomatis* MOMP genes diverged from a common ancestor.

The genus *Chlamydia* contains two species, *Chlamydia trachomatis* and *Chlamydia psittaci* (15). These obligate intracellular procaryotes exhibit a similar, biologically unique life cycle that distinguishes them from all others. Although *C. trachomatis* and *C. psittaci* resemble one another closely in these biological characteristics, the two species differ in host tropisms. *C. trachomatis* is strictly a human pathogen, while *C. psittaci* is primarily a pathogen of lower animals.

The chlamydial surface likely plays several important roles in the process of parasitization of host cells. Previous data indicate that the most abundant major outer membrane protein (MOMP) of *C. trachomatis* is a prominent determiner of both virulence and seroreactivity for the organism (3, 12, 19, 26). The MOMP genes from several *C. trachomatis* serotypes have been sequenced, and their sequences are generally conserved except in four symmetrically spaced variable domains (VDs) (1, 13, 17, 18). Three of the VDs are dominant immunologic determinants and account for the major neutralizing and serotyping sites of MOMP (1, 21). Tryptic cleavage within two VDs of intact elementary bodies of the B serotype correlates with reduced attachment of chlamydiae to HeLa cells, suggesting strongly that these particular portions of MOMP influence pathogenicity (22).

Detailed sequence information is not available for the MOMP of *C. psittaci*. Although MOMPs of *C. psittaci* and *C. trachomatis* are similar in several gross features, including their apparent sizes and abundance, they exhibit little serological relatedness. To enhance the definition of chlamydial surface sites that dictate the pathogenicity and virulence of the organisms, we cloned and sequenced MOMP genes from two *C. psittaci* strains, guinea pig inclusion conjunctivitis (GPIC) strain 1 (10) and meningopneumonitis (Mn) strain Cal 10 (5).

The sequences at the 5′ ends of MOMP genes have been shown to be highly conserved in *C. trachomatis* (1, 13, 17, 18) and have been used here as specific probes to identify MOMP gene sequences in *C. psittaci* strains by Southern blot analysis (16). Such a probe is the subclone containing 5′ sequences of the A MOMP gene, a 280-base-pair (bp) *Xho*I-*Eco*RI fragment of clone AM-11 (1). This ^{32}P-labeled fragment hybridized to both GPIC and Mn genomic DNA. The hybridization pattern suggested that 2.1-kilobase (kb) *Hin*dIII and 3.9-kb *Pst*I fragments of GPIC genomic DNA and 8-kb *Bam*HI, 7.5-kb *Eco*RI, 3.8-kb *Bam*HI-*Eco*RI, and 2.1-kb *Xba*I fragments of Mn genomic DNA harbor the 5′ ends of *C. psittaci* MOMP structural genes. The shortest fragments, a 2.1-kb *Hin*dIII fragment of GPIC (GPM-1) and a 2.1-kb *Xba*I fragment of Mn (MNM-1), were cloned into pUC vectors (24). The sequences of these inserts were determined by the chain termination method (14) with double-stranded sequencing (4) and modified T7 polymerase (23).

DNA sequence analysis of the cloned fragments showed that both GPM-1 and MNM-1 contained 1 kb of 5′ untranslated sequences but lacked the 3′ ends of their respective genes. The 610-bp *Pst*I-*Hin*dIII fragment of GPM-1 was a suitable probe to identify MOMP gene fragments harboring 3′-end sequences. In *Pst*I digests of GPIC genomic DNA, this probe hybridized to a 2.6-kb fragment but not to the 3.9-kb fragment, whereas all the fragments of Mn genomic DNA digest that hybridized with the 5′-end probe also hybridized with this 3′ probe. Hence, the 2.6-kb *Pst*I fragment of GPIC DNA (GPM-2) appeared to contain the 3′-end sequence of the MOMP gene, and the *Bam*HI-*Eco*RI fragment of Mn DNA (MNM-2) was suggested to contain both 5′- and 3′-end sequences of the MOMP gene. A 3′-end probe derived from the A MOMP gene of *C. trachomatis* did not hybridize to GPIC or to Mn genomic DNA under the same conditions. Sequence analysis of cloned GPM-2 and MNM-2 confirmed that the 3′ ends of the genes were contained in the GPM-2 and MNM-2 fragments.

