Microvirus of Chlamydia psittaci strain Guinea pig Inclusion Conjunctivitis: isolation and molecular characterization

Ru-ching Hsia, Li-Min Ting and Patrik M. Bavoil

Author for correspondence: Ru-ching Hsia. Tel: +44 20 7927 2290. Fax: +44 20 7612 7871. e-mail: r.hsia@lshtm.ac.uk

1 Department of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK
2 Department of Medicine, University of California at San Francisco, Box 0868, San Francisco, CA 94143-0868, USA

The authors report the isolation and molecular characterization of a bacteriophage, φCPG1, which infects Chlamydia psittaci strain Guinea pig Inclusion Conjunctivitis. Purified virion preparations contained isometric particles of 25 nm diameter, superficially similar to spike-less members of the φX174 family of bacteriophages. The single-stranded circular DNA genome of φCPG1 included five large ORFs, which were similar to ORFs in the genome of a previously described Chlamydia bacteriophage (Chp1) that infects avian C. psittaci. Three of the ORFs encoded polypeptides that were similar to those in a phage infecting the mollicute Spiroplasma melliferum, a pathogen of honeybees. Lesser sequence similarities were seen between two ORF products and the major capsid protein of the φX174 coliphage family and proteins mediating rolling circle replication initiation in phages, phagemids and plasmids. Phage φCPG1 is the second member of the genus Chlamydiamicrovirus, the first to infect a member of a Chlamydia species infecting mammals. Similarity searches of the nucleotide sequence further revealed a highly conserved (75% identity) 375 base sequence integrated into the genome of the human pathogen Chlamydia pneumoniae. This genomic segment encodes a truncated 113 residue polypeptide, the sequence of which is 72% identical to the amino-terminal end of the putative replication initiation protein of φCPG1. This finding suggests that C. pneumoniae has been infected by a phage related to φCPG1 and that infection resulted in integration of some of the phage genome into the C. pneumoniae genome.

Keywords: Chlamydia, bacteriophage, capsid protein, Rep protein

INTRODUCTION

Chlamydia trachomatis infections are among the most prevalent sexually transmitted and ocular infections in the world. Chronic chlamydial disease frequently leads to sequelae such as pelvic inflammatory disease, ectopic pregnancy, infertility and trachoma, the world’s leading cause of preventable blindness (Schachter, 1988). Chlamydia pneumoniae causes acute respiratory illness (Grayston et al., 1993), and C. pneumoniae infection has been associated with cardiovascular disease, lung cancer and Alzheimer’s disease in humans (Balin et al., 1998; Kuo et al., 1993; Laurila et al., 1997; Saikku et al., 1992). Other members of the genus Chlamydia cause infections and disease in a variety of vertebrate hosts, and some are significant pathogens of sheep, cattle and poultry (Storz, 1988).

The apparent diversity in host and associated disease pathology belies common elements that are central to the chlamydial pathogenic process at both the molecular and cellular levels. All members of the Chlamydiaceae share a common obligate intracellular life cycle, and comparative genomic analysis of two members of the family (C. trachomatis and C. pneumoniae) reveals a core set of shared genes whose putative functions encompass most activities known or presumed to be required for intracellular survival, growth and differentiation (Kalman et al., 1999). A counterpoint to this hypothesis is that differential properties, such as host

Abbreviations: EB, elementary body; GPIC, Guinea Pig Inclusion Conjunctivitis; MOMP, major outer-membrane protein; RB, reticulate body; RCR, rolling circle replication; Rep, replication initiation protein.

The GenBank accession number for the sequence reported in this paper is U41758.
and tissue tropism, the ability to persist, and the capacity of certain strains to cause disseminated infection, are more likely to be specified by factors that are unique to specific strains of the *Chlamydiaceae*, may not be readily annotated, and may ultimately be difficult to identify. Hence the wealth of genomic information currently gathered, however useful in allowing predictions, is unlikely to provide immediately satisfying answers to many questions regarding chlamydial infection and disease. This is particularly true where differential pathogenic outcomes are concerned, and this difficulty will be enhanced by the continuing lack of genetic methodologies for these bacteria. It is therefore essential that efforts to identify new virulence factors and tools to manipulate these factors genetically be developed concurrently with genomic and post-genomic advances to enable the new information to be exploited optimally.

