Microbiology (2002), **148**, 985–1001 Printed in Great Britain

Sequence analysis of the lactococcal bacteriophage bIL170: insights into structural proteins and HNH endonucleases in dairy phages

Anne-Marie Crutz-Le Coq,[1] Bénédicte Cesselin,[2] Jacqueline Commissaire[2] and Jamila Anba[1]†

Author for correspondence: Anne-Marie Crutz-Le Coq. Tel: +33 1 34 65 21 06. Fax: +33 1 34 65 21 05.
e-mail: lecoq@jouy.inra.fr

Laboratoire de Génétique Microbienne[1] and Unité de Recherches Laitières et de Génétique Appliquée[2], INRA, 78352 Jouy-en-Josas cedex, France

The complete 31 754 bp genome of bIL170, a virulent bacteriophage of *Lactococcus lactis* belonging to the 936 group, was analysed. Sixty-four ORFs were predicted and the function of 16 of them was assigned by significant homology to proteins in databases. Three putative homing endonucleases of the HNH family were found in the early region. An HNH endonuclease with zinc-binding motif was identified in the late cluster, potentially being part of the same functional module as terminase. Three putative structural proteins were analysed in detail and show interesting features among dairy phages. Notably, gpl12 (putative fibre) and gpl20 (putative baseplate protein) of bIL170 are related by at least one of their domains to a number of multi-domain proteins encoded by lactococcal or streptococcal phages. A 110- to 150-aa-long hypervariable domain flanked by two conserved motifs of about 20 aa was identified. The analysis presented here supports the participation of some of these proteins in host-range determination and suggests that specific adsorption to the host may involve a complex multi-component system. Divergences in the genome of phages of the 936 group, that may have important biological properties, were noted. Insertions/deletions of units of one or two ORFs were the main source of divergence in the early clusters of the two entirely sequenced phages, bIL170 and sk1. An exchange of fragments probably affected the regions containing the putative origin of replication. It led to the absence in bIL170 of the direct repeats recognized in sk1 and to the presence of different ORFs in the *ori* region. Shuffling of protein domains affected the endolysin (putative cell-wall binding part), as well as gpl12 and gpl20.

Keywords: *Lactococcus lactis*, *Siphoviridae*, host specificity, transglycosylases, HNH endonucleases

INTRODUCTION

Phages of *Lactococcus lactis* disturb fermentation processes in the dairy industry. Bacterial strains have developed resistance mechanisms to prevent initial phage DNA penetration or phage intracellular development. However, probably due to dynamic evolution of the phages, phage attacks remain a persistent problem (Forde & Fitzgerald, 1999). Lactococcal phages isolated from the dairy environment essentially belong to the *Siphoviridae* superfamily and fall into three prevalent groups of DNA homology, designated c2, 936 and P335 (Jarvis *et al.*, 1991). bIL170 belongs to the 936 group of lactococcal phages, known to contain only virulent phages that are closely related as observed by DNA hybridization experiments (Jarvis *et al.*, 1991). The c2 group has the same features, whereas the P335 group

---

† **Present address:** U.E.P.S.D., INRA, 78352 Jouy-en-Josas cedex, France.

Abbreviation: indel, insertion/deletion.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are bIL170, AF009630; bIL120, AY054975; bIL15, AY054976; bIL191, AY054977; bIL77, AY054978.

contains mainly temperate phages that are not necessarily closely related at the DNA sequence level (Chopin *et al.*, 2001). In the 936 group, only a few genomic segments (essentially encoding the holin, lysin, major structural protein, structure-specific endonuclease and *ori* fragment) have been studied experimentally (Bidnenko *et al.*, 1998; Chandry *et al.*, 1997; Chung *et al.*, 1991; Kim & Batt, 1991a). In the c2 and P335 groups, a number of structural proteins have been experimentally characterized as well as endolysins, origins of replication and transcriptional regulators, or the lysogeny module in the case of temperate phages (see Table 2 for selected references).

Important biological characteristics of the phages, such as host range or sensitivity to bacterial resistance mechanisms cannot be drawn simply from the group they belong to. The prediction of such traits will come from a better understanding of the molecular biology of the phage life cycle, phage–host interactions and comparative sequence analysis. Indeed, the acquisition and the analysis of sequence data have an impact on our understanding of phage biology at different levels: population structure, evolution, taxonomy and prediction of gene function. Past years have seen an accumulation of complete genome sequences from lactococcal phages and also from a large set of phages infecting other hosts. Since the first proposal of the modular theory of phage evolution in 1980, comparative genomics has highlighted a common scheme of modular genetic organization among various phages infecting Gram-positive or Gram-negative bacteria (Casjens *et al.*, 1992; Hendrix *et al.*, 1999): genes are organized in functional modules (carrying out particular biological functions) that can be exchanged between phages having access to a common gene pool. The modules can be entire sets of genes or single genes, but also gene segments encoding distinct protein domains as already stated for streptococcal phages (Neve *et al.*, 1998). Brussow & Desiere (2001) proposed a lambda supergroup of *Siphoviridae*, to which the lactococcal phages of the 936 group belong, as well as some other dairy phages occupying the ecological niche of the dairy environment, and for which a relative conservation of gene order was noted, especially in the late cluster. The relative conservation of gene maps can be useful for extrapolating and predicting gene function for distantly related phages even showing little or no regions of homology. However, individual variation among closely related phages is also a key element in sequence analysis, likely to help in determining functional modules, especially those involved in host-range specificity.

