INFECTION AND IMMUNITY, Apr. 1989, p. 1040–1049
0019-9567/89/041040-10$02.00/0

Vol. 57, No. 4

Nucleotide and Deduced Amino Acid Sequences for the Four Variable Domains of the Major Outer Membrane Proteins of the 15 *Chlamydia trachomatis* Serovars

YING YUAN,† YOU-XUN ZHANG,† NANCY G. WATKINS, AND HARLAN D. CALDWELL*

Laboratory of Microbial Structure and Function, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, Hamilton, Montana 59840

Received 17 October 1988/Accepted 8 November 1988

The amino acid sequences of major outer membrane proteins (MOMPs) from *Chlamydia trachomatis* serovars A, B, C, L1, and L2 are predominantly conserved but have four variable domains (VDs) in which major neutralizing and serotyping antigenic determinants are located. Because these MOMP VDs are primarily responsible for antigenic differences between serovars and are associated with important immunological and biological properties, we undertook studies focused on defining these sequences within the MOMPs of all 15 *C. trachomatis* serovars. We used oligonucleotide primer extension sequencing of MOMP mRNA to determine the nucleotide and deduced amino acid sequences of the four MOMP VDs of the 15 *C. trachomatis* serovars. Comparative amino acid sequence homologies of all four domains separated the serovars into three groups: group 1, serovars B, Ba, D, E, L1, and L2; group 2, serovars G and F; and group 3, serovars A, C, H, I, J, K, and L3. Hydrophilicity and charge values for each domain were determined. The MOMP VDs of given serovars with the greatest total hydrophilicity and charge values were found to be the location of antigenic determinants recognized by MOMP-specific monoclonal antibodies. These findings should be useful for predicting MOMP antigenic determinants and testing the antigenic properties of these VDs by using synthetic peptides corresponding to each MOMP VD. The potential usefulness of the VD sequence information is discussed in relation to the development of defined synthetic peptides and oligonucleotides that may be used to develop new serological and diagnostic assays for *C. trachomatis* infections.

Chlamydia trachomatis isolates occur as 15 distinct serovars. On the basis of serological relatedness, these 15 serovars are divided into three serogroups: B (serovars B, Ba, D, E, L1, and L2), intermediate (serovars F, G, K, and L3), and C (serovars A, C, H, I, and J). The antigen that confers serovar and serogroup specificities to chlamydiae is the major outer membrane protein (MOMP) (2, 4, 15). Protective immunity developed during chlamydial ocular infection is thought to be directed at serovar and serogroup MOMP determinants (5). These determinants are therefore considered to be potential target antigens for the development of a subunit or recombinant *C. trachomatis* vaccine (20).

The genes encoding the MOMPs of *C. trachomatis* serovars A, B, C, L1, and L2 have recently been cloned and sequenced (1, 11, 13, 14). Comparative analysis of their amino acid sequences showed that the MOMP genes encode highly conserved protein structures that contain four evenly spaced domains whose sequences vary among the different serovars. The locations of these variable domains (VDs) in the MOMP amino acid sequences are: VDI, residues 64 to 83; VDII, residues 139 to 160; VDIII, residues 224 to 237; and VDIV, residues 288 to 317. Epitope mapping has shown that three domains (VDI,-II, and -IV) contain contiguous antigenic determinants that elicit the formation of either serovar, subspecies- (MOMP determinants common to three or more serovars within a serogroup), serogroup-, or species-specific antibodies (1, 16). VDI and VDII, which show the greatest amount of interserogroup sequence variation, are the locations of serovar-specific determinants. VDIV, the largest of the domains, is located near the C terminus of the protein and is the location of subspecies-, serogroup-, and a highly conserved species-specific antigenic determinant. Antigenic determinants have not been mapped to VDIII, the smallest and least variable domain of MOMP genes.

VDI, -II, and -IV protrude toward the external environment at the chlamydial surface, as shown by their susceptibility to cleavage by trypsin (17) and accessibility to antibody binding (20). Tryptic cleavage in both VDII and -IV, but not within VDIV alone, decreases chlamydial attachment to HeLa cells, which suggests that these domains, or conformational MOMP structures that are dependent on the integrity of these domains, may function as a chlamydial ligand (17). Because the structures of MOMP VDs have important immunological and biological functions, we undertook to determine the sequences of the four VDs for all 15 *C. trachomatis* serovars.

MATERIALS AND METHODS

Chlamydiae. The following *C. trachomatis* serovars were used: A/Har-13, B/TW-5/OT, Ba/AP-2, C/TW-3/OT, D/UW-3/Cx, E/Bour, F/IC-Cal-13, G/UW-57/Cx, H/UW-4/Cx, I/UW-12/Ur, J/UW-36/Cx, K/UW-31/Cx, LGV/L1-440, LGV/L2-434, and LGV/L3-404. Serovars A, E, and F were obtained from the American Type Culture Collection, Rockville, Md. All other serovars were obtained from Cho-Chou Kuo, Department of Pathobiology, University of Washington, Seattle. All serovars used were immunotyped by the microimmunofluorescence procedure (18) in the laboratory of San-Pin Wang, University of Washington.