We describe in this report a microvirus that infects *Chlamydia psittaci* strain Guinea pig Inclusion Conjunctivitis (GPIC). This bacteriophage, designated φCPG1, is the second *Chlamydia* phage ever isolated and the first known to infect a member of the *Chlamydiaceae* infecting a mammalian species.

**METHODS**

**Bacterial strains.** The *C. psittaci* GPIC strain that became phage contaminated was obtained in 1987 from the laboratory of Professor Julius Schachter, University of California at San Francisco, and was passaged continuously in our laboratory. *C. psittaci* GPIC that are identified as phage-free in this report were obtained from the laboratory of Professor Roger Rank, University of Arkansas for Medical Sciences. All chlamydiae were grown in HeLa cells and purified as described previously (Schachter & Wyrick, 1994).

**Cloning of the φCPG1 genome.** Supernatants from φCPG1-infected GPIC culture (48 h post-infection) were collected and cleared by centrifugation at 30000 *g* for 30 min at 4 °C and filtration through a 0.22 µm membrane. Following treatment with DNase I (20 µg ml⁻¹) at 37 °C for 1 h, φCPG1 particles were sedimented by centrifugation at 100000 *g* for 3 h at 4 °C and treated with proteinase K (200 µg ml⁻¹) at 37 °C for 1 h. Single-stranded phage genomic DNA was recovered by phenol/chloroform extraction and ethanol precipitation. Conversion to double-stranded DNA was achieved using Klenow fragment and random hexamer priming, and second-strand synthesis was confirmed by mobility shift upon agarose gel electrophoresis. The resulting double-stranded DNA template was digested with *Bam*HI, generating a major linearized molecule migrating at approximately 4-5 kb on a TAE-agarose gel. The whole *Bam*HI digest was shotgun cloned into Bluescript pSK+.

All transformants generated contained identical plasmid inserts of 4-5 kb. Partial nucleotide sequence analysis of a representative clone (pCPG1) revealed sequence encoding the amino terminus of the capsid protein (VP1), thus confirming the viral origin of the insert. In order to ensure that plasmid pCPG1 contained the complete phage genome, i.e. that no genomic sequence was missing as could result from more than one *Bam*HI site, a fragment spanning 135 bp upstream and 254 bp downstream of the *Bam*HI cloning site was amplified using purified phage DNA as template. Sequence analysis of the amplified fragment revealed no extra *Bam*HI site and confirmed that pCPG1 contains the whole genome of φCPG1.

**Nucleotide sequence analysis.** For each strand, nucleotide sequence analysis was performed on overlapping segments, either by subcloning small fragments into Bluescript pSK+, and subsequently sequencing using T3 and T7 primers, or by subcloning larger fragments into pMOB, generating nested β inserts and sequencing using transposon-specific primers as described by Strathmann et al. (1991). A combination of manual (Sanger et al., 1977) and automated (Applied Biosystems) sequencing methods was used for the analysis of the complete φCPG1 genome.

**Bioinformatics.** Sequence analysis was performed using software of the Wisconsin Package Version 10.0, Genetics Computer Group, Madison, WI, through the UK Human Genome Mapping Project Resource Centre in Cambridge (PEPTIDEMAP, GAP, BLASTN, STEMLOOP), software from Le Centre de Ressources Infobiogen, Université d’Evry-Val d’Essonne (clustalw 1.8), and from the National Center for Biotechnology Information, National Library of Medicine, US National Institute of Health (PSI-BLAST).

**Electron microscopy.** Samples were fixed in 2 % (v/v) glutaraldehyde and 0.5 % (w/v) paraformaldehyde for processing and embedding in Epon-Araldite resin as outlined by Wyrick et al. (1989). Ultrathin sections were stained with uranyl acetate and lead citrate and examined using a Zeiss EM900 transmission electron microscope operated at 60 kV. Negatively stained [2 % (w/v) uranyl acetate] specimens were observed on a Philips CM12 electron microscope.

**Other methods.** SDS-PAGE was performed according to Laemmli (1970).

**RESULTS**

**Co-purification of phage structural proteins with infectious GPIC elementary bodies**

The gradual appearance of a 62 kDa band in SDS-PAGE of purified GPIC elementary bodies (EBs) was observed coincidental with a chronic lowering of infectious yield and purified EB yield (data not shown). On SDS-PAGE of purified GPIC EBs (Fig. 1, lanes 2 and 3), the 62 kDa band was more abundant than all other proteins, including the abundant major outer-membrane protein (MOMP). Retrospective analysis of previous SDS-PAGE of GPIC-infected HeLa cell lysates revealed the presence of this polypeptide for at least the two preceding years in amounts varying roughly in inverse proportion with those of MOMP (Fig. 1, lanes 5–13), consistent with a lower yield of EBs.

