Nucleotide sequence and taxonomic value of the major outer membrane protein gene of Chlamydia pneumoniae IOL-207

M. W. Carter,1* S. A. H. Al-Mahdawi,2 I. G. Giles,3 J. D. Treharne,4 M. E. Ward1 and I. N. Clarke1

1 Department of Microbiology, University of Southampton, Faculty of Medicine, Southampton General Hospital, Tremona Road, Southampton SO9 4XY, UK
2 Institute of Child Health, Department of Paediatric Cardiology, 30 Guilford Street, London WC1N 1EH, UK
3 Department of Biochemistry, University of Southampton, School of Biological Sciences, Bassett Crescent East, Southampton SO1 3TU, UK
4 Institute of Ophthalmology, Section of Virology, Judd Street, London WC1H 9QS, UK

(Received 16 July 1990; revised 10 October 1990; accepted 25 October 1990)

Chlamydia pneumoniae IOL-207 genomic DNA was hybridized with a 1.5 kb labelled DNA probe containing the 3' region of the coding sequence for the major outer membrane protein (MOMP) of C. trachomatis serovar L1. An 8.5 kb BglII fragment containing the complete MOMP gene was cloned into 1ZAP 11 cloning vector. Two hybridizing EcoRI fragments were sub-cloned into the 1ZAP II cloning vector and the resulting plasmids were used as templates for sequencing both strands of the C. pneumoniae MOMP gene. Computer taxonomic studies using the nucleotide and inferred amino acid sequence of the MOMP of C. pneumoniae IOL-207 and all known chlamydial MOMP sequences supported the designation of C. pneumoniae as a new species, but electron microscope studies suggested that the presence of pear-shaped elementary bodies (EBs) may not be a reliable taxonomic criterion.

Introduction

The Chlamydiae are obligate intracellular bacteria characterized by a unique growth cycle and responsible for a wide variety of diseases in man and animals. Three species of the genus Chlamydia are recognized: C. trachomatis, C. psittaci and the newly designated species C. pneumoniae (Grayston et al., 1989a).

C. pneumoniae is an increasingly identified cause of pneumonia and acute respiratory disease in man (Grayston et al., 1989b; Marrie et al., 1987). A high proportion of adults from different countries had antibodies to C. pneumoniae (Forsey et al., 1986; Grayston et al., 1989b), implying a high prevalence of these infections. Justification for designating the TWAR agent as a new species of Chlamydia was based on the differing ultrastructural morphology of the infectious elementary bodies (EBs) (Chi et al., 1987; Kuo et al., 1988), the lack of DNA homology (<10%) with either C. trachomatis, or C. psittaci (Cox et al., 1988), the unique DNA restriction endonuclease patterns (Campbell et al., 1987) and the demonstration by monoclonal antibodies (mAbs) of C. pneumoniae species-specific antigens (Kuo et al., 1986).

Although C. pneumoniae is a common human respiratory pathogen (Grayston et al., 1989b; Marrie et al., 1987), very little is known about the immunochemistry of this organism. One problem is that C. pneumoniae is difficult to isolate in tissue culture. Immunochemical and genetic studies should lead to simple and non-culture-dependent diagnostic methodologies for this important organism, as has been the case for C. trachomatis. Moreover, comparative studies of the nucleotide sequences of key C. pneumoniae antigens with other Chlamydia species might give insights into the evolutionary and structural diversity arising within this genus.

The chlamydial antigen for which the structure is best known is the major outer membrane protein (MOMP). MOMP is the most abundant protein at the surface of both the infectious EB and the intracellular replicative reticulate body (RB). Disulphide bridging between

Abbreviations: BGMK, Buffalo Green Monkey Kidney; CDC, Center for Disease Control, Atlanta, GA, USA; EB, elementary body; mAb, monoclonal antibody; micro-IF, micro-immunofluorescence; MOMP, major outer membrane protein; ORF, open reading frame; PCR, polymerase chain reaction.

The nucleotide sequence data reported in this paper have been submitted to GenBank, and have been assigned the accession number M34942.

0001-6384 © 1991 SGM
MOMP and other envelope proteins is considered important for maintaining the structural integrity of the infectious EB in the absence of cross-linking by peptidoglycan (Bavoil et al., 1984). Disulphide interlinking between MOMP polypeptide chains is thought to play an important role in regulating MOMP porin function, nutrient ingress through the EB outer envelope and ultimately the initiation of chlamydial differentiation itself (Bavoil et al., 1984). MOMP is an immunodominant antigen and antibodies to it form the basis of the serological classification of C. trachomatis (Stephens et al., 1982). Moreover, antibody to MOMP has been reported to neutralize chlamydial infection in vitro and in vivo (Zhang et al., 1987), making MOMP the principal candidate for the development of defined vaccines against C. trachomatis and C. psittaci and perhaps, ultimately, C. pneumoniae.