RNA extraction. Approximately 2×10^8 HeLa 229 cells

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

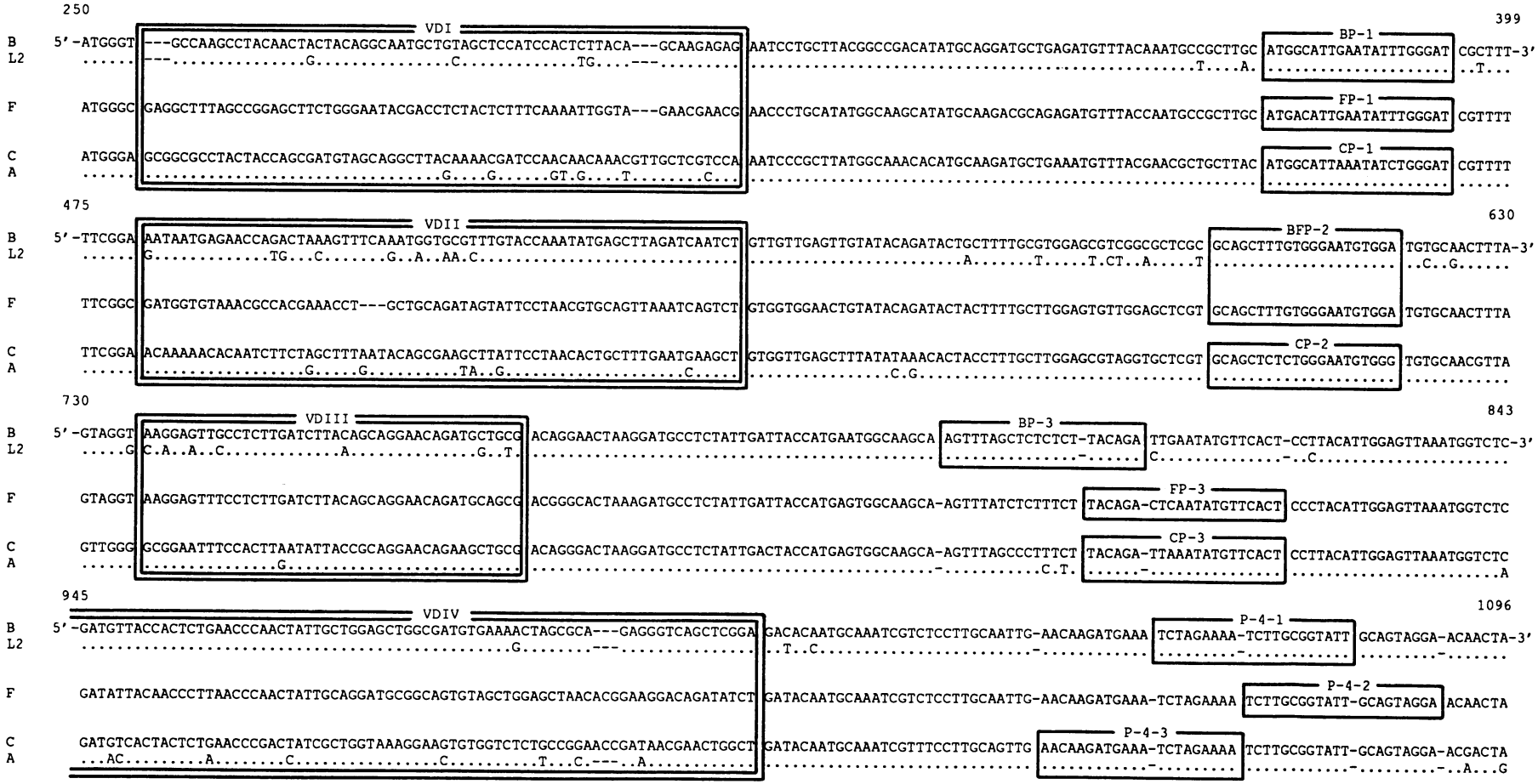


FIG. 1. Locations of MOMP gene sequences to which synthetic complementary oligonucleotides were constructed and used for primer extension mRNA sequencing of MOMP VDs. The MOMP gene sequences identified by single-lined boxes are those to which synthetic oligonucleotides were constructed. Double-lined boxes identify the four VD regions of B, L2, A, C, and F MOMP genes. Only regions of the MOMP gene sequences relevant to this study are shown.

TABLE 1. Oligonucleotides used in primer extension mRNA sequencing of *C. trachomatis* MOMP VDs

Oligonucleotide	Oligonucleotide sequence	Serovars and their VDs sequenced with oligonucleotide primers
BP-1	5'-ATCCCAAATATTCAATGCCAT-3'	VDI of B, Ba, D, E, G, L1, and L2
CP-1	ATCCCAGATATTTAATGCCAT	VDI of A, C, H, I, J, K, and L3
FP-1	ATCCCAAATATTCAATGTCAT	VDI of F and G
BFP-2	TCCACATTCCCACAAAGCTGC	VDII of B, Ba, D, E, F, G, L1, and L2
CP-2	CCCACATTCCCAGAGAGCTGC	VDII of A, C, H, I, J, K, and L3
BP-3	TCTGTAAGAGAGAGCTAAACT	VDIII of B, Ba, D, E, L1, and L2
CP-3	AGTGAACATATTTAATCTGTA	VDIII of A, C, H, I, J, K, and L3
FP-3	AGTGAACATATTGAGTCTGTA	VDIII of F and G
P-4-1	AATACCGCAAGATTTTCTAGA	VDIV of A, B, Ba, C, D, E, F, H, I, J, K, L1, and L2
P-4-2	TCCTACTGCAATACCGCAAGA	VDIV of L3 and G[a]
P-4-3	TTTTCTAGATTTCATCTTGTT	VDIV of J[a]

[a] Not done for other serovars.

grown in stationary culture were infected with each chlamydial serovar. Infected HeLa cells were removed from culture flasks with trypsin at 24 h postinfection for serovars L1, L2, and L3 and at 36 h postinfection for all other serovars. Trypsinized infected cell suspensions were pooled, centrifuged, and washed twice in cold (4°C) 0.02 M sodium phosphate–0.15 M NaCl, pH 7.2. Cells were lysed in 5 ml of cold 4 M guanidine isothiocyanate lysis buffer (3). The suspension was gently refluxed through an 18-gauge needle 15 to 20 times to shear the DNA. Total RNA was extracted twice with hot acidic phenol (65°C, pH 5.0), followed by two extractions with chloroform-isoamyl alcohol (24:1, vol/vol). The RNAs were precipitated with 2 vol of cold ethanol (−20°C), washed once with cold 75% ethanol, and evaporated to dryness in a Speed-Vac (Bachofer, Reutlingen, Federal Republic of Germany). The RNAs were suspended in glass-distilled water at a concentration of 20 to 40 mg/ml and stored at −70°C.

Synthetic oligonucleotides. Oligonucleotides (21-mers) complementary to the plus-strand MOMP DNAs flanking the 3' end of each VD were synthesized by using a SAM1 automated synthesis instrument (Biosearch, San Rafael, Calif.) by the *N*-methylimidazole phosphotriester method (6) according to the directions of the manufacturer. Deblocked oligonucleotides were evaporated to dryness in a Speed-Vac. They were then dissolved in distilled water at a concentration of 10 to 30 mg/ml and stored at −20°C. Oligonucleotides were constructed according to the sequences of A, B, C, L2 (1, 13, 16), and F (Y.-X. Zhang, unpublished data) MOMP genes. The MOMP gene sequences relevant to this work and the locations of the sequences in the MOMP genes to which synthetic oligonucleotides were constructed are shown in Fig. 1. Each oligonucleotide number, oligonucleotide sequence, and MOMP serovar and VD for which the oligonucleotide was used for primer extension mRNA sequencing are summarized in Table 1.

