Analysis of the Complete Nucleotide Sequence of Chp1, a Phage which Infects Avian *Chlamydia psittaci*

By CHRISTOPHER C. STOREY, MERYL LUSHER AND SHIRLEY J. RICHMOND*

Department of Medical Microbiology, Medical School, University of Manchester, Oxford Road, Manchester M13 9PT, U.K.

(Accepted 29 August 1989)

SUMMARY

We report the complete nucleotide sequence of bacteriophage Chp1. The genome was found to be 4877 bases long and it potentially codes for 11 proteins. Open reading frames (ORFs) 6 and 7 lie within ORFs 2 and 1 respectively but are in a second reading frame. No significant DNA homology was found when Chp1 was compared to the EMBL database. The N-terminal amino acid sequences of the three structural proteins VP1, VP2 and VP3 were determined and it was found that they were encoded by ORFs 1, 2 and 3 respectively. Amino acid homology studies revealed that VP1 has homology with the major structural protein of bacteriophages φ174 and S13, and that the protein inferred from ORF 4 shows homology to the A proteins of φ174, S13 and G4. The genome of Chp1 has an organization similar to that of φ174 although it is 509 bases smaller. We propose that Chp1 is a member of the *Microviridae* but that it is sufficiently different to warrant its own subfamily which we have called the *Chlamydiavirinae*.

INTRODUCTION

Bacteriophage infection in *Chlamydia psittaci* was first described in 1982 when two isolates from domestic ducks were found to contain phage. This phage, Chp1, replicates in *C. psittaci* reticulate bodies during the course of the chlamydial growth cycle in eukaryotic cells. Chp1 was found to be a 22 nm icosahedral virus with a buoyant density of 1.37 g/ml in CsCl (Richmond et al., 1982). We have since shown that it has three structural polypeptides, VP1, VP2 and VP3 with apparent M₉ values of 73K, 30K and 16.5K respectively and that its nucleic acid is single-stranded, circular DNA of about 4800 bases. This led us to speculate that Chp1 is either an aberrant member of the *Microviridae* or is the first member of a new phage group (Storey et al., 1989). We now present and analyse the complete DNA sequence of Chp1. This analysis suggests that Chp1 has evolved from an ancestral member of the *Microviridae*.

METHODS

Source of phage. Bacteriophage Chp1 was grown in *C. psittaci* strain N352, then phage virions and phage DNA were extracted as described previously (Storey et al., 1989).

DNA sequencing. Double-stranded DNA was synthesized *in vitro* from native Chp1 single-stranded DNA using the Klenow fragment of DNA polymerase I and priming with random hexanucleotides as described in Storey et al. (1989). It was then cleaved with a range of restriction endonucleases to give fragments of suitable sizes for sequencing, ligated into Smal-cleaved and dephosphorylated M13mp18 and then transformed into *Escherichia coli* JM103Y. This is a derivative of JM103 which has been cured of the P1 lysogen (Messing et al., 1981). Three recombinant plasmids (Chp1 E1, Chp1 E4 and Chp1 E5) that contained inserts of Chp1 DNA (Storey et al., 1989) were also cleaved with restriction enzymes and the fragments subcloned into M13mp18 as described above. The DNA sequence was determined by the dideoxynucleotide chain termination method of Sanger et al. (1977). Sequences were analysed on 6% polyacrylamide buffer gradient gels and the reactions visualized by incorporation of [³²P]dATP followed by autoradiography of the dried gels (Biggin et al., 1983). The DNA sequence was analysed using the PCGENE (Genofit) computer programs.

0000-9143 © 1989 SGM
**N-terminal amino acid sequencing.** Chp1 virions were purified on potassium tartrate/glycerol gradients (Ashley & Caul, 1982). Phage polypeptides were separated by SDS-PAGE and then transferred by electroblotting to Immobilon-P membranes (Millipore) in CAPS buffer (10 mM-3-(cyclohexylamino)-1-propanesulphonic acid, 10% v/v methanol, pH 11.0). The phage polypeptides were stained with 0.1% Coomassie Brilliant Blue R-250, excised and then subjected to N-terminal amino acid sequence determination (Matsudaira, 1987) with an Applied Biosystems 470A protein sequencer and a 120 PTH analyser. Amino acid sequencing was carried out by the Bioscience Department, ICI Pharmaceuticals.

