JOURNAL OF BACTERIOLOGY, Sept. 1987, p. 3879–3885
0021-9193/87/093879-07$02.00/0
Copyright © 1987, American Society for Microbiology
Vol. 169, No. 9

# Diversity of *Chlamydia trachomatis* Major Outer Membrane Protein Genes

RICHARD S. STEPHENS,[1,2]* RAY SANCHEZ-PESCADOR,[3] ELIZABETH A. WAGAR,[2] CARLA INOUYE,[1] AND MICKEY S. URDEA[3]

*Department of Biomedical and Environmental Health Science, University of California, Berkeley, California 94720[1]; Department of Laboratory Medicine, University of California, San Francisco, California 94143[2]; and Chiron Corporation, Emeryville, California 94608[3]*

Received 24 April 1987/Accepted 9 June 1987

**Genomic DNA libraries were constructed for *Chlamydia trachomatis* serovars B and C by using *Bam*HI fragments, and recombinants that contained the major outer membrane protein (*ompl*) gene for each serovar were identified and sequenced. Comparisons between these gene sequences and the gene from serovar $L_2$ demonstrated fewer base pair differences between serovars $L_2$ and B than between $L_2$ and C; this finding is consistent with the serologic and antigenic relationships among these serovars. The translated amino acid sequence for the major outer membrane proteins (MOMPs) contained the same number of amino acids for serovars $L_2$ and B, whereas the serovar C MOMP contained three additional amino acids. The antigenic diversity of the chlamydial MOMP was reflected in four sequence-variable domains, and two of these domains were candidates for the putative type-specific antigenic determinant. The molecular basis of *ompl* gene diversity among *C. trachomatis* serovars was observed to be clustered nucleotide substitutions for closely related serovars and insertions or deletions for distantly related serovars.**

*Chlamydia trachomatis* is a procaryote that is the cause of a wide spectrum of human diseases affecting hundreds of millions of people worldwide. The most notable diseases are trachoma, a blinding eye disease, and genitourinary tract diseases that often result in sterility (7). Chlamydiae are obligate intracellular bacteria that have a unique biphasic growth cycle which facilitates their survival in two discontinuous habitats. The major outer membrane protein (MOMP) of chlamydiae is one of the principal cell wall surface components that is responsible for the structural integrity of the extracellular infectious elementary body and the developmental conversion to the plastic and fragile intracellular reticulate body (8). This protein also has pore-forming capabilities that permit exchange of nutrients for the reticulate body form (2). The structural and porin functions of the MOMP are mediated by disulfide bond interactions within and between MOMP molecules and other components (16).

Surface components of chlamydiae are presumed to modulate the essential biological events of attachment, induced phagocytosis, inhibition of phagolysosomal fusion, infectivity, toxicity, and the host immune responses that contribute to immunity and pathogenesis (21). The immunodominant MOMP has been implicated in some of these important functions, primarily by association of many of these processes with a type-specific antigen. The predominant type-specific determinant for chlamydiae has been attributed to the MOMP by using monospecific (4) and monoclonal (25, 27) antibodies.

*C. trachomatis* has been extensively characterized serologically, with more than 15 serovars defined by polyvalent antisera (28) and monoclonal antibodies (29). Monoclonal antibody specificities to MOMP define species-, subspecies-, and type-specific determinants; thus, the MOMP represents a serological matrix of epitopes such that a single molecule possesses both constant and variable domains (25). The serological relationships separate the serovars into C- and B-complex groups. Within each group, the serovars can be arranged in a hierarchy of antigenic complexity (28). In addition to these antigenic variations among serovars, the MOMP molecular weights differ among serovars. Notably, the MOMPs of B-complex serovars have lower molecular weights than those of the C-complex serovars (15).

To address the molecular basis of MOMP antigenic diversity and the structural and functional elements of this multifaceted protein, we have compared the MOMP genes from three serovars: serovar $L_2$ and serovar B, which are closely related antigenically, and serovar C, which is the most distantly related. Sequence comparisons among these prototype serovars provided insights into the structural and functional mechanisms of these components as well as the potential genetic mechanisms that account for the antigenic diversities among these organisms.

## MATERIALS AND METHODS

**Bacterial strains.** *C. trachomatis* B/TW5/OT and C/TW3/OT have been described previously (11) and were kindly provided by C.-C. Kuo and S.-P. Wang of the University of Washington. Bacteriophage λ1059 and its host strain *Escherichia coli* Q359 (10), the bacteriophage λgt11/L2/33 recombinant (23), and the pUC plasmids and M13 phage systems have been described previously (13).

**Construction of genomic libraries for *C. trachomatis* serovars B and C.** DNA from each of the *C. trachomatis* serovars was isolated as previously described (23). Standard procedures were used for enzymatic reactions and for isolation of λ phage DNA (12). Chlamydial genomic DNA was digested with *Bam*HI, ligated into λ1059 DNA, and packaged into phage by using commercially prepared extracts (Amersham Corp., Arlington Heights, Ill.). Recombinant phage were plated on *E. coli* Q359 for screening (10).

**Screening of recombinant phage.** The λ phage gt11/L2/33 was digested with *Eco*RI, and the chlamydial insert DNA was subcloned into pUC18. The resulting clone was desig-

* Corresponding author.

