Genome structure of the *Lactobacillus* temperate phage ϕg1e: the whole genome sequence and the putative promoter/repressor system

Ken-Ichi Kodaira [a], Masaya Oki [a], Makiko Kakikawa [a], Nobukatsu Watanabe [a], Machiko Hirakawa [a], Kazuyo Yamada [a], Akira Taketo [b]

[a] *Molecular Biology Group, Chemical and Biochemical Engineering, Faculty of Engineering, Toyama University, 3190, Gofuku, Toyama 930, Japan*

[b] *Department of Biochemistry I, Fukui Medical School, Matsuoka, Fukui 910-11, Japan*

Abstract

The complete genome sequence of a *Lactobacillus* temperate phage ϕg1e was established. The double-stranded DNA is composed of 42 259 bp, and encodes for sixty-two possible open reading frames (ORF) as well as several potential regulatory sequences. Based on comparative analysis with other related proteins of the *Lactobacillus* and *Lactococcus* phages as well as the *Escherichia coli* phages (such as lambda), functions were putatively assigned to several ϕg1e ORFs: *cng* and *cpg* (encoding for repressors), *hel* (helicase), *ntp* (NTPase), and several ORFs (e.g., minor capsid proteins). An about 1000-bp DNA region of ϕg1e containing *cpg* and *cng* was inferred to function as a promoter/repressor system for the ϕg1e lysogenic and lytic pathway.

Keywords: Bacteriophage ϕg1e; Genome; DNA sequence; DNA-binding protein

1. Introduction

Prevalence of lysogeny in various lactic acid bacteria has been reported (Davidson et al., 1990), and several temperate phages have been investigated in detail: e.g. the *Lactobacillus gasseri* phage ϕadh (Fremaux et al., 1993), the *Lactobacillus bulgaricus* phage mv1 and mv4 (Dupont et al., 1995), and the *Lactococcus lactis* phage ϕLC3 (Lillehaug and Birkeland, 1993), Tuc2009 (Van de Guchte et al., 1994), BK5-T (Boyce et al., 1995), and r1t (Van Sinderen et al., 1996). Unlike the *Escherichia coli* temperate phages such as lambda (Campbell, 1994), informations on the phages of lactic acid bacteria are still insufficient at molecular level; structures, functions, and expressions of their gene products mostly remain to be elucidated.

Recently, we isolated a new *Lactobacillus* temperate phage ϕg1e, which has a small isometric head with a diameter of 63 nm and a non-contractile tail of 260 nm, and have sequenced its several genes, specifying four major capsid proteins (Kakikawa et al., 1996), lysis proteins (holin, and lysin; Oki et al., 1996a), and recombination proteins (excisionase and integrase; Kakikawa et al., 1997).

In this study, we established the whole genome sequence (42 259 bp) of ϕg1e. In the ϕg1e double-stranded DNA, sixty-two possible ORFs were detected. Based on comparative analysis with other related proteins of the *Lactobacillus* and *Lactococcus* phages as well as the *Escherichia coli* phages, possible functions were assigned to several ϕg1e ORFs encoding for repressors (Cpg and Cng), a helicase, an NTPase, and minor capsid proteins. An about 1000-bp DNA region of ϕg1e, containing the two repressor ORFs, was inferred to be involved in a genetic switch for the ϕg1e lysogenic and lytic pathway.

Abbreviations: aa, amino acids; *attP*, phage attachment site; bp, base pair(s); *cng*, ORF encoding cro-like repressor (Cng); *cpg*, ORF encoding CI-like repressor (Cpg); *hel*, ORF encoding helicase-like protein (Hel); *hol*, ORF encoding holin (Hol); *int*, ORF encoding integrase (Int); kb, kilobase(s) or 1000 bp; kDa, kilodalton; LB, Luria-Bertani (medium); MMC, mitomycin C; *lys*, ORF encoding lysin (Lys); *ntp*, ORF encoding NTPase-like protein (Ntp); ORF, putative open reading frame; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; *Ti*, possible secondary structure; *xis*, gene encoding excisionase (Xis); YT, Yeast-Triptone (medium); UV, ultraviolet.

2. Materials and methods

2.1. Bacteria, phage, and plasmids

The *Lactobacillus* strain G1e and its temperate phage φg1e were from our laboratory stock. They were propagated in M17 medium, and the φg1e DNA was extracted from virion particles as described previously (Kakikawa et al., 1996, 1997). The *E. coli* vector plasmids (pUC18, pUC19, pUC118, and pUC119) and phages (M13mp18 and M13mp19), and their host XL1-Blue were from our laboratory stock, and were propagated using LB or 2YT medium as described by Oki et al. (1996a,b).

2.2. Analysis of DNA

Random libraries of φg1e DNA fragments derived with several restriction enzymes (e.g. *Eco*RI) were constructed using *E. coli* vector plasmids (e.g. pUC18) as described previously (Kakikawa et al., 1996, 1997; Oki et al., 1996a,b). Recombinant plasmids thus constructed were introduced into *E. coli* XL1-Blue by Ca^{2+}-dependent transformation or electroporation (Taketo, 1996). Using these random libraries, various deletion clones were generated by exonuclease digestion, and were sequenced as described previously (Kodaira et al., 1996). All other procedures were performed as described by Kodaira et al. (1996).

2.3. Enzymes and biochemicals

Restriction enzymes, phage T4 DNA ligase, and alkaline phosphatase (calf intestine) were purchased from Takara Shuzo (Kyoto) and Nippon Gene (Toyama). [α-^{32}P]dCTP was from NEN (USA). All other materials were as described previously by Kodaira et al. (1996).