Among lactococcal phages of the 936 group, the complete genome sequence of phage sk1 has been determined (Chandry *et al.*, 1997) and DNA portions larger than 5 kb have been sequenced for two other phages, F4-1 and bIL41 (Chung *et al.*, 1991; Kim & Batt, 1991b; Parreira *et al.*, 1996b). To further our knowledge of the genetics and evolution of this group of virulent lactococcal phages, we undertook the sequence analysis of phage bIL170. We focused on gene function prediction by analysis of protein families, identification of domains of proteins potentially involved in host specificity and on the main points of divergence with the closely related phages of the 936 group.

METHODS

Phages, bacterial strains, DNA and media. Lactococcal phages used in this study (bIL170, bIL15, bIL77, bIL120, bIL191) are from our collection and belong to the 936 group as determined by DNA hybridization (B. Cesselin, unpublished results). They were propagated on appropriate *Lactococcus lactis* strains (IL1403 for bIL170) at 30 °C in M17 broth supplemented with 5 g glucose l^{-1} and 10 mM $CaCl_2$ (Bidnenko *et al.*, 1995). *Escherichia coli* XL1-Blue MRF′ (Stratagene), transformed with pBluescript derivatives, was cultivated in Luria–Bertani broth (Maniatis *et al.*, 1982) with 100 µg ampicillin ml^{-1}.

Phage DNA was obtained from purified phage preparations essentially as described for lambda (Maniatis *et al.*, 1982). Phages were concentrated by PEG precipitation from 0·45 µm-filtered cell lysates and purified by CsCl gradient (Maniatis *et al.*, 1982). For use in sequencing reactions, phage DNA was dialysed against water. PCR products for sequencing were purified using the QIAquick Purification Kit (Qiagen).

Library construction and DNA sequencing. A two-step strategy was used for sequencing the complete bIL170 genome. First, a random library of phage DNA fragments was constructed in *E. coli*. Sequencing approximately 120 clones with inserts larger than 0·35 kb led to the determination of about 70 % of the genome. The sequence was then completed by direct sequencing of the missing parts on the phage DNA. The library of bIL170 DNA fragments was constructed as follows. Phage DNA was partially digested with *Tsp*509I and fragments from 0·4 to 0·8 kb were purified and cloned into pBluescript-II SK+ linearized with *Eco*RI and dephosphorylated. After electroporation of *E. coli* XL1-Blue MRF′, transformants were selected on LB agar plates with 100 µg ampicillin ml^{-1} and 8 µg X-Gal ml^{-1}.

DNA was sequenced with the Dye Terminator Cycle Sequencing Ready Reaction (Taq FS) Kit or the Big Dye terminator Prism Ready Reaction Kit from PE-Applied Biosystems and the Applied Biosystems sequencers ABI-373 or ABI-3700. Sequencing reactions were performed on Thermal Cyclers 2400 or 9600 (Perkin Elmer) according to the manufacturer's protocol with the following modifications for direct sequencing on phage DNA. An initial denaturation step at 95 °C for 3 min was added and the primer concentration was raised to 0·5 µM. Synthetic oligonucleotides complementary to portions of the multiple cloning site of pBluescript were used for sequencing cloned inserts. Primers designed from phage DNA sequence were between 18 and 21 nt long with an estimated T_m above 52 °C. Sequence through the *cos* site was determined after ligation of the phage template. The degenerate primers L12N and L12C (see below) used for PCR amplification of *l12*-like fragments were also used in sequencing reactions at a concentration of 1·5 µM.

PCR amplification. DNA fragments corresponding to V domains of gpl12-like proteins were amplified using degenerate oligonucleotides complementary to conserved flanking regions, L12N (5′-AAYGCWATGGCWAARGCDAC-3′) and L12C (5′-CATHGGWARCCAYTTRTARTC-3′) in domains A and B of gpl12 of bIL170 respectively (see Fig. 6a). Promega A *Taq* polymerase was used in standard reactions with 0·4 µM