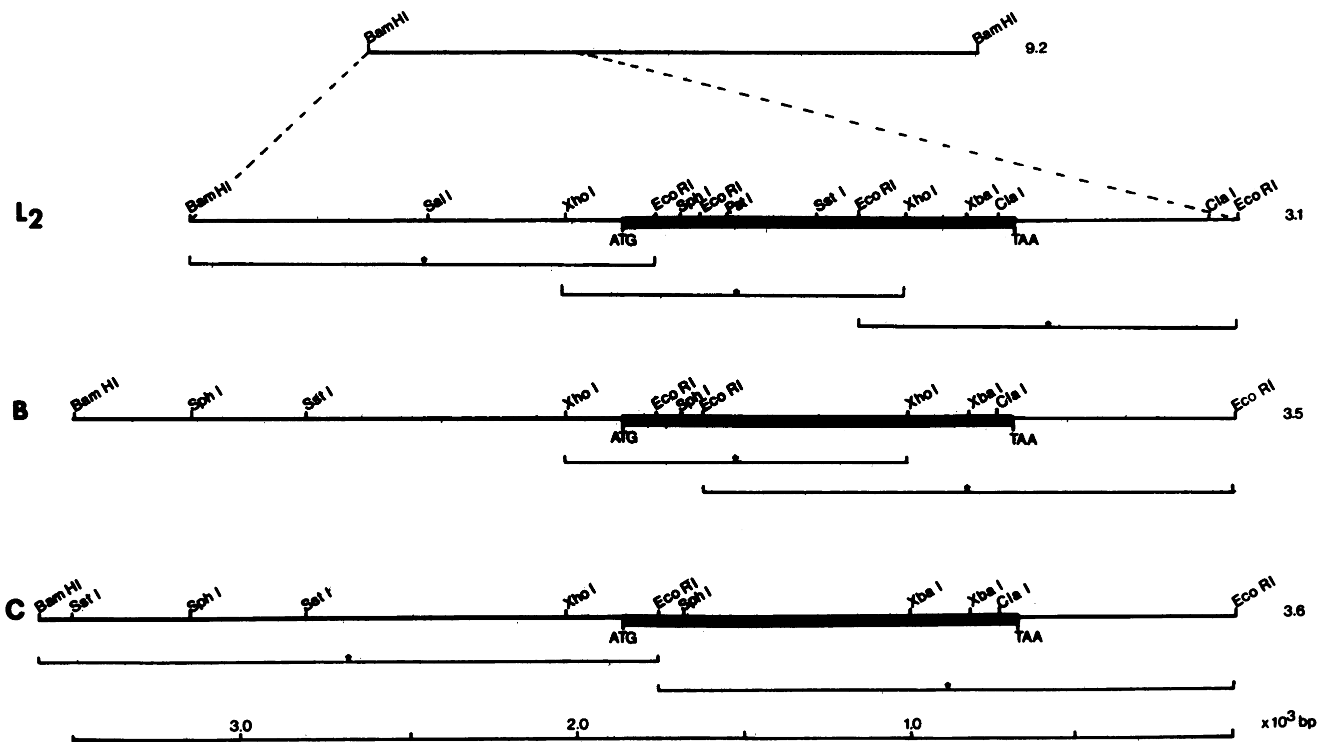


FIG. 1. Endonuclease restriction map of ~9.2-kilobase *Bam*HI fragments from serovars $L_2$, B, and C. Fragments are illustrated 5′ to 3′ and aligned with the common 3′ *Eco*RI sites. ▬, MOMP gene ORF. The sequencing strategy has been previously described (24). Brackets with asterisk beneath each map show constructions used for sequencing. bp, Base pairs.

nated pCt33, and this clone was used as the substrate for nick translation. λ 1059 recombinants were plated and overlaid with nitrocellulose disks, and the disks were probed with pCt33 at 60°C as previously described (12). Two or three clones for each serovar were mapped by restriction endonuclease analysis and Southern blotting as previously described (24).

**DNA sequencing.** The sequencing strategy used for serovars B and C was the same as that employed for the MOMP gene for serovar $L_2$ (*omp*1L2) as previously described (24). Briefly, overlapping restriction endonuclease fragments from representative λ1059 clones for each serovar were gel purified and ligated into pUC18, M13 mp18, and M13 mp19 (13). The pUC recombinants were used for cross-hybridization evaluations, and the M13 clones were sequenced by the dideoxy chain termination method (13, 20) by using oligonucleotide primers for extended sequence. Sequence discrepancies between complementary strands were rectified by synthesizing oligonucleotide primers adjacent to the region in question and resequencing. Oligonucleotide primers were synthesized by using phosphoramite chemistry as previously described (30).