**RESULTS**

The complete DNA sequence of bacteriophage Chp1 (Fig. 1) was found to be 4877 bases long. This is smaller than any previously described single-stranded circular DNA phage and is 10% smaller than the genome of φX174 (5386 bases; Sanger et al., 1978). The GC content was very low at 36.6%. Most of the sequence was determined from templates obtained by cloning native Chp1 DNA directly into M13mp18. However, to complete the sequence, recombinant plasmids Chp1 E1, Chp1 E4 and Chp1 E5 were subcloned into M13mp18 to produce further templates. A directed sequencing strategy based on the restriction endonuclease cleavage map (Storey et al., 1989) was used and the whole genome was sequenced in both directions with each residue determined four to 10 times.

**Open reading frames**

These are shown in Fig. 1. Five major open reading frames (ORFs) (ORFs 1 to 5) were identified. These all have an ATG start codon which is preceded by a region that can function as a ribosome-binding site (Shine & Dalgarno, 1974) and so all potentially code for a protein. Seven other ORFs occur in the DNA sequence. ORFs 6 and 7 occur within ORFs 2 and 1 respectively but are in a different reading frame. ORF 8 potentially codes for a 36 amino acid polypeptide but 50% of this putative protein consists of arginine. It is very unlikely that such a protein would be functional and we interpret the region between ORFs 3 and 4 as non-coding. In addition, there are secondary start sites within ORFs 2, 4 and 5 designated 2a, 2b, 4a and 5a respectively, which potentially allow Chp1 to code for four extra proteins. The positions of all ORFs except ORF 8 are shown on the genetic map of Chp1 (Fig. 2); the sizes of their inferred protein products and the sequences of the ribosome-binding sites and start codons are listed in Table 1.

**Promoters**

No region was found in the Chp1 sequence that had homology to the *E. coli* consensus promoter (Hawley & McClure, 1983). This was not unexpected as the *Chlamydia* promoters that have been determined, i.e. the *C. trachomatis* major outer membrane protein gene promoter (Stephens et al., 1988) and the promoter for the r-RNA gene (Engel & Ganem, 1987), do not show strong homology to *E. coli* promoters.

**Hairpins and repeats**

A number of regions were found capable of forming hairpin-like structures in Chp1 DNA and the four most interesting are described. The intergenomic region between ORFs 1 and 2 (Igl) is capable of forming a large hairpin (h1, Fig. 2 and 3). The structure of this hairpin is similar to that of the proposed 44 nucleotide hairpin that occurs in φX174 between the genes for proteins F and G and is the site of primosome assembly (Shlomai & Kornberg, 1980). Downstream of h1 a smaller structure (h2, Fig. 2 and 3) was found that would contain the start codon for VP2. Similarly, a potential hairpin (h4, Fig. 2 and 3) occurs at the start of ORF 4 and contains the start codon for its predicted product. The hairpin structures h2 and h4 may therefore regulate the synthesis of VP2 and the protein encoded by ORF 4 respectively. A further hairpin, at the end of ORFs 2 and 6 (h3, Fig. 1 and 2), is followed by four thymine residues; this region may therefore function as a ρ-independent terminator (Rosenberg & Court, 1979). No tandem repeats were found in the Chp1 sequence.
Sequence analysis of chlamydiophage Chp1

Fig. 1. The complete 4877 base sequence of Chp1 DNA. The sequence is numbered from the G of the unique AccI site. ATG denotes start codons, GAGG denotes proposed ribosome-binding sites and arrows denote hairpins. The underlined amino acid sequences were confirmed by amino acid sequencing of the structural polypeptides.