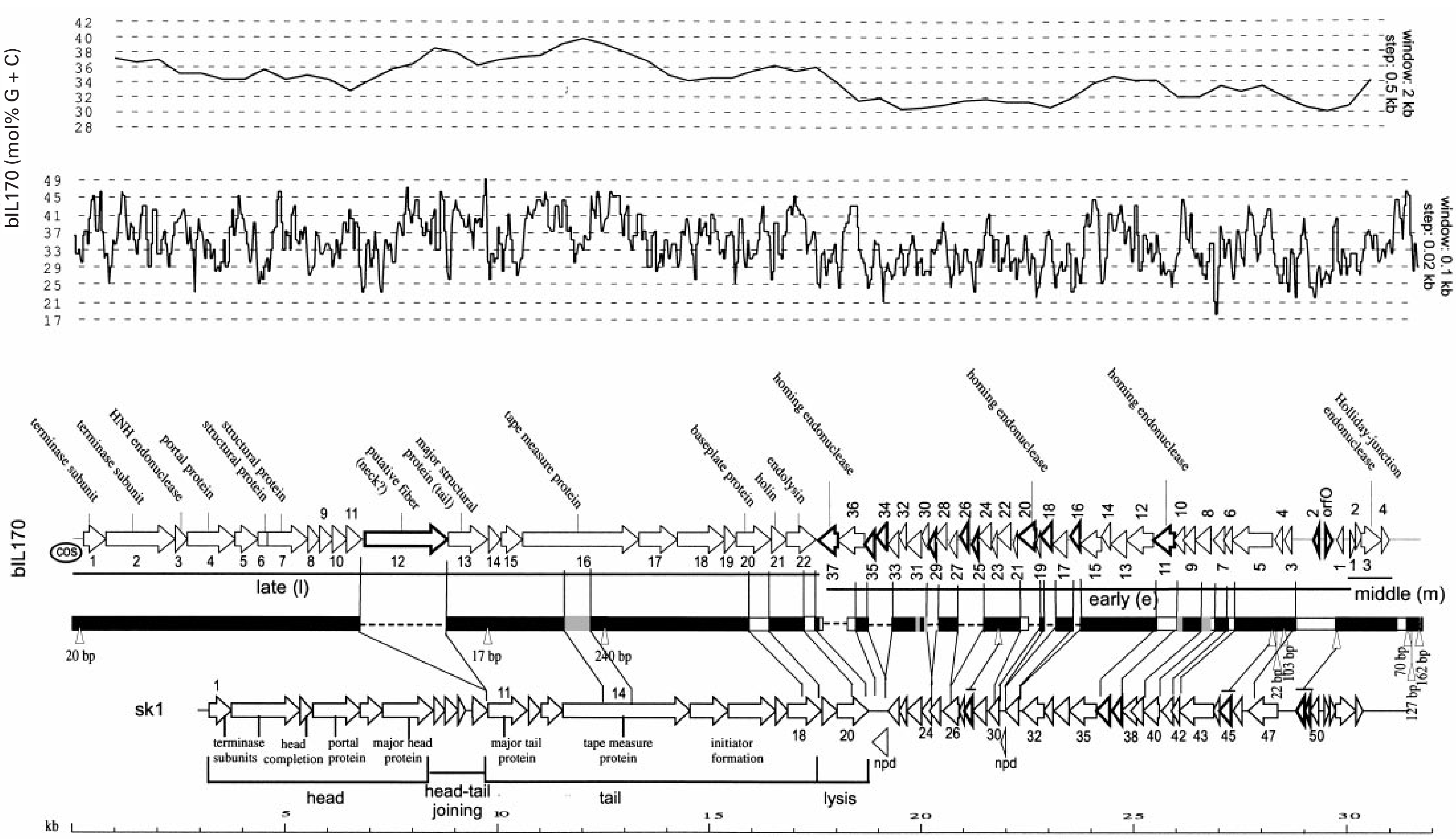

Fig. 1. Genome organization of phage bIL170, G+C content and comparison at the nucleotide level to the closely related sk1 phage. The genome of bIL170 (31754 bp) is represented linearized at the *cos* site. ORFs are depicted by leftward or rightward oriented arrows according to the direction of transcription in the late (*l1–l22*), early (*e1–e37*), middle (*m1–m4*) and *ori* (*orfO*) regions. Putative functions of gene products are indicated. G+C content calculated with two different parameters of windows is reported on the graphs at the top. The bar below the ORFs shows the identity of bIL170 at the nucleotide level with sk1 belonging to the same 936 group: in black, regions sharing more than 80% identity with sk1; in grey, those sharing 65–80% identity; in white, regions with less than 65% identity. Indels are indicated as follows: portions of DNA absent in sk1 are represented by dotted lines, additional DNA (more than 15 bp) in sk1 are indicated by triangles at the relevant positions. The genome of sk1 (28451 bp) is represented at the bottom as entered in the GenBank reference, at the same scale as that of bIL170. ORFs not previously described (npd) have been included. ORFs with no homologous counterparts in the other genome are represented with thick lines.

