Sequence and organization of the lactococcal prolate-headed blL67 phage genome

Catherine Schouler, S. Dusko Ehrlich and Marie-Christine Chopin

bIL67 is a broad-host-range prolate-headed phage that is active against Lactococcus cells. The complete phage genome sequence of 22 195 bp was established. Thirty-seven open reading frames (ORFs) organized in two clusters were identified. Functions were assigned to the putative products of six of the ORFs on the basis of comparison of the deduced amino acid sequences to known proteins, analysis of structural features of the proteins and search for conserved motifs. These were a DNA polymerase, a protein involved in recombination, a lysin, a terminase subunit, a structural protein and a holin.

Keywords: bacteriophage, DNA sequence, lactic acid bacteria

INTRODUCTION

Lactococcal bacteriophages remain a major problem in the dairy industry, since they delay or completely prevent acidification by the starter culture. Many virulent lactococcal bacteriophages have been isolated from unsuccessful fermentations and are classified into 12 species according to morphology and DNA homology (Jarvis et al., 1991). They have a double-stranded DNA genome ranging from 18 to 54 kb and carry either cohesive or redundant ends.

Information about genome organization is scarce and concerns only the most prevailing species of lactococcal phages. These species are usually designated according to their morphology and life cycle, and referred to as prolate-headed phages, which are always virulent, and small-isometric-headed phages, which are either virulent or temperate. The genome of virulent small-isometric-headed phages has most probably a modular organization. Such an organization, originally described for lambdoid phages (Campbell & Botstein, 1983), was suggested by Loof & Teuber (1986) and Jarvis & Meyer (1986) from heteroduplex studies that revealed homologous regions covering from 81 to 94% of the genome and distinct regions of non-homology. Certain regions of small-isometric-headed virulent bacteriophage genomes have been characterized, such as the lysin gene (Platteuw & de Vos, 1992), the mcp gene for the major coat protein (Chung et al., 1991), three genes encoding minor structural proteins (Kim & Batt, 1991), the ilaIM gene encoding a type-II methyltransferase (Hill et al., 1991), the epi gene encoding a putative regulator (Lakshmidevi et al., 1990), the per locus presumed to be an origin of replication (Hill et al., 1990), the region surrounding the cos site (Chandry et al., 1994; E. Biüenko, S. D. Ehrlich & M.-C. Chopin, unpublished) and a cluster of genes encoding structural proteins (R. Parreira & others, unpublished). Certain genes from temperate small-isometric-headed phages have also been identified: the int gene encoding the integrase (Lillehaug & Birkeland, 1993), the repressor (van de Guchte et al., 1994), the lysin gene and genes encoding structural proteins (Arendt et al., 1994). In the case of virulent prolate-headed phages less information is available, since only the lysin genes from three related phages (Skearman et al., 1989; Ward et al., 1993; Geis, 1992) and the N-terminal part of certain structural proteins (Schouler et al., 1992) have been characterized.

A better understanding of the bacteriophage life cycle is required for designing strategies to prevent phage infection. As a step towards this goal, the whole genome (22 195 bp) of the prolate-headed phage bIL67 has been sequenced.

METHODS

Bacterial strains and media. Escherichia coli JJCl28F', araD139 & (ara-lex7696 galE15 galK16 A(lac)X74 hsdR' hsdM' StrF' F'[lacF A(lacZ)M15 truD36], was grown at 37°C in LB broth (Miller, 1972). Lactococcus lactis IL1403 (Chopin et al., 1984) was grown at 30°C in M17 broth (Terzaghi & Sandine, 1975) in which lactose was replaced by glucose.

Phage bIL67 is from our collection and was propagated on L. lactis IL1403. It belongs to the c2 phage group (Jarvis et al.,

The GenBank accession number for the nucleotide sequence data reported in this paper is L33769.
Phage DNA was extracted as described for phage λ (Sambrook et al., 1989). Double-stranded DNA of phage M13mp18 (Yanisch-Perron et al., 1985) was purchased from Boehringer.

**Molecular cloning and DNA sequencing.** *E. coli* was transformed by electroporation (Dower et al., 1988). Digestion by restriction enzymes, alkaline phosphatase treatment, ligation, luminescence labelling, electrophoresis and hybridization were carried out according to the suppliers' protocols.