The sequence of the amino-terminal 20 residues of the gel-purified 62 kDa band was chemically determined as N-VRNRLPSVMHSHFQVPSA-C. Comparison with sequence databases revealed 50 % identity with the amino terminus of VP1 (MAKGRKLPSVMKR- FSEYPTA; conserved residues underlined), the proposed capsid protein of Chp1 (Storey et al., 1989a), a bacteriophage previously isolated by Richmond and collaborators from *C. psittaci* strain N352, which was isolated from ducks in East Anglia, UK (Richmond et al., 1982).
Isolation and structural analysis of φCPG1

Confirmation of the presence of a contaminating phage in GPIC cultures was initially obtained by direct visualization. GPIC-infected HeLa lysates were examined by transmission electron microscopy of ultrathin sections. Particles similar in size and shape to Chp1 virions were observed in abundance in GPIC-infected HeLa lysates, in particular in association with membranes of lysed chlamydiae (Fig. 2a). GPIC-infected HeLa culture supernatants were fractionated according to Storey et al. (1989b) and phage-enriched fractions were observed by electron microscopy of negatively stained samples (Fig. 2b) and by SDS-PAGE. The particles observed were consistent with an isometric phage of approximately 25 nm diameter, superficially similar to that of the Escherichia coli bacteriophage φX174 (Hayashi et al., 1988). Some of the particles had a central electron lucency and may represent viral particles empty of their nucleic acid complement. SDS-PAGE of purified phage fractions (Fig. 3) revealed three major polypeptide bands of apparent molecular masses 62, 26 and 13 kDa, similar to the SDS-PAGE profile of purified Chp1 (Storey et al., 1989b).

These data constituted strong evidence of the occurrence in our laboratory of a Chlamydia phage, herein termed φCPG1 (for phage of C. psittaci GPIC no. 1), in GPIC-infected HeLa cell cultures over at least the previous 2 years.

Analysis of the φCPG1 genome

The φCPG1 genome was isolated from culture supernatant fractions and cloned in Bluescript, generating plasmid pCPG1. Analysis of the 4529 bp included within the pCPG1 insert revealed five major non-overlapping ORFs, in one transcriptional orientation (Fig. 4). These were similar to those in Chp1, with the sequence identity of the predicted products ranging from 25.6% to 54.3% (Table 1). As the five predicted gene products of φCPG1 were highly homologous to their Chp1 counterparts, we adopted the nomenclature already in place for Chp1 for the three proposed structural proteins (i.e. VP1–3) and further propose to name the last two VG4 and VG5 for virus genes 4 and 5 (corresponding to ORF4 and ORF5 in Chp1, respectively).

The G+C content of the φCPG1 genome was 41 mol %, higher than that of Chp1 (36.5 mol %) but similar to that of Chlamydia spp. (41–44 mol %) (Moulder et al., 1984), possibly reflecting long term co-evolution of φCPG1 and its bacterial host. Nine high-quality (> 20 bonds per stem) stem–loop structures were found in the VP3–VG4 intergenic segment and within the 5′ end of the VG4 ORF of the φCPG1 genome (data not shown), while five were found in the corresponding segment of the Chp1 genome. Although the resulting hairpin structures were different between the two phages, their high quality and relatively conserved position suggest that the segment encompassing VG4 and the upstream sequence contains signals that are important for either replication initiation, regulation of phage gene expression, or both.

A search of the databases revealed similarities between φCPG1, Chp1 and SpV4, a bacteriophage infecting Spiroplasma melliferum, a helical mollicute pathogenic for honey bees (Renaudin et al., 1987). However, the similarity is restricted to VP1, VP2 and VG4 and the genetic organization is different in SpV4 (Fig. 4). Moreover, while Chp1 and φCPG1 share similar G+C contents and codon usage, the G+C (32 mol %) and
codon usage of SpV4 closely reflect its *Spiroplasma* host (Renaudin et al., 1987).