2.4. Induction of φg1e

Induction by UV light was performed as described by Mckay et al. (1973). Cells of G1e were grown in M17 broth at 30°C for 10 h. The bacteria were collected by centrifugation at 5000 × *g* for 10 min and resuspended in 1/20 volume of 0.85% saline. The cell suspension was irradiated for a period (from 10 to 50 s) using 15 W germicidal UV lamp held 37.5 cm (220 μW) above the cell suspension. The UV-irradiated cells were inoculated into M17 broth and incubated in darkness at 30°C. Lysis of the cells were monitored by measuring the change in absorbance at 660 nm (A_{660}). For MMC induction, cells of G1e were grown at 30°C in M17 broth. When the A_{660} of the culture had reached 0.25, mitomycin C (MMC) was added to a final concentration of 5.0 μg/ml. After incubation for 20 min at 30°C, the cells were harvested by centrifugation and then suspended into fresh medium. Lysis of the cells were measured as in UV-induction.

3. Results and discussion

3.1. Cloning and sequencing of the φg1e genome

φg1e has a double-stranded DNA genome of about 42.2 kb, and whether or not φg1e has a cohesive or redundant end is presently under investigation. In the φg1e genome, several regions (15 127 bp, see above) have been already sequenced (Kakikawa et al., 1996, 1997; Oki et al., 1996b; see Fig. 1). In this study, the φg1e genome was completely covered by random libraries with various restriction enzymes (e.g. *Eco*RI) cloned into several *E. coli* vector plasmids such as pUC118 (Section 2). After constructing various deletion clones from the random libraries (data not shown), the φg1e whole genome sequence was established as described in Section 2. The sequence analyses (this and previous studies) have revealed that the φg1e genome consists of 42 259 bp (EMBL accession No. X98106), and its total G+C content is 43.1%, which is higher than that (35.5%) of the *Lactococcus lactis* phage r1t (Van Sinderen et al., 1996).

3.2. ORFs encoded by φg1e

As summarized in Table 1 (see also Fig. 1), φg1e has sixty-two possible ORFs. Eight ORFs of them are encoded on one strand (designated as L-strand), and other fifty-four ORFs are located on the complementary strand (R-strand). These predicted ORFs (except three ORFs, *Lorf100*, *Rorf125*, and *Rorf232*) are preceded by a potential Shine and Dalgarno (SD) sequence (Table 1), which is in good agreement with SD consensus found in *Lactobacillus* (5′-AGGAGG-3′) and *Lactococcus* (5′-AGAAAGGAGGT-3′) species (Pouwels and Leer, 1993; Schouler et al., 1994; Matern et al., 1994). In the two ORFs, *Rorf167* and *Rorf281*, UUG may be used as a start codon, as predicted in the *Lactococcus* virulent phage bIL67 (Schouler et al., 1994). Based on comparative analysis with other putative proteins of the *Lactobacillus* and *Lactococcus* phages as well as the *Escherichia coli* phage lambda, functions were assigned to several φg1e putative ORFs (Table 1 and Fig. 1): *cpg* and *cng* (encoding for repressors), *hel* (helicase), *ntp* (NTPase), and several other ORFs (e.g. minor capsid proteins). Properties of these φg1e proteins thus deduced will be discussed below.

3.3. cpg *and* cng

The putative ORF, *cpg*, specifies a 132 amino acids product (termed Cpg) (Fig. 2), and is preceded by