Table 1. Features of bIL170 ORFs and their deduced products

ORF	Start posn*	Stop posn	Size of product (aa)	Homology with sk1†	Representative similarity to databases‡	E value§	% aa identity (overlap)§	Nature of protein/putative function attributed by sequence analysis
l1	254	778	174	orf 1, 98 (174)	–			Small subunit of terminase
l2	779	2401	540	orf 2, 96 (540)	A series of putative terminase large subunits among which are			Large subunit of terminase, ATP binding motif
					Staphylococcus phage SLT (AB045978)	3 e-32	25 (537)	
					gp2 in *S. aureus* phage PVL (AB009866)	4 e-16	21 (472)	
					gpl12 in *L. lactis* phage c2	2 e-11	20 (548)	
					gp5 in *Lactobacillus* phage A2	8 e-9	21 (492)	
					Pseudomonas phage D3 (AF198256)	9 e-6	19 (450)	
					Lactobacillus phage adh	3 e-5	20 (533)	
					gp22 in *S. thermophilus* phage 7201	0·009	21 (470)	
l3	2391	2675	94	orf 3, 97 (94)	gp63 from *S. aureus* phage PVL (AB009866)	4 e-5	38 (72)	HNH endonuclease with zinc finger motif
					gp170 in *Lactobacillus* phage adh	7 e-5	35 (84)	
					gp39 in *L. lactis* prophage bIL285	0·004	38 (67)	
l4	2688	3824	378	orf 4, 95 (378)	gp25 in *Bacillus* phage phi-105 (AB016282)	1 e–05	26 (204)	Portal protein
					Portal protein (gp412) in phage SLT	0·006	21 (259)	
l5	3805	4341	178	orf 5, 97 (178)	–			
l6	4334	5515	393	orf 6, 97 (393)	Putative structural protein P43 in *L. lactis* phage F4-1 (not in databases)			Structural protein
l7	4559	5515	318	NPD‖, 97 (318)	Structural protein P35 in *L. lactis* phage F4-1 (not in databases)			Structural protein, 3 coiled coils
l8	5536	5799	87	orf 7, 100 (87)	–			
l9	5799	6113	104	orf 8, 96 (104)	–			
l10	6103	6441	112	orf 9¶, 98 (56)	–			
l11	6432	6797	121	orf 10, 95 (121)	–			
l12	6854	8815	653	–	(1) Neck passage structure: NPS (gp51) in *L. lactis* phage TP901-1	0	68 (672)	Putative accessory fibre (neck?), two coiled coils
					(2) A series of putative structural proteins among which are			Possibly involved in host range determination
					gp55 (mip6) in *L. lactis* phage Tuc2009	0	71 (676)	
					gp47 in *L. lactis* phage r1t	0	68 (672)	
					gp59 in *L. lactis* prophage bIL286	1 e-142	69 (398)	
					Putative tail fibre gp1276 in Sfi21	7 e-47	48 (195)	
					Host specificity in *L. lactis* prophage bIL286	6 e-41	54 (162)	
					Minor structural protein gpl15 in c2	6 e-38	42 (215)	
					gp1904 in *L. lactis* phage BK5-T	1 e-37	48 (184)	
					Putative anti-receptor gp695 in Sfi11	6 e-16	34 (210)	
					gp52 (mip5) in *L. lactis* phage Tuc2009	2 e-15	37 (221)	
					gp46 in *L. lactis* prophage bIL309	3 e-14	36 (199)	
l13	8839	9744	301	orf 11, 96 (301)	MCP in *L. lactis* phage F4-1	e-165	95 (301)	Major structural protein (tail)
l14	9781	10056	91	orf 12, 99 (91)	–			
l15	10076	10588	170	orf 13, 96 (170)	–			
l16	10588	13338	916	orf 14, 74 (999)	(1) A series of phage proteins generally annotated as tail proteins or tape measure, among which are			Tail tape measure protein, ATP binding site (motif A of Walker)
					Tail protein gp50 in *L. lactis* prophage bIL309	3 e-18	20 (682)	
					Tail tape measure TMP in TP901-1	5 e-17	18 (765)	
					gp17 in mycobacteriophage TM4 (AF068845)	2 e-14	19 (750)	
					gp15 in phage DT1	1 e-13	19 (570)	
					gp1560 in phage Sfi21	2 e-10	17 (612)	
					gp1487 in phage adh	1 e-7	21 (319)	
					gp42 in phage r1t	3 e-6	20 (545)	
					(2) Tail adsorption protein			
					gp31 in *L. lactis* phage bIL67	2 e-14	22 (491)	
					gpl10 in *L. lactis* phage c2	2 e-9	21 (588)	
l17	13338	14234	298	orf 15, 92 (298)	–			
l18	14234	15361	375	orf 16, 94 (375)	–			
l19	15351	15644	97	orf 17, 98 (97)	–			
l20	15634	16437	267	orf 18, 69 (190)	Putative tail component (host specificity)			Putative baseplate protein
					gp53 in *L. lactis* prophage bIL309	4 e-22	44 (110)	
					gp57 in *L. lactis* prophage bIL286	2 e-21	43 (110)	
l21	16459	16812	117	orf 19, 94 (117)	Holin in *L. lactis* phage sk1	4 e-51		Holin
l22	16809	17510	233	orf 20, 78 (194)	Lysins			Endolysin (amidase)
					In *L. lactis* phage sk1	1 e-85	78 (194)	
					In *S. aureus* phage 80 alpha (U72397)	1 e-28	49 (136)	
					gp49 in *L. lactis* phage r1t	1 e-15	55 (69)	
					gpl3 in *L. lactis* phage c2	2 e-14	55 (74)	
					Autolysin CWLA in *B. subtilis* (P14892)	8 e-5	26 (136)	
e37	18092	17616	158	–	Endodeoxyribonuclease as in sp82	1 e-9	37 (91)	HNH homing endonuclease
e36	18709	18086	207	NPD, 74 (113)	Hypothetical protein *S. pyogenes* (AE006636)	4 e-6	26 (184)	
e35	18969	18712	85	–	–			
e34	19246	18998	82	–	–			
e33	19536	19267	89	orf 21, 95 (89)	–			
e32	19712	19533	59	orf 22, 90 (59)	–			
e31	20056	19709	115	orf 23, 67 (115)	–			
e30	20223	20056	55	orf 24, 32 (65)	–			

Table 1 (*cont.*)