The random library of phage genome fragments was established as follows. bIL67 DNA was partially digested with DNaseI (Anderson, 1981). For this purpose, 500 ng of phage DNA, in 100 μl 50 mM Tris/HCl (pH 7.5)/1 mM MnCl₂, was digested with 2 × 10⁻⁴ or 2 × 10⁻⁵ units of enzyme for 5–30 min. The reaction was arrested by addition of EDTA to 50 mM followed by a phenol/CHCl₃ (1:1, v/v) extraction. Fragments obtained were made blunt-ended by treatment with T4 DNA polymerase (Boehringer) and were separated in 1:2% (w/v) agarose gel. The fragments ranging from 300 to 500 bp, 500 to 700 bp and 700 to 900 bp were excised from the gel and recovered by the Freeze-Squeeze procedure using Spin-X tubes as described by the manufacturer (Costar). The isolated fragments were then cloned in M13mp18 which was cut with *SalI* and dephosphorylated.

Single-stranded DNA from recombinant phages grown in *E. coli* J1C128F⁻ was isolated by the perchlorate method (Zimmermann et al., 1989) as adapted for the Beckman BIOMEK 1000 Laboratory Workstation (Sorokin et al., 1993). Single-stranded DNA was sequenced by using Taq polymerase (Applied Biosystems) on CATALYST (Applied Biosystems), or by the Sequenase (Applied Biosystems) procedure, manually. The dye-labelled standard sequencing primers of Applied Biosystems and Applied Biosystems Sequencer ABI-373 were used. Complementary DNA strands of M13 were obtained according to the protocol supplied with Dynabeads (strapavidin-coated magnetic beads) template preparation kit (Dynal).

Unclonable regions were synthesized by PCR. Sequencing of PCR products was performed essentially according to the protocol supplied with the Applied Biosystems PCR sequencing kit, on the DNA Thermal Cycler from Perkin-Elmer.

**Data handling and computer analysis.** DNA sequence assembly was performed using Staden's programmes (Staden, 1982). ORF predictions were performed with Genmark software (Borodovsky & McIninch, 1993). Sequence data were analysed using the University of Wisconsin Genetics Computer Group (UWGCG) package (Devereux et al., 1984). The search for sequence homology was carried out using FASTA (Pearson & Lipman, 1988) (in Swissprot release 27) and BLAST (Altschul et al., 1990) (in GenBank release 80). Search for blocks of homology was done using MACAW software (Schuler et al., 1991).

**PCR.** Reactions were performed using the Gene ATAQ Controller (Pharmacia-LKB) or the DNA Thermal Cycler 9600 (Perkin Elmer). The standard Taq polymerase buffer of Promega, 0.2 mM of each dNTP, 0.1 μM primers and 2.5 units Taq polymerase were used. The final volume of the reactions was 100 μl. Phage bIL67 DNA (50 ng) or phage M13 supernatant (2 μl) was used as starting material. The conditions of a cycle were: 94 °C for 1 min, 50 °C for 2 min, 72 °C for 3 min. In total, 30 cycles of amplification were run. After the PCR reactions were completed, DNA was loaded in 1% low-gelling agarose gel and purified using the Freeze-Squeeze procedure (Costar).

**RESULTS AND DISCUSSION**

**Determination of the bIL67 DNA sequence**

The DNA sequencing strategy used follows closely that developed for sequencing of the entire yeast artificial chromosome, within the context of the Bacillus subtilis genome sequencing project (P. Serrot, A. Sorokin & S. D. Ehrlich, unpublished). Briefly, a random library of 960 clones in M13mp18, covering the phage genome 27 times, was established. Seventy-five percent of the bIL67 sequence was determined by shotgun sequencing of 149 clones. This sequence was organized in 17 non-overlapping clusters of a size ranging from 270 to 2700 bp. To fill the gaps between the clusters, PCR-synthesized inserts of M13 phages corresponding to the ends of each cluster were hybridized with the random library. The positive M13 clones were sequenced. Close to 96% of the sequence was thus determined. This sequence, determined on only one strand in some regions, was completed by synthesis and sequencing of the complementary strand of appropriate clones (see Methods). The remaining 4% of the genome was apparently absent from the random library, probably because of toxicity and/or of structural instability in the cloning system used. The sequence of this last part was determined directly from PCR products spanning the gaps. Thus, a linear bIL67 DNA sequence of 22,195 bp was established, with an average redundancy of 6.2, each region being sequenced at least once on both strands. Restriction sites determined from the sequence analysis were in agreement with those found by endonuclease digestions. The total G+C content of the bIL67 phage genome is 36%, which is similar to that of the lactococcal hosts (Kilpper-Bälz et al., 1982).