Primer extension and DNA sequencing. The dideoxynucleotide chain termination method of Sanger et al. (12) modified for the use of reverse transcriptase and RNA templates as described by Lane et al. (9) was used for sequence determinations. Oligonucleotides were 5'-end labeled with[γ-^{32}P] dATP (3,000 Ci/mmol; Dupont, NEN Research Products, Boston, Mass.), using T4 polynucleotide kinase (10,000 U/ml; New England BioLabs, Beverly, Mass.) as described by Maniatis et al. (10). Labeled oligonucleotides were isolated on Sephadex G-25 columns (Boehringer Mannheim Biochemicals, Indianapolis, Ind.). Oligonucleotide primers were hybridized to MOMP mRNA templates in 12.5 μl of reaction mixture containing 2.5 μl of 5× annealing buffer (250 mM Tris hydrochloride, 300 mM NaCl, 50 mM dithiothreitol, 5 mM EDTA, pH 8.3), 7.5 μl of total RNA (3 to 10 μg/μl), and 2.5 μl of kinased primer (2 to 2.5 μM). The mixtures were heated at 90°C for 3 min, centrifuged, and then allowed to cool slowly over 30 min to 45°C. A 2-μl volume of the hybridized mixture was then added to each of four tubes containing 3 μl of reaction mixture (250 mM Tris hydrochloride [pH 8.3], 300 mM NaCl, 50 mM dithiothreitol, 30 mM $MgCl_2$, 2 mM deoxynucleotides 0.4 mM dideoxynucleotide triphosphates, 0.6 to 0.8 U of avian myeloblastosis virus reverse transcriptase per μl [20,000 U/ml; Pharmacia Fine Chemicals, Piscataway, N.J.]). The mixtures were incubated at 45°C for 40 min. The reactions were stopped by evaporation to dryness in a Speed-Vac. Samples were suspended in 3 μl of 100% formamide containing 0.8% xylene cyanol and 0.8% bromophenol blue and heated at 95°C for 3 min immediately before the sequencing gels were loaded. Samples were separated on 8% polyacrylamide gels. The gels were fixed with methanol, dried, and then subjected to autoradiography.

RESULTS

Primer extension of MOMP mRNA and DNA sequencing of VDs. With oligonucleotide primers constructed to the desired MOMP gene sequence (Fig. 1), the VD-encoding portions of MOMP mRNA were easily sequenced. Figure 2 shows a representative sequence gel for MOMP VDII of serovars B and A when oligonucleotides BFP-2 and CP-2 were used as primers. The mRNA-derived MOMP sequences are identical to previously published genomic DNA sequences for MOMP VDII of these two serovars (Fig. 1), thereby demonstrating both the feasibility and accuracy of this sequencing method.

Sequence analysis and characteristics of MOMP VDs. The synthetic oligonucleotides used for sequencing the MOMP VDs of all 15 *C. trachomatis* serovars are described in Table 1. Locations of the complementary sequences in the MOMP genes of serovars B, L2, A, C, and F are shown in Fig. 1. All oligonucleotides were 21-mers and were complementary to conserved nucleotide sequences 40 to 60 nucleotides downstream (3') from the VD to be sequenced. Interserogroup but not intraserogroup nucleotide variation occurs within these regions; therefore, oligonucleotides complementary to serogroup-specific MOMP sequences were constructed and used to sequence VDs of serovars within a given serogroup. The

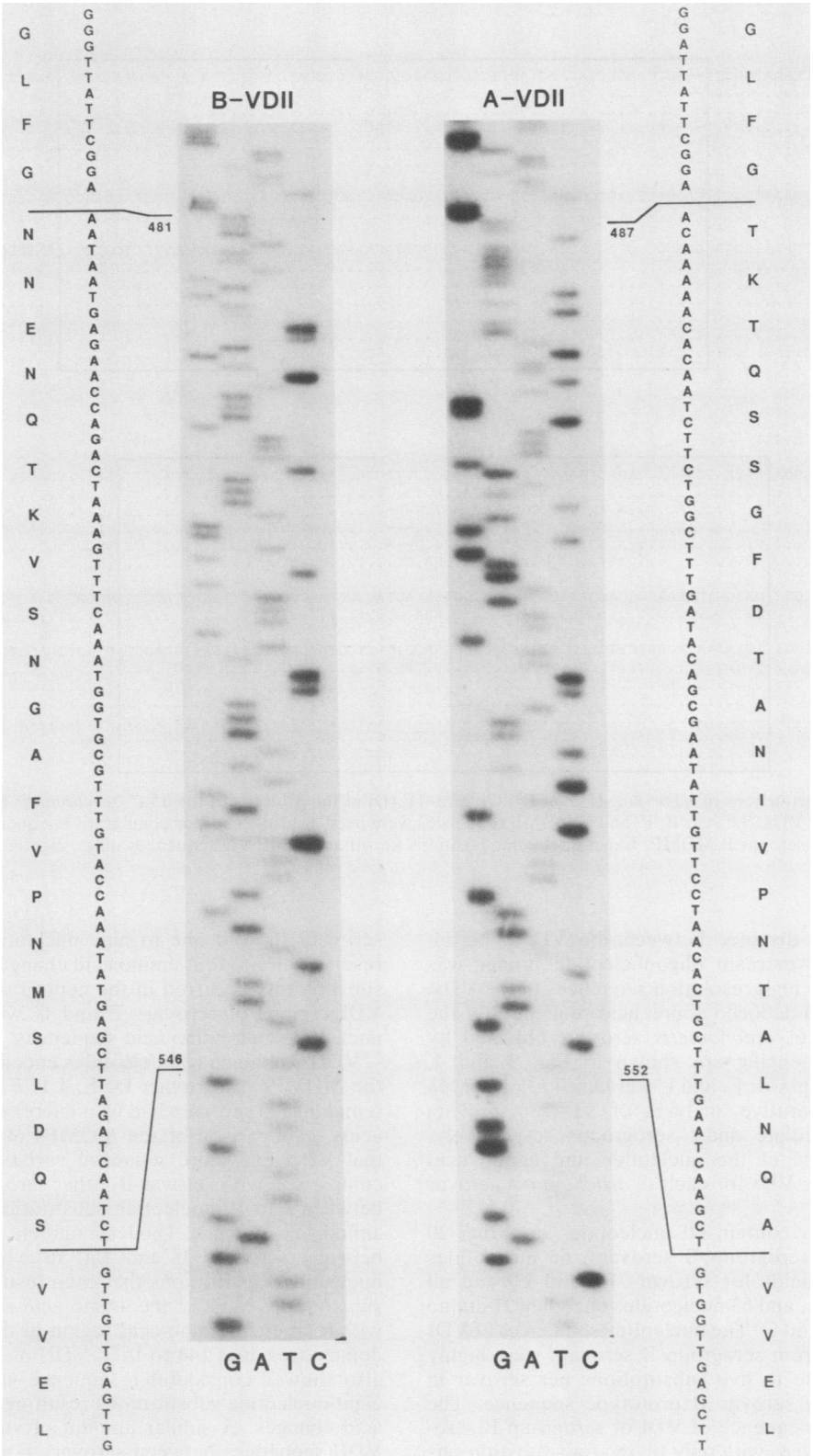


FIG. 2. Sequencing of serovar B and A MOMP VDII by primer extension of MOMP mRNA. Total RNA was isolated from HeLa 229 cells infected with serovars B and A and annealed to kinased oligonucleotides BFP-2 and CP-2 (Table 1), respectively. The oligonucleotides were identified to MOMP gene sequences that are 50 nucleotides downstream 3′ of VDII. The RNA templates were transcribed with reverse transcriptase, and the sequences of the transcribed DNAs were determined by the dideoxynucleotide chain termination method. Audioradiographs were read to depict plus-strand MOMP DNA sequences. Nucleotide numbers shown for each MOMP are the beginning and end of the sequence encoding VDII of each protein. Results are representative of sequencing results obtained for VDs of all other serovars.