Fig. 2. Genetic map of Chp1 showing all the open reading frames except ORF 8. The map is numbered in kb from the AccI site. ORFs 6 and 7 are in a second reading frame contained within ORFs 2 and 1 respectively. ORFs 2a, 2b, 4a and 5a are initiated from secondary start sites. The two intergenic regions are labelled Ig1 and Ig2 and h1 to h4 are regions capable of forming hairpin-like structures.

DNA homology

When compared to the EMBL DNA database (issue 17) no significant DNA homologies were detected.

Polypeptide sequence determination

Three structural proteins, VP1, VP2 and VP3, with apparent Mr values of 73K, 30K and 16.5K respectively were previously identified for Chp1 (Storey et al., 1989). These match the predicted Mr of the proteins potentially encoded by ORFs 1, 2 and 3. All three structural proteins were purified by SDS–PAGE and the amino acid sequence was obtained by a microsequencing method. The actual sequences matched the predicted sequences of VP1, -2 and -3 and so confirmed the locations of the genes for these proteins. The N-terminal amino acid sequences of VP1, -2 and -3 that were determined by microsequencing are shown underlined in Fig. 1. Neither VP1 nor VP2 contains an N-terminal methionine.
Fig. 3. Possible secondary structure of the DNA sequence of (a) Igl and (b) the start of ORF 4. Start codons are boxed and proposed ribosome-binding sites denoted by dots.

Table 1. Open reading frames of Chp1

<table>
<thead>
<tr>
<th>Open reading frame</th>
<th>Ribosome-binding site*</th>
<th>Position of start codon†</th>
<th>Number of amino acids‡</th>
<th>Predicted Mr, kDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF 1</td>
<td>GAAGAGGTTTAATATG</td>
<td>4460</td>
<td>595</td>
<td>66.850</td>
</tr>
<tr>
<td>ORF 2</td>
<td>TAGAGGAATTATG</td>
<td>1518</td>
<td>262</td>
<td>28.417</td>
</tr>
<tr>
<td>ORF 2a</td>
<td>ATAGGCAAATG</td>
<td>1782</td>
<td>175</td>
<td>19.310</td>
</tr>
<tr>
<td>ORF 2b</td>
<td>TCAGGAGCGAATG</td>
<td>1800</td>
<td>169</td>
<td>18.547</td>
</tr>
<tr>
<td>ORF 3</td>
<td>ATGGAGAAGGTAAATG</td>
<td>2309</td>
<td>144</td>
<td>16.681</td>
</tr>
<tr>
<td>ORF 4</td>
<td>ATGGAGATCTCTTTTAAATG</td>
<td>2960</td>
<td>399</td>
<td>47.560</td>
</tr>
<tr>
<td>ORF 4a</td>
<td>AAAGCTCGTATG</td>
<td>3539</td>
<td>206</td>
<td>24.665</td>
</tr>
<tr>
<td>ORF 5</td>
<td>TTAGGGAGATATCGTGATG</td>
<td>4168</td>
<td>96</td>
<td>11.469</td>
</tr>
<tr>
<td>ORF 5a</td>
<td>TAGAGATGTTGTGATG</td>
<td>4282</td>
<td>58</td>
<td>6.881</td>
</tr>
<tr>
<td>ORF 6</td>
<td>AAAGGAACTATG</td>
<td>2071</td>
<td>82</td>
<td>9.426</td>
</tr>
<tr>
<td>ORF 7</td>
<td>GCGGGATATG</td>
<td>1087</td>
<td>31</td>
<td>3.575</td>
</tr>
<tr>
<td>ORF 8</td>
<td>AAAGGATGTTTTCCGATG</td>
<td>2823</td>
<td>36</td>
<td>4.676</td>
</tr>
</tbody>
</table>

* ---, predicted ribosome-binding sites; ---, start codons.
† Position of the A of the start codon in the sequence shown in Fig. 1.
‡ The predicted number of amino acids and their Mr, include an N-terminal methionine except for ORFs 1 and 2 where amino acid sequence data show it is removed.
Sequence analysis of chlamydiophage Chp1

