

The major outer membrane protein nucleotide sequence of *Chlamydia trachomatis*, serovar E

Ellena M.Peterson*, Brian A.Markoff and Luis M.de la Maza
Pathology Department, University of California Irvine, Irvine, CA 92717, USA

Submitted April 3, 1990 EMBL accession no. X52557

The major outer membrane protein (MOMP) of *C. trachomatis* is surface exposed and accounts for 60% of the outer membrane protein of this widespread pathogen. This protein has been shown to contain conserved regions among the 15 serovars of *C. trachomatis* as well as variable domains (VD) that are partially responsible for antigenic differences among the serovars. In addition antibodies raised to this protein have the ability to neutralize this pathogen both *in vivo* and *in vitro*. Therefore these characteristics of MOMP make it an excellent candidate for incorporation into a subunit vaccine against this pathogen. In order to construct such a vaccine it is necessary to have sequence information on all the serovars so that variable as well as conserved regions can be explored. We present here the complete sequence of serovar E which is the most common serovar causing sexually transmitted infections in the Western World.

The gene for the MOMP protein from *C. trachomatis*, serovar E was sequenced from polymerase chain reaction products (PCR). The PCR was performed using the following oligonucleotide primers which contained SalI and BamHI sites, respectively, at the 5′ ends: TCG GTC GAC ATG AAA AAA CTC TTG AAA TCG and ATA GGA TCC TTA GAA GCG GAA TTG TGC ATT. Primers used, as underlined below, were to the previously published leader sequence (1) and the carboxy end sequence of an EcoRI fragment from a λgt11 expression library. This EcoRI fragment was detected by identifying a λgt11 plaque that reacted with a monoclonal antibody to the VD IV of MOMP. This fragment contained the 447 bases of the C-terminus of the E MOMP.

Compared to a previously published sequence of L2 (1), the E MOMP sequence had 3 bases missing (451−453 in L2 sequence). There were 88 base changes in E compared to L2 which accounted for the 23 amino acid substitutions in serovar E. All except for 6 of these (aa# 53, 102, 120, 168, 170 and 173 of the E sequence) amino acid changes were in the VDs. The sequences of the VD regions were the same as those previously reported for serovar E by Yuan *et al.* (2).

REFERENCES

1. Stephens,R.S., Mullenbach,R., Sanchez-Pescador,R and Agabian,N. (1986) *J. Bacteriol.* **168**, 1277−1282.
2. Yuan,Y., Zhang,Y.-X., Watkins,N.G. and Caldwell,H.D. (1989) *Infect. Immun.* **57**, 1040−1049.

```
-66.    ATGAAAAAACTCTTGAAATCGGTATTAGTATTTGCCGCTTTGAGTTCTGCTTCCTCCTTGCAAGCT
1.    CTGCCTGTGGGGAATCCTGCTGAACCAAGCCTTATGATCGACGGAATTCTGTGGGAAGGTTTCGGCGGAG
71.   ATCCTTGCGATCCTTGCACCACTTGGTGTGACGCTATCAGCATGCGTATGGGTTACTATGGTGACTTTGT
141.  TTTCGACCGTGTTTTGAAAACAGATGTGAATAAAGAATTCCAAATGGGTGACAAGCCTACAAGTACTACA
211.  GGCAATGCTACAGCTCCAACCACTCTTACAGCAAGAGAGAATCCTGCTTACGGCCGACATATGCAGGATG
281.  CTGAGATGTTTACAAATGCCGCTTGCATGGCATTGAATATTTGGGATCGCTTTGATGTATTCTGTACACT
351.  AGGAGCCTCTAGCGGATACCTTAAAGGAAACTCTGCTTCTTTCAATTTAGTTGGATTGTTTGGAGATAAT
421.  GAAAATCAAAGCACGGTCAAAACGAATTCTGTACCAAATATGAGCTTAGATCAATCTGTTGTTGAACTTT
491.  ACACAGATACTGCCTTCTCTTGGAGCGTGGGCGCTCGAGCAGCTTTGTGGGAGTGCGGATGTGCGACTTT
561.  AGGGGCTTCTTTCCAATACGCTCAATCTAAACCTAAAGTCGAAGAATTAAACGTTCTCTGTAACGCAGCT
631.  GAGTTTACTATCAATAAGCCTAAAGGATATGTAGGGCAAGAATTCCCTCTTGCACTCATAGCAGGAACTG
701.  ATGCAGCGACGGGCACTAAAGATGCCTCTATTGATTACCATGAGTGGCAAGCAAGTTTAGCTCTCTCTTA
771.  CAGATTGAATATGTTCACTCCCTACATTGGAGTTAAATGGTCTCGAGCAAGTTTTGATGCCGATACGATT
841.  CGTATAGCCCAGCCAAAATCAGCTACAGCTATCTTTGATACTACCACGCTTAACCCAACTATTGCTGGAG
911.  CTGGCGATGTGAAAGCTAGCGCAGAGGGTCAGCTCGGAGATACCATGCAAATCGTCTCCTTGCAATTGAA
981.  CAAGATGAAATCTAGAAAATCTTGCGGTATTGCAGTAGGAACGACTATTGTAGATGCAGACAAATACGCA
1051. GTTACAGTTGAGACTCGCTTGATCGATGAGAGAGCTGCTCACGTAAATGCACAATTCCGCTTCTAA
```

* To whom correspondence should be addressed