**Sequence analysis of VP1, the major capsid protein**

Analysis of the nucleotide sequence of the φCPG1 genome revealed a hypothetical gene downstream of the *Bam*HI site, the translated amino terminal sequence of which was identical to the amino-terminal sequence determined for the 62 kDa band, confirming the high homology to the sequence of VP1 of Chp1 (Table 1). VP1 displayed the highest sequence conservation between φCPG1, Chp1 and SpV4, reflecting its important role as the major structural component of the capsid (Chipman et al., 1998; Renaudin et al., 1987; Storey et al., 1989a).

As reported by Storey et al. (1989a), the major structural protein, VP1, is similar to the bacteriophage capsid F protein of φX174. Alignment of VP1 from Chp1, VG1 from SpV4 and F protein sequences from coliphages φX174, φK, α3 and G4 allowed the identification of seven insertion loops, IN1–7, in VP1/VG1 proteins, which are absent in the F protein family (Chipman et al., 1998). Overlay of the VG1 structure on the three-dimensional structure of the F protein of φX174 further suggested that the larger insertion loop (IN5) lies on the exposed surface of the capsid protein and potentially trimerizes at each of the twenty three-fold axes of symmetry of the capsid, forming protrusions at the phage surface. Chipman et al. (1998) demonstrated such protrusions on purified SpV4 particles using cryo-electron microscopy and three-dimensional image reconstruction. Sequence alignment with VP1 of φCPG1 likewise identified an IN5 insertion sequence (Fig. 5). The IN5 insertion of φCPG1 consisted of 78 residues, with a predicted molecular mass of 8209 Da and a pI of 9-3. These gross properties are similar to those of IN5 of *Spiroplasma* phage (71 residues, 7422 Da and pI of 10-5), but are relatively dissimilar to those of IN5 of Chp1 (104 residues, 10836 Da and pI of 4-25). Alignment of the IN5 sequences revealed comparable similarity in the three phages (29-5% and 31-3% identity for φCPG1 vs Chp1

**Fig. 2.** Ultrastructural analysis of φCPG1. (a) φCPG1 phage particles seen by transmission electron microscopy in association with chlamydial envelopes in a lysate of GPIC-infected HeLa cells. Bar, 0-25 µm. (b) Negatively stained (2% uranyl acetate) purified φCPG1 particles. Centrally electron-lucent particles may be capsids lacking DNA. Bar, 0-2 µm. The inset shows an enlargement revealing the isocahedral configuration of φCPG1.

**Fig. 3.** Coomassie-blue-stained SDS-PAGE of highly enriched φCPG1. The positions of the 62 kDa (VP1), 26 kDa and 13 kDa polypeptide bands are indicated.
and SpV4, respectively), with positionally conserved gaps in φCPG1 and SpV4 (Fig. 5).

### Analysis of VG4, involved in DNA replication

The predicted VG4 protein of φCPG1 is a polypeptide whose counterpart in Chp1 (ORF4) is weakly homologous to gene A proteins of phages φX174, S13, z3, G4 and φK (BLAST E values ranging between $2 \times 10^{-3}$ and $0.075$), which are involved in DNA replication in these phages (Storey et al., 1989a). However, BLAST analysis revealed that the relationship with the coliphage protein family was not significant for VG4 of φCPG1 and the corresponding VG2 of SpV4 (E values $>3.6$ and $>0.17$, respectively). Amino-terminal extensions of 167 residues in gene A protein (φX174), and of 107 residues in ORF4 of Chp1 relative to VG4 of φCPG1, further contribute to the dissimilarity between VG4 and ORF4 of Chp1 and the coliphage gene A protein family.

Position-specific iterated BLAST (PSI-BLAST) analysis revealed weak similarities of VG4 and ORF4 with the gene A protein family of coliphages (E values ranging from 0.003 for phage z3 to 0.55 for phage G4, and from $1 \times 10^{-3}$ for phage G4 to 0.002 for phage z3, respectively). However, this analysis indicated that an internal segment conserved in VG4 and ORF4 displays high homology (E values $2 \times 10^{-28}$ and $2 \times 10^{-29}$, respectively) with a segment of the plasmid-encoded replication initiation protein (Rep) protein of Brevibacillus borstelensis, which is involved in rolling circle replication (RCR) (Ebusu et al., 1995). Consensus motifs characteristic of the superfamily I of proteins mediating RCR initiation, including two (putative) DNA-linking tyrosine residues, are conserved in both VG4 and ORF4 of the Chlamydia phages, in VG2 of SpV4 and in the gene A/A* protein family (Fig. 6) (Ilyina & Koonin, 1992). Moreover, these motifs are conserved in a number of bacteriophage replication initiation proteins and