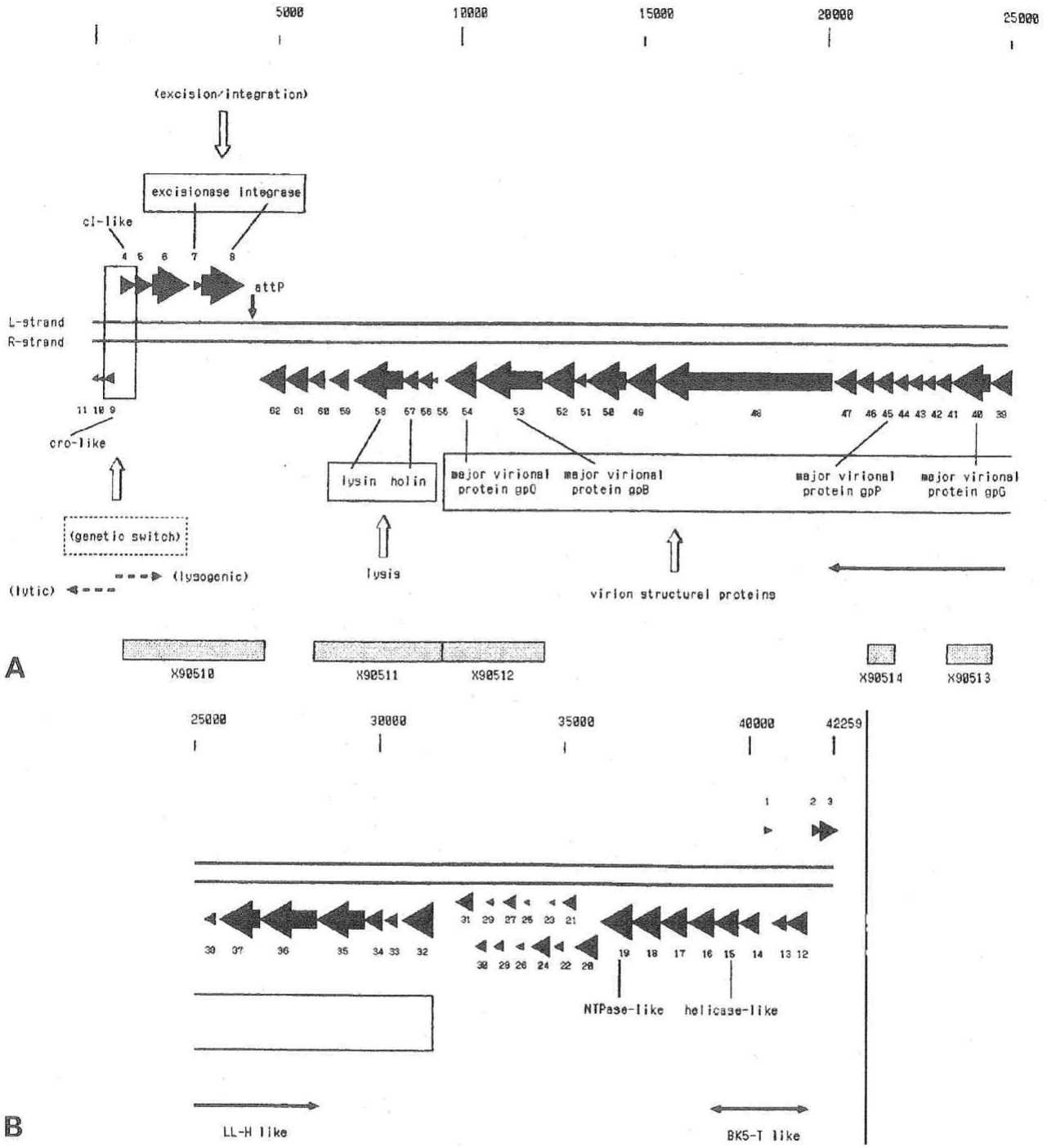

Fig. 1. Genome organization of ϕgle. The double-stranded DNA of the ϕgle genome (double thick lines) was shown linearizing at one site of *Eco*RI together with ϕgle ORFs (in the details, Table 1 and text) predicted from sequence analysis (EMBL accession No. X98106). The putative ϕgle ORFs were represented by arrows or arrowheads above or below the ϕgle DNA depending on their orientation. A putative promoter region required for the lysogenic and lytic pathway was shown by a broken box (see text and Fig. 2). A vertical arrow shown by *attP* points out a ϕgle attachment site (Kakikawa et al., 1997). DNA regions homologous to those of the *Lactococcus* phage BK5-T (Boyce et al., 1995) and the *Lactobacillus* phage LL-H (Mikkonen and Alatossava, 1994) were indicated by 'BK5-T-like' and 'LL-H-like', respectively. Five DNA regions reported previously (see text) were shown by dotted boxes with their EMBL accession numbers (X90510, X90511, X90512, X90513 and X90514).

Table I
ORFs of the *Lactobacillus* phage φgle[1]