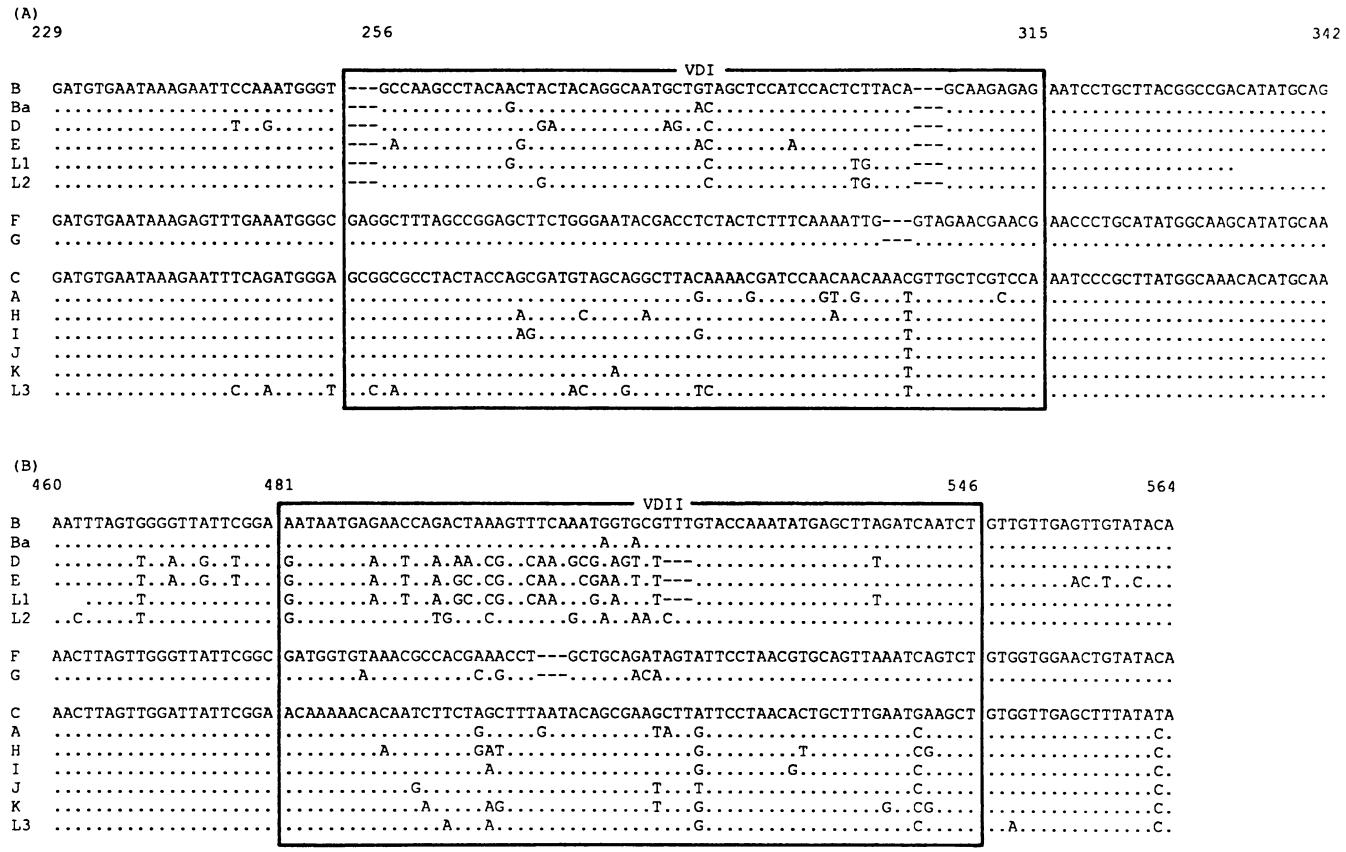


FIG. 3. Nucleotide sequences of VDI (A), -II (B), -III (C), and -IV (D) of the MOMPs of the 15 *C. trachomatis* serovars. Boxed regions identify each of the four VDs. Serovar B, F, and C MOMP sequences were used as prototypes for comparative sequence analysis. Nucleotide positions shown are for serovar B MOMP. Exact nucleotide positions for other MOMP VD sequences differ slightly because of insertions or deletions within VDs.

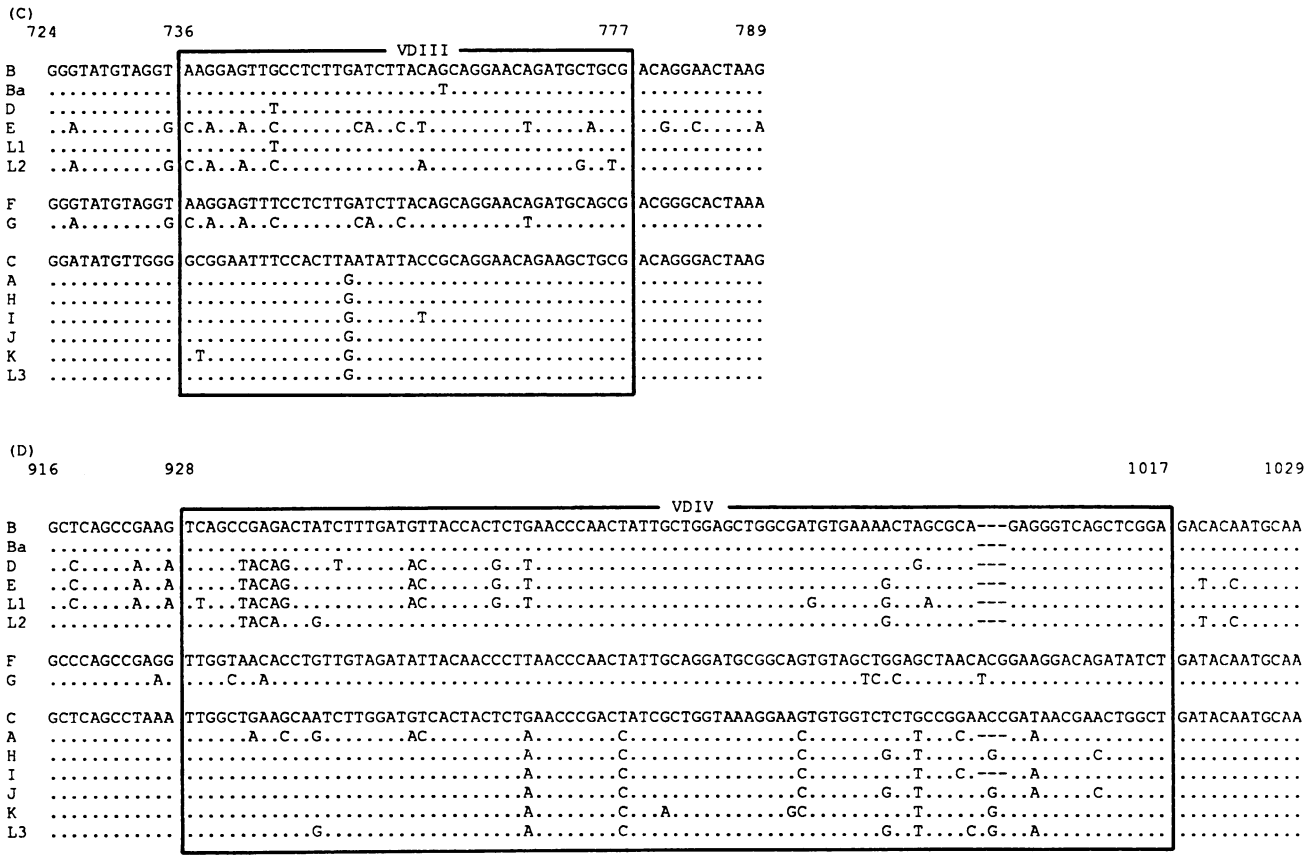
40- to 60-nucleotide distance between the VD to be sequenced and its downstream oligonucleotide primer was necessary to produce high-resolution sequences of the VDs.