## RESULTS

**Cloning and mapping of MOMP genes.** In a previous study, the chlamydial DNA insert from a λgt11 recombinant (gt11/L2/33) that expresses a β-galactosidase fusion protein representing the carboxyl-terminal portion of serovar $L_2$ MOMP was used to probe Southern blots of chlamydial genomic DNA. This probe hybridized to an approximately 9-kilobase *Bam*HI fragment from each of the serovars $L_2$, B, and C (23). Recently, Stephens et al. (24) determined the DNA sequence for the structural gene for serovar $L_2$ MOMP by locating the gene with the λgt11 recombinant. We used the same approach to identify the MOMP genes for serovars B and C. Briefly, the insert obtained from the λgt11 recombinant or, in some experiments, that obtained from the pCt33 subclone was used to probe λ1059 libraries constructed from *Bam*HI-digested genomic DNA obtained from serovars B and C. DNA was prepared from several identified plaques representative of each serovar and mapped with restriction endonucleases (Fig. 1). The restriction endonuclease maps were verified by generating subclones of two or three fragments for each serovar in pUC18, and predicted cross-hybridizations between clones were observed in Southern blots (data not shown). The MOMP gene for serovar $L_2$ has been designated *omp*1L2 (24); thus, the genes for serovars B and C have been designated *omp*1B and *omp*1C, respectively.

**DNA sequence.** Overlapping fragments were cloned into M13 mp18 and M13 mp19 and sequenced (Fig. 1). Although the entire 3.1-kilobase *Bam*HI-*Eco*RI fragment has been sequenced for serovar $L_2$ (24), the sequences for serovars B and C were extended approximately 100 base pairs 5′ and 3′ to the MOMP gene open reading frame (ORF). Figure 2 shows the differences in ORF sequences of *omp*1B and *omp*1C as compared with the *omp*1L2 ORF sequence.