ORF[2]		Ribosomal binding site	Predicted product (aa)		Mass (kDa)		Predicted features
1	*Lorf59*	gtttcGGAGta..caataaaaATC		(59)	6.7		
2	*Lorf100*	.....ATG		(100)	11.2		
3	*Lorf166*	cAtgAGGAGGT..gaataataTTG		(166)	19.2		
4	*cpg*	AGAAAGGAGcg.....gtactATG	Cpg	(132)	15.1		Helix-turn-helix
5	*Lorf143*[3]	AaAAAGtAGGT....gtttacATG		(143)	16.3		
6	*Lorf304*[3]	AatttGGAGGa...gtaaataATG		(304)	33.5		
7	*xis*[3]	cGAAAGGAtGT....gtttgtATG	Xis	(66)	7.6		Exicisionase
8	*int*[3]	AGtttGGAGGg..ataaaaacATG	Int	(391)	45.5	[41.0]	Integrase
9	*cng*	gaAttGGAGGT....gcaaaaATG	Cng	(88)	10.1		Helix-turn-helix
10	*Rorf49*	AcAAAGGAGGc....agcaaaATG		(49)	5.3		
11	*Rorf58*	gacgAGGAaGc.tttatatcaATG		(58)	7.1		
12	*Rorf184*	tGAAAGGtGGT....gatggaATG		(184)	20.7		
13	*Rorf117*	AGgAAGGAGtg....atataaATG		(117)	13.7		
14	*Rorf167*	cctActGAGGa..agagctgaTTG		(167)	19.6		
15	*hel*	caAAAGGATga....taaaccATG	Hel	(220)	24.9		NTP-binding
16	*Rorf224*	gcAAAGGAGca...aaataaaATG		(224)	23.6		
17	*Rorf242*	gcAcAGGAGtg.....attagATG		(242)	28.3		
18	*Rorf262*	gtAAAGGAGGT....gtagatTTG		(262)	29.7		
19	*ntp*	tcAAgGGAGG..cacaaacgtATG	Ntp	(280)	31.2		Zn-finger, NTP-binding
20	*Rorf166*	ttAgcGGgGGT..gctgtctcATG		(166)	19.1		
21	*Rorf125*	.....ATG		(125)	14.3		
22	*Rorf76*	gttAcGGAGGg....tgttaaATG		(76)	9.0		
23	*Rorf43*	gacAAGGgGGa....ttagatATG		(43)	4.8		
24	*Rorf161*	tcgAtGGAGGg.......cgaATG		(161)	18.9		
25	*Rorf51*	ttcggGGAGGa....ttaaatATG		(51)	5.4		
26	*Rorf52*	ttcggGGAGGa....ttgaaaATG		(52)	5.7		
27	*Rorf115*	tGAttGGAGaT...ggcgacgATG		(115)	13.0		
28	*Rorf73*	ctActGGAGGc....ggacaaATG		(73)	8.4		
29	*Rorf72*	AtAcAGGAGGc....gttgaaATG		(72)	8.1		
30	*Rorf95*	ctAtAGGAGGc....aacaacATG		(95)	10.9		
31	*Rorf147*	AGcgAGGAGtg..gcggggtaATG		(147)	17.2		
32	*Rorf281*	AagcccGgGGTcggaaccgggcttTTG		(281)	32.6		
33	*Rorf92*	tAcAAGGAcaa..gaagcgtcGTG		(92)	10.4		
34	*Rorf172*	AaAAcGGAGGTgtGGtGGTatgtaATG		(172)	19.2		
35	*Rorf447*	AtgtgGGAGGc...atggacgATG		(447)	52.5		
36	*Rorf508*	AGAtgGtAGGT....gattaaATG		(508)	56.9		(Minor virional protein)
37	*Rorf347*	taAAcGGtGGT..gatggtgaATG		(347)	39.2	[41.2]	(Minor virional protein)
38	*Rorf84*	AatAtGGgGGT.....gcacgATG		(84)	9.2		
39	*Rorf204*	cgttAGGAGcg....attgacATG		(204)	22.4		(Minor virional protein)
40	*G*[4]	AaAAAGGAGGc...catatcaATG	gpG	(351) [350]	37.7	[43.0]	Major virional protein
41	*Rorf142*	gaAttGGAaGT....gattagATG		(142)	15.7		(Minor virional protein)
42	*Rorf117a*	AGtAgGGcGGT.gagctaccgATG		(117)	13.2		(Minor virional protein)
43	*Rorf117b*	gaActGGAGGT....gctgtgATG		(117)	13.5		(Minor virional protein)
44	*Rorf135*	ActAAGGgGaT....gaagtgATG		(135)	14.9		(Minor virional protein)
45	*P*[4]	AGgAAGGAaGTagttacaaaaATG	gpP	(178) [177]	18.8	[26.0]	Major virional protein
46	*Rorf143*	taAtAGGAGGa...atttattATG		(143)	16.9		(Minor virional protein)
47	*Rorf198*	AcgAAGaAGaa.gtgatgtgaATG		(198)	22.5		(Minor virional protein)
48	*Rorf1608*	gcAAAGGAGGg...atagtccATG		(1608)	172.8		(Minor virional protein)
49	*Rorf269*	AGAAgGGAGGT..ttgcattaGTG		(269)	30.2		
50	*Rorf372*	tactAGGAGGc...ttttgacATG		(372)	41.1		
51	*Rorf100*	ctggAcaAGGT........agGTG		(100)	10.6		
52	*Rorf282*	AcgAAGGAaGg..ggatagcgATG		(282)	30.5		
53	*B*[4]	AaAAAGGAGag...tgattgaATG	gpB	(602)	67.0	[64.0]	Major virional protein
54	*O*[4]	tactAGGAGGT.gatttaaatATG	gpO	(261) [260]	28.6	[29.0]	Major virional protein
55	*Rorf50*[5]	cAAAGGAGGca...gttaataATG		(50)	5.4		
56	*Rorf118*[5]	AtAAtGGAGat..gatggaatGTG		(118)	13.1		
57	*hol*[5]	ttAAAGGAGtg....gaagtaATG	Hol	(142)	14.2		Holin
58	*lys*[5]	cGgAAGGAGGc...agtacgaGTG	Lys	(442)	48.4	[48.0]	Muramidase
59	*Rorf175*[5]	tGAAAGGAGac....gtcacaATG		(175)	20.1		(Membrane-binding)
60	*Rorf148*	AGAAAGGAaGT...cttaaaaATG		(148)	17.0		Arginine-rich
61	*Rorf192*	AaAAgGGAGaT....tataaaATG		(192)	21.6		
62	*Rorf232*	.....ATG		(232)	25.8		
16s rRNA		3'- UCUUUCCUCCA -5'					

[1]See the text for the details; [2]ORFs were numbered from 1 to 62 (see Fig. 1);[3]Kakikawa et al. (1997); [4]Kakikawa et al. (1996); [5]Oki et al. (1996a). In Kakikawa et al. (1996), ORFs *Rorf204*, *Rorf142*, *Rorf135*, *Rorf143*, *Rorf282*, *Rorf50* have been named *RorfU1*, *RorfU2*, *RorfU3*, *RorfU4*, *RorfU5*, *RorfU6*, respectively. aa, amino acids. Mass, molecular weight from DNA sequence. [ ], deduced from protein sequence or SDS PAGE (Kakikawa et al., 1996, 1997; Oki et al., 1996a).