The nucleotide and deduced amino acid sequences of the four VDs of all 15 *C. trachomatis* serovars obtained by MOMP mRNA sequencing are shown in Fig. 3 and 4. MOMP VDs of serovars B, F, and C were used as prototype sequences for comparative analyses of VDs of serovars within the B, intermediate, and C serogroups, respectively. Comparative analyses of the nucleotide and amino acid sequences of the four VDs for each *C. trachomatis* serovar are presented below.

VDI was found to contain 60 nucleotides encoding 20 amino acids for all serogroup B serovars, 66 nucleotides encoding 22 amino acids for serovars K, and L3 and all serogroup C serovars, and 63 nucleotides encoding 21 amino acids for serovars F and G. The nucleotide sequences of VDI regions of MOMPs from serogroup B serovars were highly conserved, with three to five substitutions per serovar in comparison with the serovar B prototype sequence. The deduced amino acid sequence of VDI of serogroup B serovars was also conserved, with only two to four substitutions per serovar; most of the substitutions occur within a 12-amino-acid region (residues 68 to 79) that resides in the central portion of the domain. In contrast, serovars in serogroup C and serovars K and L3 showed greater nucleotide and amino acid variation within VDI. In comparison with results for VDI of serovar C, the other serogroup C serovars showed one to nine nucleotide substitutions that result in zero to four amino acid changes. Again, most of the substitutions occurred in the central region of the domain. VDI regions of serovars F and G were identical in their nucleotide and amino acid sequences.

VDII contained 63 nucleotides encoding 21 amino acids in the MOMPs of serovars D, E, L1, F, and G. VDII of the remaining 10 serovars had 66 nucleotides encoding 22 amino acids. Comparison of the MOMPs of all serovars showed that VDII had more sequence variation than did VDI. In comparison with serovar B, other serogroup B serovars had between 2 to 19 nucleotide substitutions resulting in 2 to 8 amino acid changes. The least amount of variation occurred between serovars B and Ba, in which there were two nucleotide substitutions that each resulted in an amino acid substitution. Most of the amino acid substitutions occurred within an eight-amino-acid region in the central part of the domain (residues 144 to 151). VDII of serogroup C serovars also showed considerable sequence variation, with four to eight nucleotide substitutions resulting in three to six amino acid changes. A similar amount of variation was found in VDII sequences between serovars F and G of the intermediate serogroup. In comparison with serovar F VDII sequences, the sequence of serovar G VDII had 6 nucleotide substitutions resulting in 3 amino acid changes.

VDIII was the smallest and least variable domain among the serovars. It contained 42 nucleotides encoding 14 amino acids. The VDIII sequences of serovars A, H, J, and L3



were identical to each other. Serovars D, L1, and F had sequences identical to each other, differing from that of serovar B by three amino acid substitutions. The remaining serovars had variation in VDIII that does not group serovars into serogroups.

VDIV was the largest domain, containing 96 nucleotides encoding 32 amino acids in serovars A and I and all serogroup B serovars and 99 nucleotides encoding 33 amino acids for the remaining serovars. This domain could be separated into three distinct regions on the basis of sequence homology: (i) the N-terminal region (residues 288 to 295), (ii) the central region (residues 296 to 306), and (iii) the C-terminal region (residues 307 to 317). Within serogroup B, serovars B and Ba had identical amino acid sequences in the N-terminal regions, as did serovars D and E. There were three amino acid substitutions between serovars B and Ba and serovars D and E. Serovars L1 and L2 contained one and two, respectively, additional amino acid substitutions in this region. Within serogroup C, serovars C, H, I, J, and K had identical sequences in this region of VDIV. Serovar L3 contained a single amino acid substitution. Serovar A had the most variable sequence in this region, containing four sequence substitutions in comparison with the prototype serovar C sequence. Serovars F and G differed by two amino acids within this region.

The central region of VDIV was the most highly conserved sequence among the VDs of the chlamydial serovars. A nine-amino-acid sequence (TTLNPTIAG) was found in all serovars except K, which contained a threonine instead of an alanine at residue 303. The central sequence (Asn-Pro-Thr) in VDIV was present in all 15 serovars.

The C-terminal region of VDIV was highly conserved among the serogroup B serovars. In contrast, this region of VDIV in C-complex serovars showed a significant amount of amino acid variation, although the amino acid compositions were similar, being predominately S, T, E, and A. Serovars F and G differed by three amino acid residues in this region.

Inter- and intraserogroup amino acid homologies of MOMP VDs. The percentage of amino acid homology and the total number of amino acid substitutions for the VDs of each serovar are shown in Table 2. The *C. trachomatis* serovars could be separated into three groups on the basis of the amino acid homologies of their VDs. Group 1 consisted of serovars B, Ba, D, E, L1, and L2; group 2 consisted of serovars F and G; and group 3 consisted of serovars A, C, H, I, J, K, and L3. Intergroup VD amino acid homology was between 33 and 46%. Intragroup VD amino acid homology was between 74 and 94% for groups 1 and 3 and 85% for group 2. Therefore, classification of *C. trachomatis* isolates on the basis of MOMP VD amino acid homology correlates extremely well with immunotyping classification using MOMP-specific monoclonal antibodies (19). The exception is that immunotyping with monoclonal antibodies places serovars K and L3 within the intermediate serogroup (serovars F and G), whereas VD amino acid homology shows that serovars K and L3 are more closely related to serovars within serogroup C.