ORF	Start posn*	Stop posn	Size of product (aa)	Homology with sk1†	Representative similarity to databases‡	E value§	% aa identity (overlap)§	Nature of protein/putative function attributed by sequence analysis
e29	20411	20223	62	–	–			
e28	20658	20404	84	orf 25, 73 (82)	–			
e27	20909	20721	62	orf 26, 88 (34)	–			
e26	21180	20983	65	–	–			
e25	21419	21258	53	–	–			
e24	21724	21416	102	orf 26, 83 (99)	gpe11 in *L. lactis* phage c2	6 e-6	43 (67)	
e23	21843	21724	39	orf 27, 82 (39)	–			
e22	22177	21854	107	orf 29, 94 (107)	Unknown gp31 in *L. lactis* prophage bIL286	3 e-47	85 (107)	
					gp1 in *Enterococcus hirae* (Z50854)	5 e-4	35 (45)	
e21	22334	22233	33	orf 30, 87 (32)	–			
e20	22767	22345	140	–	Endodeoxyribonucleases among which are			HNH homing endonuclease
					Bacillus subtilis phage sp82 (q38137)	2 e-7	28 (164)	
					gp41 in *L. lactis* phage r1t	5 e-5	27 (148)	
e19	22878	22768	36	NPD, 86 (36)	–			
e18	23129	22878	82	–	–			
e17	23480	23166	104	orf 31, 86 (105)	–			Potential zinc-binding motif
e16	23793	23569	74	–	–			
e15	24299	23790	169	orf 32, 95 (169)	gpe18 in *L. lactis* phage c2	2 e-12	29 (168)	
e14	24511	24296	71	orf 33, 97 (71)	–			
e13	24886	24527	119	orf 34, 92 (119)	gpe16 in *L. lactis* phage c2	8 e-10	33 (118)	
e12	25526	24891	211	orf 35, 93 (211)	gp14 in *L. lactis* prophage bIL285	4 e-53	55 (203)	
					gp11 in *L. lactis* phage TP901-1	1 e-52	54 (203)	
e11	26023	25523	166	–	Endodeoxyribonucleases			HNH homing endonuclease
					gp41 in *L. lactis* phage r1t	4 e-10	35 (113)	
					In *B. subtilis* phage sp82 (q38137)	4 e-10	29 (177)	
					In *B. subtilis* phage spp1 (q38145)	3 e-9	39 (87)	
					In *B. subtilis* phage SP01 (p34081)	3 e-9	28 (166)	
					In *B. subtilis* phage phiE (q37970)	2 e-8	26 (171)	
					gp168 in *Lactobacillus* phage LLH	5 e-6	35 (98)	
e10	26276	26010	88	orf 38, 82 (83)	–			
e9	26520	26230	96	orf 39, 97 (95)	gpe13 in *L. lactis* phage c2	5 e-6	32 (82)	
e8	26894	26517	125	orf 40, 51 (132)	–			
e7	27210	26947	87	orf 41, 89 (84)	–			
e6	27391	27182	69	orf 42, 33 (69)	–			
e5	28334	27384	316	orf 44, 91 (45)	Putative DNA polymerase gpe5 in *L. lactis*			
				orf 43, 96 (270)	Phage c2 (gp3 in phage bIL67)	3 e-57	40 (314)	
e4	28577	28392	61	orf 46, 95 (61)	–			Potential zinc-binding motif
e3	28782	28612	56	orf 47, 87 (53)	–			
e2	29421	29302	39	–	–			
orfO	29559	29717	52	–	–			
e1	30001	29834	55	orf 50, 91 (55)	–			
m1	30152	30280	42	orf 51, 98 (42)	–			
m2	30285	30416	43	orf 52, 93 (43)	–			
m3	30413	30895	160	orf 53, 92 (160)	gp3 (M nuclease) in *L. lactis* phage bIL66	6 e-80	92 (159)	Holliday junction endonuclease
					gp23 in *L. lactis* phage bIL67	5 e-27	42 (161)	
m4	30892	31059	55	orf 54, 96 (55)	–			

* Non-ATG codons were as follows. TTG; *l4*, *e15*. GTG; *e8*. ATA; *e16*, *e32*. ATC; *e36*.

† ORF, % aa identity (aa overlap). Determined with the FASTA program (GCG package).

‡ Hits with lactococcal phages of the 936 group are not reported in this column unless the function has been reported.

§ Determined with the BLAST2P program. Selected hits are shown.

‖ Not previously described.

¶ Presence of a premature stop codon.

each primer, 0·2 mM each dNTP, 2·2 mM MgCl$_2$ and 1·25 U *Taq* polymerase. Phage lysates (0·5 µl) were used as templates. Reactions were performed using DNA Thermal Cycler 2400 or 9600 (Perkin-Elmer) in a total volume of 50 µl.

Computer analysis. DNA sequence was assembled with the Development Assembly Program of Staden (1996) and analysed mainly with the GCG software (Wisconsin Package, Genetics Computer Group, Madison, WI, USA). Nucleotide composition of the linear genomes was analysed with the program used for *L. lactis* (Bolotin *et al.*, 1999) with a window of 2 kb and a step of 0·5 kb except where otherwise indicated in Fig. 1. Criteria applied for identifying putative genes were length of ORFs (>30 codons), translation initiation signals [start codons ATG, TTG, GTG, ATA or ATC, preceded by a ribosome-binding site at least partly complementary to the 3′ end of 16S rRNA of *L. lactis* (3′-UCUUUCCUCCA-5′)] (Ludwig *et al.*, 1985) and minimizing intergenic regions or overlap of genes. The start codon for *l12* was assigned to nt 6854, even though no ribosome-binding site nor potential translational coupling were recognized at that position, because of the homology of the gene product to other phage-encoded proteins. *l7* is included within *l6*, deduced from an alternative downstream start codon, which was suggested by the analysis of structural proteins of the closely related phage F4-1 (Kim & Batt, 1991b) and confirmed by indirect experimental evidence (Parreira, 1996). *orfO*, although in an unexpected orientation relative to surrounding genes, was retained as a putative gene because of the presence of a putative promoter for its expression and its estimated coding probability with the GeneMark program. Note that *e2*, *e10*, *e19*, *e21* and *e29* showed a coding probability under 0·5 with the *L. lactis* matrix of the GeneMark gene prediction software (Borodovsky & McIninch, 1993).

Predicted proteins were compared to sequences in the PDB, SWISS-PROT and PIR protein databases as well as the GenBank translations with the aid of the BLAST network service at the NCBI. Prodom and Pfam protein domain databases were also used. Coiled coils were predicted by the COILS program at ISREC (http://www.ch.embnet.org/software/COILS_form.html).

RESULTS AND DISCUSSION

Sequence and organization of the phage bIL170 genome

Phage bIL170 was shown by restriction analysis to possess a genome of 32 kb with cohesive ends (data not shown). By a combination of shotgun cloning and genomic DNA sequencing, a linear sequence of 31754 bp was established with a mean redundancy of five, each region being sequenced at least once on both strands. The global G+C content of the bIL170 genome (34·3 mol%) was similar to that of the lactococcal host (35·4 mol%) (Bolotin *et al.*, 1999) and of the other sequenced lactococcal bacteriophages (see Table 2 for a non-exhaustive list).