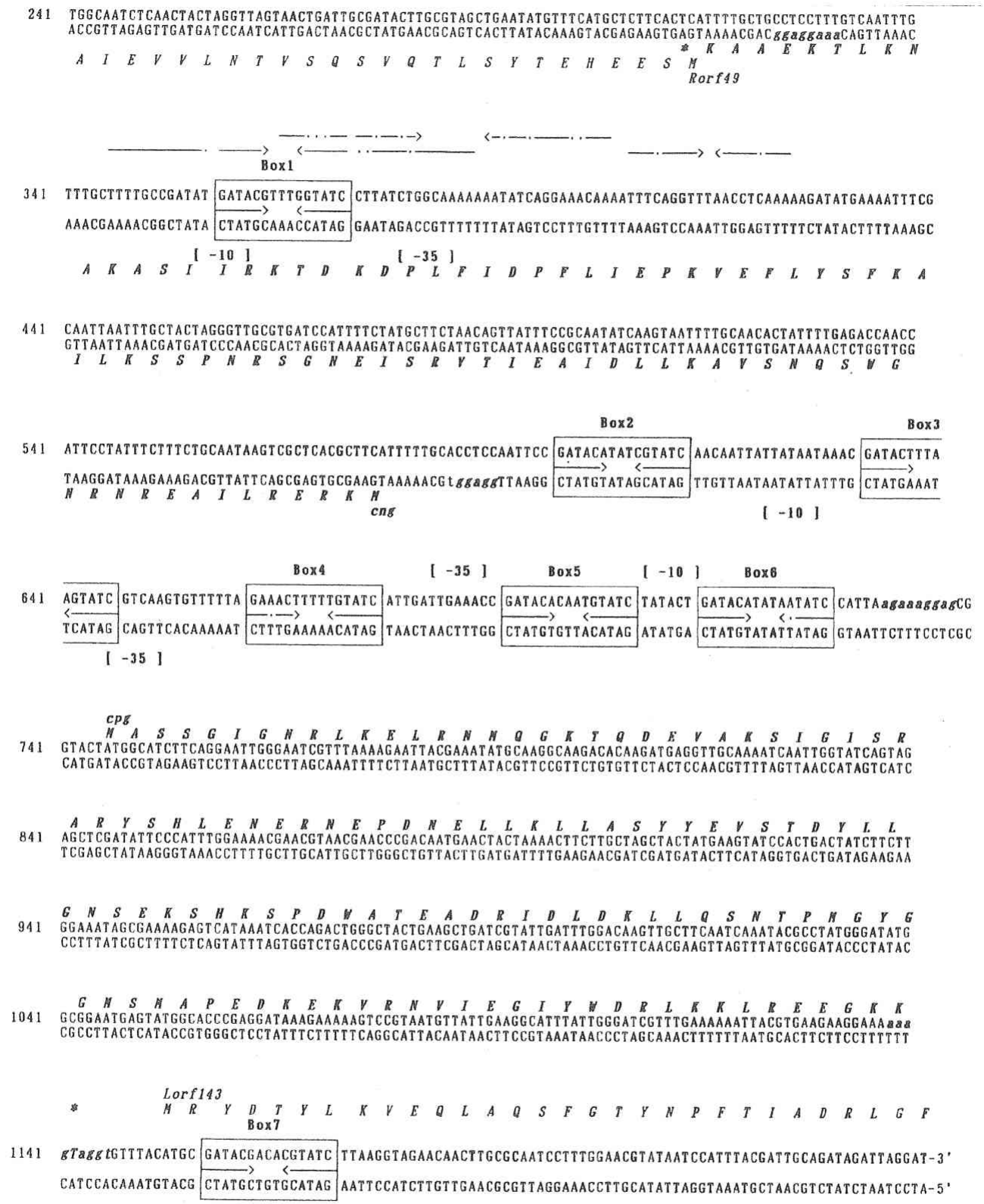

Fig. 2. DNA sequence surrounding the ϕg1e *cng* and *cpg* ORFs. A DNA region of 1000 bp containing *cng* and *cpg* (see Fig. 1) was shown with their deduced amino acid sequences (EMBL accession No. X98106). Asterisks, stop codons; bold italic lower letters, putative ribosome binding sites (R. B.); arrows, inverted repeats. The seven GATAC-boxes (for the details, see the text) were enclosed rectangularly. −35 and −10, potential mRNA promoter sequences.

promoter/operator-like sequences (in the details, see below). Cpg bears a significant resemblance to other presumed SOS-related repressors of the *Lactococcus* phages Tuc2009 (Van de Guchte et al., 1994), BK5-T (Boyce et al., 1995), and r1t (Van Sinderen et al., 1996), the lambdoid phages lambda, P2, P22, and 434 as well as the *E. coli* proteins LexA (Ljungquist et al., 1984; Saha et al., 1987) and LacR (Pabo and Sauer, 1984), although the size of φg1e Cpg is shorter than those of other repressors.

The N-terminal domain of φg1e Cpg resembles that of the lambda repressor family, and contains a potential helix-turn-helix motif implicated in DNA binding (Fig. 3). In the well characterized repressor of lambda CI or *E. coli* LexA, its C-terminal region has two key amino acids, Ser and Lys, which have been proposed to function in RecA-mediated cleavage (Sauer et al., 1990; Kim and Little, 1993). φg1e Cpg contains two corresponding residues, Ser-101 and Lys-131 (see Fig. 2). The C-terminal halves of LexA and CI contain one more common amino acid stretch, Ala-Gly, constituting the site of RecA-mediated cleavage (Sauer et al., 1990; Kim and Little, 1993). Unlike other repressors, φg1e Cpg is devoid of such Ala-Gly residues in its C-terminal half (Fig. 2). As shown in Fig. 4, φg1e was induced by UV as well as MMC. Like lambda, the φg1e induction by UV or MMC seems to be dependent on a host SOS-related pathway(s), but its molecular mechanism may be different from that of lambda, as predicted in 186 (Lamont et al., 1989).

The deduced *cng* protein (Cng) is composed of 88 amino acids (Fig. 2). Its N-terminal half shows similarity (38% identity) to that of Cpg (see above), and contains a potential helix-turn-helix motif (see Fig. 3). In addition, Cng resembles (33% overall identity) the lambda Cro repressor (Anderson et al., 1981), as well.

3.4. hel *and* ntp

Four putative proteins encoded by *Rorf184-Rorf117-Rorf167-hel*, localized about 1000 bp downstream of *cng* (Fig. 1), show similarity (28, 37, 30, and 34% overall identities, respectively) to those of four putative ORFs (ORF266-ORF111-ORF169-ORF234) of the *Lactococcus* phage BK5-T (Boyce et al., 1995).