The 2.1-kb GPM-1 and MNM-1 fragments were completely sequenced by using random deletion clones generated by exonuclease III digestion (8). The larger GPM-2 and MNM-2 clones were partially sequenced with specific oligonucleotide primers. The complete sequences of GPIC and Mn MOMP structural genes and their predicted proteins are presented in Fig. 1.

The translation initiator ATGs of the GPIC and Mn

* Corresponding author.
† Permanent address: Beijing Institute of Ophthalmology, Beijing 100005, Beijing, People's Republic of China.

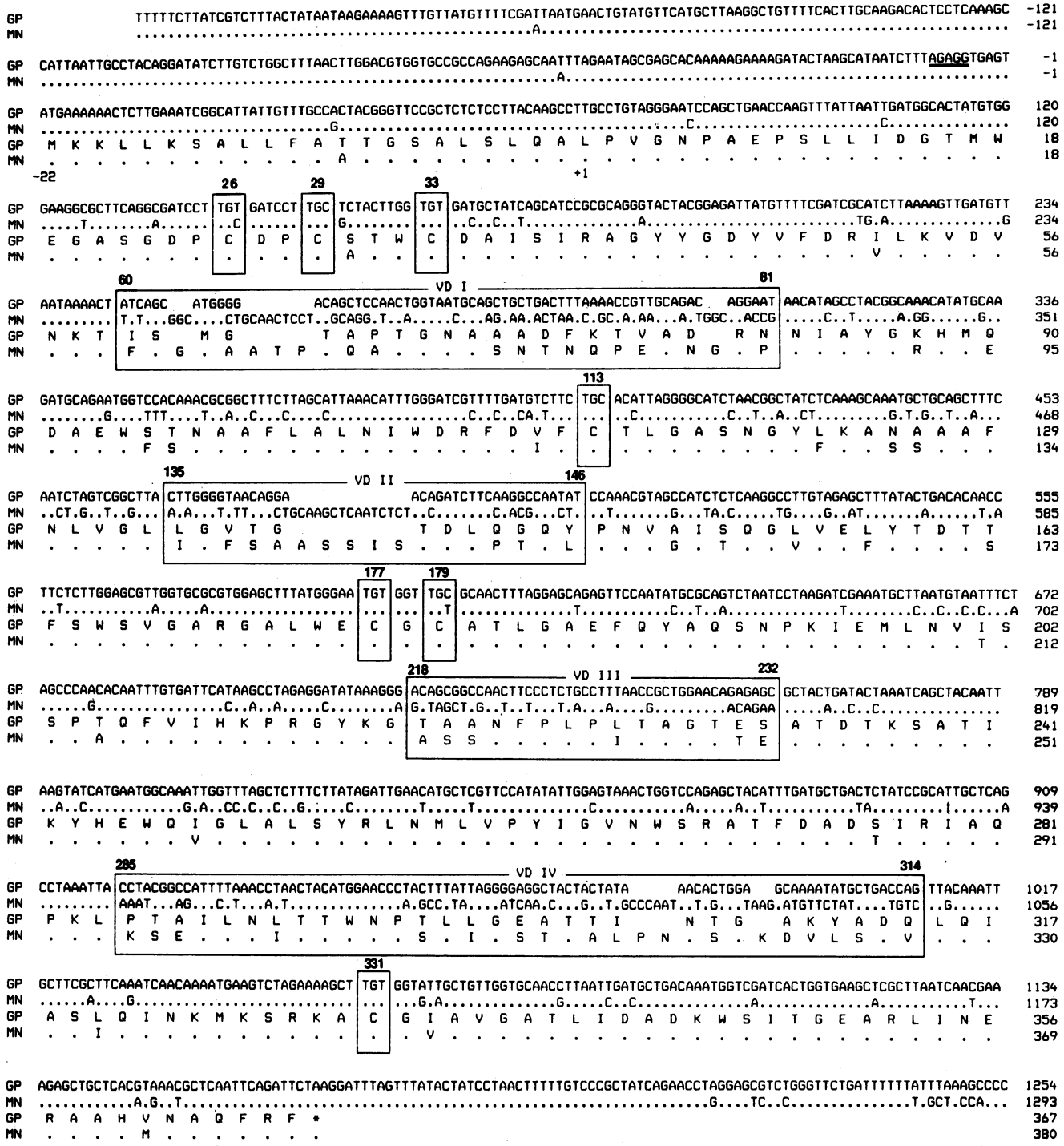

FIG. 1. MOMP gene and deduced peptide sequences of GPIC (GP) and Mn. The complete nucleotide and amino acid sequences of GPIC MOMP are shown. Nucleotides and amino acids that deviate in Mn MOMP are indicated below the GPIC sequences; dots indicate identical residues. A +1 above the gene sequence marks the translational initiation codon ATG; a +1 below the peptide sequence marks the N terminus of mature MOMP. The VDs and conserved cysteines are boxed. Numbers above boxes denote amino acid residues of GPIC MOMP. A putative Shine-Dalgarno sequence is underlined.

MOMP genes were identified by alignment of the genes with homologous *C. trachomatis* MOMP sequences. In each *C. psittaci* MOMP gene, ATG was preceded by a Shine-Dalgarno-like ribosome-binding site, as shown for other MOMP genes (13, 17, 18). The 5′-flanking areas of Mn and GPIC MOMP genes were essentially identical, suggesting evolutionary pressure to strictly conserve regulation of MOMP mRNA transcription. Only two nucleotides were substituted between position −2 (the base preceding ATG) and position −227. In a segment of comparable length in the coding sequence, many more substitutions were observed (Fig. 1). Preliminary results with primer extension indicate that GPIC