---

### Table 1. Physical properties of predicted gene products of φCPG1, Chp1 and SpV4

Values in this table were generated for the VP1, VP2, VP3, VG4 and VG5 predicted structural genes starting and ending at nt 156–1817, 1923–2483, 2523–2930, 3602–4393 and 4428–148 of the φCPG1 genome, respectively. VP1, VP2 and VG4 of φCPG1 correspond to VG1, VG4 and VG2 of SpV4, respectively (Renaudin et al., 1987), while VG4 and VG5 correspond to ORF4 and ORF5 of Chp1 (Storey et al., 1989a).

<table>
<thead>
<tr>
<th>Molecular mass (Da)*</th>
<th>Sequence similarity†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>φCPG1 vs Chp1</td>
</tr>
<tr>
<td>VP1 61935</td>
<td>66850</td>
</tr>
<tr>
<td>VP2 20234</td>
<td>28548</td>
</tr>
<tr>
<td>VP3 16846</td>
<td>16681</td>
</tr>
<tr>
<td>VG4 31624</td>
<td>47560</td>
</tr>
<tr>
<td>VG5 9512</td>
<td>11469</td>
</tr>
</tbody>
</table>

* Molecular masses predicted using PEPTIDESORT.
† Percentage similarity/identity of the full-length protein sequence using GAP.
Fig. 5. Sequence alignment of the IN5 loop of the VP1 capsid protein family. Sequences of gene F protein of φX174 (P03641), VP1 of φCPG1 (this work), Chp1 (P19192) and SpV4 (P11333), were aligned using CLUSTAL W. The IN5 loop was defined by subtraction of the gene F protein sequence (Chipman et al., 1998). Residues that are identical in at least three of the sequences are boxed. Gaps are indicated by dashes.

Fig. 6. Sequence alignment of VG4 of φCPG1, ORF4 of Chp1, VG2 of SpV4 and gene A/A* protein of φX174, and comparison with motifs of Rep superfamily I. Partial sequences from φCPG1 (this work), Chp1 (P19189), SpV4 (P11334) and φX174 (P03631) were aligned using CLUSTAL W starting at position 33 of the φCPG1 sequence. Conserved Rep motifs are boxed and consensus sequence (derived from comparison of 59 Rep sequences) is indicated below the alignment [upper case, conserved in all sequences; lower case, conserved in half the sequences; u, bulky hydrophobic residue (I, L, V, M, F, Y, W); x, no consensus] (Ilyina & Koonin, 1992). Boxed positions correspond to residues that are identical in at least three of the sequences. Italicized sequence upstream of position 33 of the φCPG1 sequence was not subjected to alignment. The position of the start methionine of gene A* protein of φX174 is double underlined. Dashes indicate gaps in sequence.

related cyanobacterial and archaeal plasmid Rep proteins (Ilyina & Koonin, 1992).

Insertion of highly conserved phage sequence in the genome of C. pneumoniae strains

A 375 nt fragment of the genome of C. pneumoniae CWL029 (Kalman et al., 1999) displayed high similarity to a segment of the φCPG1 genome sequence, encompassing 274 nt at the 5′ end of VG4 and 101 nt of upstream sequence (Fig. 7). The C. pneumoniae sequence, Cpn0222, has been annotated as encoding a truncated polypeptide with weak homology to Chp1 ORF4 (33% sequence identity) (Kalman et al., 1999). Accordingly, nucleotide sequence similarity is not perceptibly above background between the Chp1 and C. pneumoniae genomes. In contrast, VG4 of φCPG1 was 73% identical to the predicted Cpn0222 truncated product of C. pneumoniae CWL029. Moreover, high identity at the nucleotide sequence level of the 375 base segment (75% identity) unequivocally demonstrated the relationship of the C. pneumoniae genome insert to φCPG1 genomic sequence.