Comparative analysis of deduced amino acid sequences for several C. trachomatis MOMPs shows that the protein has four variable sequences (VS 1–4) forming loops between five highly conserved regions. The serovar- and subspecies-specific epitopes, which have been associated with neutralization (Zhang et al., 1987), are located in VSs 1, 2 or 4 (Baehr et al., 1988; Conlan et al., 1988; Stephens et al., 1988). Nucleotide sequences have recently been published for the variable segments of 15 C. trachomatis serovars (Yuan et al., 1989). In addition, complete nucleotide sequences are available for the MOMP genes of a smaller number of C. trachomatis serovars (Baehr et al., 1988; Hamilton & Malinowski, 1989; Hayes et al., 1990; Pickett et al., 1987; Stephens et al., 1986, 1987; Zhang et al., 1990) and for four strains of C. psittaci (Herring et al., 1989; Pickett et al., 1988; Zhang et al., 1989) but not for C. pneumoniae. This presumably reflects the difficulty of working with C. pneumoniae and the paucity of C. pneumoniae isolates now available.

We present here: (i) the complete nucleotide sequence of the MOMP gene of C. pneumoniae IOL-207; (ii) computer-derived phylogenetic maps of the evolutionary inter-relationships of MOMP nucleotide sequences in the genus Chlamydia; and (iii) a re-evaluation of the ultrastructural morphology of C. pneumoniae.

**Methods**

*Bacterial strains and growth conditions.* Chlamydia pneumoniae IOL-207, isolated in 1967 from the conjunctiva of an Iranian child thought to be suffering from mild trachoma (Dwyer et al., 1972), was propagated in the yolk sacs of 6-8-d-old embryonated hens' eggs. The inoculated eggs were incubated at ≧66% relative humidity at 37 °C and candled daily. Embryos were killed 20 d after infection by approved chilling at −20 °C for 1 h. C. pneumoniae ATCC VR 1310 (CDC/CWL 029) was originally isolated from the throat of a patient with pneumonia. This strain was obtained from the American Type Culture Collection and was grown in Buffalo Green Monkey Kidney (BGMK) cells (Flow Laboratories) in Dulbecco's Minimum Essential Medium (DMEM) supplemented with 1 μg emetine hydrochloride ml⁻¹ (Sigma). Escherichia coli strain JM109 [recA1 endA1 gyrA46 thi hsdR17 (rC mK supE44, relA1) Δ(lac-proAB), (F’ traD36 proAB lacP2ΔM15)], carrying a PsI–EcoRI insert in pUC9 coding for β of the MOMP gene from C. trachomatis L1/440/LN, was grown in Luria-Bertani (LB) medium (Sambrook et al., 1989) containing 50 μg ampicillin ml⁻¹ (Sigma) or on plates of ampicillin supplemented LB with 1.5% (w/v) agar (Difco). E. coli strain Q358 (had rC mK supF 880b) was used as a host for λEMBL3 and was grown in either TB broth (5 g NaCl and 10 g Bacto-Tryptone per litre) supplemented with 0.2% (w/v) maltose and 10 mM-MgSO₄, or on NZY (Sambrook et al., 1989) 1.5% (w/v) agar plates. Growth and amplification of phage in E. coli Q358 was in NZY 0.7% (w/v) agarose overlays upon NZY 1.5% (w/v) agar plates. E. coli strain BB4, LE392.23, (F’ lacP2ΔM15 proAB Tn10 (reFl3) obtained from Stratagene (Northumbria Biologicals) was used as the host for AZAP II and was grown in either LB broth supplemented with 0.2% (w/v) maltose, 10 mM-MgSO₄ and 15 μg tetracycline ml⁻¹ as recommended by the manufacturer; or on NZY tetracycline 1.5% (w/v) agar plates. Growth and amplification of recombinant phage in E. coli BB4 was in NZY 0.7% (w/v) agarose overlays upon NZY 1.5% (w/v) agar plates. E. coli strain J109 was grown either in TB broth supplemented with 0.2% (w/v) maltose and 10 mM-MgSO₄ or on NZY 1.5% (w/v) agar plates. E. coli J109 cells carrying Bluescript SK (−) recombinant plasmids were grown on LB 1.5% (w/v) agar plates containing 50 μg ampicillin ml⁻¹.

*Preparation of C. pneumoniae from infected yolk sacs.* Embryonated eggs were used to overcome the difficulties associated with the bulk growth of C. pneumoniae in tissue culture. Yolk sacs were harvested aseptically and examined for the presence of chlamydial EBs. Four positive yolk sacs were suspended in Dulbecco's phosphate buffered saline (PBS) to give an approximate 20% (w/v) suspension. This mixture was incubated at 37 °C for 1 h to digest remaining eukaryotic DNA. DNase I (9.5 mg ml⁻¹) (Sigma, molecular biology grade) was added and the mixture incubated at 37 °C for 1–2 h or until clear. The nucleic acids were precipitated with 5 M-ammonium acetate (pH 7.5) and 200 μg proteinase K ml⁻¹ at 37 °C for 1 h. The EBs were then re-dissolved in PBS and dispersed by vortexing with glass beads, cooled on ice and briefly ultra-sonicated. The chlorofluorocarbon Arklone (ICI) was added to the suspension at a ratio of 1:3 and the mixture homogenized for 2 min in a Sorvall Omnimix homogenizer and the resulting homogenate was centrifuged in a Sorvall HB4 rotor for 10 min at 10000 r.p.m. at 4 °C. Eighteen millilitres of the middle layer was removed and centrifuged for 30 min. at 18000 r.p.m. and 4 °C in a Sorvall SS34 swing-out rotor. Pellets were resuspended in 18 ml PBS and dispersed by vortexing with glass beads, cooled on ice and briefly ultra-sonicated. The chlorofluorocarbon Arkalone (ICI) was added to the suspension at a ratio of 1:3 and the mixture homogenized for 2 min in a Sorvall HB4 rotor for 10 min at 4 °C. The upper white opalescent layer was removed and placed on ice whilst the sonication and extraction procedure was repeated on the residue. The final suspension was centrifuged at 10000 r.p.m. in a Sorvall HB4 rotor for 5 min at 4 °C and the upper opalescent layer was removed and frozen at −20 °C.