TABLE 1. Nucleotide and amino acid identities of MOMPs of *C. psittaci* and *C. trachomatis* strains

Strain or serovar	% Nucleotide (amino acid) identity with:					
	GPIC	Mn	A	C	B	L2
GPIC	100 (100)	78.9 (80.6)	68.1 (67.3)	68.9 (67.0)	68.3 (66.6)	68.2 (65.6)
Mn		100 (100)	67.5 (65.3)	67.7 (65.1)	67.9 (64.7)	67.8 (63.9)
A			100 (100)	96.3 (93.2)	83.4 (85.6)	83.4 (85.4)
C				100 (100)	82.6 (85.2)	82.5 (84.4)
B					100 (100)	94.1 (93.9)
L2						100 (100)

and Mn MOMP genes may have several transcriptional start sites (results not shown). Differential transcription of *C. trachomatis* MOMP genes linked to the chlamydial life cycle has been suggested earlier by Stephens et al. (20), who observed two tandemly arranged promoters in *C. trachomatis* serovar L2 MOMP genes.

The Mn MOMP structural gene (from the translational start ATG to the stop codon TAA) is 39 bp longer than the GPIC MOMP gene. Dot-matrix comparisons of *C. trachomatis* and *C. psittaci* genes (not shown) indicate that, as observed earlier (1, 18), GPIC and Mn genes are interspersed symmetrically with four VDs (VDs I to IV) and that all insertions and deletions have occurred in these VDs. Computer alignment (Table 1) of the MOMP structural genes of the two *C. psittaci* strains revealed 78.9% nucleotide identity, arguing for strong sequence conservation outside the VDs. This degree of similarity is significantly lower than that observed between *C. trachomatis* MOMP genes of two serovars of the same serogroup (95%) and also lower than that observed between two serovars of different serogroups (83%). The degree of interspecies conservation between *C. trachomatis* and *C. psittaci* MOMP genes, however, is surprisingly high, ranging from 67.5 to 68.9% (Table 1).