The phage bIL170 genome map is presented in Fig. 1. Sixty-four putative genes, listed in Table 1, were determined by taking into account the criteria listed in Methods. Sixty-three are organized in three clusters corresponding to the early (37 ORFs), middle (4 ORFs) and late (22 ORFs) regions determined by transcriptional studies (Parreira *et al.*, 1996a). They are named according to the cluster they belong to (e, m and l for early, middle and late genes, respectively). The last ORF, *orfO*, is located in the region containing the putative origin of replication, on the opposite strand to the early genes.

Nucleotide comparison with other phages of the 936 group

As expected, since they belong to the 936 group that is known to be homogeneous at the DNA level, the two entirely sequenced phages, bIL170 and sk1 (Chandry *et al.*, 1997), share the same overall genetic organization and are highly similar at the nucleotide level (more than 84% nt identity over 80% of the shortest genome length). However, regions of high levels of DNA homology are interspersed with regions of lower or no similarity (Fig. 1). Genetic diversity is mainly due to insertions/deletions (indels) and to likely exchanges of fragments covering part of ORFs, entire ORF(s) or non-coding regions. Analysis of G+C content throughout the bIL170 genome revealed a globally higher G+C content in the late region than in the early region (Fig. 1, first line) but did not make it possible to identify sets of 'exogenous' genes (Fig. 1, second line).

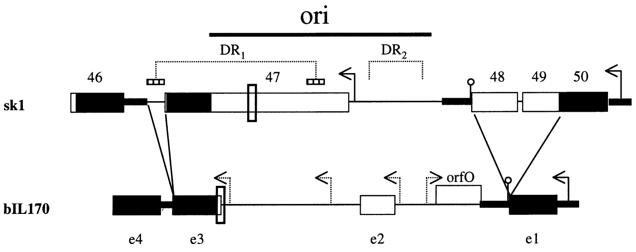

Fig. 2. Divergence in the putative *ori* regions of bIL170 and sk1. The 800 bp minimal region of sk1 able to function as a *L. lactis* plasmid origin of replication (Chandry *et al.*, 1997) is overlined. Rectangles and lines represent coding and non-coding regions respectively. Homologous regions between sk1 and bIL170 (>85% nt identity) are highlighted in black (heavy lines for non-coding regions). An identical stretch of 28 bp is boxed. The other regions show no detectable homology between the two phages. Reported direct repeats in sk1 (Chandry *et al.*, 1997) are indicated (DR_1, 67 bp imperfect repeats; DR_2, region of three pairs of small direct repeats) as well as transcription signals (dotted-line and plain-line arrows for putative and experimentally determined promoters, circles for putative transcription terminators).

The functions encoded by the late gene cluster seem to be particularly well conserved since 18 ORFs out of 22 share more than 90% aa identity over their whole length (Table 1). Such a high level of conservation is also observed with phages bIL41 and F4-1 (not shown), for which 10 kb and 6 kb sequences, covering part of the late cluster, are available respectively. The middle gene cluster is also highly conserved. Divergence is higher in the early gene cluster, where 15 bIL170 ORFs out of 37 share less than 90% aa identity with their homologues in sk1 and 11 ORFs have no homologous counterpart in sk1. Among them, three correspond to putative homing endonucleases and seven are likely to be encoded within indels. Excluding putative homing endonucleases, a total of 13 ORFs are probably encoded within indels among sk1 and bIL170. Direct repeats were observed at the boundaries of two large indels (CTTTCATT in sk1 *orf28*; imperfect 80 bp repeat in bIL170 *e16*). The target of insertion seems to be within an ORF (sk1 *orf26*) in one case (see Fig. 1).

A fragment containing a functional origin of replication in *L. lactis* has been identified in sk1, at the beginning of the early region (Chandry *et al.*, 1997). Comparison with the putative corresponding *ori* region of bIL170 provides an example of an exchange fragment, leading to different structural features in the two phages. The fragment is clearly identified by a sharp transition from high to no sequence similarity (Fig. 2). The main reported features of the minimal sk1 fragment able to provide a *L. lactis* plasmid origin of replication are absent from the bIL170 fragment, namely the 5′ end of *orf47*, tentatively suggested to be the origin-binding protein, one unit of the large DR_1 direct repeats (67 bp) and a series of small direct repeats (DR_2) (Fig. 2). It is noteworthy that the other unit of DR_1 repeats, which is outside the minimal sk1 *ori* fragment, is also absent from bIL170 because of a small deletion just encom-