*Preparation and extraction of C. pneumoniae DNA.* The EBs from the thawed suspension were pelleted by centrifugation, resuspended in 5 ml of 0.2 M-sodium acetate, pH 6.9 plus 10 mM-MgSO₄ and dispersed by vortexing with glass beads. To this suspension 50–100 μl DNAase I (9.5 mg ml⁻¹) (Sigma, molecular biology grade) was added and the mixture was incubated at 37 °C for 1 h to digest remaining eukaryotic DNA. The reaction was stopped by the addition of EDTA to 50 mM and the EBs pelleted by centrifugation in a Sorvall SS34 rotor at 10000 r.p.m. at 4 °C for 1 h. The pellet was resuspended in 2 ml saline Tris/EDTA (STE: 0.015 M-NaCl, 10 mM-Tris, pH 7.4, 1 mM-EDTA); RNase A (Sigma, molecular biology grade) was added to 20 μg ml⁻¹ and the mixture incubated at 37 °C for 45 min. The EBs were then pelleted and lysed in 5 ml of 50 mM-Tris/HCl, pH 7.5, 20% (w/v) sucrose, 20 mM-EDTA, 25 mM-DTT, 1% (w/v) Sarkosyl and 200 μg proteinase K ml⁻¹ at 55 °C for 1–2 h or until clear. The nucleic acids...
were extracted with phenol/chloroform and the aqueous phase dialysed for 24 h against STE and stored at 4°C.

**Enzymes and vectors.** Restriction endonucleases, T4 DNA ligase, calf intestinal phosphatase and the digoxigenin DNA labelling kit were obtained from Boehringer Mannheim. rTag DNA polymerase was obtained as Amplitag from Perkin Elmer Cetus. The \( \lambda EMBL3 \) and undigested \( \lambda ZAP II \) cloning kits and Gigapack II Gold packaging extract were obtained from Stratagene. Sequence Version 2.0, containing the additional nucleotide analogue dITP, and 7-deaza-dGTP sequencing kits were obtained from United States Biochemical Corporation.

**DNA probe production. (a) 5' MOMP gene probe.** Oligonucleotides were synthesized by \( \beta \)-cyanoethyl phosphoramidite chemistry on a model 381A automated DNA synthesizer (Applied Biosystems). Separate oligonucleotide primers starting at 37 bases upstream, 5' CCGCCGAAAAAGATAG 3', on the coding strand and at 403 bases downstream, 5' ACATCAAACGATCCA 3', on the complementary strand from the methionine start codon of the \( C. \) trachomatis \( L1/440/LN \) MOMP gene sequence, plus an \( L1/440/LN \) MOMP DNA sequence, plus rTag DNA polymerase, digoxigenin nucleotide and a \( C. \) trachomatis A/SA/1/OT genomic DNA template were used to prepare a labelled, 460 bp polymerase chain reaction (PCR) product, using a Perkin Elmer Cetus DNA thermal cycler.

(b) 3' MOMP gene probe. A large-scale alkaline-lysed preparation of the recombinant plasmid pMP148 was purified (Sambrook et al., 1989). Following complete \( PstI \) and \( EcoRI \) digestion of the plasmid the \( \approx 1.5 \) kb insert was separated by agarose gel electrophoresis. The liberated fragment, containing 898 bp of the coding sequence of \( C. \) trachomatis \( L1/440/LN \) MOMP, was excised from the gel, purified by GeneClean (Stratech Scientific) and labelled with digoxigenin.

**Genomic library construction.** Restriction endonuclease digests of the \( C. \) pneumoniae \( IOL-207 \) DNA were separated by gel electrophoresis. Fragments were transferred from the gel to Hybond-N nylon membrane (Amersham) and baked at 80°C for 90 min. A single 8.5 kb \( BglII \) fragment hybridized, at 65°C under standard conditions, with the digoxigenin-labelled \( PstI \)-EcoRI 3' MOMP gene probe. Therefore, \( BglII \)-digested DNA was ligated into \( BamHI \)-digested \( \lambda EMBL3 \) and packaged using Gigapack II Gold with \( E. \) coli Q358 as host. Plaques were transferred to Hybond-N membranes, fixed, then hybridized with the \( PstI \)-EcoRI 3' MOMP probe.