(a) 16  73
    SEVPTAEIRSSSFORS_G1ATTFEMLFLVFPVGVDEVFCSST6RSET_HLCHLVTTLVQVP
    VP1

    3

    2 3

(b) 181  237
    FTVEQGDKKGRNIIW|IHIVFGWKPKSEEQLEPYLGGKYRTDVRYRSRKLKWLKFG
    ORF 4 protein

    248  302

(c) 250  259
    FYVARNP
    ORF 4 protein

    342  351

Fig. 4. Amino acid homology between Chp1 and φX174 proteins. (a) Homologous region between Chp1 VP1 and φX174 F protein; (b and c) homologous region between the inferred Chp1 ORF 4 protein and the φX174 A protein region 1 and (c) region 2; : denotes same amino acid; · denotes similar amino acid.

Fig. 5. Comparison of the genetic maps of (a) Chp1 and (b) φX174 drawn to the same scale and aligned by the start of the genes for each of the major capsid proteins. M, Major structural protein; P, packaging; R, replication; S, structural protein; U, unknown. Shaded boxes show areas of local amino acid homology between VP1 of Chp1 and the F protein of φX174, and the two areas of homology between the inferred Chp1 ORF 4 protein and the A protein of φX174. Frameshifted genes, secondary start sites and intergenomic regions of Chp1 are labelled as in Fig. 2. The φX174 map is drawn from Sanger et al. (1978) and Godson et al. (1978a). The φX174 genes B, K and E are frameshifted and the A* gene is from a secondary start site.

Amino acid homology

Amino acid sequence comparisons were initially performed using an AMT 510-4 Distributed Array Processor by the Biocomputing research group at Edinburgh University (Coulson et al., 1987). Several local homologies were detected. VP1 was found to have a region of homology to the F protein (Sanger et al., 1978), the major capsid protein of φX174 (Tessman & Tessman, 1978). In addition, the inferred amino acid sequence of ORF 4 had two regions of homology with
Table 2. Significant homologies between the proteins of Chp1 and those of \( \Phi X174 \), S13 and G4

<table>
<thead>
<tr>
<th>Phage</th>
<th>Protein</th>
<th>VP1</th>
<th>VP2</th>
<th>ORF 4 protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Phi X174 )</td>
<td>A</td>
<td>-*</td>
<td>-</td>
<td>5.2\†</td>
</tr>
<tr>
<td>S13</td>
<td>A</td>
<td>-</td>
<td>-</td>
<td>5.9</td>
</tr>
<tr>
<td>G4</td>
<td>A</td>
<td>-</td>
<td>-</td>
<td>4.0</td>
</tr>
<tr>
<td>( \Phi X174 )</td>
<td>F</td>
<td>4.2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>S13</td>
<td>F</td>
<td>4.2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>F</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>K</td>
<td>3.7</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

* Alignment score less than 3.0.
† The alignment scores are in s.d. units and were calculated using the Dayhoff\(_{78}\) matrix with a gap penalty and bias parameter of 60, and 99 random permutations.

the A protein (Sanger et al., 1978), which is the nicking/closing protein of \( \Phi X174 \) (Tessman & Tessman, 1978). The homologous regions of the proteins are shown in Fig. 4 and the positions of the homologies are indicated on the genetic maps of Chp1 and \( \Phi X174 \) (Fig. 5).

Homologies between Chp1 proteins and the proteins of \( \Phi X174 \), G4 (Godson et al., 1978b) and S13 (Lau & Spencer, 1985) were further investigated using the algorithm of Needleman & Wunsch (1970) with the Dayhoff MDM\(_{78}\) matrix (Dayhoff et al., 1978). The gap penalty and bias parameter were both set at 60 (Dayhoff et al., 1978) and the number of random permutations used was 99. The computations were carried out using the program PCOMPARE (a part of the PCGENE suite of analysis programs). All of the potential Chp1 proteins were compared with all of the \( \Phi X174 \), G4 and S13 proteins. Those with a significant homology [an alignment score of greater than 3.0 s.d. units as suggested by Dayhoff et al. (1978) and Feng et al. (1985)] are shown in Table 2. VP1 has significant homology with the F proteins of \( \Phi X174 \) and S13, but not with the F protein of G4 (alignment score of 2.6). The ORF 4 protein of Chp1 has significant homology with the A proteins of \( \Phi X174 \), S13 and G4, but again the homology with G4 is less than is shown by \( \Phi X174 \) or S13. In addition VP1 had a significant homology (3.7 s.d. units) with the K protein of G4.

