APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 1995, p. 4348–4356 Vol. 61, No. 12
0099-2240/95/$04.00+0

Sequencing and Analysis of the Prolate-Headed Lactococcal Bacteriophage c2 Genome and Identification of the Structural Genes

MARK W. LUBBERS,[1]* NICK R. WATERFIELD,[2] TOM P. J. BERESFORD,[1]†
RICHARD W. F. LE PAGE,[2] AND AUDREY W. JARVIS[1]

Biological Science Section, New Zealand Dairy Research Institute, Palmerston North, New Zealand,[1] *and Department of Pathology, University of Cambridge, Cambridge CB2 1QP, England*[2]

Received 2 June 1995/Accepted 20 September 1995

The 22,163-bp genome of the lactococcal prolate-headed phage c2 was sequenced. Thirty-nine open reading frames (ORFs), early and late promoters, and a putative transcription terminator were identified. Twenty-two ORFs were in the early gene region, and 17 were in the late gene region. Putative genes for a DNA polymerase, a recombination protein, a sigma factor protein, a transcription regulatory protein, holin proteins, and a terminase were identified. Transcription of the early and late genes proceeded divergently from a noncoding 611-bp region. A 521-bp fragment contained within the 611-bp intergenic region could act as an origin of replication in *Lactococcus lactis*. Three major structural proteins, with sizes of 175, 90, and 29 kDa, and eight minor proteins, with sizes of 143, 82, 66, 60, 44, 42, 32, and 28 kDa, were identified. Several of these proteins appeared to be posttranslationally modified by proteolytic cleavage. The 175- and 90-kDa proteins were identified as the major phage head proteins, and the 29- and 60-kDa proteins were identified as the major tail protein and (possibly) the tail adsorption protein, respectively. The head proteins appeared to be covalently linked multimers of the same 30-kDa gene product. Phage c2 and prolate-headed lactococcal phage bIL67 (C. Schouler, S. D. Ehrlich, and M.-C. Chopin, Microbiology 140:3061–3069, 1994) shared 80% nucleotide sequence identity. However, several DNA deletions or insertions which corresponded to the loss or acquisition of specific ORFs, respectively, were noted. The identification of direct nucleotide repeats flanking these sequences indicated that recombination may be important in the evolution of these phages. Several poorly conserved ORFs and a poorly conserved module containing several structural genes that might be involved in phage-specific properties, such as host range determination, were identified.

Phage attack on lactococcal starter cultures was identified six decades ago as a cause of fermentation failure in the production of dairy products (52), and it remains the major cause of fermentation failure today. Lactococcal phages are classified into 12 species on the basis of morphology and DNA homology (28). Most lactococcal phage species conform to two morphotypes: B1, small isometric head and long noncontractile tail, and B2, prolate head and long noncontractile tail. Nearly all small isometric phages fall into two species groups, designated 936 and P335, represented by type phages P008 and P335, respectively (28). The prolate phages form one highly homologous species group, c2, represented by type phage c6A (28). All members of a species share extensive DNA homology and also usually have similar structural protein profiles, at least for the major protein bands.

Despite the apparently close relationship of the prolate lactococcal phages (26), they exhibit some important phenotypic differences, such as different host ranges. The complete genome sequence of a prolate phage, bIL67, has recently been reported (45), and genes for a putative DNA polymerase, holin, recombination protein, structural gene, and terminase have been identified. We report here the complete genome sequence of prolate phage c2 and the identification of putative genes involved in regulation of transcription, a putative origin of replication, early and late promoters, genes for the 11 structural proteins detected by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and head, tail, and possible tail adsorption proteins. Comparison of the sequences of phages bIL67 and c2 also revealed genes which may encode phage-specific properties and revealed possible modes of lactococcal phage evolution.

* Corresponding author. Mailing address: New Zealand Dairy Research Institute, Private Bag 11 029, Palmerston North, New Zealand. Phone: +64 6 350 4649. Fax: +64 6 350 4740. Electronic mail address: lubbers@nzdri.org.nz.

† Present address: National Dairy Products Research Centre, Moorepark, Fermoy, Co. Cork, Ireland.

MATERIALS AND METHODS

Strains, media, and culture conditions. Bacteriophage c2 (40) was propagated in *Lactococcus lactis* subsp. *lactis* MG1363 (17). Preparation of phage stocks and concentration and purification of phages by CsCl gradient centrifugation were carried out as previously described (26). *Escherichia coli* JM109 (55) was used for routine cloning and was cultivated in Luria-Bertani broth as described by Sambrook et al. (43). The following antibiotics incorporated into the media were used at the concentrations indicated (in micrograms per milliliter): for *L. lactis*, erythromycin, 5; streptomycin, 200; for *E. coli*, chloramphenicol, 25; tetracycline, 12.5.

DNA methods. Small-scale preparation of plasmids from *E. coli* was done as described by Holmes and Quigley (24) or He et al. (21). Isolation of large-scale plasmids from *E. coli* was done by a modification of the method of Ish-Horowicz and Burke (25). Plasmid DNA was isolated from lactococci as described by O'Sullivan and Klaenhammer (39). Bacteriophage c2 DNA was isolated as described by Jarvis (26). Restriction endonuclease digests, agarose gel electrophoresis, and DNA ligations were done by standard procedures described by Sambrook et al. (43).

DNA sequencing and sequence analysis. Double-stranded DNA (plasmid or c2 genome) was sequenced by the dideoxy chain termination method using the f-mol DNA cycle sequencing system (Promega, Madison, Wis.) according to the manufacturer's protocols for direct incorporation of α-^{35}S-dATP or primers end labelled with [γ-^{32}P]ATP. Annealing temperatures were between 50 and 70°C, depending on the primer-template combination. The Genetics Computer Group Inc. sequence analysis package (18) was used to assemble and analyze sequence data.

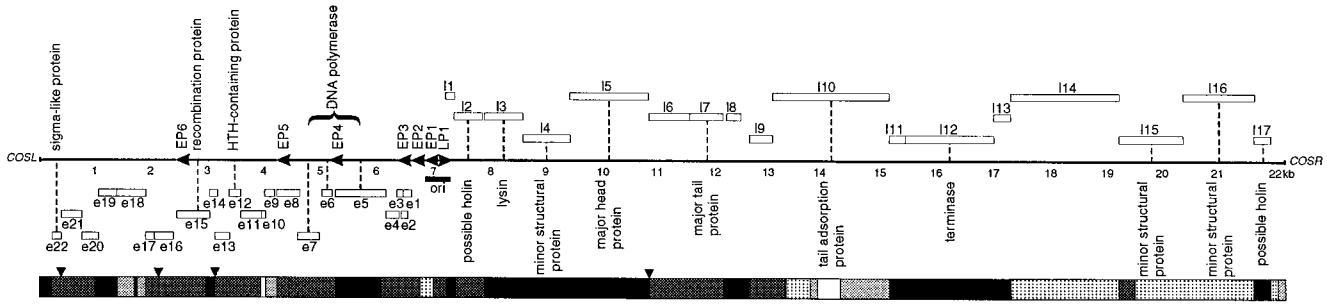

FIG. 1. Organization of genome of phage c2. The numbers directly beneath the genome indicate the distance from *cosL* (in kilobases). Open boxes below or above the genome indicate leftward (*e1* to *e22*)- or rightward (*l1* to *l17*)-reading ORFs in the early and late regions, respectively. The different box heights indicate the three different reading frames. Putative gene functions are indicated. Arrowheads indicate the leftward-directed early promoters (EP1 to EP6) and the rightward-directed late promoter (LP1). An origin of replication (*ori*) region is indicated by the solid bar. The shaded bar beneath the genome indicates percent nucleotide identity with the genome of phage bIL67 and DNA deletions: ■, >90%; ▩, 80 to 90%; ▨, 70 to 80%; ⬚, <70%; □, DNA deleted from phage bIL67; ▼, location of deletions of DNA from phage c2.