**A phage DNA production and cloning of the \( C. \) pneumoniae \( IOL-207 \) MOMP gene into \( \lambda ZAP II \).** Using \( E. \) coli Q358 as host, an amplified preparation of a single hybridizing recombinant plaque was used for the large-scale production of the phage DNA (Sambrook et al., 1989) of the phage DNA. DNA was purified and concentrated with an Elutip-d Minicolumn from Schleicher & Schüll (Anderman & Co.) then digested with restriction endonucleases specific to the unique cloning sites of \( \lambda ZAP II \). Using duplicate agarose gels, restriction fragments were separated, blotted and hybridized at 65°C with either the PCR-generated 5' MOMP or the \( PstI \)-EcoRI 3' MOMP probe.

Hybridizing EcoRI fragments corresponding to the 5' and 3' ends of the \( C. \) pneumoniae MOMP gene were excised from a preparative agarose gel and purified by GeneClean. Each of these fragments was ligated into separate EcoRI-digested, dephosphorylated preparations of \( \lambda ZAP II \).

Separate preparations of \( \lambda \) phage containing the EcoRI fragments were packaged using Gigapack II Gold with \( E. \) coli BB4 as the host. Amplified preparations of recombinant plasmids hybridizing with either the PCR or \( PstI \)-EcoRI MOMP probes were made in \( E. \) coli BB4. Recombinant PhBluescript SK(−) plasmids were prepared by in vitro excision with R408 helper phage. \( E. \) coli JM109 cells carrying plBluescript SK(−) recombinant plasmids were grown on LB 1.5% (w/v) agar plates containing 50 μg ampicillin ml\(^{-1}\).

**DNA sequencing.** Large-scale preparations of the recombinant plasmids carrying EcoRI restriction fragments were extracted by alkaline lysis from \( E. \) coli JM109, and purified by chromatography on NACS-PREPAC columns (BRL). Templates were prepared (Zhang et al., 1988) and sequenced by the dideoxynucleotide chain-termination method using [\( \alpha \)-32P]dATP and modified T7 DNA polymerase (Sequenase). Nucleotide analogues dITP and 7-deaza-dGTP were used to resolve adjacent band compression areas.

**Computer taxonomy.** Nucleotide and inferred amino acid sequences of the MOMP of \( C. \) pneumoniae \( IOL-207 \) were compared with those of the other \( Chlamydia \) strains listed in Table 1, using PHYLIP version 3.2 computer software (Felsenstein, 1980). Programs used within PHYLIP were: the protein parsimony program PROTPARS; the DNA parsimony program DNA-PARS and the program DNADIST, which computes the 'distance' between pairs of species from DNA sequences which are then utilized by the distance matrix programs FITCH and KITSCH. Sequences were first aligned with the multiple alignment program CLUSTAL (Higgins & Sharp, 1988) so that they contained the same number of characters including padding characters. A small program was written to convert the CLUSTAL output into the format required for the PHYLIP programs. Alignment of the 15 protein sequences required 403 sequence positions (including padding characters); the 15 nucleotide sequences required 1220 residues. In general the sites of insertion of padding characters were similar for both protein and DNA sequences.

**Electron microscopy.** Infected cell monolayers on 24-well tissue culture trays were processed as previously described (Ward & Murray, 1984). Briefly, infected cell monolayers in 24-well trays were fixed in glutaraldehyde, post-fixed in osmium tetroxide, stained with 1% (w/v) aqueous uranyl acetate and dehydrated in a graded ethanol series. The intact cell sheet was then detached from the plastic substratum by brief treatment with 2% epoxy propan and flat-embedded in Spurr low-viscosity epoxy resin. After ultra-thin sectioning with a diamond knife, the preparations were examined in a Hitachi H7000 transmission electron microscope.

**Results**

**Molecular cloning of \( C. \) pneumoniae \( IOL-207 \) MOMP gene**

The \( \approx 1.5 \) kb fragment liberated from pMP148 by digestion with \( PstI \) and EcoRI was labelled with digoxigenin and used as a probe (Fig. 1, line a) for the 3' region of the \( C. \) pneumoniae MOMP gene. The hybridization of digested \( C. \) pneumoniae \( IOL-207 \) DNA with the 3' \( PstI \)-EcoRI probe suggested that the MOMP gene was most likely contained within an 8.5 kb \( BglII \) fragment (Fig. 2a). This fragment was therefore cloned into \( BamHI \)-digested \( \lambda EMBL3 \) DNA. Successful cloning was confirmed by hybridization of the resulting plasmids with the 3' \( PstI \)-EcoRI probe. Digestion of the recombinant \( \lambda \) phage DNA with various restriction endonucleases was followed by separate hybridizations with both the digoxigenin-labelled 3' \( PstI \)-EcoRI (Fig. 1, line a) and the 5' PCR-generated (Fig. 1, line b) probes. A 1.4 kb EcoRI fragment hybridized with the 3' \( PstI \)-EcoRI probe
Fig. 1. Digoxigenin-labelled hybridization probes constructed from the MOMP genes of C. trachomatis: (a) 3' 1.5 kb PstI–EcoRI restriction fragment of L1/440/LN; (b) 5' 460 bp PCR-synthesized fragment of A/SAl/OT. The hatched box marks the position of the complete MOMP gene. P1, primer 1; P2, primer 2.