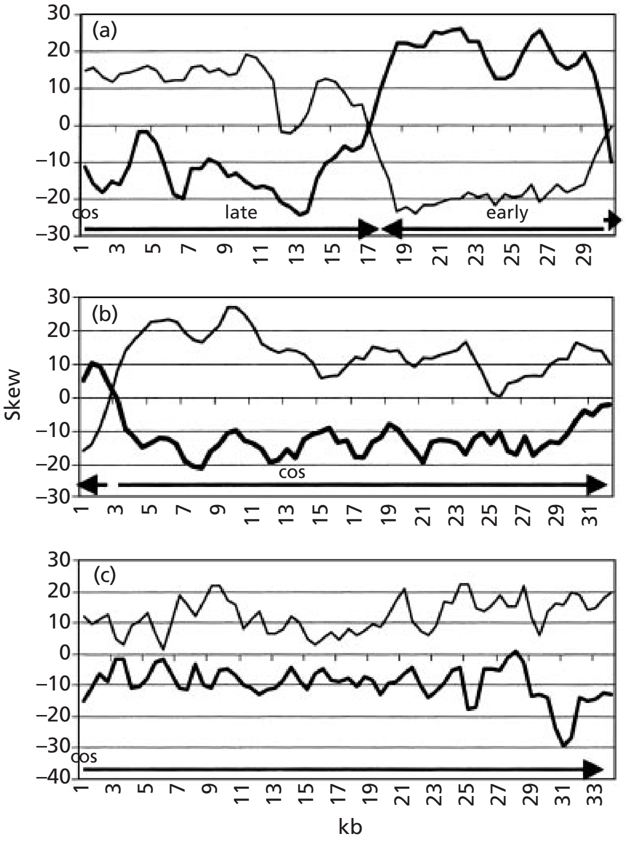

Fig. 3. Intra-strand compositional bias in dairy phage genomes. The genomes of the phages [(a) bIL170, (b) r1t, (c) DT1] are presented as entered in GenBank (see Table 2 for accession numbers). The direction of transcription is indicated by arrows at the bottom of each graph. GC and AT skews were calculated as %(C−G/C+G) (shown as thick lines) and %(A−T/A+T) (shown as thin lines), respectively. Excess of G over C in one part of the strand leads to negative value along that part.

passing it. Additional ORFs were predicted in bIL170, among which is *orf*O, which is putatively transcribed in the opposite direction to the early genes. However, the fragment does retain an A+T-rich composition (71 mol%) expected for an *ori* segment as well as a stretch of 28 bp that is perfectly conserved between the two phages but is of unknown function. Differences in the *ori* region have also been noted for lactococcal phages of the c2 group (bIL67/c2) (Waterfield *et al.*, 1996). Interestingly, exchange of the *ori* region was obtained in mutant phages of the P335 group challenged with an abortive infection mechanism (Bouchard & Moineau, 2000).

Intra-strand compositional bias in the genome of dairy phages

We examined the bIL170 genome for GC and AT skews (deviation from G = C and A = T frequencies for each strand of the DNA). This type of intra-strand compositional bias and its sharp transition in two regions of the genome has been used to locate origins and termini of replication in bacteria including *L. lactis* (Bolotin *et al.*, 1999; Lobry, 1996). The genome of bIL170 divides into two parts (Fig. 3 top panel) with predominance of G over C as well as A over T along the late and middle gene cluster (coding strand), and predominance of C over G as well as T over A in the early gene cluster (non-coding strand). GC and AT skews switch polarity in two regions, around coordinates 17 kb and 30 kb. From our analysis of other lactococcal and streptococcal phages (two are shown on Fig. 3, lower two panels), we conclude that these switches are likely to correspond to a coding strand switch. Indeed, preference of G over C has also been observed to correlate with coding strands in the genome of bacteriophages lambda and T7 (Mrazek & Karlin, 1998).

Database similarity searches for assignment of putative functions to the ORFs

Sixty-four ORFs listed in Table 1 were detected in the genome of bIL170. To attribute a function to the gene products, they were compared to sequences from databases using the BLAST program. Synteny relationships are mentioned when they emphasize the function assignment suggested by homology. The results of such an analysis focused on gene order, made by Chandry *et al.* (1997) as well as Desiere *et al.* (2001) with sk1, are summed up in Fig. 1 (under the sk1 genome line). Names of dairy phage ORFs cited throughout this paper are those attributed by the authors in the GenBank references listed in Table 2. The products of the putative genes are given the prefix gp.

Forty-eight bIL170 ORFs remain of unknown or uncertain function. We left *e5* and *e12* among the latter. Genes *e5*, *e6* and *e7* in bIL170 are organized similarly to *orf3*, *orf4* and *orf5* in bIL67, which were suggested to encode the phage DNA polymerase on the basis of conserved aa motifs (Schouler *et al.*, 1994). gpe5 of bIL170 shares 41% identity with gp3 of bIL67. However, the divergence between gpe6-like proteins in the closely related phages bIL170 and sk1 (only 33% identity) is unexpected for a subunit of the core polymerase itself. gpe12 seems to be a two-domain protein and shares 70% identity over a large N-terminal portion with gp11 from TP901-1, itself reported to be slightly similar to topoisomerase I from *Mycoplasma* (Madsen & Hammer, 1998). However, gpe12 does not appear to be related to topoisomerases in our database searches and topoisomerase I is a much longer protein than the phage proteins. Interestingly, the two phage proteins diverge in the C-terminal parts: a coiled-coil (often involved in dimerization or interaction with other partners) is predicted only in the bIL170 gene product and not in gp11 of TP901-1 (not shown).