Sequence database analyses. Nucleotide and predicted amino acid sequences were compared to sequences in protein (SwissProt, PIR, and Brookhaven Protein Data Bank), nucleotide (GenBank and EMBL), translated nucleotide (GenPept), and protein motif (BLOCKS) databases by using the BLAST (2), MPsrch (47), and BLOCKS SEARCHER (23) programs.

SDS-PAGE of phage structural proteins. The concentrated phage sample was dialyzed against water before fractionation by SDS-PAGE according to the method of Laemmli (31). Samples were heated for 10 min at 100°C in 1% SDS–2.5% β-mercaptoethanol–6.25 mM Tris-HCl (pH 6.8)–5% glycerol before loading, and 12 or 8% acrylamide gels were used. The gels were stained with Coomassie blue.

Antiserum production. Polyclonal rabbit antiserum to phage was prepared as follows. Rabbits were inoculated with a total of 0.5 ml of 10^{12} PFU/ml plus 0.5 ml of Freund's incomplete adjuvant, given as two subcutaneous and two intramuscular injections. Two successive injections of phage were given, as described above, after 21 and 49 days. Rabbits were bled 7 days after the last injection, and antiserum was collected (43).

Antibodies which recognized specific phage structural proteins were eluted from Western blots (immunoblots) of whole phage proteins as follows. Proteins fractionated on SDS-polyacrylamide gels were electrophoretically transferred with a Trans-Blot semidry electrophoretic transfer cell (Bio-Rad Laboratories Pty. Ltd., Sydney, Australia) onto polyvinylidene difluoride (PVDF) sheets by using Towbin transfer buffer (25 mM Tris, 192 mM glycine, 10% methanol [pH 8.3]). Proteins were identified by comparison to a sample lane stained with Coomassie blue. Regions of the PVDF sheets which contained specific transferred proteins were cut from the sheet and incubated with the antisera, and then the bound antibodies were eluted, as previously described (43).

Western blotting. Proteins were fractionated on SDS-polyacrylamide (12% acrylamide) gels and electrophoretically transferred to PVDF sheets by using CAPS transfer buffer (10 mM 3-cyclohexylamino-1-propanesulfonic acid, 10% methanol [pH 11]) and the Mini-V (Bethesda Research Laboratories, Gaithersburg, Md.) or Trans-Blot (Bio-Rad Laboratories) apparatus. Blotted proteins were incubated with the antisera, and the bound antibodies were visualized with horseradish peroxidase-labelled anti-rabbit immunoglobulin G antibodies by using the enhanced chemiluminescence system (Amersham International plc, Little Chalfont, Buckinghamshire, England) according to the manufacturer's instructions. Proteins were identified by comparison to a sample lane stained with Coomassie blue.

N-terminal sequencing of proteins. Proteins were fractionated by SDS-PAGE and blotted onto PVDF sheets as described for Western blotting. Protein bands were stained with Coomassie blue, cut from the PVDF sheets, and sequenced directly by Edman degradation on an Applied Biosystems (Scoresby, Victoria, Australia) model 470A protein sequencer.

Immunogold electron microscopy. Concentrated CsCl-purified phage samples were applied to copper grids (200 mesh, carbon-coated Formvar) and then rinsed with TAM (10 mM Tris-HCl [pH 7.0], 10 mM sodium azide, 1 mM magnesium chloride). The grid was then incubated with antiserum for 30 min to 1 h at 37°C, rinsed with water, incubated for 1 h at room temperature in anti-rabbit immunoglobulin G-gold conjugate (5-nm particle size, Sigma G-7277) which was diluted fivefold in 0.5 M NaCl (pH 6 to 8)–0.1% bovine serum albumin–0.05% Tween 20–5% fetal bovine serum, and then rinsed with water. Stain (1% phosphotungstate, 1% ammonium molybdate [pH 4.0]) was applied to the grid for 2 min, and then the grid was dried and examined.

Nucleotide sequence accession number. The GenBank accession number for the nucleotide sequence data is L48605.

RESULTS

Cloning and sequencing of phage c2 genome. A shotgun approach was used to clone fragments of the phage c2 genome. Genomic DNA was digested with various restriction endonucleases, ligated into the *E. coli-L. lactis* shuttle vectors pSA3 (14) or pFX3 (54), and transformed into *E. coli*. Approximately 70% of the c2 genome was cloned in this way. Those regions not cloned included almost all of the early transcribed region, suggesting that many of the early genes may function to subvert host metabolism for the production of phage proteins and so are lethal to the host cell. In contrast, the entire late region was readily cloned.

The entire genome was sequenced directly from c2 genomic DNA by using primers designed from the sequence data obtained from the cloned fragments. A linear sequence of 22,163 bp was obtained. The predicted restriction endonuclease map of the sequence corresponded to the experimentally determined restriction map (40).

Identification and organization of ORFs. Putative open reading frames (ORFs) in the c2 genomic sequence were identified by using the following criteria: (i) the ORF could encode ≥20 amino acids; (ii) the ORF began with an AUG, GUG, UUG, or AUA start codon; and (iii) either the ORF was preceded by a Shine-Dalgarno (SD) sequence or the possibility of translational coupling existed. In all, 39 ORFs were identified: 22 reading from right to left and located in the early gene region (5) and 17 reading from left to right and located in the late gene region (Fig. 1; Table 1). Between the divergently oriented ORFs was a 611-bp intergenic region. The free energies (ΔG) of binding of the SD sequences to the 3′ end of the *L. lactis* 16S rRNA (Table 1) and the spacing between the SD sequences and the putative start codons of all ORFs corresponded well to those observed for *L. lactis* genes (−8.4 to −17.8 kcal/mol [~−35 to −74.5 kJ/mol] and 5 to 9 nucleotides, respectively [15]), suggesting that many ORFs are likely to be expressed genes.