Hydrophilicity and charge values of MOMP VDs. Major antigenic sites that elicit the formation of protective serotyping antibodies have been mapped to VDI, VDII, and VDIV of the MOMPs of serovars A, B, C, and L2 (1). However, the antigenic properties of the VDs of other serovars have

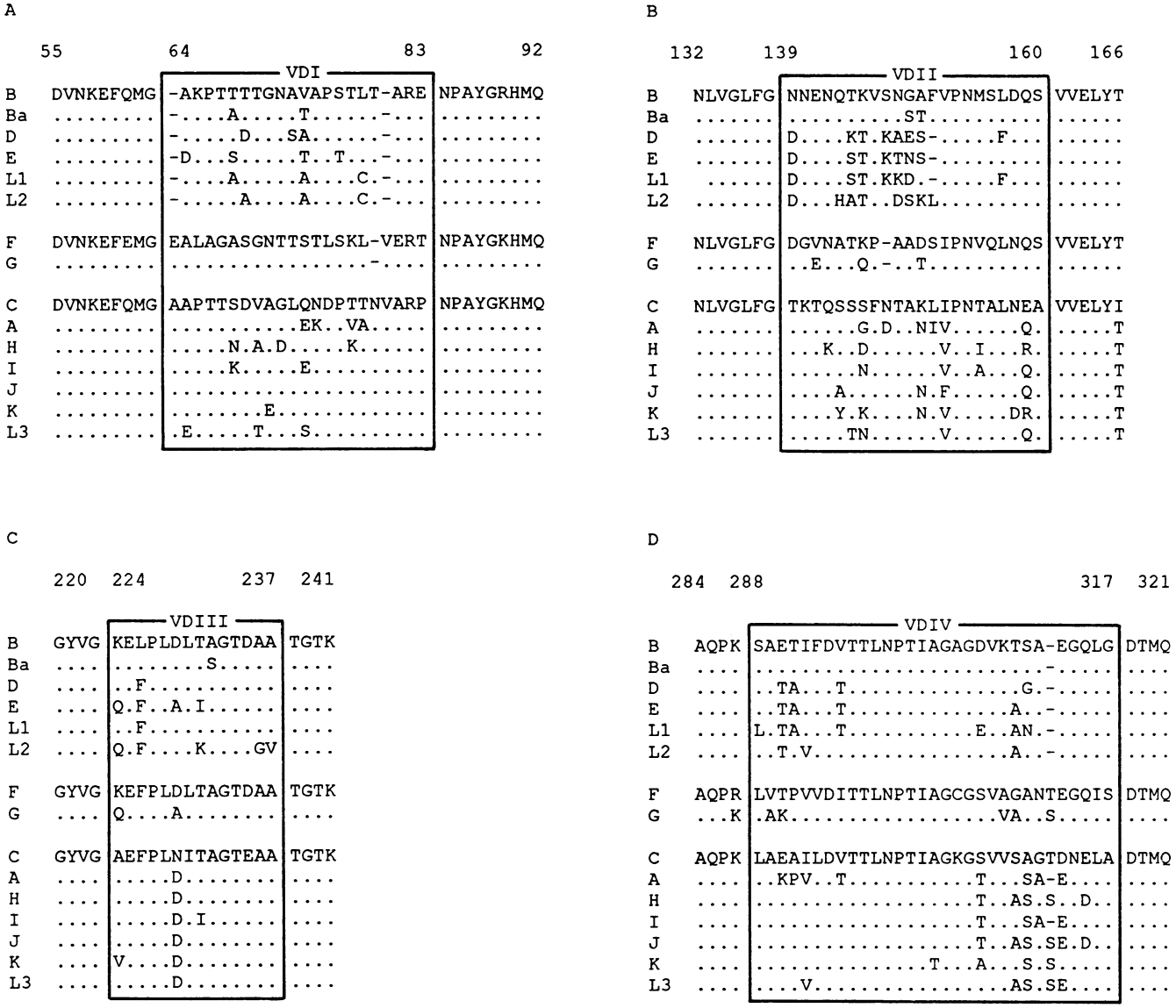


FIG. 4. Deduced amino acid sequences of VDI (A), -II (B), -III (C), and -IV (D) of the MOMPs of the 15 *C. trachomatis* serovars. Boxed regions identify each of the four VDs. Serovar B, F, and C MOMP amino acid sequences were used as prototypes for comparative sequence analysis.

not been described. We are interested in defining the antigenic properties of these domains, since we believe that this is necessary to the logical design of a chlamydial subunit vaccine. Because the immunogenicity of protein determinants is largely influenced by both hydrophilicity and charge (8), we analyzed the MOMP VDs of all chlamydial serovars for these properties to select sequences to be synthesized and tested as immunogens to define the antigenic characteristics of individual VDs. The number of charged amino acids and the hydrophilicity values (8) of sequences within each MOMP VD for all 15 serovars are shown in Table 3. The VDs had a broad range of both hydrophilicity and charge values. VDI of serovars in serogroup B had an average hydrophilicity value of 4.47 and contained an average of 2 positive charges and 1.3 negative charges. VDII of this group had an average hydrophilicity value of 6.95 and contained an average of 1.35 positive charges and 3.2 negative charges. VDII of serovars within serogroup B would be predicated to have the greater immunogenic potential. Indeed, epitope-mapping studies have located immunodominant serovar-specific epitopes in VDII for both serovar B and serovar L2 (Table 4). In contrast, the VDI regions of the MOMP serogroup C serovars were more hydrophilic and charged than were the VDII regions. The average hydrophilicity for VDI was 5.3, with 1.4 positive and 2.7 negative charges, whereas the hydrophilicity value for VDII was 2.7, with 2 positive and 0.7 negative charges. These data suggest that VDI of C-complex serovars has a greater immunogenic potential than does VDII. Interestingly, the epitope recognized by monoclonal antibody A-20 (1), which is specific for serovar A MOMP, was mapped to VDI (Table 4). Similar to results for serogroup C, the MOMP VDI of the intermediate serovars F and G was more hydrophilic and charged than was VDII. The MOMPs of several serovars (D, E, and H)

TABLE 2. Comparative amino acid homologies of MOMP VDs of the 15 *C. trachomatis* serovars[a]