**Genome organization**

The sequence was translated in all six frames and the criterion chosen for ORFs was the presence of a start codon (AUG, UUG or GUG) preceded by a ribosome binding site (RBS) complementary to the 3' end of the 16S rRNA of Lactococcus lactis (3'CUUUUCUCCA 5') (Ludwig et al., 1985). When no length limitation was set, 37 ORFs could be identified, 11 of which were shorter than 300 bp. The protein-coding regions were checked all along the sequence, using a programme developed for gene recognition in newly sequenced DNA (Borodovsky & McIninch, 1993), which is based on non-homogeneous Markov chains. The presence of the 37 ORFs was thus confirmed. The ORFs were organized in two divergent clusters (Fig. 1). It is tempting to correlate this organization with that deduced from transcriptional studies of the prolacte-phage e2 (Beresford et al., 1993), where early and late regions were identified. As discussed below, putative genes encoding lysis and a structural protein, which were expected to be expressed late, are present in the leftward orientated cluster (Fig. 1). We therefore speculate that this cluster corresponds to the late transcription region and, accordingly, the other cluster to the early transcribed region (Fig. 1).

Early and late regions are separated by two intergenic regions, the structure of which is detailed below.
Genome organization of phage bIL67

Fig. 1. Genetic organization of the bIL67 genome. The 22195 bp linear genome of bIL67 is represented as an open bar with vertical 1 kb markers. ORFs are shown in boxes either above or below the genome. They are at three different heights, depending on which of the three frames on each strand is used. Direction of transcription is indicated by arrows. Gene functions are indicated when assigned.

Upstream of the early and the late regions, putative promoters were identified, having an extended −10 consensus box (TATGCTATAAT and TTTTTGATATATT, respectively), and no significant resemblance to the −35 consensus box. Promoters of this type (A/TntnTGa/tTATAAT, lacking a clear −35 sequence) are known to allow σ70-dependent transcriptional initiation in E. coli (Kumar et al., 1993). One such promoter was recently shown to be functional in another lactococcal phage, bIL66 (E. Bidnenko, S. D. Ehrlich & M.-C. Chopin, unpublished). Fourteen of the 21 early genes overlap by 1–8 bp and might, therefore, be transcriptionally coupled. A putative terminator sequence (CCACTCAATCAAGGTGGTTTTTGTTT; underlined sequences can form a stem structure) is present 30 bp downstream of the last early gene. A striking feature of the early genes is their small average size, about 350 bp. Late genes are, on average, longer and eight of them overlap by 1–50 bp. Translational signals of the 37 identified genes are compiled in Table 1.

Intergenic regions

The cohesive ends (cos site) of the prolate-headed phage P001 have the sequence 5’ TCAAGGCCTA 3’ (Billard et al., 1992). Since the same sequence is present in the intergenic region of bIL67, in which transcription of early and late genes converge, we propose that it corresponds to the cos site of this phage. This cos site contains a single-stranded 3’ overhang, as already observed for several phages active on Gram-positive bacteria. Six different types of direct repeats and one inverted repeat have been identified in the vicinity of the cos site (Fig. 2). The repeats denoted R2 present a dyad symmetry, evocative of those found at the cos site of phage λ. Some of the structural features near the cohesive ends could be involved in recognition by terminase and host factors.