(Fig. 2b, panel B) and a 3.3 kb fragment hybridized with both probes (Fig. 2b, Panels A and B). This suggested that these two restriction fragments carried the whole MOMP gene between them; the 3.3 kb fragment containing the region bounded by the 5' start codon and the EcoRI site, whilst the 1.4 kb fragment contained the region between the EcoRI site and the 3' stop codon. These two restriction fragments were therefore subcloned into separate EcoRI-digested, dephosphorylated preparations of λZAP II. Successful cloning of the two EcoRI inserts was confirmed by hybridization of their respective plaques in E. coli BB4 with either the 3' PstI–EcoRI or the 5' PCR-generated probe. λZAP II recombinants were then converted to pBluescript SK(-) plasmids using R408 helper phage.

The C. pneumoniae IOL-207 MOMP gene sequence

Plasmids carrying the 1.4 and 3.3 kb EcoRI inserts were purified and used as templates for DNA sequencing. Forward Universal primer, used in the first sequencing reactions of both plasmid templates, provided open reading frames (ORFs) of the MOMP gene. Alignments were found between the translated product of the ORFs of both inserts and the MOMP gene sequence of C. trachomatis L1. Orientation of the aligned inserts confirmed that the 5' and 3' regions of the C. pneumoniae IOL-207 MOMP gene were contained within the 3.3 and the 1.4 kb fragments, respectively. Both strands of the C. pneumoniae IOL-207 MOMP gene were then sequenced, giving an ORF of 1167 bp (389 amino acids). This sequence is shown in Fig. 3. A conserved XbaI restriction site was then used to align and compare the restriction maps of the MOMP genes of C. pneumoniae IOL-207 and C. trachomatis serovar L1 (Fig. 4).
The deduced amino acid sequences of the MOMP gene of *C. pneumoniae* IOL-207 and of 10 strains representing six serovars of *C. trachomatis* and four strains of *C. psittaci* (Table 1) were aligned. For clarity, Fig. 5 shows the alignments of the MOMP gene sequences for only *C. pneumoniae*, *C. trachomatis* A/SA1/OT, B/Jali20/OT and L1/440/LN, *C. psittaci* ovine abortion agent A22/M and guinea pig inclusion conjunctivitis strain 1. The signal peptide of *C. pneumoniae* MOMP consisted of 23 amino acids in contrast to the 22 amino acids of the other Chlamydiae. Homology with other amino acid sequences shows the characteristic four variable regions for *C. pneumoniae* and 71.5–72.4% homology with the *C. pneumoniae* sequences (Table 1). Overall, there was 67.2–67.9% homology between the cysteine residue at position 333 and there was an additional cysteine at position 212 which was also conserved cysteine residues.

### Amino acid sequence

The deduced amino acid sequences of the MOMP gene of *C. pneumoniae* IOL-207 and of 10 strains representing six serovars of *C. trachomatis* and four strains of *C. psittaci* (Table 1) were aligned. For clarity, Fig. 5 shows the alignments of the MOMP gene sequences for only *C. pneumoniae*, *C. trachomatis* A/SA1/OT, B/Jali20/OT and L1/440/LN, *C. psittaci* ovine abortion agent A22/M and guinea pig inclusion conjunctivitis strain 1. The signal peptide of *C. pneumoniae* MOMP consisted of 23 amino acids in contrast to the 22 amino acids of the other Chlamydiae. Homology with other amino acid sequences shows the characteristic four variable regions for *C. pneumoniae*. Cysteine residues are considered crucial for MOMP porin function; *C. pneumoniae* had a unique cysteine residue at position 333 and there was an additional cysteine at position 212 which was also present in *C. trachomatis* but not *C. psittaci* (Fig. 5). Overall, there was 67.2–67.9% homology between the *C. pneumoniae* MOMP nucleotide sequence and *C. trachomatis* sequences and 71.5–72.4% homology with the *C. psittaci* sequences (Table 1).
### Table 1. Nucleotide and amino acid homology of the MOMP gene of C. pneumoniae IOL-207 with strains of C. trachomatis and C. psittaci