All ORFs were closely linked and intergenic regions (gaps between ORFs) were small (Table 1), with a total of only 1.5 kb (7%) of the genome being a noncoding sequence. Several of the ORFs overlapped, suggestive of the potential for translational coupling of these ORFs on a polycistronic mRNA (38). The putative protein products were also generally very small, especially in the early gene region. It is possible that ribosomal frameshifting may occur, resulting in some larger protein products whose sequence is determined jointly by two overlapping

TABLE 1. Identification of putative proteins, SD sequences, sequence identity, and possible functions for phage c2 ORFs

ORF	Size (kDa)	Coordinates[a]	SD-binding energy (ΔG, kcal/mol [1 kcal = 4.184 kJ])[b]	SD start codon distance (no. of nucleotides)	Size of overlap or intergenic region (in nucleotides)[c]	Phage bIL67 homolog and amino acid identity (%)[d]	Possible function
Early genes							
e1	5.4	6634 (AUG)–6494 (UGA)	−18.0	8		ORF 1, 82	
e2	4.4	6551 (AUG)–6438 (UGA)	−9.4	9	Overlap 58	Stop codon present	
e3	4.1	6493 (UUG)–6386 (UGA)	−19.0	7	Overlap 56	Not previously described, 86	
e4	7.8	6389 (AUG)–6189 (UAA)	−14.4	8	Overlap 4	ORF 2, 76	
e5	32.9	6187 (AUG)–5288 (UAA)	−11.2	9	IR 1	ORF 3, 97	DNA polymerase
e6	7.9	5218 (AUG)–5024 (UGA)	−14.8	7	IR 69	ORF 4, 95	DNA polymerase
e7	14.8	5034 (AUG)–4654 (UAA)	−18.4	6	Overlap 11	ORF 5, 80	DNA polymerase
e8	15.4	4654 (AUG)–4265 (UGA)	−15.4	9	Overlap 1	ORF 6, 79	
e9	6.5	4207 (AUG)–4034 (UGA)	−14.8	10	IR 57	ORF 7, 94	
e10	2.6	4037 (AUG)–3969 (UAA)	−9.4	7	Overlap 4	Deleted from bIL67	
e11	14.6	3962 (AUG)–3594 (UAA)	−15.4	10	IR 6	ORF 8, 83	
e12	7.4	3616 (AUG)–3434 (UAA)	−15.2	9	Overlap 23	ORF 9, 65	Transcription regulation
e13	10.0	3462 (AUG)–3211 (UGA)	−13.3	6	Overlap 29	ORF 10, 95	
e14	6.1	3217 (AUG)–3065 (UAA)	−14.4	9	Overlap 7	ORF 12, 94	
e15	20.0	3065 (AUG)–2526 (UAA)	−14.0	9	Overlap 1	ORF 13, 85	Recombination
e16	12.8	2460 (AUG)–2116 (UAA)	−16.0	11	IR 65	ORF 14, 96	
e17	6.5	2103 (AUG)–1939 (UGA)	−16.6	7	IR 12	ORF 15, 92	
e18	19.6	1942 (AUG)–1442 (UAA)	−11.4	14	Overlap 4	ORF 16, 75	
e19	12.5	1441 (AUG)–1118 (UAA)	−16.6	5	Adjacent	ORF 17, 94	
e20	11.8	1125 (AUG)–823 (UAA)	−9.8	8	Overlap 8	ORF 18, 96	
e21	14.7	800 (AUG)–426 (UAA)	−10.4	9	IR 22	ORF 19, 83	
e22	6.6	423 (AUA)–253 (UAG)	−14.4	7	IR 2	ORF 21, 76	Transcription regulation
Late genes							
l1	5.7	7246 (AUG)–7395 (UAA)	−14.4	7		ORF 22, 96	
l2	18.3	7406 (AUG)–7891 (UAG)	−9.8	6	IR 10	ORF 23, 93	Holin
l3	25.4	7937 (AUA)–8617 (UAG)	−14.4	7	IR 45	ORF 24, 94	Lysin
l4	31.8	8628 (AUG)–9467 (UGA)	−16.6	6	IR 11	ORF 25, 97	Minor structural protein
l5	53.0[e]	9418 (UUG)–10860 (UAA)	−19.0	5	Overlap 50	ORF 26, 97	Major head protein
l6	25.8	10880 (AUG)–11596 (UAA)	−16.0	7	IR 19	ORF 27, 87	
l7	22.1	11603 (AUG)–12220 (UAA)	−16.6	10	IR 6	ORF 28, 89	Major tail shaft protein
l8	9.6	12290 (AUG)–12544 (UAG)	−12.8	8	IR 69	ORF 29, 98	
l9	15.5	12732 (AUG)–13136 (UAA)	−14.4	7	IR 187	ORF 30, 87	
l10	75.3	13129 (AUG)–15249 (UGA)	−12.8	10	Overlap 8	ORF 31, 69	Tail adsorption protein
l11	11.2	15246 (UUG)–15533 (UAA)	−15.4	15	Overlap 4	Not previously described, 96	
l12	59.6	15546 (AUG)–17102 (UAA)	−12.8	7	IR 12	ORF 32, 98	Terminase
l13	11.1	17102 (AUG)–17401 (UGA)	−20.4	13	Overlap 1	ORF 33, 99	
l14	73.3	17398 (AUG)–19314 (UGA)	−14.4	13	Overlap 4	ORF 34, 43	
l15	43.2	19311 (AUG)–20456 (UAG)	−16.0	6	Overlap 4	ORF 35, 53	Minor structural protein
l16	48.2	20449 (AUG)–21696 (UAA)	−9.4	8	Overlap 8	ORF 36, 38	Minor structural protein
l17	10.7	21719 (AUG)–22009 (UAG)	−9.0	8	IR 22	ORF 37, 95	Holin

[a] AUG, UUG, GUG, and AUA start codons have been proposed for *L. lactis* (46, 49).
[b] The SD sequence binding energy to the 3′ end of *L. lactis* 16S rRNA (35) was calculated as described by Tinoco et al. (48).
[c] Size of the overlap or the intergenic region (IR) with the preceding ORF.
[d] Comparison of predicted amino acid sequences encoded by phage c2 ORFs and the phage bIL67 homologs described by Schouler et al. (45). Some homologs of phage c2 ORFs were not previously identified but are present in the bIL67 genome, as indicated. The start codons of bIL67 ORFs 9, 10, and 24 were reassigned to alternative start codons upstream to allow better alignment with the c2 ORFs. Apparent frameshift mutations or sequencing errors were noted in bIL67 ORFs 6 and 16, which were adjusted to allow maximum alignment. Deletion of the entire ORFs 11 and 20 occurred (data not shown). Deletions were also noted within phage bIL67 ORFs 8, 14, and 31. A stop codon in the bIL67 homolog of phage c2 ORF *e2* would prevent translation of a protein product.
[e] The experimentally determined N terminus of the protein products from *l5* is 205 codons downstream from the start codon (see the text).

ORFs. Potential slippery sequences consisting of a run of ≥5 nucleotides, at which ribosome slippage could occur (4), were observed in many c2 ORFs. Frameshifting has been shown for several viral proteins (32).

Codon usage. The codon usage in the c2 genome was found to be generally similar to the codon usage of *L. lactis* genes (11, 50). However, the c2 early genes showed a very strong stop codon bias. The UAA codon was strongly favored among the early ORFs >80 codons in size (Table 2), and the UAG codon was not used by any of these ORFs. This is a striking deviation from the expected frequency, based on the nucleotide composition of the genome, and from the observed lactococcal stop codon frequencies. However, the stop codons were used as frequently as expected in the late gene region. Stop codon usage has been shown to be important in controlling the efficiency of translation termination and frameshifting in *E. coli* (1).