	B	Ba	D	E	L1	L2	F	G	C	A	H	I	J	K	L3
B	100[b]	94.2	80.2	76.7	77.9	77.9	43.7	42.5	38.2	40.9	40.4	43.2	42.7	42.7	41.6
	(0)[c]	(5)	(17)	(20)	(19)	(19)	(49)	(50)	(55)	(52)	(53)	(50)	(51)	(51)	(52)
Ba		100	77.9	76.7	76.7	76.7	43.7	43.7	37.1	39.8	39.3	42	41.6	41.6	40.4
		(0)	(19)	(20)	(20)	(20)	(49)	(49)	(56)	(53)	(54)	(51)	(52)	(52)	(53)
D			100	82.4	85.9	76.7	44.8	43.7	38.2	42.7	39.3	42	41.6	40.4	41.6
			(0)	(15)	(12)	(20)	(48)	(49)	(55)	(51)	(54)	(51)	(52)	(53)	(52)
E				100	82.4	76	39.1	43.7	40.4	43.2	40.4	42	43.8	41.6	43.8
				(0)	(15)	(21)	(53)	(49)	(53)	(50)	(53)	(51)	(50)	(52)	(50)
L1					100	77.3	44.8	44.8	40.4	43.2	42.7	44.3	46.1	41.6	42.7
					(0)	(20)	(48)	(48)	(53)	(50)	(51)	(49)	(48)	(52)	(51)
L2						100	40.2	42.5	36	39.8	39.3	42	40.4	38.2	43.8
						(0)	(52)	(50)	(57)	(53)	(54)	(51)	(53)	(55)	(50)
F							100	85.1	34.8	37.5	36	34.1	37.1	32.6	39.3
							(0)	(13)	(58)	(55)	(57)	(58)	(56)	(60)	(54)
G								100	39.3	40.9	38.2	37.5	40.4	36	43.8
								(0)	(54)	(52)	(55)	(55)	(53)	(57)	(50)
C									100	77.5	83.1	86.4	88.8	85.4	85.4
									(0)	(20)	(15)	(12)	(10)	(13)	(13)
A										100	74.2	81.8	84.3	75.3	79.8
										(0)	(23)	(16)	(14)	(22)	(18)
H											100	83	86.5	83.1	83.1
											(0)	(15)	(12)	(15)	(15)
I												100	87.6	81.8	86.4
												(0)	(11)	(16)	(12)
J													100	86.5	87.6
													(0)	(12)	(11)
K														100	80.9
														(0)	(17)
L3															100
															(0)

[a] Amino acid sequences of all four VDs were used for calculating homologies.
[b] Percent homology.
[c] Number of amino acid substitutions.

TABLE 3. Hydrophilicity values and charges of *C. trachomatis* MOMP VDs

Serovar	VDI			VDII			VDIII			VDIV								
										N′ end			Middle			C′ end		
	Basic[a]	Acidic[b]	HV[c]	Basic	Acidic	HV	Basic	Acidic	HV	Basic	Acidic	HV	Basic	Acidic	HV	Basic	Acidic	HV
B	2	1	+1.8	1	2	+1.8	1	3	+4.3	1	2	2.6	0	0	−5.1	1	3	+7.6
Ba	2	1	+2.8	1	2	+2.2	1	3	+5.1	1	2	+2.6	0	0	−5.1	1	3	+7.6
D	2	2	+8.8	2	4	+9.7	1	3	+3.6	1	1	+0.2	0	0	−5.1	1	3	+7.3
E	2	2	+6.4	1	3	+7.8	0	2	−3.8	1	1	+0.2	0	0	−5.1	1	3	+7.5
L1	2	1	+3.5	2	4	+12.5	1	3	+3.6	1	1	−1.9	0	0	−5.1	1	3	+7.9
L2	2	1	+3.5	1	4	+7.7	1	3	+3.7	1	1	−0.5	0	0	−5.1	1	3	+7.5
Avg[d]	2	1.3	+4.47	1.35	3.2	+6.95	0.8	2.8	+2.75	1	1.3	+0.53	0	0	−5.1	1	3	+7.57
F	2	2	+3.5	1	2	+2.1	1	3	+3.6	1	2	−2.5	0	0	−5.1	0	2	+1.3
G	2	2	+3.5	0	2	0	0	2	−2.8	2	1	+1.9	0	0	−5.1	0	2	+0.6
Avg[e]	2	2	+3.5	0.5	2	+1.05	0.5	2.5	+4.4	1.5	1.5	−0.3	0	0	−5.1	0	2	+0.95
C	1	2	+1.5	2	1	−0.3	0	2	−2.7	1	2	+1.1	0	0	−5.1	1	3	+6.6
A	2	3	+5.9	1	1	−3.1	0	3	+0.1	2	1	+3.0	0	0	−5.1	1	3	+6.6
H	2	3	+8.8	4	1	+5.7	0	3	+0.1	1	2	+1.1	0	0	−5.1	1	3	+6.6
I	2	3	+7.0	2	0	−0.1	0	3	−1.2	1	2	+1.1	0	0	−5.2	1	3	+6.6
J	1	2	+1.5	1	0	−7.4	0	3	+0.1	1	2	+1.1	0	0	−5.1	1	3	+6.6
K	1	3	+6.5	2	2	+0.1	0	3	−0.9	1	2	+1.1	0	0	−5.0	1	3	+7.3
L3	1	3	+6.2	2	0	−3.6	0	3	+0.1	1	2	+1.4	0	0	−5.1	1	3	+7.3
Avg[f]	1.4	2.7	+5.3	2	0.7	−1.24	0	2.9	−0.63	1.1	1.9	+1.4	0	0	−5.1	1	3	+6.8

[a] Total number of histidines, arginines, and lysines.
[b] Total number of aspartic acids and glutamic acids.
[c] HV, Hydrophilicity value assigned by Hopp and Woods (8).
[d] Average charge and hydrophilicity value for serogroup B.
[e] Average charge and hydrophilicity value for the intermediate serogroup.
[f] Average charge and hydrophilicity value for serogroup C.

TABLE 4. Charge and hydrophilicity value of epitopes on MOMPs mapped by monoclonal antibodies

Monoclonal antibody	Sequence of epitope[a]	No. of amino acids		Hydrophilicity value	Location of epitope
		Basic	Acidic		
A-20	D-V-A-G-L-E-K-D-P-V	1	3	+4.1	A, VDI
L21-45	D-N-E-N-H-A-T-V-S-D-S-K-L-V	2	3	+6.8	L2, VDII
B-B6	N-N-E-N-Q-T-K-V-S-N-G-A-F-V	1	1	+0.9	B, VDII
B-B5	P-T-I-A-G-A-G-D-V-K-T-S-A-E-G	1	2	+3.8	B VDIV C terminus
L21-5	A-E-G-Q-L-G	0	1	+0.9	B and L2, VDIV C terminus
L21-10	T-T-L-N-P-T-I-A-G	0	0	−5.1	VDIV of all serovars except K

[a] As determined by lambda gt11 epitope mapping (1).

contained both VDI and VDII which were extremely hydrophilic and charged, which suggests that in MOMPs of these serovars both domains may be immunogenic.

Although antigenic sites have not yet been mapped to VDIII of MOMP, its hydrophilicity and charge values implicate this domain as a potential immunogenic site.