<table>
<thead>
<tr>
<th>Species</th>
<th>Serovar</th>
<th>Strain</th>
<th>Percentage homology</th>
<th>Reference of strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. trachomatis</td>
<td>A</td>
<td>A/Har13/OT</td>
<td>Nucleotide: 67-2</td>
<td>Baehr et al. (1988)†</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>A/SA1/OT</td>
<td>Amino acid: 63-5</td>
<td>Hayes et al. (1990)</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>B/TW5/OT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>B/TW5/OT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>B/Jali20/OT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>C/TW3/OT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>C/TW3/OT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>IC-Cal-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>L1/440/LN</td>
<td>67-6</td>
<td>Pickett et al. (1987)</td>
</tr>
<tr>
<td></td>
<td>L2</td>
<td>L2/434/Bu</td>
<td>67-6</td>
<td></td>
</tr>
<tr>
<td>C. psittaci</td>
<td>A22/M</td>
<td></td>
<td>71-8</td>
<td>Pickett et al. (1988)</td>
</tr>
<tr>
<td></td>
<td>S26/3</td>
<td></td>
<td>72-4</td>
<td>Herring et al. (1989)</td>
</tr>
<tr>
<td></td>
<td>GPIC 1</td>
<td></td>
<td>72-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mn-Cal-10</td>
<td></td>
<td>71-5</td>
<td>Zhang et al. (1989)</td>
</tr>
<tr>
<td>C. pneumoniae</td>
<td>IOL-207</td>
<td></td>
<td>100-0</td>
<td></td>
</tr>
</tbody>
</table>

* Strain unknown. † Personal communication.

Fig. 5. Comparative alignment of the amino acid sequence of the MOMP gene of C. pneumoniae IOL-207 with the MOMP genes of C. trachomatis A/SA1/OT; B/Jali 20/OT; L1/440/LN, C. psittaci ovine abortion agent A22/M and guinea pig inclusion conjunctivitis strain 1 (GPIC 1). Dots indicate identity and dashes indicate deletions. Conserved cysteine residues and VS 1–4 are boxed. The signal peptide is marked as SP. Numbers refer to C. pneumoniae IOL-207 MOMP.
Phylogenetic classification

When protein and DNA sequences are used to develop a phylogenetic relationship the initial result is an unrooted tree, i.e. a tree where the branching order, intermediate nodes and the relative distances between each node have been identified but where the ancestral node (the root of the tree) has not been identified. In order to identify the root, additional assumptions must be made. Usually it is sufficient to assume that there is a constant evolutionary clock which affects all sequences equally. Protein parsimony studies in which the order of presentation of the complete protein sequences was randomized always gave the same results: an unrooted tree with three minor variations in one location. These variations arose through the uncertainty of determining whether there are one or two intermediate nodes between the sequences for the MOMPs of serovars H and C.

More detailed studies were conducted using the nucleotide data as it is possible to estimate the relationships (distance) between nucleotide sequences with less ambiguity than for amino acid sequences. The program DNAPARS confirmed that the unrooted trees obtained using the nucleotide sequences were equivalent to those obtained using the protein sequences. Two trees were identified, again differing as to whether there are one or two intermediate nodes between the MOMPs of C. trachomatis serovars H and C.

In order to produce a rooted tree, the program DNADIST was used to calculate the distances between each pair of DNA sequences. The Kimura '2-parameter' model (Kimura, 1980) was used; this allows a difference between transition and transversion rates. Unfortunately, this program could not be used for all 15 sequences, as two of the C. trachomatis serovar B sequences had only four nucleotide differences between them; this degree of similarity caused a numerical error condition and the program DNADIST terminated prematurely. However, removing one of these sequences permitted calculation of the distance matrix for the remaining 14 sequences. The program KITSCH, which employs a Fitch–Margoliash least-squares approach, was used to identify a single rooted tree. KITSCH assumes that the sequences available were all obtained at the same relative evolutionary time from the origin (i.e. that they are all present-day isolates). This implies, therefore, since there is also an assumption of a constant evolutionary clock, that the distance from the root to each tip representing an extant species must be the same. With these assumptions a rooted tree was obtained (Fig. 6).

Electron microscopy

One of the taxonomic criteria suggested for C. pneumoniae is the presence of pear-shaped EBs (Grayston et al., 1989a; Kuo et al., 1988). However, we have also observed occasional aberrant pear-shaped EBs in preparations of C. trachomatis and C. psittaci and were thus concerned to establish whether pear-shaped EBs are a reliable characteristic of all C. pneumoniae strains. Figs 7(a) and 7(b) are transmission electron micrographs of C. pneumoniae strains IOL-207 and ATCC VR 1310 respectively. Representative C. trachomatis and C. psittaci strains whose ultrastructural morphology is well
known were also processed in parallel (data not shown) as a control for possible fixation artifacts. Fig. 7(a) shows that excellent growth of \textit{C. pneumoniae} was achieved in BGMK cells treated with 1 µg emetine ml$^{-1}$. However, although large mature inclusions were obtained, remarkably few infectious EBs were present. Those \textit{C. pneumoniae} EBs which were observed were round rather than pear-shaped (Fig. 7b), occasionally crenellate, and could not reliably be distinguished from EBs of \textit{C. trachomatis} or \textit{C. psittaci}. 