The *omp*1 gene sequence for serovar B was identical to that of $L_2$ except for 70 nucleotide changes. Sequences 5′ to the ORFs were identical for all three serovars except for an extra base (cytosine) for serovars B and C between a putative promoter sequence and the Shine-Dalgarno complementarity (22) (Fig. 2). The sequence for *omp*1C had more than 70 nucleotide differences as compared with *omp*1L2 and *omp*1B, with, significantly, three codon insertions.

```
L2  TAATTATACAAT•TTAGAGGTAAGA   ATG AAA AAA CTC TTG AAA TCG GTA TTA GTG TTT GCC GCT TTG AGT TCT GCT TCC TCC TTG CAA GCT CTG
                            -22Met Lys Lys Leu Leu Lys Ser Val Leu Val Phe Ala Ala Leu Ser Ser Ala Ser Ser Leu Gln Ala Leu 1
B               C                                               A
C               C                                               A

L2  CCT GTG GGG AAT CCT GCT GAA CCA AGC CTT ATG ATC GAC GGA ATT CTA TGG GAA GGT TTC GGC GGA GAT CCT TGC GAT CCT TGC ACC ACT
    Pro Val Gly Asn Pro Ala Glu Pro Ser Leu Met Ile Asp Gly Ile Leu Trp Glu Gly Phe Gly Gly Asp Pro Cys Asp Pro Cys Thr Thr 31
B                                                              G
C                                                              G               T

L2  TGG TGT GAC GCT ATC AGC ATG CGT ATG GGT TAC TAT GGT GAC TTT GTT TTC GAC CGT GTT TTG CAA ACA GAT GTG AAT AAA GAA TTC CAA
    Trp Cys Asp Ala Ile Ser Met Arg Met Gly Tyr Tyr Gly Asp Phe Val Phe Asp Arg Val Leu Gln Thr Asp Val Asn Lys Glu Phe Gln 61
B       GT                                                                    C         A
        Val                                                                             Lys
C                                   G T           C   A                                 A    T                       T   G
                                    Val                                                 Lys

L2  ATG GGT GCC AAG CCT ACA ACT ··· ··· GCT ACA GGC AAT GCT GCA GCT CCA TCC ACT TGT ACA GCA AGA GAG AAT CCT GCT TAC GGC CGA
    Met Gly Ala Lys Pro Thr Thr ··· ··· Ala Thr Gly Asn Ala Ala Ala Pro Ser Thr Cys Thr Ala Arg Glu Asn Pro Ala Tyr Gly Arg 91
B                               ··· ··· A                   T               CT
                                ··· ··· Thr                 Val             Leu
C         A   G GC        T   C AGC GAT  TA G       TTA CAA AAC  A      A A  TA AAC GTT    T CGT CCA       C        T      AA
                Ala             Ser Asp Val Ala     Leu Gln Asn Asp     Thr Thr Asn Val          Pro                       Lys

L2  CAT ATG CAG GAT GCT GAG ATG TTT ACA AAT GCT GCT TAC ATG GCA TTG AAT ATT TGG GAT CGT TTT GAT GTA TTC TGT ACA TTA GGA GCC
    His Met Gln Asp Ala Glu Met Phe Thr Asn Ala Ala Tyr Met Ala Leu Asn Ile Trp Asp Arg Phe Asp Val Phe Cys Thr Leu Gly Ala 121
B                                             C    G                                C                              C
                                                   Cys
C     C       A         A         G  C                          A      C              T                            G      A

L2  ACC AGT GGA TAT CTT AAA GGA AAT TCA GCA TCT TTC AAC TTA GTT GGC TTA TTC GGA GAT AAT GAG AAC CAT GCT ACA GTT TCA GAT AGT
    Thr Ser Gly Tyr Leu Lys Gly Asn Ser Ala Ser Phe Asn Leu Val Gly Leu Phe Gly Asp Asn Glu Asn His Ala Thr Val Ser Asp Ser 151
B   T T   C       C               C   T   T           T      G   G                A              G A   A             A  G
    Ser                                                                           Asn            Gln Thr Lys         Asn Gly
C     T  CC    T     T A          C   T   T   C                  A                ACA   A ACA C.A TC  T    GC T   AAT ACA GCG
      Thr                                                                         Thr Lys Thr Gln Ser Ser Ser Phe Asn Thr Ala

L2  AAG CTT GTA CCA AAT ATG AGC TTA GAT CAA TCT GTT GTT GAG TTG TAT ACA GAT ACT ACT TTT GCT TGG AGT GCT GGA GCT CGT GCA GCT
    Lys Leu Val Pro Asn Met Ser Leu Asp Gln Ser Val Val Glu Leu Tyr Thr Asp Thr Thr Phe Ala Trp Ser Ala Gly Ala Arg Ala Ala 181
B   GC  T                                                                     G           G           TC   C   G
    Ala Phe                                                                   Ala                     Val
C           A T    T    C  CT GCT    G A   G   G      G        C T       T A C       C               C  TA    T
            Ile             Thr Ala    Asn Glu Ala                       Ile Asn                        Val

L2  TTG TGG GAA TGT GGA TGC GCG ACT TTA GGC GCT TCT TTC CAA TAC GCT CAA TCC AAG CCT AAA GTC GAA GAA TTA AAC GTT CTC TGT AAC
    Leu Trp Glu Cys Gly Cys Ala Thr Leu Gly Ala Ser Phe Gln Tyr Ala Gln Ser Lys Pro Lys Val Glu Glu Leu Asn Val Leu Cys Asn 211
B                         T   A           A                   T           T   A           A                          C   T
C   C C              G    T   A    G      A                   T           T   A           A       G       T       T      T

L2  GCA GCT GAG TTT ACT ATC AAT AAG CCT AAA GGA TAT GTA GGG CAA GAA TTC CCT CTT GAT CTT AAA GCA GGA ACA GAT GGT GTG ACA GGA
    Ala Ala Glu Phe Thr Ile Asn Lys Pro Lys Gly Tyr Val Gly Gln Glu Phe Pro Leu Asp Leu Lys Ala Gly Thr Asp Gly Val Thr Gly 241
B         A                 T      A           G           T A G       G                  C                     C   C
                                                           Lys       Leu                  Thr                 Ala Ala
C       T C   A             T          G                T     GCG       T  A     A   A   CC                 A  C   C       G
        Ser                                                   Ala                Asn Ile Thr                Glu Ala Ala

L2  ACT AAG GAT GCC TCT ATT GAT TAC CAT GAA TGG CAA GCA AGT TTA GCT CTC TCT TAC AGA CTG AAT ATG TTC ACT CCC TAC ATT GGA GTT
    Thr Lys Asp Ala Ser Ile Asp Tyr His Glu Trp Gln Ala Ser Leu Ala Leu Ser Tyr Arg Leu Asn Met Phe Thr Pro Tyr Ile Gly Val 271
B                                                                                 T                   T
C                             C         G                                         T A                 T

L2  AAA TGG TCT CGA GCA AGT TTT GAT GCA GAC ACG ATT CGT ATT GCT CAG CCG AAG TCA GCT ACA ACT GTC TTT GAT GTT ACC ACT CTG AAC
    Lys Trp Ser Arg Ala Ser Phe Asp Ala Asp Thr Ile Arg Ile Ala Gln Pro Lys Ser Ala Thr Thr Val Phe Asp Val Thr Thr Leu Asn 301
B                         C                                                   C GAG     A
                                                                                Glu     Ile
C                 A   T                   C           C       C          T   A  TG    GA  G A A     G        C   T        A
                      Val                                                       Leu   Glu Ala Ile Leu

L2  CCA ACT ATT GCT GGA GCT GGC GAT GTG AAA GCT AGC GCA GAG GGT ··· CAG CTC GGA GAT ACC ATG CAA ATC GTT TCC TTG CAA TTG AAC
    Pro Thr Ile Ala Gly Ala Gly Asp Val Lys Ala Ser Ala Glu Gly ··· Gln Leu Gly Asp Thr Met Gln Ile Val Ser Leu Gln Leu Asn 331
B                                         A                     ···             C               C
                                          Thr
C    G   C   C        T AAA   A AG    GTC T   GC   G ACC  A  AAC G A   G  CT       A                        G
                        Lys     Ser   Val Ser Ala Gly Thr Asp Asn Glu     Ala

L2  AAG ATG AAA TCT AGA AAA TCT TGC GGT ATT GCA GTA GGA ACA ACT ATT GTG GAT GCA GAC AAA TAC GCA GTT ACA GTT GAG ACT CGC TTG
    Lys Met Lys Ser Arg Lys Ser Cys Gly Ile Ala Val Gly Thr Thr Ile Val Asp Ala Asp Lys Tyr Ala Val Thr Val Glu Thr Arg Leu 361
B
C                                                  G        A                                                G
                                                                                                             Ala

L2  ATC GAT GAG AGA GCT GCT CAC GTA AAT GCA CAA TTC CGC TTC TAA  TTAATTGTATAATTTTGTTAA
    Ile Asp Glu Arg Ala Ala His Val Asn Ala Gln Phe Arg Phe STOP 375     STOP      STOP
B
C                    A                         G          •STOP           A A   G
```

FIG. 2. DNA and translated amino acid sequences for the complete MOMP of serovars $L_2$, B, and C. Sequences for B and C were identical to those for $L_2$ (24) except for the differences noted.