Comparison of lactococcal prolate-headed phage c2 and bIL67 sequences. Prolate phages have a high degree of DNA homology and appear to be closely related (26). Phage bIL67 is a lactococcal prolate-headed phage that has been recently sequenced (45). The sequences of phage c2 and bIL67 were

TABLE 2. Termination codon usage in phage c2 early genes and in *L. lactis* genes

Termination codon	Codon usage in:			
	Phage c2 early genes[a]		*L. lactis* genes[b]	
	Mean	Expected	Mean	Expected
UAA	75	49	39	46
UAG	0	25	20	27
UGA	25	25	41	27

[a] Mean codon usage was calculated from all ORFs of >80 codons in the early region of the phage c2 genome, with the CODONFREQUENCY program (18). The ORFs used were *e4*, *e7*, *e8*, *e11*, *e13*, *e15*, *e16*, *e18*, *e19*, *e20*, and *e21*. Expected codon usage was calculated by attributing the individual probabilities of occurrence of 0.17, 0.33, 0.33, and 0.17 to G, A, T, and C, respectively, on the basis of the 34 mol% G+C content determined for the early gene region. The expected frequencies of the synonymous codons specifying each amino acid were determined from the product of the probability of occurrence of each individual nucleotide in the codon.

[b] Mean codon usage values are from Ward et al. (50), calculated from 57 *L. lactis* genes. Expected codon usage was calculated as described above on the basis of the 37 mol% G+C content of the *L. lactis* genome (11).

compared in order to determine the conservation of the genomic structure and to identify conserved and nonconserved sequences, which may correspond to functions universal to the phage species and specific to an individual phage, respectively. The organization of the two phage genomes was conserved, with the putative proteins encoded by the c2 ORFs aligning with the corresponding putative proteins of bIL67 (Table 1). Overall, the genomes of phages c2 and bIL67 shared 80% nucleotide identity; however, the homology was not uniformly distributed along the entire genome. Rather, some regions were highly conserved while other regions shared poor sequence identity (Fig. 1). The boundaries between highly conserved and poorly conserved regions often corresponded almost exactly to the boundaries of ORFs. The putative gene products (gp; e, early, and l, late) of phage c2 ORFs *e4*, *e8*, *e12*, *e18*, and *e22* shared ≤80% amino acid identities with the bIL67 homologs, and gpl10, gpl14, gpl15, and gpl16 were particularly poorly conserved, with 69, 43, 53, and 38% identities, respectively. The most conserved putative proteins of phage c2, all with over 90% amino acid identities with the corresponding bIL67 proteins, were gpe5, gpe6, gpe9, gpe13, gpe14, gpe16, gpe17, and gpe19 and all of the late gp except gpl10, gpl14, gpl15, and gpl16.

Several DNA regions, ranging from 11 to 271 bp in length, were absent from one of the genomes of bIL67 and c2. It is not possible to distinguish if these were deletions or insertions of DNA or a combination of these events; however, for convenience, they will be called deletions. Some of these deletions corresponded precisely to an entire ORF: bIL67 ORFs 20 and 11 were precisely deleted from phage c2, and phage c2 ORF *e10* was precisely deleted from bIL67. Direct repeats were observed flanking each of the deleted DNA regions; for example, a 23-bp sequence (TTTACACCTCTTTCATATCTTCG) was repeated either side of bIL67 ORF 20, exactly at the boundaries of the deleted DNA. Apart from the deletions affecting entire ORFs (as described above), the following ORFs had internal deletions or were truncated: bIL67 ORFs 8 and 14 and c2 ORFs *e2*, *e16*, and *l10*.

Identification of phage c2 gene functions. (i) *e15* and *l12*. gpe15 and gpl12 share 85 and 98% amino acid identity with phage bIL67 gp13 and gp32, respectively (45). Phage bIL67 gp13 may be a recombination protein, on the basis of the similarity to the essential recombination function (Erf) protein of *Salmonella* phage P22, and gp32 was proposed to be a terminase subunit because it contains putative ATP-binding domains and terminases are ATP-binding and hydrolyzing enzymes (45). Therefore, *e15* and *l12* of phage c2 may encode a recombination protein and a terminase subunit, respectively. A possible function for a recombination protein in prolate phage replication has been previously discussed (45).

(ii) *e5*, *e6*, and *e7*. gpe5, gpe6, and gpe7 share a high degree of identity with phage bIL67 gp3, gp4, and gp5, respectively (45), which have been proposed to be subunits of a DNA-dependent DNA polymerase. Therefore, *e5*, *e6*, and *e7* might be genes for a DNA polymerase of phage c2.

(iii) *e22* and *e12*. It is probable that an early gene of phage c2 encodes a protein, such as a sigma factor or transcription activator, which is responsible for controlling transcription of the late genes. gpe22 (56 amino acids) shared 60% similarity and 35% identity with sigma B and sigma C proteins from *Stigmatella aurantiaca* and *Myxococcus xanthus* (SwissProt accession no. Q01624, P19433, and Q07083, respectively). The regions of similarity involved the sigma factor-conserved regions 2.4 and 3.1 (Fig. 2).

Transcriptional regulatory proteins and sigma factors often contain a helix-turn-helix motif (8, 19, 22) which is responsible for sequence-specific DNA binding of the protein. The method of Dodd and Egan (16) was used to examine each ORF for the potential to encode a helix-turn-helix domain. This analysis indicated with a confidence of 80 to 90% (SD score of 3.8 [16]) that one ORF, *e12*, encoded a helix-turn-helix domain.

(iv) *l2*, *l3*, and *l17*. Lysis of bacterial hosts by bacteriophages requires the action of a phage-encoded lysin or lysozyme enzyme on the cell wall. A phage-encoded holin is also usually required to disrupt the cell membrane, allowing the cytoplasmic lysin to gain access to the cell wall substrate (56). Intracellular lysin activity is detected in phage c2-infected cells at the onset of late gene expression, 10 min after infection is initiated; however, it does not appear extracellularly until the cells begin to lyse, at about 25 min postinfection, indicating that a holin may be required for lysin release to the cell wall (29). Phage holins share little sequence similarity and cannot be identified by sequence comparisons. However, they do share several conserved features: the holin gene is usually located immediately upstream of the lysin gene, and the holin protein is usually about 80 to 130 amino acids in size, contains two transmembrane domains separated by a β turn, and has a highly charged C terminus. Gene *l3* has been identified as the lysin gene of phage c2 (50). ORF *l2* is located immediately upstream of *l3* and could encode a 161-amino-acid protein with two weakly predicted transmembrane regions, separated by a β turn and a charged C terminus. Another ORF, *l17*, located

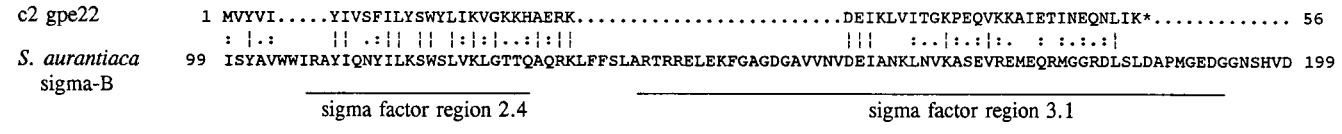

FIG. 2. Amino acid sequence alignment of putative product of phage c2 *e22* and the sigma B protein of *Stigmatella aurantiaca*. Vertical bars between residues indicate identities, colons indicate highly conservative replacements, and periods indicate less conservative replacements identified by the GAP program (18).