Fig 7. Transmission electron micrographs of \textit{C. pneumoniae} in tissue culture. (a) Low-power picture of IOL-207 in BGMK cells. The EBs appear as small dark dots within the inclusions. Note the relatively small number of infectious EBs despite the maturity of the inclusions. Bar, 5 µm. (b) ATCC VR 1310. Higher magnification picture to show the morphology of the EBs. The EBs of both \textit{C. pneumoniae} strains were round, occasionally crenellated but rarely pear-shaped. These EBs could not reliably be distinguished from EBs of \textit{C. trachomatis} or \textit{C. psittaci} in material processed in parallel. Bar, 0.5 µm.
Discussion

Strains of *C. pneumoniae* were formerly known as TWAR agents, an acronym derived from the fact that the first isolate was from Taiwan and that subsequent isolates were associated with acute respiratory infection (Grayston *et al.*, 1986). Species designation was attained as recently as 1989 (Grayston *et al.*, 1989). The type strain, TW 183, was not available to us at the commencement of this study. We therefore chose to work with IOL-207, the second known *C. pneumoniae* isolate (Dwyer *et al.*, 1972; Grayston *et al.*, 1989). Both strains were originally isolated from the conjunctivae of children in trachoma-endemic areas; TW 183 was isolated in Taiwan in 1965 whilst IOL-207 was isolated in London from an Iranian child with mild trachoma in 1967. Both organisms share the same fastidious growth characteristics of *C. pneumoniae* (Grayston *et al.*, 1989). In 1972 TW 183 and IOL-207 were found to be immunologically identical using a two-way micro-immunofluorescence (micro-IF) typing test (Dwyer *et al.*, 1972). During the course of this study a *C. pneumoniae* strain was deposited by CDC in the American Type Culture Collection (ATCC VR 1310) and thus became available to us. mAb to IOL-207 reacts with TW 183 in the micro-IF test (J. D. Treharne, unpublished observation) and with ATCC VR 1310 on dot blot. Conversely, *C. pneumoniae*-specific mAb to TW 183 also reacts on micro-IF with IOL-207 (J. D. Treharne, unpublished observation) and with ATCC VR 1310 (Carolyn Black, CDC, personal communication). Neither of these mAbs reacts with *C. psittaci* or *C. trachomatis*. We have also sequenced the gene for the 60 kDa cysteine-rich, outer membrane protein of IOL-207; oligonucleotide probes derived from this sequence reacted in a species-specific manner in the PCR reaction with genomic DNA from *C. pneumoniae* ATCC VR 1310 (M. W. Watson and others, unpublished observations). Finally our computer study (see below) confirmed that the IOL-207 MOMP sequence is taxonomically distinct from known MOMP sequences for *C. trachomatis* or *C. psittaci*. The ecletic association of the above findings would suggest that there can be little doubt that IOL-207 is a genuine *C. pneumoniae* isolate.

The MOMP gene of *C. pneumoniae* shows all the characteristics of a chlamydial MOMP. The signal peptides for MOMP in all three species of *Chlamydia* are virtually identical, with a classic Leu-Glu-Ala protease cleavage site. The *C. pneumoniae* signal peptide has 23 rather than 22 amino acids and there are a few minor substitutions for amino acids of similar properties. The N-terminus of the MOMP for all three species is also similar, with the first six amino acids identical (LPVGVPN). We speculate that this region of the protein is conserved in order to ensure correct anchorage in the membrane. The other conserved feature between the MOMP sequences is the position of seven conserved cysteine residues, which are probably important in the structural and functional role of this protein.

The four variable regions of MOMP are of major interest for vaccine development (Taylor-Robinson & Ward, 1989) and are also present in *C. pneumoniae*. VS 1, 2 and 4 of *C. trachomatis* are of particular interest because they are hydrophilic, have high beta turn potential, and are thought to form loops at the surface of chlamydial EBs accessible to neutralizing serovar- or subspecies-specific antibodies (Zhang *et al.*, 1987). VS 2 and VS 4 are topographically juxtaposed on the chlamydial surface, forming a discontinuous epitope which can be spanned by a single mAb paratope (Conlan *et al.*, 1990). Both these exposed loops may function in a co-ordinated manner to promote chlamydial adhesion (Su *et al.*, 1990). VS 1 of the C complex of *C. trachomatis* contains immunodominant subspecies- and type-specific epitopes (Baehr *et al.*, 1988; Stephens *et al.*, 1988) whose critical binding sites, in the case of A/SA 1/OT, have been localized to the peptides VAGL and EKD respectively (Hayes *et al.*, 1990). As might be expected, there are no analogous peptides to these in the VS 1 of *C. pneumoniae*. In the B complex of *C. trachomatis*, by contrast, it is VS 2 which so far appears to contain the dominant type epitope (Baehr *et al.*, 1988; Conlan *et al.*, 1988; Stephens *et al.*, 1988). In particular in *C. trachomatis* B/Jali20/OT and L1/440/LN the critical binding sites for type-specific antibody are the peptides GTFVP and DA-VP (Conlan *et al.*, 1988; Hayes *et al.*, 1990). Interestingly, in all known chlamydial MOMP sequences there are two or three variable amino acids followed by a hydrophobic amino acid (L, Y or V) and proline, a computer-predicted turn-generating amino acid. We speculate that the unique peptide ANELP in the equivalent position in *C. pneumoniae* MOMP is also surface exposed and this may represent a type-specific region. In VS 4 of the B complex serovars of *C. trachomatis* there are two known subspecies- and one species-specific epitope (Baehr *et al.*, 1988; Conlan *et al.*, 1988, 1989). We are unable to identify a clear counterpart in *C. pneumoniae*.