The conservation of MOMP genes between members of the chlamydial species, *C. psittaci* and *C. trachomatis*, described here does not agree with earlier observations which showed by DNA cross-hybridization that members of the two species had 12% or less genomic similarity (9). There are two possible explanations for these apparent discrepancies. (i) There are only a limited number of conserved chlamydial genes, as reflected by the low amount of homology observed previously. For example, perhaps only those chlamydial outer membrane proteins whose function is critical to the common biological and molecular pathogenetic properties, such as MOMP, have been conserved during evolution. (ii) Earlier genomic hybridization studies, for reasons not currently understood, do not adequately reflect the MOMP gene homologies between the two chlamydial species. Considering the findings reported here, a more complete evaluation of genomic homology between chlamydial species is needed to differentiate between these two possibilities. Recent 16S rRNA sequence analysis of *C. psittaci* 6BC and *C. trachomatis* serovar L2, however, demonstrated more than 95% sequence similarity (25), indicating a much closer relationship than was previously assumed.

The translational reading frames consisted of 1,167 bp encoding a 389-amino-acid pre-MOMP of GPIC and of 1,206 bp encoding a 402-amino-acid pre-MOMP of Mn. By comparison with known *C. trachomatis* MOMP sequences and by inference, the mature N terminus of *C. psittaci* MOMP is designated leucine (position 1). The first 11 residues of *C. trachomatis* (serovars A, B, C, and L2) and *C. psittaci* (GPIC and Mn) MOMPs are identical. The mature MOMPs of GPIC and Mn contain 367 and 380 amino acid residues, respectively. The predicted molecular weights were 39,677 (GPIC) and 41,022 (Mn), consistent with their mobilities on sodium dodecyl sulfate gels (data not shown). In all MOMPs, the N-terminal leucine is preceded by a leader peptide of 22 residues. The leader peptide displays significant heterogeneity between the two *Chlamydia* species (Fig. 2). The characteristic hydrophobic core and the basic amino terminus, however, have been preserved.

GPIC and MN MOMPs are interspersed with four VDs (VD I to VD IV) located at exactly the same positions as those observed in *C. trachomatis* MOMPs. VD I, VD II, and VD IV have little sequence similarity to corresponding domains in other MOMPs (Fig. 2). VD III is the least-variable domain within the same species. The lengths of VDs differ considerably, the shortest being 13 to 15 residues (VD III) and the longest being 29 to 32 residues (VD IV). In *C. trachomatis* serovars, VD IV had been shown to contain an invariant nonapeptide, TTLNPTIAG, which is recognized by species-specific monoclonal antibodies (1). This sequence has been mutated in five positions in the two *C. psittaci* strains, resulting in a loss of the antigenicity of this epitope. However, the substitutions are conserved hydrophobic residues, arguing that this central, uncharged hydrophobic domain of VD IV, which is flanked at both its N- and C-terminal ends by serovariable sequences (1), contributes an important function to MOMP.

The *C. psittaci* MOMPs contain 29 basic and 32 or 33 acidic amino acid residues, predicting a weakly acidic pI of the proteins. In *C. trachomatis* MOMPs, the number of acidic residues is significantly higher (39 to 42), predicting a much lower pI. The excess of acidic residues in the MOMP sequence is reflected in VDs which are also predominantly acidic. Considering only the VDs, the only parts of a MOMP thought to be exposed in viable elementary bodies, the number of exposed charged amino acid residues in *C. trachomatis* MOMP is nearly twice as high as that in GPIC and Mn MOMP. The numbers of aromatic and hydrophobic residues in MOMPs of both species are essentially identical. It is tempting to postulate that these cell surface differences play a critical role in defining the distinct tissue or host tropisms of these two *Chlamydia* species.

MOMP inter- and intramolecular disulfide bonds are thought to be involved in membrane stability of chlamydiae and in the regulation of chlamydial differentiation (2, 6, 7, 11). It is currently not known which cysteines are involved in intermolecular cross-links to stabilize the outer membrane and which cysteines are involved in stabilizing the structure of the MOMP. We observed seven cysteines in *C. psittaci* MOMPs at precisely the same positions as in *C. trachomatis* MOMPs. Three of them are located near the N terminus, one is located between VD I and VD II, two are located between VD II and VD III, and the last one is located between VD IV and the C terminus. One cysteine which is conserved in all