In the intergenic region upstream of the early and late genes (Fig. 1), which is 680 bp long, several short direct (7 bp) and inverted repeats are present. The role of these repeats is not known, but similar structures are often found in the DNA replication origin. If bi-directional replication were initiated in this region, it would be co-directional with phage transcription. Such co-directionality is observed, for instance, in the chromosome of E. coli and B. subtilis (Burland et al., 1993; Ogasawara et al., 1994; Sorokin et al., 1993).

Putative function of the ORFs

Using the BLAST (Altschul et al., 1990) and FASTA (Pearson & Lipman, 1988) programs, similarities between proteins specified by bIL67 ORFs and protein sequences present in databases were searched for. Significant homology was detected for translation products of three ORFs, encoding a recombinase, a lysin and a structural protein. The putative function of three additional proteins can be deduced from the presence of conserved domains and structural features in common with known proteins (a DNA polymerase, a terminase subunit and a holin). In the following sections the relevant ORFs are discussed.

ORFs 3, 4, 5 and 6

Five domains typical of DNA polymerases have been identified in the protein specified by ORF3 (Fig. 3). Although the overall similarities among distantly related DNA polymerases are weak, three exonuclease domains
Table 1. General features of the blL67 ORFs

<table>
<thead>
<tr>
<th>Gene designation</th>
<th>Start*</th>
<th>Stop*</th>
<th>Product size (aa)</th>
<th>Similarities</th>
<th>Translation start†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>15132(AUG)</td>
<td>15278(UGA)</td>
<td>48</td>
<td>AAAGAGGUGuuuaau AUG UUU AUU</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>15379(AUG)</td>
<td>15579(UAA)</td>
<td>66</td>
<td>GGAGGUGuuuucca AUG ACA AUU</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>15581(AUG)</td>
<td>16480(UAA)</td>
<td>299</td>
<td>AAAGGGuugauuuuca AUG CAG GAC</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16550(AUG)</td>
<td>16738(UGA)</td>
<td>62</td>
<td>GAGGUuaccca AUG GAU AA</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>16735(AUG)</td>
<td>17115(UAA)</td>
<td>126</td>
<td>AAAGGAcuacaaa AUG AAA AUU</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>17115(AUG)</td>
<td>17483(UGA)</td>
<td>122</td>
<td>GAGGGuaccaauu AUG GCG AUU</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>17562(AUG)</td>
<td>17735(UAA)</td>
<td>57</td>
<td>GAGGgacaaaaa AUG AUU AA</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>17735(AUG)</td>
<td>18085(UAA)</td>
<td>116</td>
<td>GAGGGgaaaagaa AUG GCU AA</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>18138(AUG)</td>
<td>18263(UAA)</td>
<td>41</td>
<td>GAGGau AUG GCG AA</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>18295(AUG)</td>
<td>18480(UAA)</td>
<td>61</td>
<td>AAGGguuacuaua AUG CUU CGG</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>18480(AUG)</td>
<td>18683(UAA)</td>
<td>67</td>
<td>GAGGGuuuaaaau AUG GCA UUA</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>18751(AUG)</td>
<td>18903(UAA)</td>
<td>50</td>
<td>AAGGGuuuaaaau AUG UUG ACU</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>18903(AUG)</td>
<td>19442(UAA)</td>
<td>179</td>
<td>AAAGGGuagcc auu AUG GAA AGC</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>19508(AUG)</td>
<td>19879(UAA)</td>
<td>123</td>
<td>GAGGgaacaa aaaa AUG AAA AUU</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>19892(AUG)</td>
<td>20056(UGA)</td>
<td>54</td>
<td>AAAGGGuucagga AUG GGAUAU</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>20053(AUG)</td>
<td>20532(UAG)</td>
<td>159</td>
<td>GAAAGGgaucuaacacau AUG AAC AUU</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>20554(AUG)</td>
<td>20887(UAA)</td>
<td>107</td>
<td>AAAGGggaacgaa AUG UUU AAC</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>20870(AUG)</td>
<td>21172(UAA)</td>
<td>100</td>
<td>GAGGGuacucu AUG A UU AAU</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>21194(AUG)</td>
<td>21568(UAA)</td>
<td>124</td>
<td>GAGGacuuaaau AUG AAA C UU</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>21573(AUG)</td>
<td>21779(UAA)</td>
<td>68</td>
<td>GAGGguuaaaau AUG GAA UAA</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>21780(AUG)</td>
<td>21953(UAG)</td>
<td>57</td>
<td>GAGGGuuaau AUG GUU UA U</td>
<td></td>
</tr>
<tr>
<td>Late</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>14448(AUG)</td>
<td>14299(UAA)</td>
<td>49</td>
<td>GGAGGGuuaagc AUG GCA GAA</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>14288(AUG)</td>
<td>13803(UAA)</td>
<td>161</td>
<td>abi-105 target</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>13757(AUC)</td>
<td>13077(UAG)</td>
<td>226</td>
<td>GagGGuugauaau AUG AUC AA</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>13067(AUG)</td>
<td>12288(UAA)</td>
<td>279</td>
<td>AAAAGGgagcuaaua AUG GCA UAA</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>12277(UUG)</td>
<td>10832(UAA)</td>
<td>481</td>
<td>AAAGGGA gcuaauaaua AUG A UU AA</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>10801(UUG)</td>
<td>10097(UAG)</td>
<td>234</td>
<td>AAGAAAGG gaguaauaaua AUG GCU GAA</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>10079(UAG)</td>
<td>9462(UAA)</td>
<td>205</td>
<td>AGGAGG Gauuaauuuaua AUG GCA UAA</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>9390(AUG)</td>
<td>9136(UAA)</td>
<td>84</td>
<td>AGGAGGauuagaaa AUG AAG AA</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>8948(AUG)</td>
<td>8544(UAA)</td>
<td>134</td>
<td>GAGGAGUuaauuuu AUG GCA AAC</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>8551(AUG)</td>
<td>6689(UAG)</td>
<td>620</td>
<td>Tail protein</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>6390(AUG)</td>
<td>4834(UAA)</td>
<td>518</td>
<td>AAGAGG uaaauuuaua AUG GCU AAA</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>4834(AUG)</td>
<td>4535(UAA)</td>
<td>99</td>
<td>AAGAGGuuaauuuuaua AUG GCU A UG</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>4542(UUG)</td>
<td>2656(UAA)</td>
<td>628</td>
<td>GAGG Guuccggca UUG AAU AAG</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>2663(AUG)</td>
<td>1725(UAG)</td>
<td>312</td>
<td>AGAAAGG Gauauaauuuuuu AUG AAU AA</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>1670(AUG)</td>
<td>477(UAA)</td>
<td>397</td>
<td>GAGGGuuaaauu AUG GAU AGU</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>458(AUG)</td>
<td>165(UAG)</td>
<td>97</td>
<td>AAAGGAG aagagag AUG A UU AU</td>
<td></td>
</tr>
</tbody>
</table>