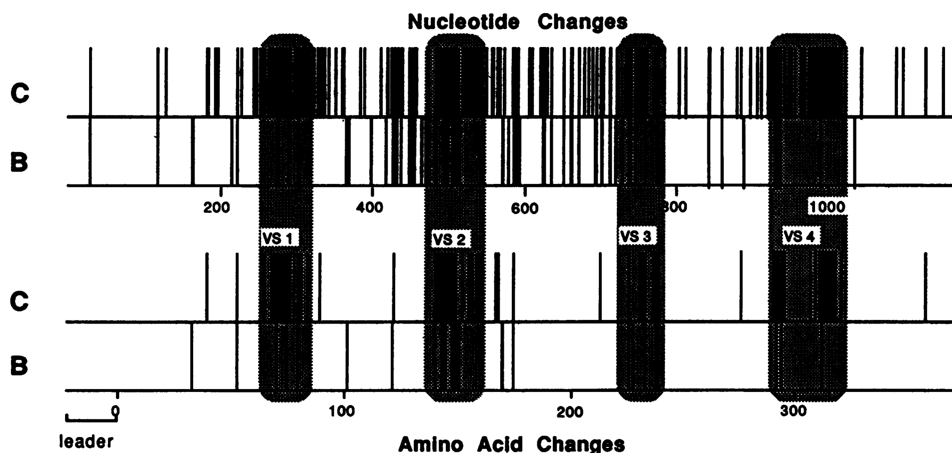


FIG. 3. Frequency and distribution of nucleotide and amino acid differences for *ompl*B and *ompl*C as compared with *ompl*L2. Each line represents one nucleotide or amino acid change from the *ompl*L2 sequence. B, Differences for serovar B; C, differences for serovar C. Shaded areas delineate variable sequence (VS) domains.

When the additonal codons were arbitrarily aligned for maximum homology within these heterogeneous segments, the serovar C gene had 216 nucleotide differences as compared with $L_2$; nevertheless, only 30 of these base pair changes were identical to serovar B changes from serovar $L_2$ (Fig. 2).

**Amino acid sequence.** The inferred amino acid sequence of each of the *ompl* ORFs was translated from the DNA sequences. The first 54 amino acids in each of the three serovars were identical. Nano et al. have determined the 26 amino-terminal amino acids by Edman degradation of purified $L_2$/434/Bu MOMP (14). The first 21 amino acids of their sequence matched translated amino acid residues 23 through 44; thus, the MOMP has a 22-amino-acid leader or signal sequence for translocation of the protein through the outer membrane (Fig. 2). The ORF for *ompl*L2 and *ompl*B encoded 394 amino acids, or 372 amino acids with the 22-amino-acid signal sequence removed. Thus, the calculated MOMP molecular weight for serovars $L_2$ and B was 40,282 without the signal sequence, and this agreed with estimates of 39,500 to 40,500 obtained by polyacrylamide gel electrophoresis (3, 15). The *ompl*C gene encoded three additional amino acids with a calculated molecular weight of 40,607 (without the signal sequence), which was less than the polyacrylamide gel electrophoresis estimates of 41,000 to 44,000 molecular weight (3, 15). In our assessments, the molecular weight of serovar C MOMP was approximately 1,000 higher than that of serovar $L_2$ MOMP as judged by polyacrylamide gel electrophoresis (data not shown).

There was a notable homogeneous distribution of charged and hydrophobic amino acids; thus, the chlamydial MOMP is similar in this respect to other porin proteins such as OmpF of *E. coli* (6). Cysteine residues have been shown to be structurally (8, 16) and functionally (2) important for the MOMP, and comparison of the three sequences demonstrated that there were seven conserved cysteine residues, although serovar $L_2$ had a total of nine cysteine residues and serovars B and C had eight.

**Diversity of MOMP genes.** Comparison of the three DNA sequences revealed relatively few nucleotide differences between serovars $L_2$ and B, with only one triple- and six double-base-pair changes within codons. Furthermore, most of the nucleotide changes occurred in the third base position of a codon, which did not result in an amino acid change. Serovar C had many more multiple nucleotide changes within codons and, notably, three additional codons. It was interesting, however, that serovars C and B shared many of the same mutational events, with 43% of the nucleotide changes in serovar B being identical to changes observed in serovar C (Fig. 2).

The frequency of nucleotide and amino acid changes is illustrated in Fig. 3. Long stretches were relatively invariant, and most of the resulting changes in amino acids in these regions were conservative. However, there were four domains that demonstrated the most nucleotide divergence and often resulted in nonconservative amino acid substitutions (Fig. 3). Since many of these changes may translate into the antigenic polymorphisms observed among these serovars, it was reasoned that comparisons of the hydrophilicity profiles used for the prediction of antigenic determinants, as refined by Hopp and Woods (9), would also reflect nonconservative amino acid changes that may be important for the observed antigenic diversity. Conversely, conserved areas could be identified that may be essential to the structure and function of these proteins.