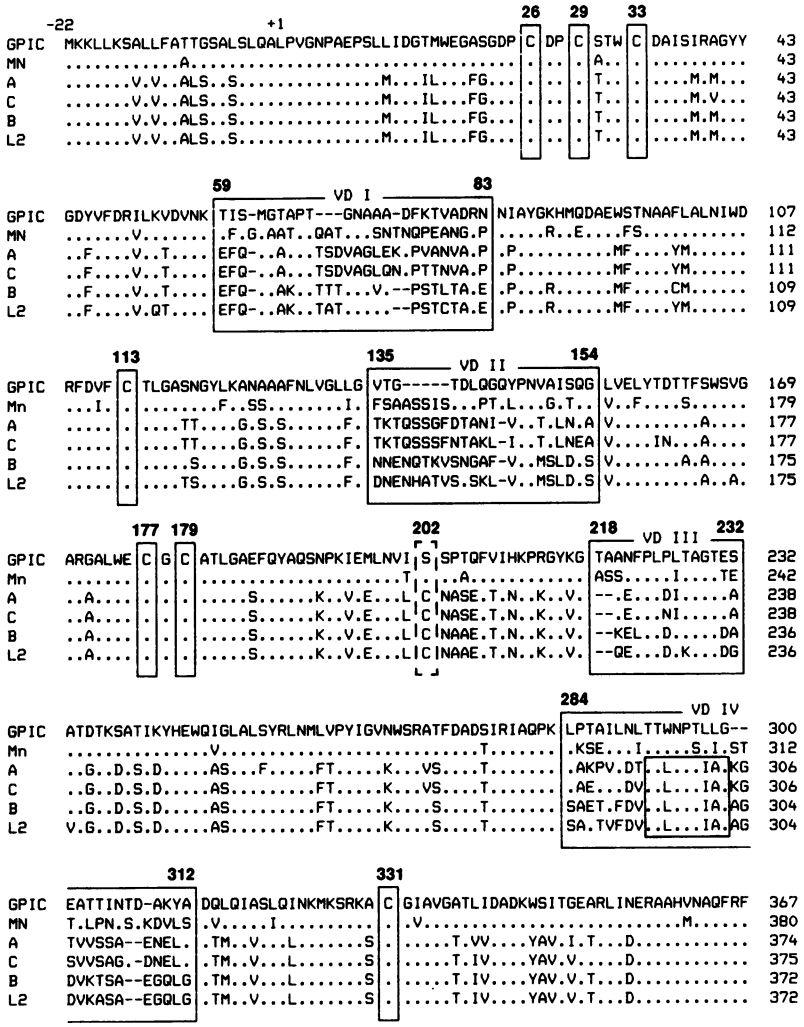

FIG. 2. Comparison of *C. psittaci* and *C. trachomatis* MOMPs. The *C. psittaci* GPIC and Mn MOMP sequences are shown, along with those of *C. trachomatis* serotypes A, C, B, and L2. Residues that are identical to those in GPIC MOMP are represented by dots. The four VDs (VD I to VD IV) and the conserved cysteines are boxed by solid lines. The conserved position where cysteine is located in all known *C. trachomatis* MOMP sequences, but is replaced by serine in GPIC and Mn MOMPs, is boxed by broken lines. Numbers above boxes denote amino acid residues of the GPIC MOMP only. The species-specific epitope of *C. trachomatis* is indicated by an inner box in VD IV.

known *C. trachomatis* MOMP sequences (A, B, C, L1, and L2) has been mutated to serine in GPIC and MN MOMPs (position 202 in GPIC MOMP) (Fig. 2). It appears that this cysteine is dispensable for the structure and function of *C. psittaci* MOMPs.

The MOMPs are critically important for many of the common biological and molecular properties of the two *Chlamydia* species. We found by gene analysis that depending on the serotype compared, MOMPs of various chlamydiae are 64 to 67% identical, the lengths of the proteins are very similar, and N- and C-terminal sequences as well as the positions of seven cysteines have been precisely conserved. In summary, the sequence and structure similarities of the MOMP genes from two *C. psittaci* strains and four serotypes of *C. trachomatis* demonstrate convincingly that these proteins and, by inference, the procaryotic organisms from which they derive are evolutionary homologs. The conservation of these proteins between chlamydial species argues strongly for a divergent evolutionary relationship among strains of the two species and descent from a common ancestor.