In φg1e, the deduced proteins of Hel (220 amino acids) and Ntp (280 amino acids) contain so called NTP-binding-like sequences. The NTP-binding proteins reported so far have two highly conserved regions (known as domains A and B) accompanied by an additional motif(s) such as C, 1a, II, and/or IV, and have been classified putatively into several superfamilies (Ilyiana et al., 1992): e.g., UvrA-related ATPases (containing A-II-B-IV motifs); DNA and RNA helicases such as T antigen of SV40 (A-B-C); DnaB-associated helicases such as bacterial DnaB protein (A-1a-B). φg1e Hel (as well as BK5-T gpORF234) carries A-1a-B-like regions in the N-terminal half (Fig. 5A). On the other hand, φg1e Ntp has A-B-C-like sequence in the C-terminal half, in addition to a Zn-finger-like motif, located in its N-terminus (Fig. 5B). Hel (as well as gpORF234) may be related to a DnaB-helicase for DNA replication, and Ntp may be involved in morphogenesis, as a subunit of a terminase.

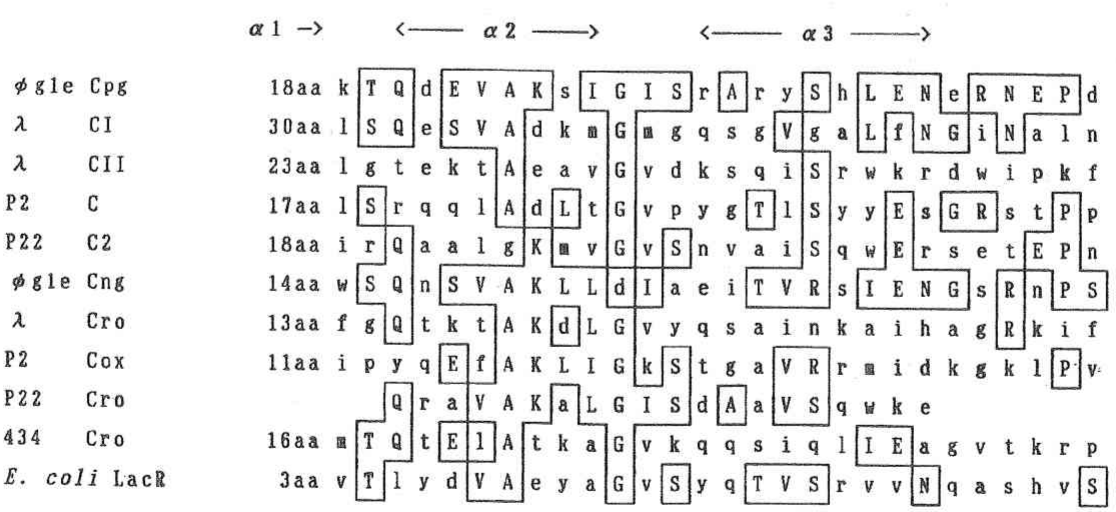

Fig. 3. Comparison of the putative helix-turn-helix structures of the φg1e *cpg* and *cng* products with those of other DNA-binding proteins. The N-terminal amino acid sequences of φg1e Cpg and Cng (this study) were compared with those of other DNA-binding proteins of the lambdoid phages, lambda, P2, P22, and 234 as well as the *E. coli* LacR repressor protein (see for a paper, Pabo and Sauer, 1984): CI, CII and Cro, from lambda; C and Cox, from P2; C2 and Cro, from P22; Cro, from 434. The helix-turn-helix motif was indicated by α 2-turn-α 3 (Pabo and Sauer, 1984). The amino acids of the putative φg1e proteins homologous to those of other DNA-binding proteins were boxed.

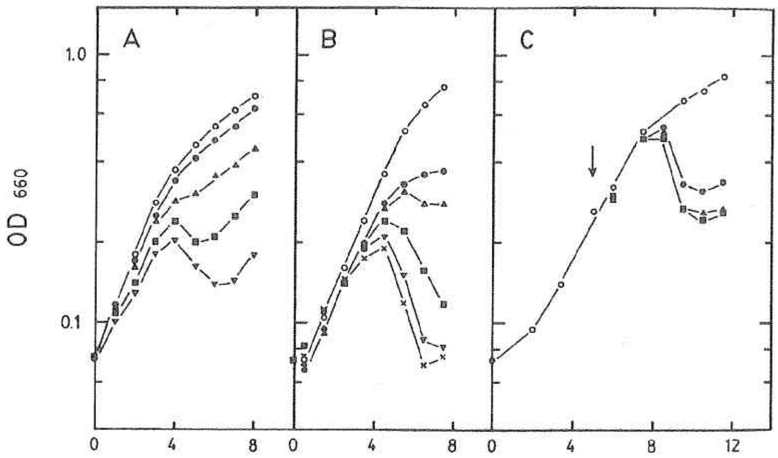

Fig. 4. Induction of ϕg1e by UV or MMC. (A) G1e cells were irradiated with an UV lamp, and their changes in absorbance at A_{660} was measured as described in Section 2. Irradiation periods were as follows: ○, 0 s; ●, 5 s; ▲, 15 s; ■, 25 s; ▼, 35 s. (B) MMC was added to early-log-phase G1e cells (A_{660}, 0.1), and the change in A_{660} was measured as in (A). Final concentrations of MMC (μg/ml) were as follows: ○, 0; ●, 0.5; ▲, 0.75; ■, 1.5; ▼, 3.0; ×, 5.0. (C) Early-log-phase G1e cells (A_{660}, 0.2) were exposed to MMC for 20 min. The treated cells were collected, washed, and then suspended in fresh medium (Section 2). The change in A_{660} was measured as in (A). Final concentrations of MMC (μg/ml): ○, 0; ●, 5.0; ▲, 7.5; ■, 10.0.