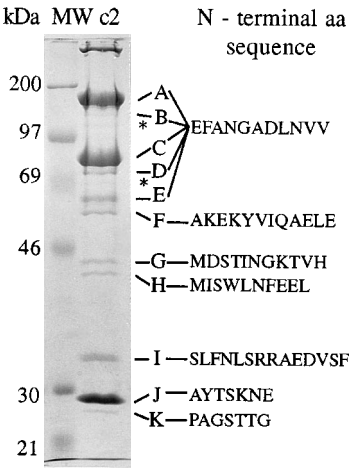

FIG. 3. SDS–12% polyacrylamide gel profile of total phage c2 proteins. MW, molecular mass markers (indicated on the left); c2, phage c2 protein bands (A to K) and corresponding N-terminal amino acid (aa) sequences; *, positions at which additional minor bands were sometimes visible.

nearly 14 kb downstream from *l3*, has two strongly predicted transmembrane domains and a charged C terminus. Either or both of these genes might encode a c2 holin. Schouler et al. (45) have suggested that ORF 36 of bIL67, which is the homolog of *l17*, is a holin gene on the basis of the structural features discussed above.

(v) *l4*, *l5*, *l7*, *l10*, *l15*, and *l16*. SDS-PAGE of phage c2 indicated that the phage is composed of three major proteins (Fig. 3, labelled A, C, and J) and at least eight minor proteins (Fig. 3, labelled B, D to I, and K). Comparison of the N-terminal sequences of the proteins to the DNA sequence of the phage enabled the identification of the corresponding structural genes. The amino acid sequence data indicated that proteins A to E had the same N terminus and were all encoded by gene *l5*. Proteins J and K had different N-terminal sequences, but both were encoded by gene *l7*. Proteins F to I were encoded by genes *l10*, *l16*, *l15*, and *l4*, respectively. A putative start codon adjacent to the experimentally determined N terminus of the proteins encoded by *l5* was not observed; however, a UUG start codon and an excellent SD sequence were located 205 codons upstream. Similarly, the N terminus of protein K was located 21 codons downstream from the start codon of *l7*, although the N terminus of protein J, also encoded by *l7*, began at the start codon. Alternative start codons for *l5* and *l7* corresponding to the N-terminal sequences of the proteins were not observed. These data indicate that proteins A to E and K may result from proteolytic cleavage of larger protein precursors; for example, protein K might be cleaved from protein J.

The experimentally determined (SDS-PAGE) molecular masses of the proteins were compared with those calculated from the corresponding gene sequences (Table 3). The experimentally determined and the calculated sizes of proteins G to K were in reasonable agreement. However, protein F was smaller than expected, and the ladder of protein bands encoded by *l5* (proteins A to E) ranged from about two to six times the expected size. There were also some high-molecular-mass proteins that barely penetrated the gel matrix. The protein profile was not changed by additional pretreatments of the samples before loading, i.e., by heating at 100°C for 10 min in the presence of 2.5% β-mercaptoethanol and any of the following denaturants: 3 M urea, 20% formic acid, 20% acetonitrile, and 2% cetyltrimethylammonium bromide. The major proteins A and C were approximately six and three times the expected sizes, respectively, and could correspond to hexamer and trimer forms of gpl5. The minor protein B could be a pentamer of gpl5.

Identification of major head and tail proteins and a possible tail adsorption protein. The locations of proteins A, C, J, and F in the phage c2 structure were determined by immunogold electron microscopy by using antibodies eluted from the corresponding protein bands on Western hybridization filters. Antibodies against whole phage localized the gold label to both phage heads and tails (Fig. 4A). Antibodies against major proteins A and C localized gold label to phage heads but not tails (Fig. 4B), indicating that these proteins are the major head proteins. Antibodies against major protein J localized the gold label along the length of tail structures but not heads (Fig. 4C), indicating that protein J is the major tail protein. Antibodies against minor protein F localized the gold label specifically to the tip of the tail and not to the shaft of the tail or to the heads (Fig. 4D), suggesting that protein F is located at the end of the tail and, therefore, could be the tail adsorption protein. The homologous protein of phage bIL67, gp31, has also been shown to be localized to the end of the tail (45). Western blots of total phage c2 proteins probed with the antibodies described above confirmed the specificities of the antibodies and also showed that the head proteins (A to E) had low antigenicity and that proteins J and F were highly antigenic, relative to the abundancy of these proteins. This explains the relatively small amount of gold label localized to the phage heads by immunogold electron microscopy (Fig. 4). The Western blots also showed that the high-molecular-mass protein near the wells of the SDS-polyacrylamide gels contained the head protein.

An EF-hand calcium-binding domain signature, which includes a 12-residue calcium-binding loop followed by a hydrophobic residue (BLOCKS database accession no. BL00616), was identified in protein F (residue 499 to 511, DKNKDG KVSNEEM) by BLOCKS SEARCHER analysis. Calcium is a general requirement for the replication of lactococcal phages (33) and is required for efficient adsorption of phage c2 to the host cell (34). The presence of a calcium-binding domain in protein F, and not in any other phage c2 protein, is a further indication that this protein might be involved in adsorption to the host cell.

TABLE 3. Experimentally determined masses of phage c2 structural proteins compared with sizes predicted from the gene sequence

Protein	Gene	Protein size (kDa)	
		SDS-PAGE	Calculated from gene sequence[a]
A	*l5*	175	31
B	*l5*	143	31
C	*l5*	90	31
D	*l5*	82	31
E	*l5*	66	31
F	*l10*	60	75
G	*l16*	44	48
H	*l15*	42	43
I	*l4*	32	32
J	*l7*	29	22
K	*l7*	28	20

[a] Molecular mass predicted from the gene sequence, taking into account the experimentally determined N terminus of the proteins and the stop codon of the gene.

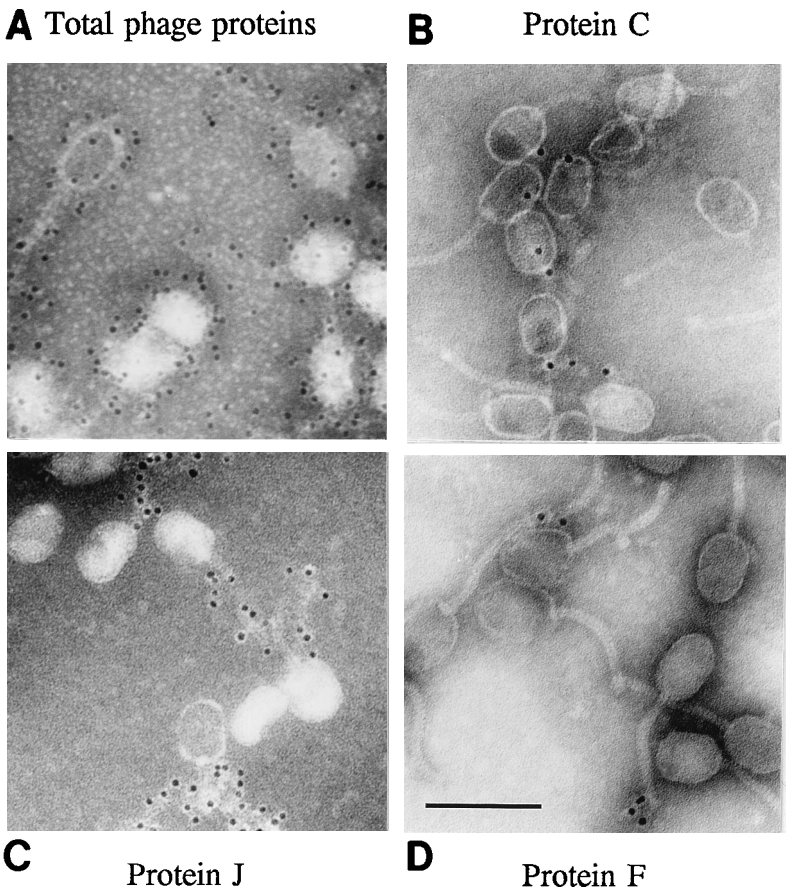

FIG. 4. Immunogold electron microscopy of phage c2. Antibodies were specific for the indicated phage proteins. Black particles indicate immunoglobulin G-gold conjugates. Many grids were examined, and representative views are shown. Bar, 100 nm; magnification was the same for all micrographs.