We thank Jim Simmons for technical assistance, Ken Robbins for synthesis of oligonucleotide primers, Susan Smaus for secretarial assistance in the preparation of the manuscript, and Gary Hettrick and Bob Evans for photography. The Rocky Mountain Laboratories review committee had many helpful suggestions.

This work was supported in part by a grant from the Edna McConnell Clark Foundation.

LITERATURE CITED

1. **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.
2. **Bavoil, P., A. Ohlin, and J. Schachter.** 1984. Role of disulfide bonding in outer membrane structure and permeability in *Chlamydia trachomatis*. Infect. Immun. **44**:479–485.
3. **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.
4. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.

5. **Francis, T., Jr., and T. O. Magill.** 1938. An unidentified virus producing acute meningitis and pneumonia. J. Exp. Med. **68:** 147–160.
6. **Hackstadt, T., W. J. Todd, and H. D. Caldwell.** 1985. Disulfide-mediated interactions of the chlamydial major outer membrane protein: role in the differentiation of chlamydiae? J. Bacteriol. **161:**25–31.
7. **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.
8. **Hoheisel, J., and F. M. Pohl.** 1986. Simplified preparation of unindirectional deletion clones. Nucleic Acids Res. **14:**3605.
9. **Kingsbury, D. T., and E. Weiss.** 1968. Lack of deoxyribonucleic acid homology between species of the genus *Chlamydia*. J. Bacteriol. **96:**1421–1423.
10. **Murray, E. S.** 1964. Guinea pig inclusion conjunctivitis virus. I. Isolation and identification as a member of the psittacosis-lymphogranuloma-trachoma group. J. Infect. Dis. **114:**1–12.
11. **Newhall, W. J., V, and R. B. Jones.** 1983. Disulfide-linked oligomers of the major outer membrane protein of chlamydiae. J. Bacteriol. **154:**998–1001.
12. **Newhall, W. J., V. P. Terho, C. E. Wilde, III, B. E. Batteiger, and R. B. Jones.** 1986. Serovar determination of *Chlamydia trachomatis* isolates by using type-specific monoclonal antibodies. J. Clin. Microbiol. **23:**333–338.
13. **Pickett, M. A., M. E. Ward, and I. N. Clarke.** 1987. Complete nucleotide sequence of the major outer membrane protein gene from *Chlamydia trachomatis* serovar L1. FEMS Microbiol. Lett. **42:**185–190.
14. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
15. **Schachter, J., and H. D. Caldwell.** 1980. Chlamydiae. Annu. Rev. Microbiol. **34:**285–309.
16. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
17. **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 L_2. J. Bacteriol. **168:**1277–1282.
18. **Stephens, R. S., R. Sanchez-Pescador, E. A. Wagar, C. Inouye, and M. S. Urdea.** 1987. Diversity of *Chlamydia trachomatis* major outer membrane protein genes. J. Bacteriol. **169:**3879–3885.
19. **Stephens, R. S., M. R. Tam, C.-C. Kuo, and R. C. Nowinski.** 1982. Monoclonal antibodies to *Chlamydia trachomatis*: antibody specificities and antigen characterization. J. Immunol. **128:**1083–1089.
20. **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.
21. **Stephens, R. S., E. A. Wagar, and G. K. Schoolnik.** 1988. 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.
22. **Su, H., Y.-X. Zhang, O. Barrera, N. G. Watkins, and H. D. Caldwell.** 1988. Differential effect of trypsin on infectivity of *Chlamydia trachomatis*: loss of infectivity requires cleavage of major outer membrane protein variable domains II and IV. Infect. Immun. **56:**2094–2100.
23. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84:**4767–4771.
24. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.
25. **Weisburg, W. G., T. P. Hatch, and C. R. Woese.** 1986. Eubacterial origin of chlamydiae. J. Bacteriol. **167:**570–574.
26. **Zhang, Y.-X., S. Stewart, T. Joseph, H. R. Taylor, and H. D. Caldwell.** 1987. Protective monoclonal antibodies recognize epitopes located on the major outer membrane protein of *Chlamydia trachomatis*. J. Immunol. **138:**575–581.