(A)

```
                                     Motif A                                       Motif Ia
           1
Hel        MKFYE--DGNIPEIPNMYFIY GDGGTGKT SVVKQFVGHKL-LFSFDMSSNVLIGDKD VDVIMFEH RDMPNIQAMVEQYVMQGIQDVKYRII-VLD
gpORF234   MKITKATDINRTQYWRVLL-Y GKPGLGKT SAVKGLTGRTLVL-SLDNSHKVLSGIPN IDVRTIDD EGLESFNRNEPIEDINIFLKELDVVIGDYD
           1
Consensus         <hydrophobic>G--G-GKT                                 V-----E

                     Motif B
Hel        NITALQNLV-LENI DD AA------KDN-R---QNYQKLQLWF-RDLGTIL-KESGKT--VYATAHQLDNGSSGLSGE-GRYQADMNEKTFNAFTSMF
gpORF234   NL-AIDN-VTSFQS DW LIERGRSSKGGIRNEIQDYGDWTNYFLR----IMTKIYGLPINVYVTA-WED-TQEI-SLEDGRVITQFVPKIRKQVLSEL
Consensus    <hydrophobic>DD

                                                          280
Hel        DL-----VGRIYL---TGGERMIDLDPEKGNHAKNRIDGRKLIKANELIQTSKGAK
gpORF234   -LGWTDVVGRIKVNPNTG-NRGAILEGNDGIYAKNRIDSRTACPIDEL-FKFEGEK
                                                          234
```

(B)

```
                                         Z-finger
           1
Ntp        MSETTSKSARGISFPELQRLKTSDQV CPRHGVNMVYMQGHQPFCMVC AKEKIEQQNHKIIDHANDCWHKRRTSDVLAMDSIFDDPTLMDANFDNFRP
Consensus                                C--H------------C--C

                                       Motif A
Ntp        NSSESANNLKLARKIAGEYLNPKTTYNTILT GLPGRGKS HLALSIAKAVNDHADKPMACLFVSVNELFRLIKGSFGHPDSRYTEQNMVQLLGDAD
Consensus                       <hydrophobic>G--G-GKS                                                     <hydro-

             Motif B                                   Motif C
                                                                                                      280
Ntp        LLVL DD LGSEATFQSDQSKNRKEASDYVQNVLFGIVNNRQR TIITTNL GSAGLANVYNPKIISRLYRGINGHVISFTAATPDTREVSF
Consensus  phobic>DD                       <hydrophobic>T---tN
```

Fig. 5. Putative DNA-binding motifs of the ϕg1e *hel* and *ntp* products. The NTP-binding motifs of Hel and Ntp were shown together with those found in the putative ORF234 product of the *Lactococcus* phage BK5-T (Boyce et al., 1995). The NTP-binding motifs and the Zn-finger referred in this study were derived from Gorbalenya et al. (1992), Koonin and Gorbalenya (1992) and Ilyiana et al. (1992).

3.5. *ORFs for minor capsid proteins*

The φgle virion particles consist of four major and more than sixteen minor proteins (Kakikawa et al., 1996). As shown in Fig. 1 (see also Table 1), an about 22-kb region of φgle genome contains twenty-three consecutive ORFs from *Rorf281* to *O*, including the four major proteins (Kakikawa et al., 1996). Immediately upstream of *Rorf281*, there is a noncoding region of about 600 bp rich in secondary structure, repeated sequence, and promoter-like sequence (data not shown), whereas *O* is followed by a potential rho-independent mRNA terminator (Kakikawa et al., 1996; Oki et al., 1996a).

In twelve consecutive ORFs from *Rorf508* to *Rorf198* (Fig. 1; Table 1), their putative proteins show similarity (from 29 to 49% overall identity) to other predicted capsid proteins of the *Lactobacillus* virulent phage LL-H as well as the *Lactobacillus* temperate phage mv4 (Mikkonen and Alatossava, 1994). For example, the identity between gp*Rorf117a* (φgle) and gpORF118 (LL-H) is 49%. These comparative results suggest that the φgle proteins specified by the 22-kb region are capsid components.

3.6. *Putative promoter/repressor system*

As shown in Fig. 2, a φgle genome region of about 1000 bp, which contains *cng* and *cpg* (see above), carries seven homologous inverted sequences with a 15-bp consensus (termed GATAC-box) of 5′-GATAC(a/t)t(a/t)(a/t)(a/t)GTATC-3′. The box1 is situated within a 3′-region of *cng* rich in complicated secondary structure, and its R-strand is accompanied by an mRNA-promoter-like sequence. Other five consecutive boxes, box2-(19 bases)-box3-(14)-box4-(13)-box5-(6)-box6, are positioned in an intergenic region between *cng* and *cpg*. The two boxes 3 and 5 are accompanied by a possible mRNA promoter, on the R-strand and the L-strand, respectively. On the other hand, the box7 is localized just downstream of *cpg*. Like the lambda CI/Cro system (Campbell, 1994), these GATAC-boxes of φgle may function as an operator, and may be recognized by Cpg and/or Cng (see above).

Further studies on structure, expression and function of the φgle genes are currently in progress.

References

Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981) Structure of the cro repressor from bacteriophage λ and its interaction with DNA. Nature 290, 754–758.

Boyce, J.D., Davidson, B.E. and Hillier, A.J. (1995) Sequence analysis of the *Lactococcus lactis* temperate bacteriophage BK5-T and demonstration that the phage DNA has cohesive ends. Appl. Environ. Microbiol. 61, 4089–4098.

Campbell, A. (1994) Comparative molecular biology of lambdoid phages. Annu. Rev. Microbiol. 48, 193–222.

Davidson, B.E., Powell, I.B. and Hillier, A.J. (1990) Temperate bacteriophages and lysogeny in lactic acid bacteria. FEMS Microbiol. Rev. 87, 79–90.