Identification of phage c2 promoters, a transcription terminator, and a possible origin of replication. Six leftward reading putative promoters (EP1 to EP6, Fig. 1) were identified by their similarities to the consensus −10 and −35 promoter sequences recognized by the *L. lactis* σ^{39} transcription factor: TTGACA-17 bp-TATAAT. A divergently oriented putative late promoter with a consensus −10 sequence, but no recognizable −35 sequence, was also identified immediately upstream of the late region ORFs. The activity and temporal expression of each of the promoters were verified by primer extension experiments (51). The data were consistent with previous findings (5) which showed that early gene transcription occurs within the first 5 min of infection while late gene transcripts only appear 10 min after infection is initiated. A putative transcription terminator sequence (CCACTCAATCAAGAGTGG TTTTTTTGTTT; underlined nucleotides can form a stem structure) was noted at the end of the early gene region 29 bp 3′ to the stop codon of *e22*. A homologous terminator, nearly identical in sequence and position, was previously described for phage bIL67 (45).

The 611-bp intergenic region between the divergently oriented early and late ORFs contained no obvious ORFs. It has been suggested that the corresponding region of phage bIL67 might contain an origin of replication because of the presence of distinctive nucleotide repeat sequences (45). A 521-bp fragment from this region was shown to act as an origin of DNA replication when cloned in the origin screening vector pVA891 (51). These data suggest that the intergenic region between the divergently oriented early and late genes contains a c2 origin of replication.

DISCUSSION

The analysis of phage c2 structural proteins and comparison of the phage c2 genome sequence to that of another prolate phage, bIL67, allows a number of insights into the molecular biology and evolution of lactococcal prolate phages. Protein and DNA sequence data and immunogold electron microscopy experiments suggest that the protein products translated from the head protein gene, *l5*, may form covalently linked multimers, which could include hexameric and trimeric forms (major proteins) and a minor pentameric form. Alternatively, the proteins may not be homomultimers but may instead be covalently linked to other proteins that were not detected when the protein bands were N-terminally sequenced. The putative hexameric and pentameric forms might correspond to the hexamers of the head facets and the pentamers of the vertices which have been proposed to make up the icosahedral phage head (10). The relative abundancies of the protein forms are consistent with these roles. A function for the proposed trimeric form, which is a major protein component, is not clear; however, trimeric proteins (not covalently linked) have been observed in phage head structures (53). The minor dimer and 2.7-mer forms also observed do not present an obvious function. Covalent cross-linking of phage subunits has previously been reported only for phage HK97 (41), a lambdoid phage of

E. coli, and phage L5 (20), which infects *Mycobacterium* spp. For both of these phages, head protein monomers were linked into hexamers and less abundant pentamers. For HK97, a large amount of head protein was also present in a very high-molecular-mass form, which was presumed to consist of hexameric units joined together, similar to our observations for phage c2. Therefore, it appears that while subunit cross-linking has rarely been reported, the process can be found in widely diverse phage groups that include a lytic phage (c2) and a lysogenic phage (L5) of two very different gram-positive hosts, *Lactococcus* and *Mycobacterium* spp., respectively, as well as a lysogenic phage (HK97) of a gram-negative organism (*E. coli*).

Covalent cross-linking of head subunits has not yet been described for the most common small isometric phage species, P335 and 936, and might distinguish these phages from the prolate lactococcal phages. P335 and 936 phage species have one major structural protein with a size of about 30 to 40 kDa (7, 13, 28, 42). The relatively small size of the major proteins of the 936 phage species indicates that head protein multimers are unlikely in this phage species. For phage Tuc2009 (P335-type phage), it has been shown that the major protein is the head protein and that its mass corresponds to that predicted from the gene sequence (3, 36). This indicates that the head proteins for phage Tuc2009 are not covalently linked into multimers. However, the P335 phage species in particular form a very diverse group, which includes both lytic and temperate phages (28), and further studies are required to determine if covalent linkage of head subunits is involved in the assembly of other P335-type phages.

A second difference between the prolate and small isometric lactococcal phages may be the degree to which their structural proteins are processed. Proteolytic cleavage of structural proteins during phage assembly is often observed, and it has been suggested that it may be important to make the assembly irreversible, to cause conformational changes required for further assembly, and/or to remove the cleaved protein or portion of it from the phage structure (10). Such proteolytic activity is phage encoded, and in some cases the protein cleavage is autocatalytic. Phage c2 proteins A to E, which are encoded by *l5*, and protein K, which is encoded by *l7*, do not have N termini that correspond to the start codon of their genes and therefore may be produced from a larger precursor by proteolytic cleavage. Other proteins (F and G) which were smaller than the corresponding predicted gene products were also observed. A BLOCKS SEARCHER analysis revealed that a region of gpl6 had some similarity to conserved domains of autocatalytically spliced proteins, suggesting a possible role for gpl6 in proteolytic processing. In contrast, the major capsid proteins of the small isometric phages Tuc2009 (P335 species) and F4-1 (unknown species) do not appear to be cleaved from precursor proteins (3, 12). Posttranslational processing of the structural proteins of phage c2 may be an important control point in the development of prolate lactococcal phages.

The evolution of phages has been proposed to involve the exchange of functional modules and loss or acquisition of genes by recombination between phages and also between phages and their hosts (6, 9, 37). Heteroduplex studies have shown that the genomes of small isometric lactococcal phages consist of modules of conserved and divergent sequences (30). These sequences may correspond to modules encoding phage-specific functions (divergent sequences) and functions common to all members of the phage species (conserved modules). Although the genomes of phages c2 and bIL67 shared a high degree of nucleotide identity (80%), a particularly poorly conserved region, which included *l14*, *l15*, and *l16*, was identified in the late gene region. The points at which the homology changed from highly conserved to poorly conserved corresponded exactly to the start of *l14* and the end of *l16*, both of which are structural genes. Within this region was a small sequence corresponding to the N-terminal codons of *l15*, which had relatively high conservation between the two phages. Therefore, the region of the prolate phages corresponding to *l14*, *l15*, and *l16* may be a module of structural proteins that has a phage-specific function, and the pattern of conservation observed for this region might have been produced by block recombination events between prolate phages. In addition, several DNA deletions or insertions which were flanked by repeat sequences were observed. In several cases, precisely a single putative gene was lost or acquired. These repeat sequences could facilitate loss, acquisition, or exchange of sequences by recombination. Thus, prolate lactococcal phage evolution may involve a modular organization and sequence loss or acquisition by recombination between repeat sequences.