The putative phage insert in C. pneumoniae is flanked by several genes that are not assigned and are also not conserved in C. trachomatis genomes (Kalman et al., 1999; Stephens et al., 1998). Moreover, no recognizable integration or recombination hot-spot is found in the sequence on either side of the insert. Identical inserts have also been found at the same genomic sites in other C. pneumoniae isolates of different geographical origin (personal communications, Drs Tim Read, TIGR, Rockville, MD, USA, and Joel Peek, Wyeth-Lederle, West-Henrietta, NY, USA), consistent with a recent acquisition of the phage insert by C. pneumoniae and subsequent clonal dispersion of these strains. Southern blot analysis using pCPG1 insert as a probe failed to reveal similar phage inserts in the GPIC genome (data not shown).

The presence of phage-derived sequence in C. pneumoniae strongly suggests that variants of a single phage or members of a group of highly related phages were recently able to infect both C. psittaci GPIC, which is isolated from guinea pigs, and C. pneumoniae, which is isolated from humans. Moreover, these findings indicate that phage genomic sequence has the capacity to integrate into the chlamydial genome.
**Fig. 7.** Alignment of a proposed *C. pneumoniae* phage genomic insert with φCPG1 genome sequence. The segment of the φCPG1 genome from nt 3397 to nt 3896 was compared to the corresponding segment of the *C. pneumoniae* CWL029 (Cpn) genome sequence [nt 62362 to nt 62861 of contig 3 (Kalman et al., 1999)], including the 375 nt proposed phage insert (highlighted in black), using GAP. Vertical lines indicate identity to the φCPG1 nucleotide sequence. Sequences of the VG4 amino terminus and of the *C. pneumoniae* VG4 truncate (Cpn0222) are displayed above and below the alignment. Residues conserved in φCPG1 and CWL029 and conserved Rep motifs are underlined and double underlined, respectively (also see Fig. 6).

### Other predicted genes

Apart from VP1 and VG4, for which function could be assigned tentatively based on sequence similarity, there were three other major non-overlapping ORFs on the genome of φCPG1 that could not be annotated based on sequence similarity alone. The predicted products, VP2, VP3 and VG5, were homologous to VP2, VP3 and ORF5 of Chp1 (Storey et al., 1989a), while only the VP2 protein was conserved in SpV4 (Renaudin et al., 1987). VP2 and VP3 correspond to abundant polypeptides of apparent molecular mass 30 and 16–19 kDa in Chp1, as determined by SDS-PAGE (Storey et al., 1989b), in close agreement with the calculated values (28–30 and 16–19 kDa). VP2 and VP3 of φCPG1 may likewise correspond to dominant bands of 26 and 13 kDa observed by SDS-PAGE (Fig. 3), although their calculated molecular masses differ somewhat (Table 1). Discrepancies in molecular mass may correspond to post-translational modifications or alternative starts in φCPG1.

Storey et al. (1989a) described several additional overlapping minor ORFs in Chp1 that do not have detectable similarity to other bacteriophage sequences. A comparison of all Chp1 minor ORFs with all minor ORFs of φCPG1 revealed only limited homology of two small ORFs, one of which lacks an AUG start, and resides in a segment which is comparatively more conserved in φCPG1 and Chp1 (not shown). Whether these ORFs are genes or artefacts due to high local sequence identity is unclear.

Hence, there is currently no firm evidence to support the existence of overlapping genes in φCPG1 or Chp1, although these do occur in φX174 and related phages.

### DISCUSSION

The origin of φCPG1 in our *C. psittaci* GPIC stock is unknown. Several possibilities can be considered. One is that φCPG1 was present in low titre in our GPIC stock and initially went undetected. As the strain was passaged in culture, the titre may have risen slowly over a period of several years, up to the point where it reached high titre as reflected by the abundance of the capsid protein (VP1) seen in Fig. 1 (lanes 2 and 3). A retrospective SDS-PAGE analysis profiling the appearance/disappearance of VP1 indicated that it was present in low amounts in our oldest samples (2 years old) and that the amount later rose and declined unpredictably (Fig. 1, lanes 5–13). This supports the hypothesis that φCPG1 infection of *C. psittaci* GPIC may maintain itself at low titre for long periods and implies that undetected low-titre phage infection of GPIC isolates may be widespread. An alternative possibility is that φCPG1 was accidentally introduced into our *C. psittaci* GPIC cultures after 1988 (after which date our *C. psittaci* GPIC inoculum stock was built from a single sample). Although this is unlikely, as GPIC was then the only *Chlamydia* strain maintained and stored in our laboratory, contamination of our stock could have occurred either from guinea pig reagents (e.g. convalescent sera), which we routinely received from Dr Rank's laboratory.
Coliphages related to ϕX174 possess a circular DNA genome that undergoes replication by means of a double-stranded replicative form and is packaged as a single-stranded molecule into virions inside the cytoplasm of infected bacteria. Phage progeny is then released in abundance upon lysis of the infected host bacterium. We took advantage of these presumed properties and of consistent previous findings with Chp1 (Storey et al., 1989b) to isolate and clone the entire genome of ϕCPG1. Sequence analysis of the cloned insert confirmed the relationship of ϕCPG1 with Chp1, in terms of general structure and predicted genetic architecture, including five conserved predicted genes (Storey et al., 1989a) (Fig. 4), encoding the major capsid protein VP1, corresponding to the 62 kDa protein observed by SDS-PAGE, VP2, VP3, VG4 and VG5.