The hydrophilicity profile for serovar $L_2$ MOMP displayed two prominent peaks; however, these fell within conserved domains of the molecule and are unlikely candidates to account for antigenic diversity (Fig. 4). The type-specific

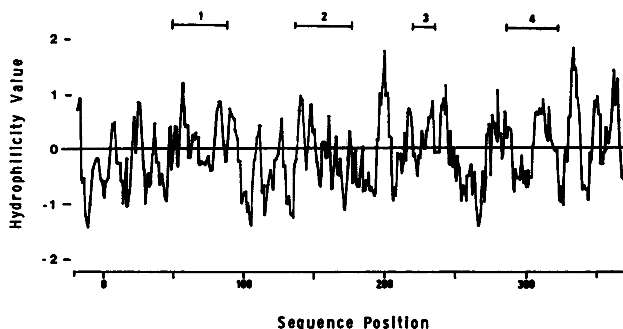


FIG. 4. Hydrophilicity plot for serovar $L_2$ MOMP sequence. Mean hexapeptide antigenicity values are plotted over the median amino acid sequence position from which they were calculated (9). Numbered bars delineate the sequence-variable domains represented in Fig. 3.

determinants are remarkably immunogenic (3, 28); thus, sequences that account for the antigenic diversity of chlamydiae might be expected to be characterized by hydrophilic regions that are also represented by differences in the profiles of each of the serovars. Variable segment 2 between mean sequence positions 136 and 144 and variable segment 4 between positions 288 and 291 showed considerable differences among all three serovars, and these segments were hydrophilic (data not shown). In contrast, the marked differences observed within variable segments 1 and 3 resulted in relatively hydrophobic segments for at least one of the serovars.

## DISCUSSION

The serologic and antigenic relationships among serovars $L_2$, B, and C were reflected in the level of DNA homology between their respective *omp*1 genes. Serovar C is the most antigenically distant and complex of the three serovars (28); hence, these comparisons span the limits of divergence among all the *C. trachomatis* serovars, and also show the finer differences between closely related serovars. There were 70 base pair differences between serovars $L_2$ and B, whereas the serovar C gene had over 200 base pair differences when compared with $L_2$. Although the two trachoma-causing serovars, B and C, shared many of the same mutations, serovars $L_2$ and C shared equivalent numbers of singular DNA homologies. Thus, within this structural gene serovar C appears to be as evolutionarily distant from B as serovar C is from $L_2$.

The translated amino acid sequence provided information concerning the potential structural and functional elements of the MOMP, and the comparisons of the three serovars highlighted the conserved and divergent domains proposed from antigenic (25) and structural (4) studies. The first 60 amino acids were highly conserved, and the MOMP amino terminus identified by Nano et al. (14) coincided with amino acid 23. Thus, the MOMP has a 22-amino-acid leader or signal sequence. The conserved nature of this portion of the chlamydial MOMP, as these sequences are found in both procaryotes and eucaryotes (18), suggests that each of the MOMPs is processed similarly and that they would share a common amino terminus. Consequently, we designated the amino-terminal amino acid of the processed protein as residue 1 and the initial methionine as residue −22.

The calculated molecular weights of MOMP from serovars $L_2$ and B were the same and correlated with estimates obtained by polyacrylamide gel electrophoresis. The three additional amino acids found in the serovar C sequence may be insufficient to account for the higher molecular weight observed for serovar C MOMP by polyacrylamide gel electrophoresis. Since the number of amino acid residues may not account for the discrepancy in molecular weight evaluations, other factors could be considered. One possibility is the stable interaction of MOMP with lipid or carbohydrate moieties.

The distributions of charged, polar, and hydrophobic residues were similar in that there were no long linear stretches of any of these groups. The even distribution of charged and polar residues and the lack of an alpha-helical hydrophobic segment long enough to span a membrane are compatible with the structures of integral membrane porin proteins (17). Disulfide bonds play an essential role in maintaining the structural integrity of the infectious elementary body. Such integrity is necessary for survival in an extracellular habitat by an organism that lacks peptidoglycan (1). Additionally, there is evidence that MOMP is capable of forming membrane pores; however, unlike the activity of porins studied in other systems, this porin activity is uniquely mediated and possibly regulated by disulfide interactions (2). Three participating cysteine residues may be sufficient for these activities (2); however, the observation that there were seven conserved cysteine residues suggests that each of these residues participates. Thus, the MOMP has the capacity for a number of disulfide-mediated interactions with itself and other components.