16S rRNA 3’ UCUUUCCUCCA 5’

*Numbers refer to position in the sequence; start and stop codons are shown in parentheses.
†The RBS and the beginning of the ORFs are shown in capital letters.

![Fig. 2. Region of cohesive ends. The direct repeats are represented by boxes. The inverted repeats are indicated as arrows below the horizontal bar representing the phage DNA. Ends of ORF 21 and ORF 37 are represented as arrows on the horizontal line. Scale, sequence of the repeats and the cos region are shown below the figure. Cohesive ends are boxed.](image-url)
located in the N-terminal region and six domains involved in the polymerization function, located in the C-terminal region, have been identified in 27 different enzymes (Blanco et al., 1991). Domains corresponding to 3'-5' exonuclease activity (I, II and III) were found in ORF 3 (Fig. 4). The Y residue in domain III, which plays an important role in catalytic activity, is replaced by an L residue in ORF 3, as previously observed for the B. subtilis DNA polymerase BsIII (Blanco et al., 1992). Two domains (region 2a and region 2b) out of the six involved in the polymerization function were also found in ORF 3. Remarkably, two additional polymerase domains, known as region 4 and region 5, are present downstream of ORF 3 in ORF 4 and ORF 5, respectively. Moreover, proteins encoded by ORFs 4 and 5 are basic (pI 10.43 and pI 10.20, respectively) and have one and two helix-turn-helix motifs, respectively, which is indicative of DNA binding capacity. In contrast, polymerase regions 1 and 3 were not identified. It is possible that the proteins encoded by ORFs 4 and 5 are subunits of a DNA polymerase. Interestingly, the cluster composed of ORFs 4, 5 and 6 is flanked by direct repeats of 25 bp (AATTAAG-TCATCAAGTTAAGAGG), which include a ribosome binding site. This organization is evocative of evolution by exchange of DNA modules, via a Campbell type recombination between two circular molecules, a
process proposed for lambdoid phages (Campbell & Botstein, 1983).