Similarity was also observed between the ϕCPG1 genome and that of SpV4, a phage of the mollicute Spiroplasma melliferum (Renaudin et al., 1987) (Fig. 4). Two of the three Chlamydia phage predicted proteins that are conserved in SpV4 (VP1 and VG4/ORF4) display limited sequence similarity to proteins of other bacteriophages.

VP1 of ϕCPG1 (like its counterparts in Chp1 and SpV4) was weakly related to the capsid gene F protein family of enteric microphages including ϕX174, ϕ2, S13, G4 and ϕK. Full-length alignment of the gene F protein family with the VPI family defines a large internal loop, IN5, in VP1 proteins, which is absent in gene F proteins (Chipman et al., 1998). Chipman et al. (1998) speculate that IN5 capsid protrusions represent evolutionary substitutes for spike-forming protein G pentamers of the coliphages, and that IN5 sequence divergence may represent adaptation to different bacterial host ranges. In this context, it is surprising that the IN5 sequences of the two Chlamydia phages are relatively dissimilar. The relative divergence of capsid surface structures in the two Chlamydia phages may reflect different receptor structures in avian and guinea pig Chlamydia, and/or different lineages of the two Chlamydia phage capsid genes, a notion consistent with the observation of the narrow (restricted to avian C. psittaci) host-range of Chp1 (Richmond et al., 1990).

Storey et al. (1989a) also reported the homology of ORF4 of Chp1 to the gene A protein of phage ϕX174, in which it is involved in phage DNA replication (Hayashi et al., 1988). However, the full-length VG4 protein of ϕCPG1 and its SpV4 counterpart (VG2) share little homology with this gene product from either member of the ϕX174 group. Moreover, the primary sequence of ϕCPG1 VG4 is marginally more similar to its counterpart in SpV4 than to that in Chp1 (Table 1). This is somewhat surprising in view of the predicted housekeeping function of these proteins. We therefore used PSI-BLAST to investigate whether smaller conserved motifs could be identified in VG4. This analysis led to two significant results: 1, a previously described gene A/A* protein zinc finger motif (Kodaira et al., 1992) was not conserved in ϕCPG1, Chp1 and SpV4; 2, a larger segment encompassing several conserved motifs of DNA-binding Rep proteins (superfamily I) involved in RCR initiation of plasmids, phages and phagemids (Ilyina & Koonin, 1992) was conserved across ϕCPG1, Chp1, SpV4 and the coliphages (Fig. 6). Alignment of the full-length amino acid sequences based on the Rep homology also revealed that VG4 of ϕCPG1 lacked the amino-terminal extension of gene A protein and, as such, was similar to gene A* protein (Fig. 6), which results from an alternative translational start within gene A of the coliphages.

Moulder (1988) suggested that phages perhaps infected ancestral Chlamydia forms, prior to their irreversible commitment to an obligate intracellular lifestyle. However, our related studies (Hsia et al., 2000) also suggest that ϕCPG1 may gain access to RBs by efficiently attaching to EBs prior to internalization. This de facto eliminates the primary physical barriers on phage infection of intracellular chlamydiae and implies that phages have had—and continue to have—the opportunity to infect Chlamydia throughout their evolution. A ‘multiple entry point’ hypothesis is consistent both with the highly conserved genetic architecture of the two Chlamydia phages, and with the contrasting finding of relative dissimilarity—implying differential lineages—between two of their gene products, VP1 and VG4/ORF4. A plausible interpretation of these results is that, as has been observed with double-stranded-DNA tailed bacteriophages (Hendrix et al., 1999), Chlamydia phage evolution has been subject to frequent horizontal transfer exchanges with a large, as yet unidentified, heterologous genetic pool. This is supported by the existence of the closely related Spiroplasma phage and implies that close relatives of the Chlamydia phages may infect other free-living bacteria.