Despite the amino acid homology between the ORF 4 protein and the \( \Phi X174 \) A protein, no significant homology was found between the Chp1 nucleotide sequence and the A protein recognition sequence 5' CAACTTGATATTAATAACACTATAGACCAC 3' which is conserved among other isometric ssDNA phages (Heidekamp et al., 1982), even when a 25% mismatch was allowed.

**DISCUSSION**

The analysis of the DNA sequence of Chp1 showed it to have a small genome of 4877 bases, which has the potential to code for 11 proteins due to overlapping genes and secondary start sites. Although we have no evidence for the production of any proteins except from the major ORFs (1, 2 and 3), other phages make proteins from frameshifted genes and secondary start sites (Fiddes & Godson, 1979; Rasched & Oberer, 1986), so it is reasonable to assume that at least some of these additional ORFs are used in Chp1. An efficient use of a minimal amount of DNA is characteristic of single-stranded, circular DNA phages.

The DNA and amino acid sequence analysis of Chp1 sheds light on its evolutionary relationship to other phages. Previous evidence based on virion size, morphology and nucleic acid species led us to speculate that Chp1 may be an aberrant member of the Microviridae (Storey et al., 1989). The amino acid homologies and the similarity in genome organization now reinforce this view and suggest that Chp1 and the other Microviridae have evolved from a common ancestor. The amino acid homologies between Chp1 and \( \Phi X174 \), G4 and S13 are unlikely to be due to convergent evolution since there are two Chp1 proteins (VP1 and the ORF 4 protein) which independently show homology with the A and F proteins of these Microviridae. The genome organization of these phages is also remarkably similar. This is demonstrated in
Fig. 5 which compares genome maps of Chp1 and φX174 orientated by the start site of the gene for their respective major capsid proteins. The positions of the genes for the structural proteins (VP1, VP2 and VP3 for Chp1 and genes F, G and H for φX174) coincide as do the two main intergenicomic gaps (lg1 and lg2). The two regions of amino acid homology in ORF 4 and gene A also correspond. The complex region of the φX174 genome between genes K and J is missing in Chp1 where only one ORF (ORF 5) is found. This could indicate that a simpler packaging system is employed in Chp1 than in φX174.

The data presented here therefore provide evidence that Chp1 is a member of the Microviridae. However, since it is only distantly related to other isometric ssDNA phages and since its host species, C. psittaci, is an extremely specialized intracellular bacterium, we suggest that Chp1 is placed in a new subfamily of the Microviridae which is called the Chlamydiavirinae.

This proposed relationship between Chp1 and the Microviridae does not contradict our present understanding of the evolution of Chlamydiaceae (Weisburg et al., 1986; Moulder, 1988). Moulder pointed out that a free-living bacterium would be more accessible to phage infection than an intracellular one and he therefore suggests that chlamydiae became infected with phage before they became host-dependent (Moulder, 1988). This implies that a chlamydial ancestor acquired phage before the species recognized today had evolved and that phage and chlamydiae have coevolved since then over many millions of years. Some chlamydiae may have lost phage in the course of their intracellular evolution. However, the description of phage particles within C. trachomatis-like organisms recovered from clams (Harshbarger et al., 1977; Meyers, 1979) makes it probable that other chlamydiophages will be discovered. The phylogenetic relationships of such phages should throw further light on the evolution of the family Chlamydiaceae.

We are grateful to J. Young and Dr H. Gregory for determination of N-terminal amino acid sequences and to Dr A. F. W. Coulson for DAP database searches. This work was supported by a grant from the Wellcome Trust.

REFERENCES


3390

C. C. STOREY, M. LUSHER AND S. J. RICHMOND


(Received 26 June 1989)