**ORF 13**

The product of ORF 13 shares 37.5% identity and 57.1% similarity to the Erf protein of *Salmonella typhimurium* phage P22 (Poteete, 1982). This protein is required for the circularization of the linear chromosome of the phage, by homologous recombination between its repetitive ends (Weaver & Levine, 1977). Since bIL67 has cohesive ends, the ORF 13 product probably plays a different role. In phage \(\lambda\), the homologous recombination directed by the red system was shown to increase the amount of packageable DNA (Smith, 1983). The homologous recombination system of phage P22 and phage \(\lambda\) are structurally unrelated (Poteete et al., 1991) but functionally interchangeable (Poteete & Fenton, 1984). We suggest that ORF 13 plays a role similar to red. Interestingly, ORF 12, located upstream of ORF 13, encodes a protein of 51 aa. The position of the gene and the size of the protein are evocative of the accessory recombination function (arf) gene of P22 (Poteete et al., 1991). Nevertheless, no significant homology has been observed between ORF 12 and arf gene products. The ORF 12/13 cluster is flanked by two direct repeats of 31 bp with only two mismatches (in bold letters) (TGTGACAAAAGAAAGCAACGT TATAAATTA and TGTGACAAAAGAAATAACGTTATAAATTA) which suggests that they may also correspond to an exchangeable module. Interestingly, a promoter consensus sequence (de Vos, 1987) is present in these repeats (underlined). No homologies with other components of the P22 or \(\lambda\) recombination systems have been observed in bIL67.

**ORF 23**

The ORF 23 product shares over 90% identity with the protein encoded by the gene located upstream of the lysin gene of the lactococcal prolate-headed phages \(\phi\)VML3 (Shearman et al., 1989) and c2 (Ward et al., 1993), respectively. The function of this gene remains unknown in both cases. Moreover, the ORF 23 product shares 41% identity with the protein encoded by ORF 3 of the virulent isometric-headed bIL66 phage (E. Bidnenko, S. D. Ehrlich & M.-C. Chopin, unpublished). ORF 3 of bIL66 is located in the region which is a target of the abortive infection mechanism \(abi-105\), active against isometric- and prolate-headed phages (including bIL67) (Gautier & Chopin, 1987). ORF 23 may thus be the target of \(abi-105\). ORF 3, located near the cohesive ends of bIL66, is a middle-expressed gene, while ORF 23, located in the middle of the genome, might be one of the first late genes of bIL67.

**ORF 24**

The ORF 24 product shares very high identity (up to 95%) with the proteins encoded by the three lysin genes already sequenced from lactococcal prolate-headed phages, \(\phi\)VML3 (Shearman et al., 1989), P001 (Geis, 1992) and c2 (Ward et al., 1993). The published \(\phi\)VML3 lysin gene sequence (Shearman et al., 1989) has been recently corrected in the database (accession number X16178) and is in good agreement with the one from phage c2. However, in both cases, the position of the start codon of the gene was unclear. This question was recently answered by the report of the purification of the phage P001 lysin protein and the determination of its N-terminal amino acid sequence, MKVSQ (A. Geis & S. Hertwig, unpublished). Comparison between this sequence and the one deduced from the nucleotide sequence, revealed that the isoleucine codon, AUC, is used as a start codon (it is translated as the first methionine). Sequence comparisons revealed that the start codon is also AUC in phage bIL67 but AUA in phages c2 and \(\phi\)VML3. AUA has already been found as a start codon in vivo in *E. coli* (Belin et al., 1979) but the use of AUC as a start codon has never been reported before. Use of a very rare start codon could be a way to decrease translation of lysin during phage multiplication. The bIL67 lysin gene shares 40.3% and 49.5% homology with those of *B. subtilis* phages PZA (Paces et al., 1986) and \(\phi\)20 (Vlcek & Paces, 1989), respectively, which were shown to exhibit muramidase activity.