Schouler et al. (44) cloned from the prolate phage ϕ197 the N-terminal 109 codons of a gene which shared 98% predicted amino acid identity with the corresponding region of gpl10, the putative tail adsorption protein of phage c2. This region of gpl10 is also highly conserved in phage bIL67. Antibodies raised against the cloned part of the ϕ197 protein recognized minor structural proteins with sizes of 45 and/or 46 kDa in Western blots of several prolate phage proteins, and it was suggested that this was a conserved protein among the prolate phages (44). However, the corresponding phage c2 protein (gpl10) was much larger than the ϕ197 protein, with an estimated molecular mass of 60 kDa by SDS-PAGE. The *l10* gene contained an additional 86 codons, potentially encoding an additional 10 kDa of protein, compared with the corresponding gene of phage bIL67. Also, the predicted molecular mass of gpl10 was 75 kDa, indicating that the mature 60-kDa protein may be cleaved from a 75-kDa precursor. Possibly, the insertion or deletion of coding sequences, and differences in the processing of these proteins from larger precursors, may be an important determinant of phage-specific properties, such as host range.

The pattern of amino acid conservation for gpl10 is similar to that observed for lambdoid phage tail fibers. The N-terminal part of some coliphage tail fiber proteins attaches to the virion, and the C-terminal part interacts with the host cell to facilitate adsorption (10). The C-terminal one-half to two-thirds of the tail fibers of many phages appears to consist of sequence segments that exchange freely among different phage groups (9). While the N-terminal 109 amino acids were highly conserved between gpl10 and the corresponding proteins of phage bIL67 (and ϕ197), the remaining approximately 600 amino acids were poorly conserved. This sequence conservation pattern fits precisely with the hypothesis that the parts of the tail adsorption protein (fiber) that contact the host cell are under evolutionary pressure to diversify as they are presented with a genetically changing target by the host cell population (9). In contrast, among the lambdoid phages there is no evidence for genetic exchange within the head or tail shaft gene regions (9). The high level of conservation of the major head and tail proteins observed between phage c2 and phage bIL67 mirrors the observations for the lambdoid phages and indicates that the head and tail proteins are not under such pressure to diversify. Indeed, these proteins may be more constrained by the need to maintain the complex contacts with other proteins in the head and tail structures, while the tail fiber peptides are in an extended conformation and make only minimal contact with other parts of the polypeptide (9).

In order to understand the natural mechanisms of lactococcal phage resistance in bacterial starter strains and to exploit

phage genes to construct novel resistance mechanisms, it is important to understand the molecular biology of these phages: what the gene functions are, how the genes interact, and what are the possible modes of phage evolution that could contribute to the emergence of phages which are immune to host resistance mechanisms. We have described here the full sequence of the lactococcal prolate phage c2 and identified potential mechanisms of prolate phage evolution. We have identified highly conserved genes which likely encode the functions common to all members of the prolate phage species and so may serve as targets for resistance mechanisms effective against all prolate phages. We also have identified several points in the phage life cycle, the targeted disruption of which could confer resistance to prolate phages. These include DNA replication, late gene activation, proteolytic processing and covalent linkage of phage structural proteins, and phage adsorption. Sufficient sequence information on phage c2 has also now been accumulated to begin investigating the mechanisms of naturally occurring phage resistance, such as the abortive infection mechanism on plasmid pAJ2074 (27) which inhibits replication of phage c2. Such studies will provide further information on the molecular biology of the phage life cycle and a firm foundation for the development of effective and long-lasting resistance mechanisms for bacterial starter strains.

ACKNOWLEDGMENTS

We thank K. M. Polzin for critical reading of the manuscript and many helpful comments. We also thank L. Collins for technical assistance.

The research was funded in part by the New Zealand Foundation for Research, Science and Technology. Collaborative work between the New Zealand Dairy Research Institute and the University of Cambridge was made possible by a Higher Education Links Award from the British Council.