Dupont, L., Boizet-Bonhoure, B., Coddeville, M., Auvray, F. and Ritzenthaler, P. (1995) Characterization of genetic elements required for site-specific integration of *Lactobacillus delbrueckii* subsp. *bulgaricus* bacteriophage mv4 and construction of an integration-proficient vector for *Lactobacillus plantarum*. J. Bacteriol. 177, 586–595.

Fremaux, C., de Antoni, G.L., Raya, R.R. and Klaenhammer, T.R. (1993) Genetic organization and sequence of the region encoding integrative functions from *Lactobacillus gasseri* temperate bacteriophage φadh. Gene 126, 61–66.

Gorbalenya, A.E., Koonin, E.V. and Wolf, Y.I. (1992) A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. FEBS Lett. 262, 145–148.

Ilyiana, T.V., Gorbalenya, A.E. and Koonin, E.V. (1992) Organization and evolution of bacterial and bacteriophage primase-helicase systems. J. Mol. Evol. 34, 351–357.

Kakikawa, M., Oki, M., Tadokoro, H., Nakamura, S., Taketo, A. and Kodaira, K.-I. (1996) Cloning and nucleotide sequence of the major capsid proteins of *Lactobacillus* bacteriophage φgle. Gene 175, 157–165.

Kakikawa, M., Oki, M., Watanabe, N., Yasukawa, H., Masamune, Y., Taketo, A. and Kodaira, K.-I. (1997) Characterization of the genes encoding integrative and excisive functions of *Lactobacillus* phage φgle: cloning, sequence analysis, and expression in *Escherichia coli*. Gene, in press .

Kim, B. and Little, J.W. (1993) LexA and λ repressors as enzyme: specific cleavage in an intermolecular reaction. Cell 73, 1165–1173.

Kodaira, K.-I., Oki, M., Kakikawa, M., Kimoto, H. and Taketo, A. (1996) The virion proteins encoded by bacteriophage φK and its host-range mutant φKhT: host-range determination and DNA binding properties. J. Biochem. 119, 1062–1069.

Koonin, E.V. and Gorbalenya, A.E. (1992) The superfamily of UvrA-related ATPases includes three more subunits of putative ATP-dependent nucleases. Protein Seq. Data Anal. 5, 43–45.

Lamont, I., Brumby, A.M. and Egan, J.B. (1989) UV induction of coliphage 186: prophage induction as an SOS function. Proc. Natl. Acad. Sci. USA 86, 5492–5496.

Lillehaug, D. and Birkeland, N.-K. (1993) Characterization of genetic elements required for site-specific integration of the temperate lactococcal bacteriophage φLC3 and construction of integration-negative φLC3 mutants. J. Bacteriol. 175, 1745–1755.

Ljungquist, E., Kockum, K. and Bertani, L.E. (1984) DNA sequences of the repressor gene and operator region of bacteriophage P2. Proc. Natl. Acad. Sci. USA 81, 3988–3992.

Matern, H.T., Klein, J.R., Henrich, B. and Plapp, R. (1994) Determination and comparison of *Lactobacillus delbrueckii* ssp. *lactis* DSM7290 promoter sequences. FEMS Microbiol. Lett. 122, 121–128.

Mckay, L.L., Cords, B.R. and Baldwin, K.A. (1973) Transduction of lactose metabolism in *Streptococcus lactis* C2. J. Bacteriol. 115, 810–815.

Mikkonen, M. and Alatossava, T. (1994) Characterization of the genome region encoding structural proteins of *Lactobacillus delbrueckii* subsp. *lactis* bacteriophage LL-H. Gene 151, 53–59.

Oki, M., Kakikawa, M., Yamada, K., Taketo, A. and Kodaira, K.-I. (1996a) Cloning, sequence analysis, and expression of the gene encoding lytic functions of bacteriophage φgle. Gene 176, 215–223.

Oki, M., Kakikawa, M., Taketo, A. and Kodaira, K.-I. (1996b) Genome structure of a *Lactobacillus* plasmid pNMO belonging to the rolling-circle type replicon. J. Gen. Appl. Microbiol. 42, 271–283.

Pabo, C.O. and Sauer, R.T. (1984) Protein-DNA recognition. Annu. Rev. Biochem. 53, 293–321.

Pouwels, P.H. and Leer, R.J. (1993) Genetics of lactobacilli: Plasmid and gene expression. Antonie van Leeuwenhoek 64, 85–107.

Saha, S., Haggard-Ljungquist, E. and Nordstrom, K. (1987) The *cox* protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. EMBO J. 6, 3191–3199.

Sauer, R.T., Jordan, S. and Pabo, C.O. (1990) λ Repressor: a model system for understanding protein-DNA interactions and protein stability. Adv. Protein Chem. 40, 1–61.

Schouler, C., Ehrlich, S.D. and Chopin, M.-C. (1994) Sequence and organization of the lactococcal prolate-headed bIL67 phage genome. Microbiology 140, 3061–3069.

Taketo, A. (1996) Electrotransformation of bacteria. In: Zimmermann, U. and Neil, G.A. (Eds.), Electromanipuration of Cells. CRC Press, pp. 107–135.

Van de Guchte, M., Daly, C., Fitzgerald, G.F. and Arendt, E.K. (1994) Identification of the putative repressor-encoding gene *c1* of the temperate lactococcal bacteriophage Tuc2009. Gene 144, 93–95.

Van Sinderen, D., Karsens, H., Kok, J., Terpstra, P., Ruiters, M.H.J., Venema, G. and Nauta, A. (1996) Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t. Mol. Microbiol. 19, 1343–1355.