Chlamydiae have no other porin proteins and therefore MOMP is presumed essential for chlamydial survival. From a taxonomic viewpoint MOMP is in many ways an ideal protein to study; there must inevitably be large regions of structure which have been conserved throughout evolution. However, there are also immunodominant variable regions which, at least for *C. trachomatis*, can be related to the serological classification and which are also epidemiologically associated with different patterns of clinical disease. Moreover, with the publication of the *C. pneumoniae* sequence, there
are now considerable data on MOMP gene structure for all Chlamydia species.

Computer taxonomy provides an objective method of applying generalized taxonomic criteria to the particular problem of evolutionary relationships amongst Chlamydiae; the program which we used was developed for plant taxonomy (Felsenstein, 1980). The same classification tree resulted irrespective of whether we used nucleotide or inferred amino acid sequence data. This implies that there are no dramatic differences in codon preference amongst the Chlamydiae. Hybridization studies using whole genomic DNA found less than 10% homology between C. pneumoniae and the other two chlamydial species (Cox et al., 1988). This led to the suggestion that the chlamydial species might be an example of convergent evolution from divergent ancestors. This has never seemed to us a convincing argument given the uniqueness of the genus Chlamydia; it is certainly not supported by the high degree of homology amongst the chlamydial MOMP genes. Comparisons of the MOMP gene sequences of C. pneumoniae IOL-207 with those of other Chlamydia species gave homologies ranging from 67-2 to 72-4% for the nucleotides and from 63-5 to 75-9% for the amino acid sequences. This is reflected in the computer taxonomy, which shows all three species arising from a common ancestor, i.e. divergent evolution. Moulder (1988) suggests that the protochlamydiae, chlamydia-like organisms parasitic in invertebrates, are present-day descendants of a possible common ancestor.

The computer analysis classifies the present known MOMP sequences into three main branches. One branch consists of the four C. psittaci sequences, with a strong similarity between the sequence of C. psittaci A22/M ovine abortion agent and C. psittaci MnCal-10. The second branch consists of the single C. pneumoniae IOL-207 sequence. The distinction made for this branch supports separate species status for this organism whilst its origin amongst C. psittaci suggests that C. pneumoniae diverged from C. psittaci. This concept is supported by the fact that neither C. psittaci nor C. pneumoniae produce inclusions with an iodine-staining reserve carbohydrate (Grayston et al., 1989b). C. trachomatis is classified by the computer analysis into two broad groups which correspond with the serogroups related to either serovar B or serovar C. Each of these serogroups is formed by the presence of a distinct subspecies epitope on MOMP (Conlan et al., 1988; Hayes et al., 1990). This subspecies epitope forms only a tiny portion of the MOMP variable sequences; it is remarkable that the computer analysis was able to consistently discriminate between them. Serovar F is shown as being distinct from either of these complexes. Similarly, in the serological classification of C. trachomatis by micro-IF, serovar F does not belong clearly to either group (Wang & Grayston, 1970).

One of the taxonomic criteria used to give species status to C. pneumoniae was the presence of unusual, pear-shaped EBs (Chi et al., 1987). It has been suggested that the tip of these EBs might be a specialized attachment structure (Kuo et al., 1988). We have seen similar structures in immature C. trachomatis EB and therefore re-examined C. pneumoniae ultrastructure ourselves. The electron micrographs clearly showed that the majority of EBs of both IOL-207 and ATCC VR 1310 were round. However, inclusions of C. pneumoniae contained a much lower proportion of fully developed EBs than those of C. trachomatis or C. psittaci. This may explain the greater difficulty in passaging C. pneumoniae in cell culture, as only mature EBs are infectious. We accept that there may be unknown technical reasons to explain our failure to observe significant numbers of pear-shaped EBs in this organism. Perhaps IOL-207 and VR 1310 differ in morphology from TW 183. In that case further work is required to ascertain the general validity of the ultrastructural criteria chosen for this new species.

Our work on the MOMP gene of IOL-207 and its taxonomic status fully supports species designation for C. pneumoniae. This is corroborated by the distinct clinical syndromes with which this organism is associated. One of the most controversial of clinical reports is the apparent association between antibody to C. pneumoniae and myocardial infarction (Saikku et al., 1988). One of the key problems in defining the clinical syndromes associated with this fastidious organism is the lack of specific diagnostic probes. We hope that the C. pneumoniae MOMP gene sequence presented here will facilitate the development of improved, non-culture-dependent, diagnostic reagents for this newly recognized human pathogen.

The authors thank Dr M. A. Pickett for synthesis of the oligonucleotides, Mr M. W. Watson for his advice, and Mr Watson and Miss S. J. Cooke for help in preparing the manuscript. M. W. Carter is in receipt of a Medical Research Council research studentship. This investigation received financial support from the Special Programme of Research, Development and Research Training in Human Reproduction, World Health Organization.

References