VDIV was divided into three regions, N terminal (residues 284 to 295), central (residues 296 to 306), and C terminal (residues 307 to 317), and the hydrophilicity and charge for each region were calculated independently. This domain showed a consistent pattern of hydrophilicity and charge in each region for each MOMP of all serovars. The N-terminal region was charged but is weakly hydrophilic; the central region was consistently uncharged and hydrophobic; and the C-terminal region was both highly charged and hydrophilic except for serovars F and G, which had only moderate charge and hydrophilicity values in this region of VDIV. On the basis of these values, the C-terminal end of VDIV has the greatest immunogenic potential. The immunogenicity of this region is supported, at least in part, by the fact that the serogroup B-specific MOMP epitope recognized by monoclonal antibody L2I-5 (Table 4) mapped to this region of VDIV. The single exception in correlating hydrophilicity and charge with known MOMP antigenic determinants was the species-specific MOMP epitope recognized by monoclonal antibody L2I-10. This epitope was located in the central region of VDIV (TTLNPTIAG) and was uncharged and strongly hydrophobic (−5.1).

DISCUSSION

We have used primer extension mRNA sequencing to determine the nucleotide and predicted amino acid sequences of the four MOMP VDs of the 15 *C. trachomatis* serovars. This technique was rapid and required minimal amounts of total RNA; 35 μg of RNA isolated from chlamydia-infected HeLa cells was needed to sequence all four MOMP VDs for a given serovar. This technique circumvented the more standard and laborious approach of molecular cloning and direct DNA sequencing of the MOMP gene(s).

The amino acid homologies among the MOMP VDs of all 15 serovars separated them into three groups. Group 1 contained serovars B, Ba, D, E, L1, and L2; group 2 contained serovars F and G; and group 3 contained serovars A, C, H, I, J, K, and L3. This classification, based on VD sequence homology, is in good agreement with the serological classification of these strains by using polyclonal antisera or monoclonal antibodies. Historically, by microimmunofluorescence with polyclonal mouse sera, serovars K and L3 were placed within serogroup C (7). More recently, by immunotyping with monoclonal antibodies specific to MOMP, serovars K and L3 have been classified within the intermediate serogroup (19). Our data, based on MOMP VD amino acid homology, clearly place serovars K and L3 within serogroup C.

The VDs are hydrophilic and charged, consistent with these sequences being associated with the chlamydial cell surface and having potential antigenic sites (1). Indeed, the antigenic properties and surface association of some MOMP VDs of a limited number of serovars have recently been reported. Stephens et al. (16) used antisera prepared against synthetic peptides corresponding to MOMP VDs to map and define antigenic properties associated with VDII and -IV. They showed that serovar-specific determinants were located in VDII of the MOMPs of serovars B, C, and L2, whereas a highly conserved species-specific MOMP determinant mapped to VDIV. Baehr et al. (1), using a lambda gt11 epitope-mapping procedure, showed that the serovar-specific MOMP epitope for serovars A and L2 mapped to VDI and -II, respectively, whereas subspecies-, serogroup-, and species-specific epitopes mapped to VDIV. The latter group of investigators concluded that VDI, VDII, and portions of VDIV protruded from the chlamydial cell surface on the basis of both the ability of these domains to bind monoclonal antibodies and their susceptibility to proteolysis by trypsin. From these limited epitope-mapping data, it appears that serovar-specific determinants reside in MOMP VDI or VDII or both, whereas common or serogroup-specific MOMP antigenic determinants reside in VDIV. Because these conclusions are based on data from a limited number of serovars, it is not known whether MOMP VDs of other serovars present similar antigenic properties and surface topographies. For example, our data show that sufficient amino acid sequence heterogeneity occurs in VDIV, especially among serovars within serogroup C, to yield sequences which may elicit type-specific antibody responses. Furthermore, although VDIII is clearly a potential immunogenic region, the antigenic properties of this domain are currently unknown, since neither monoclonal nor polyclonal antibodies have been mapped to this region. The sequence information reported here will be useful in developing reagents capable of more thoroughly defining the antigenic properties and surface topography of MOMP VDs. For example, it is now possible to synthesize peptides analogous to each MOMP VD for all serovars. Polyclonal or monoclonal antibodies generated against these peptides can be used for both defining the antigenic properties of these domains and determining their immunoaccessibility at the chlamydial surface of a single *C. trachomatis* serovar. This information will provide a more accurate description of the antigenic topography of MOMP VDs than has been possible by using antibodies specific for MOMP VDs of two different serovars (1).

Knowledge of the nucleotide and the amino acid sequences of the VDs of the MOMPs from all serovars may lead to the development of specific reagents to be used to study immunity to chlamydial infections. Synthetic peptides analogous to specific VD sequences could be used to examine the molecular specificity of the immune response of the host during chlamydial infection. This information would provide a necessary and rational approach for the selection of MOMP antigenic determinants to be used in the construction of a subunit or recombinant chlamydial vaccine.

The sequence information reported here may also prove useful for production of reagents and methodologies that are applicable to development of new diagnostic tests for *C. trachomatis* infections. For example, synthetic peptides analogous to the serovar-specific MOMP antigenic determinants may be used as antigens in serological assays to both diagnose and serotype the infecting chlamydial organism. In addition, peptides analogous to VD sequences can be used as immunogens to prepare monospecific polyclonal antibodies or monoclonal antibodies which can be used to identify the infecting serovar in either smears of clinical specimens or cell cultures of chlamydial isolates. Alternatively, the nucleotide sequences with VDI and -II are sufficiently different among all serovars to allow construction of serovar-specific synthetic oligonucleotides. Labeled with appropriate reporter molecules, these oligonucleotides could be used to develop nonimmunological probes capable of both detecting and typing *C. trachomatis* organisms.

ACKNOWLEDGMENTS

We thank Jim Simmons for technical assistance, Susan Smaus for secretarial assistance, John Swanson for suggesting the use of the mRNA-sequencing technique, and the Laboratory of Microbial Structure and Function staff for critical review of the manuscript.

Y.Y. and Y.-X.Z. are Visiting Fogarty Fellows.

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. **Bergstrom, S., K. Robbins, J. M. Koomey, and J. Swanson.** 1986. Pilation control mechanisms in *Neisseria gonorrhoeae*. Proc. Natl. Acad. Sci. USA **83**:3890–3894.
4. **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.
5. **Caldwell, H. D., S. Stewart, S. Johnson, and H. Taylor.** 1987. Tear and serum antibody response to *Chlamydia trachomatis* antigens during acute chlamydial conjunctivitis in monkeys as determined by immunoblotting. Infect. Immun. **55**:93–98.
6. **Efimov, V. A., A. A. Buryakova, S. V. Reverdatto, O. G. Chakhmakhcheva, and Y. A. Ovchinnikov.** 1983. Rapid synthesis of long-chain deoxyribo-oligonucleotides by the N-methylimidazole phosphotriester method. Nucleic Acids Res. **11**:8369–8387.
7. **Grayston, J. T., and S.-P. Wang.** 1975. New knowledge of chlamydiae and the diseases they cause. J. Infect. Dis. **132**:87–104.
8. **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.
9. **Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. USA **82**:6955–6959.
10. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
11. **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.
12. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
13. **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.
14. **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.
15. **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.
16. **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.
17. **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.
18. **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.
19. **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.
20. **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.