The molecular basis for the antigenic diversity among *C. trachomatis* serovars should be reflected in comparisons of amino acid sequences of the immunodominant MOMP. To initiate this evaluation, we chose three serovars ($L_2$, B, and C) that, based upon their antigenic relationships, are representative and exemplify the spectrum of antigenic relatedness within the species. Serovars $L_2$ and B are members of the B-complex group and are quite closely related antigenically, despite the fact that $L_2$ also represents a biovar separate from the trachoma biovar represented by serovars B and C. Sequence comparisons between *omp*1 genes for serovars $L_2$ and B should reflect potential type-specific regions as well as the antigenic commonalities between these serovars that can be demonstrated by the reactivity patterns of subspecies-specific monoclonal antibodies (25). Serovar C is the prototype for the C-complex group and is the most antigenically divergent serovar (28). Thus, *omp*1 sequence comparisons of serovar C with serovars $L_2$ and B should reflect type-specific regions as well as define the maximum potential differences likely to be encountered within the structural constraints of the MOMP, thereby delineating common domains which are probably essential for the structural and functional properties shared by all *C. trachomatis* strains.

The frequency of nucleotide and amino acid changes in *omp*1 genes for all three serovars (Fig. 3) illustrates five constant domains of the MOMP that are separated by four sequence-variable domains. The four variable domains also demonstrated divergence in the hydrophilicity comparisons, which highlighted nonconserved amino acid substitutions. The seven conserved cysteine residues each mapped within the conserved domains; this finding lends support for their essential participation in MOMP functions. It is of interest that the most silent regions are observed at the 5′ and 3′ ends of the gene (Fig. 3); thus, it is these regions where homologies may be found to *Chlamydia psittaci omp*1 genes (23). The other conserved regions among the variable segments display considerably more nucleotide changes; however, these did not translate into significant amino acid changes (Fig. 3). The even linear distribution and alternation of variable and conserved domains were remarkable and may relate to the structural symmetry observed for this protein (5).

The molecular basis of genetic variation among *C. trachomatis* serovars is founded in clustered base substitutions for closely related serovars (i.e., $L_2$ and B) and insertions or deletions for distantly related serovars. The molecular mechanisms that account for the development of *omp*1 gene-variable domains fall into two categories. The differences between $L_2$ and B are most easily explained as the result of an accumulation of point mutations. For example, of the 70 base pair differences between *omp*1L2 and *omp*1B, there was only one triple-base-pair mutation and six double-base-pair mutations within codons. The development of the C-complex gene was a much more radical departure, wherein three regions bear no sequence relationship to the

other two serovars. The coding sequence for the string of amino acids 69 to 85 (Fig. 2) within variable segment 1 had virtually no homology to sequences in serovars $L_2$ or B, and this string had two additional codons. The linear string of amino acids 141 to 151 in variable segment 2 shared no amino acid homology with either serovar $L_2$ or serovar B and displayed 30 of 33 base changes. Like the coding sequence of variable segment 1, the coding sequence for the string of amino acids at positions 307 to 318 in variable segment 4 had one additional codon, and depending on where this was aligned, this string shared a maximum of two homologous amino acids. In contrast, the changes that comprised variable segment 3 consisted of single base changes within codons.

The contrast between point mutations observed for variable segment 3 and the virtually complete changes within variable segments 1, 2, and 4 suggests that the latter resulted from dramatic molecular events such as recombination. Whether these changes are immunologically driven and the roles such changes play in pathogenesis remain obscure, but like other parasite systems that utilize antigenic variation for their survival in the mammalian host, these changes may be related to type-specific domains that function in a similar fashion. Indeed, the trachoma-causing strains display marked antigenic diversity, and infections by these strains are usually limited to mucosal surfaces where they are exposed to extracellular host immune mechanisms when they must reinfect new host cells during each growth cycle. In contrast, the lymphogranuloma venereum biovar consists of only three serovars; these are invasive, readily capable of cell-to-cell transmission, and causative of primarily systemic infections, and hence they can avoid constant exposure to extracellular host defenses. Teleologically, because of the different host defense milieus, type-specific changes may provide a selective advantage at the level of the microbial population in the niche that trachoma strains have chosen. An understanding of these components may now be tested in a more direct fashion by mapping of antigenic domains with monoclonal antibodies and testing these domains for immunological and biological activity.

Type-specific determinants have been shown to be important for induction of immunity both in animal models and in human vaccine trials (21). Inspection of the amino acid sequences demonstrated that there were four candidates represented by each of the sequence-variable domains that may account for type-specific antigenicity. Preliminary calculations of MOMP secondary structure predict β-turn potential for each of these segments, which can be a marker for antigenic segments (26). Nevertheless, of the four, variable segment 2 in the amino-terminal half of MOMP and variable segment 4 in the carboxyl-terminal half of MOMP were the most promising candidates for type-specific antigens, since each of these segments displayed hydrophilic singularities for each serovar. Variable segment 2 had the highest degree of turn and loop potential, and this potential could be stabilized by the conserved cysteine residues 117 and 185 that flank this segment. A 67-amino-acid disulfide loop that may stabilize serovar-specific determinants is reminiscent of structures observed in other systems such as gonococcal pilus antigen (19). Expression of these unique segments and assessment of their binding to monoclonal antibodies, as well as the specificity of antibody response elicited by immunization, should define these characteristics. Indeed, preliminary evaluation with type-specific and species-specific monoclonal antibody probes located these determinants to amino- and carboxyl-terminal domains respectively.