REFERENCES

1. **Adamski, F. M., B. C. Donly, and W. P. Tate.** 1993. Competition between frameshifting, termination and suppression at the frameshift site in the *Escherichia coli* release factor-2 mRNA. Nucleic Acids Res. **21:**5074–5078.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Arendt, E. K., C. Daly, G. F. Fitzgerald, and M. van de Guchte.** 1994. Molecular characterization of lactococcal bacteriophage Tuc2009 and identification and analysis of genes encoding lysin, a putative holin and two structural proteins. Appl. Environ. Microbiol. **60:**1875–1883.
4. **Atkins, J. F., R. B. Weiss, and R. F. Gesteland.** 1990. Ribosome gymnastics—degree of difficulty 9.5, style 10.0. Cell **62:**413–423.
5. **Beresford, T. P. J., L. J. H. Ward, and A. W. Jarvis.** 1993. Temporally regulated transcriptional expression of the genomes of lactococcal bacteriophages c2 and sk1. Appl. Environ. Microbiol. **59:**3708–3712.
6. **Botstein, D.** 1980. A theory of modular evolution for bacteriophages. Ann. N. Y. Acad. Sci. **354:**484–491.
7. **Braun, V., S. Hertwig, H. Neve, A. Geis, and M. Teuber.** 1989. Taxonomic differentiation of bacteriophages of *Lactococcus lactis* by electron microscopy, DNA-DNA hybridization, and protein profiles. J. Gen. Microbiol. **135:** 2551–2560.
8. **Brennan, R. G., and B. W. Matthews.** 1989. Structural basis of DNA-protein recognition. Trends Biochem. Sci. **14:**286–290.
9. **Casjens, S., G. F. Hatfull, and R. Hendrix.** 1992. Evolution of dsDNA tailed-bacteriophage genomes. Semin. Virol. **3:**383–397.
10. **Casjens, S., and R. Hendrix.** 1988. Control mechanisms in dsDNA bacteriophage assembly, p. 15–91. *In* R. Calendar (ed.), The bacteriophages, vol. 1. Plenum Press, New York.
11. **Chopin, A.** 1993. Organization and regulation of genes for amino acid biosynthesis in lactic acid bacteria. FEMS Microbiol. Rev. **12:**21–38.
12. **Chung, D. K., J. H. Kim, and C. A. Batt.** 1991. Cloning and nucleotide sequence of the major capsid protein from *Lactococcus lactis* ssp. *cremoris* bacteriophage F4-1. Gene **101:**121–125.
13. **Coveney, J. A., G. F. Fitzgerald, and C. Daly.** 1987. Detailed characterization and comparison of four lactic streptococcal bacteriophages based on morphology, restriction mapping, DNA homology, and structural protein analysis. Appl. Environ. Microbiol. **53:**1439–1447.
14. **Dao, M. L., and J. J. Ferretti.** 1985. *Streptococcus-Escherichia coli* shuttle vector pSA3 and its use in the cloning of streptococcal genes. Appl. Environ. Microbiol. **49:**115–119.
15. **de Vos, W. M., and G. F. M. Simons.** 1994. Gene cloning and expression systems in lactococci, p. 52–105. *In* M. J. Gasson and W. M. de Vos (ed.), Genetics and biotechnology of lactic acid bacteria. Blackie Academic and Professional, Glasgow, Scotland.
16. **Dodd, I. B., and J. B. Egan.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18:**5019–5026.
17. **Gasson, M. J.** 1983. Plasmid complements of *Streptococcus lactis* NCDO712 and other lactic streptococci after protoplast-induced curing. J. Bacteriol. **154:**625–629.
18. **Genetics Computer Group Inc.** 1991. Program manual for the GCG package, version 7, April 1991. University of Wisconsin, Madison.
19. **Gribskov, M., and R. R. Burgess.** 1986. Sigma factors from *E. coli*, *B. subtilis*, phage SP01, and phage T4 are homologous proteins. Nucleic Acids Res. **14:** 6745–6763.
20. **Hatfull, G. F., and G. J. Sarkis.** 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol. Microbiol. **7:**395–405.
21. **He, M., A. Wilde, and M. A. Kaderbhai.** 1990. A simple single-step procedure for small-scale preparation of *Escherichia coli* plasmids. Nucleic Acids Res. **18:**1660.
22. **Helmann, J. D., and M. J. Chamberlin.** 1988. Structure and function of bacterial sigma factors. Annu. Rev. Biochem. **57:**839–872.
23. **Henikoff, S., and J. G. Henikoff.** 1991. Automated assembly of protein blocks for database searching. Nucleic Acids Res. **19:**6565–6572.
24. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114:**193–197.
25. **Ish-Horowicz, D., and J. F. Burke.** 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. **9:**2989–2998.
26. **Jarvis, A. W.** 1984. Differentiation of lactic streptococcal phages into phage species by DNA-DNA homology. Appl. Environ. Microbiol. **47:**343–349.
27. **Jarvis, A. W.** 1993. Analysis of phage resistance mechanisms encoded by lactococcal plasmid pAJ2074. Can. J. Microbiol. **39:**252–258.
28. **Jarvis, A. W., G. F. Fitzgerald, M. Mata, A. Mercenier, H. Neve, I. B. Powell, C. Ronda, M. Saxelin, and M. Teuber.** 1991. Species and type phages of lactococcal bacteriophages. Intervirology **32:**2–9.
29. **Jarvis, A. W., M. W. Lubbers, T. P. J. Beresford, L. J. H. Ward, N. R. Waterfield, L. J. Collins, and B. D. W. Jarvis.** Molecular biology of lactococcal phage c2. Dev. Biol. Stand. **85:**561–567.
30. **Jarvis, A. W., and J. Meyer.** 1986. Electron microscopic heteroduplex study and restriction endonuclease cleavage analysis of the DNA genomes of three lactic streptococcal bacteriophages. Appl. Environ. Microbiol. **51:**566–571.
31. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
32. **Levin, M. E., R. W. Hendrix, and S. R. Casjens.** 1993. A programmed translational frameshift is required for the synthesis of a bacteriophage λ tail assembly protein. J. Mol. Biol. **234:**124–139.
33. **Lowrie, R. J., and L. E. Pearce.** 1971. The plating efficiency of bacteriophages of lactic streptococci. N. Z. J. Dairy Sci. Technol. **6:**166–171.
34. **Lubbers, M. W., and A. W. Jarvis.** Unpublished data.
35. **Ludwig, W., E. Seewaldt, R. Kilpper-Bälz, K. H. Schleifer, L. Magrum, C. R. Woese, G. E. Fox, and E. Stackebrandt.** 1985. The phylogenetic position of *Streptococcus* and *Enterococcus*. J. Gen. Microbiol. **131:**543–551.
36. **Moineau, S., D. Bernier, M. Jobin, J. Hébert, T. R. Klaenhammer, and S. Pandian.** 1993. Production of monoclonal antibodies against the major capsid protein of the *Lactococcus* bacteriophage ul36 and development of an enzyme-linked immunosorbent assay for direct phage detection in whey and milk. Appl. Environ. Microbiol. **59:**2034–2040.
37. **Moineau, S., S. Pandian, and T. R. Klaenhammer.** 1994. Evolution of a lytic bacteriophage via DNA acquisition from the *Lactococcus lactis* chromosome. Appl. Environ. Microbiol. **60:**1832–1841.
38. **Normark, S., S. Bergstrom, T. Edlund, T. Grundstrom, B. Jaurin, F. P. Lundberg, and O. Olsson.** 1983. Overlapping genes. Annu. Rev. Genet. **17:** 499–525.
39. **O'Sullivan, D. J., and T. R. Klaenhammer.** 1993. Rapid mini-prep isolation of high-quality plasmid DNA from *Lactococcus* and *Lactobacillus* spp. Appl. Environ. Microbiol. **59:**2730–2733.
40. **Pillidge, C. J., and A. W. Jarvis.** 1988. DNA restriction maps and classification of the lactococcal bacteriophages c2 and sk1. N. Z. J. Dairy Sci. Technol. **23:**411–416.
41. **Popa, M. P., T. A. McKelvey, J. Hempel, and R. W. Hendrix.** 1991. Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. J. Virol. **65:**3227–3237.
42. **Powell, I. B., P. M. Arnold, A. J. Hillier, and B. E. Davidson.** 1989. Molecular comparison of prolate- and isometric-headed bacteriophages of lactococci. Can. J. Microbiol. **35:**860–866.
43. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
44. **Schouler, C., C. Bouet, P. Ritzenthaler, X. Drouet, and M. Mata.** 1992. Characterization of *Lactococcus lactis* phage antigens. Appl. Environ. Microbiol. **58:**2479–2484.
45. **Schouler, C., S. D. Ehrlich, and M.-C. Chopin.** 1994. Sequence and organi-

zation of the lactococcal prolate-headed BIL67 phage genome. Microbiology **140:**3061–3069.
46. **Shearman, C. A., K. L. Jury, and M. J. Gasson.** 1994. Controlled expression and structural organization of a *Lactococcus lactis* bacteriophage lysin encoded by two overlapping genes. Appl. Environ. Microbiol. **60:**3063–3073.
47. **Sturrock, S. S., and J. F. Collins.** 1993. MPsrch, version 1.3. Biocomputing Research Unit, University of Edinburgh, Edinburgh.
48. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246:**40–41.
49. **van de Guchte, M., J. Kok, and G. Venema.** 1992. Gene expression in *Lactococcus lactis*. FEMS Microbiol. Rev. **88:**73–92.
50. **Ward, L. J. H., T. P. J. Beresford, M. W. Lubbers, B. D. W. Jarvis, and A. W. Jarvis.** 1993. Sequence analysis of the lysin gene region of the prolate lactococcal bacteriophage c2. Can. J. Microbiol. **39:**767–774.
51. **Waterfield, N. R., M. W. Lubbers, J. M. Wells, R. W. F. Le Page, and A. W. Jarvis.** Unpublished data.
52. **Whitehead, H. R., and G. A. Cox.** 1935. The occurrence of bacteriophage in cultures of lactic streptococci. N. Z. J. Dairy Sci. Technol. **16:**319–320.
53. **Wurtz, M.** 1992. Bacteriophage structure. Electron Microsc. Rev. **5:**283–309.
54. **Xu, F., L. E. Pearce, and P.-L. Yu.** 1991. Construction of a family of lactococcal vectors for gene cloning and translation fusion. FEMS Microbiol. Lett. **77:**55–60.
55. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
56. **Young, R.** 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. **56:**430–481.