**ORF 31**

The ORF 31 product shares 99% identity with the N-terminal part of protein POA17 from the prolate-headed phage \(\phi\)197 (Schouler et al., 1992). However, antibodies raised against POA17 were shown to react with two minor structural proteins of molecular mass 45 and 46 kDa, respectively, while ORF 31 encodes a protein of 66-2 kDa. It is possible that this 66-2 kDa protein is processed to fragments of 45 and 46 kDa. These proteins were localized at the end of the tail of phage \(\phi\)197 by electro-immuno microscopy (H. Neve & P. Ritzenthaler, personal communication). A calcium-binding domain (Nakayama et al., 1992), consisting of a 13 residue loop (DKNHDGKVSQDEM) flanked by \(z\)-helical domains, named EF-hand, was found from position 413 to 425. This is most probably related to the calcium requirement for the development of bIL67 and numerous other phages (Potter, 1970). However, it is still unknown if calcium plays a role in the lactococcal phage adsorption via a depolarization of the membrane, as observed for T4 (Letellier & Labedan, 1985), or participates in DNA injection, as observed for the PL-1 *Lactobacillus* phage (Watanabe & Takesue, 1972).

**ORF 32**

A search for conserved motifs suggests that the ORF 32 gene product could be a DNA binding protein with ATPase activity. Two ATP binding domains and a helix-turn-helix motif were found in the product of ORF 32. The ATP/GTP-binding site motif B (Walker et al., 1982) was found at position 178–190 (TLGGREQVLVID). The conserved domain which could be involved both in ATP binding and in phosphorylation of the substrate (Brenner, 1987), was found at position 459–482. This domain is observed in a
The primary sequence of protein 37. For comparison, the primary sequence of φ29 protein 14 is shown above the protein 37 sequence. Identical amino acids are represented by vertical bars and homologous amino acids by colons. The charged side chains are indicated by + or − above the sequence. Potential transmembrane domains, defined as regions of 20 generally hydrophobic residues with no net charge, are underlined. Potential β-turn regions as predicted by the Chou–Fasman prediction algorithm are indicated by ‘t’. The putative dual start motifs are indicated by arrows.

**CONCLUSIONS**

bIL67 is the first prolate-headed phage from *L. lactis* for which the complete genomic sequence has been determined. Its genome is small and very compact since non-coding regions represent only 5% of the chromosome. It is organized in two transcriptional units, the early region and the late region, which represent 31% and 64% of the genome, respectively. Searches for protein homology allowed assignment of putative functions to products of three genes among the 37 identified. However, only one of these proteins shows homology with the product of a heterologous phage (ref gene of *S. typhimurium* phage P22). The three other assignments of function were based on analysis of protein structure or on the search for conserved domains. Phage proteins commonly show little or no amino acid homology with proteins from databases. Even the proteins which have the same function on different phages show little sequence conservation, and the similarities are mainly observed at the structural level, as illustrated by the case of the holin (Young, 1992) or the connector (Donate et al., 1993).

The presence of several different direct repeats (two discussed above and a third one flanking ORF 24) suggests that, like other phages (Campbell & Botstein, 1983), lactococcal phages evolve by exchange of DNA modules. Moreover, these exchanges seem to involve essential functions such as DNA replication and recombination.

**ACKNOWLEDGEMENTS**

We thank Alexei Sorokin and Pascale Serror for introducing us to the large scale sequencing strategy, Vasco Azevedo for the *E. coli* strain JJCl28F*, Nathalie Galleron for technical advice, Christiane Nivet for the use of GeneMark software, Horst Neve and Paul Ritzenthaler for communication of unpublished results. We also thank Costa Anagnostopoulos and Alain Chopin for critical reading of the manuscript. This work was supported, in part, by EEC Grant BIOT-CT91-0263.

**REFERENCES**


Received 20 June 1994; revised 14 July 1994; accepted 20 July 1994.