The high homology of a 375 bp segment of the C. pneumoniae genome to a segment of ϕCPG1 genome has several important implications. First, the high homology leaves no doubt as to the relationship between the ϕCPG1 and C. pneumoniae sequences: either one acquired it recently from the other, or both recently acquired it from a common source. In comparison,
homology of the *C. pneumoniae* genomic segment with the corresponding Chp1 sequence does not rise above background noise, highlighting the uniqueness of the φCPG1–*C. pneumoniae* relationship. Given the biology of chlamydial and bacteriophages, it is most likely that *C. pneumoniae* acquired this sequence from a bacteriophage closely related to φCPG1.

Identical 375 bp sequences have also been found in two other *C. pneumoniae* genomes. While this reflects the clonality of these strains, it also suggests that the inserted sequence may play an important role in *C. pneumoniae* biology. Indeed, if this insert were not essential, it would likely have been lost or inactivated by mutation in at least some of the strains. If the phage insert is essential to *C. pneumoniae*, it possibly is so by virtue of the product it potentially encodes, i.e. a 113 residue carboxy-terminal truncate of VG4. Homology of the integrated phage genome segment starts approximately 100 nt upstream of the start codon and stops immediately after the stop codon, suggesting that the VG4 truncate was likely generated by integration/deletion en bloc of phage sequence and did not involve further internal rearrangements/mutations in its genesis. This suggests that expression is possible by virtue of conserved promoter elements in the upstream sequence. However, the function of the expressed peptide is unclear, as only two of the three Rep motifs are found on the *C. pneumoniae* VG4 peptide, with the proposed tyrosine-containing DNA-binding domain missing entirely (Fig. 7).

In summary, we have reported the isolation, cloning and partial characterization of φCPG1, the second phage known to infect *Chlamydiaceae* and the first known to infect a member of the *Chlamydiaceae* with a mammalian host. The phylogenetic relationship of φCPG1 with a previously described phage of avian *Chlamydia* has been established, but it is unclear based on sequence analysis alone whether these phages have co-evolved with their host bacteria from a single ancestor or whether they, or specific phage genes, have been independently ‘acquired’ at later stages of evolution. The finding that a 375 nt sequence present in the genome of *C. pneumoniae* is highly homologous to a segment of the φCPG1 genome is consistent with *C. pneumoniae* having been recently infected by a bacteriophage closely related to φCPG1. Moreover, the previous observation of phage crystalline arrays in *Chlamydia*-like forms infecting Chesapeake Bay bivalves (Harshbarger et al., 1977), the isolation of phages from avian and guinea pig *C. psittaci* (Richmond et al., 1982; this work), and the likely existence of a phage infecting *C. pneumoniae* (Kalman et al., 1999), suggest that phage infection is more widely spread in the genus than previously thought, and implies that phage infection may play an important role in the development of chlamydial infection and disease. Phage φCPG1 will provide the opportunity for the first time to evaluate the role of phage infection in chlamydial disease using the ocular and genital guinea pig model systems.

**ACKNOWLEDGEMENTS**

The authors are grateful to Drs Jane Raulston and Priscilla Wryick (University of North Carolina at Chapel Hill, USA) and to Mrs Hélène Ohayon and Dr Pierre Gounon (Institut Pasteur, Paris, France) for performing the electron microscopy. They also thank Drs Joel Peck (Wyeth-Lederle Vaccines and Pediatrics, West-Henrietta, NY, USA), Tim Read, Claire Fraser (TIGR, Rockville, USA) and Robert Brunham (University of British Columbia Centre for Disease Control, Canada) for communicating their unpublished results, and Dr Mark Pallen (The Queen’s University of Belfast, UK) for helpful suggestions. The early part of this work was performed while the authors were members of the University of Rochester (Rochester, NY, USA). This research was supported in the US by NIH grant AI26280, and by Research Career Development Award A101057 (to P.B.), and in the UK by a project grant from the Wellcome Trust (to P.B. and R.C.H.).

**REFERENCES**


Received 20 December 1999; revised 21 March 2000; accepted 7 April 2000.