The MOMP is not only a complex antigen and potent immunogen but also a structural protein that interacts with other outer membrane proteins through disulfide bonds which may provide structural integrity to an organism that lacks peptidoglycan (1). It is also a porin protein whose activity is uniquely mediated by disulfide bonds (2). The *omp*1 sequence comparisons between serovars have delimited sequence-variable domains and conserved domains. These sequence data should help consolidate the functional and structural concepts proposed for MOMP and provide a basis for modeling the significant secondary, tertiary, and quaternary interactions of this protein.

## ACKNOWLEDGMENTS

We thank Tack Kuntz, Department of Molecular Structure, University of California, San Francisco Medical Center, for his advice and calculations of secondary structure.

This research was supported by the Edna McConnell Clark Foundation, by the MacArthur Foundation, and by funds provided through Office of Naval Research contract N00014-81-C-0570. E.A.W. was supported by Sexually Transmitted Diseases Training Program grant AI07234.

## LITERATURE CITED

1. **Barbour, A. G., K. I. Amano, T. Hackstadt, L. Perry, and H. D. Caldwell.** 1982. *Chlamydia trachomatis* has penicillin-binding proteins but not detectable muramic acid. J. Bacteriol. **151:** 420–428.
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., and R. C. Judd.** 1982. Structural analysis of chlamydial major outer membrane proteins. Infect. Immun. **38:**960–968.
4. **Caldwell, H. D., and J. Schachter.** 1982. Antigenic analysis of the major outer membrane protein of *Chlamydia* spp. Infect. Immun. **35:**1024–1031.
5. **Chang, J.-J., K. Leonard, T. Arad, T. Pitt, Y.-X. Zhang, and L.-H. Zhang.** 1982. Structural studies of the outer envelope of *Chlamydia trachomatis* by electron microscopy. J. Mol. Biol. **161:**579–590.
6. **Chen, R., C. Kramer, W. Schmidmayr, and U. Henning.** 1979. Primary structure of major outer membrane protein I of *Escherichia coli* B/r. Proc. Natl. Acad. Sci. USA **76:**5014–5017.
7. **Grayston, J. T., and S.-P. Wang.** 1975. New knowledge of chlamydiae and the diseases they cause. J. Infect. Dis. **132:** 87–105.
8. **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.
9. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78:**3824–3828.
10. **Karn, J., S. Brenner, L. Barnett, and G. Cesareni.** 1980. Novel bacteriophage λ cloning vector. Proc. Natl. Acad. Sci. USA **77:**5172–5176.
11. **Kuo, C.-C., S.-P. Wang, and J. T. Grayston.** 1977. Growth of trachoma organisms in HeLa 229 cell culture, p. 328–336. *In* D. Hobson and K. K. Holmes (ed.), Nongonococcal urethritis and related infections. American Society for Microbiology, Washington. D.C.
12. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101:**20–78.
14. **Nano, F. E., P. A. Barstad, L. W. Mayer, J. E. Coligan, and H. D. Caldwell.** 1985. Partial amino acid sequence and molecular cloning of the encoding gene for the major outer membrane

protein of *Chlamydia trachomatis*. Infect. Immun. **48**:372–377.

15. **Newhall, W. J., V, B. Batteiger, and R. B. Jones.** 1982. Analysis of the human serological response to proteins of *Chlamydia trachomatis*. Infect. Immun. **38**:1181–1189.
16. **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.
17. **Osborn, M. J., and H. C. P. Wu.** 1980. Proteins of the outer membrane of gram-negative bacteria. Annu. Rev. Microbiol. **34**:369–422.
18. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391–409.
19. **Rothbard, J. B., R. Fernandez, and G. K. Schoolnik.** 1984. Strain specific and common epitopes of gonococcal pili. J. Exp. Med. **160**:208–221.
20. **Sanger, F. S., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
21. **Schachter, J., and H. D. Caldwell.** 1980. Chlamydiae. Annu. Rev. Microbiol. **34**:285–309.
22. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
23. **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.
24. **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.
25. **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.
26. **Tainer, J. A., E. D. Getzoff, H. Alexander, R. A. Houghten, A. J. Olson, R. A. Lerner, and W. A. Hendrickson.** 1984. The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. Nature (London) **312**:127–134.
27. **Terho, P., M. T. Matikainen, P. Arstila, and J. Treharne.** 1982. Monoclonal typespecific antibodies for *Chlamydia trachomatis*/LGV strains, p. 321–324. *In* P.-A. Mardh, K. K. Holmes, J. D. Oriel, P. Piot, and J. Schachter (ed.), Chlamydial infections. Elsevier Biomedical Press, Amsterdam.
28. **Wang, S.-P., and J. T. Grayston.** 1971. Classification of TRIC and related strains with micro-immunofluorescence, p. 305–321. *In* R. L. Nichols (ed.), Trachoma and related disorders caused by chlamydial agents. Excerpta Medica, Amsterdam.
29. **Wang, S.-P., C.-C. Kuo, R. C. Barnes, R. S. Stephens, and J. T. Grayston.** 1985. Immunotyping of *Chlamydia trachomatis* with monoclonal antibodies. J. Infect. Dis. **152**:791–800.
30. **Warner, B. D., M. E. Warner, G. A. Karns, L. Ku, S. Brown-Shimer, and M. Urdea.** 1984. Construction and evaluation of an instrument for the automated synthesis of oligonucleotides. DNA **3**:401–411.