The putative function of 16 ORFs of bIL170 was assigned by significant similarity to proteins with a proposed or experimentally determined function (Table 1). They are detailed in the remaining paragraphs according to their functional category. Structural genes

Table 2. General characteristics and GenBank accession numbers for dairy phages cited in this paper

Phage	Host	Phage group	GenBank accession	Experimentally characterized protein function (where applicable)
bIL170	*L. lactis*	936	AF009630*	
sk1	*L. lactis*	936	AF011378*	Lysis, origin of replication (Chandry *et al*., 1997)
bIL41	*L. lactis*	936	L35061	
F4-1	*L. lactis*	936	M37979 and Kim & Batt, 1991b	Structural proteins (Chung *et al*., 1991; Kim & Batt, 1991a)
bIL66	*L. lactis*	936	L35175	RuvC-like endonuclease (Bidnenko *et al*., 1998)
phi US3	*L. lactis*	936	M90423	Lysis (Platteeuw & de Vos, 1992)
bIL67	*L. lactis*	c2	L33769*	
c2	*L. lactis*	c2	L48605*	Structural proteins (Lubbers *et al*., 1995), origin of replication (Waterfield *et al*., 1996), lysogeny (Nauta *et al*., 1996)
r1t	*L. lactis*	P335	U38906*	Structural proteins (van Sinderen *et al*., 1996)
BK5-T	*L. lactis*	P335-like	L44593†	Structural proteins and origin of replication (Mahanivong *et al*., 2001), transcriptional regulator (Lakshmidevi *et al*., 1990)
Tuc2009	*L. lactis*	P335	AF109874*	Lysis (Sheehan *et al*., 1996)
TP901-1	*L. lactis*	P335	AF304433*	Structural proteins (Johnsen *et al*., 1996; Johnsen *et al*., 1995; Pedersen *et al*., 2000), replication (Ostergaard *et al*., 2001), lysogeny (Breuner *et al*., 1999; Christiansen *et al*., 1996)
phi 31	*L. lactis*	P335	AF022773	Transcriptional activator (Walker & Klaenhammer, 1998)
bIL285	*L. lactis*	P335	AF323668*	
bIL286	*L. lactis*	P335	AF323669*	
bIL309	*L. lactis*	P335	AF323670*	
Sfi11	*Streptococcus thermophilus*		AF057033*	
Sfi19	*S. thermophilus*		AF115102*	
Sfi21	*S. thermophilus*		AF115103*	
O1205	*S. thermophilus*		U88974*	Structural proteins (Stanley *et al*., 1997), lysis (Sheehan *et al*., 1999)
DT1	*S. thermophilus*		AF085222*	Structural proteins (Tremblay & Moineau, 1999) anti-receptor (Duplessis & Moineau, 2001)
7201	*S. thermophilus*		AF145054*	
LL-H	*Lactobacillus delbrueckii*		L37351	Lysis (Vasala *et al*., 1995), structural proteins (Mikkonen & Alatossava, 1994)
phi gle	*Lactobacillus plantarum*		X98106*	Structural proteins (Kakikawa *et al*., 1996), lysis (Oki *et al*., 1997, 1996)
phi adh	*Lactobacillus gasseri*		AJ131519*	Lysis (Henrich *et al*., 1995), structural proteins and origin of replication (Altermann *et al*., 1999)
A2	*Lactobacillus casei*		X97563	Terminase (Garcia *et al*., 1997), lysogeny (Alvarez *et al*., 1998; Garcia *et al*., 1999), origin of replication (Moscoso & Suarez, 2000)

* Complete genome sequence.

† Genome sequence complete in AF176025 (in which gp1904 becomes gp18).

were examined in view of their potential role in host-range determination.

DNA endonucleases

Phages with a genome over 20 kb are known to possess a number of *bona fide* endonucleases (e.g. lambda) and sequence analysis helps detect unknown proteins that are structurally related to them. In phage bIL170, we found four proteins with typical motifs of HNH endonucleases and one protein belonging to the RuvC family.

gpm3 of bIL170 shares more than 85 % identity with gp3 of the middle operon in the closely related phage bIL66. gp3 of phage bIL66 has been shown to express a structure-specific endonuclease activity (Bidnenko *et al*., 1998). Structure-specific endonucleases cleave branched DNA structures generated during replication or recombination. They belong to different structural families and could have diverse physiological roles in replication, recombination, repair, maturation or packaging of phage DNA. gp3 of bIL66 is homologous to the *E. coli* RuvC Holliday-junction resolvase (Bidnenko *et al*., 1998) and has been shown to be involved in phage sensitivity to the AbiD1 abortive infection mechanism (Bidnenko *et al.*, 1995). Its exact role in phage development is still unknown, but it is most probably essential for phage multiplication under certain conditions (Bidnenko *et al.*, 1995).

gpe11, gpe20 and gpe37 share homology with a number of endonucleases of the HNH family (Shub *et al.*, 1994) (Table 1 and Fig. 4a). Actually, gpe11, gpe20 and gpe37 seem to belong to a subfamily of proteins exhibiting an HNN rather than an HNH motif (Fig. 4a) and mainly found in phages infecting Gram-positive hosts (Dalgaard *et al.*, 1997; Foley *et al.*, 2000). Members of the HNH family can be found as free-standing ORFs between genes or encoded within introns or inteins. Most of them are homing endonucleases, involved in the mobility of their own genes or of the introns/inteins in which they are targeted (for a review see Chevalier & Stoddard, 2001). All phage hits obtained with the bIL170 gene products are intron-encoded homing endonucleases (Goodrich-Blair & Shub, 1994; Lazarevic *et al.*, 1998; Mikkonen & Alatossava, 1995; van Sinderen *et al.*, 1996). Their degree of similarity with the three gene products of bIL170 (25–40 % over at least half of the length of the proteins) and the absence of homologues in sk1 strongly suggest that gpe11, gpe20 and gpe37 are related to homing endonucleases. Comparing bIL170 with sk1 (see Fig. 1) revealed no interruption of the adjacent ORFs by the endonuclease genes, suggesting that they could occur in free-standing form and not as part of an intron. gpe11, gpe20 and gpe37 are not closer to each other (30–35 % identity) than to other members of the HNN subfamily (25–40 % identity), suggesting that each of the ORFs has arisen independently in the bIL170 genome. Their function in the phage cycle and/or the reason for their maintenance in such compact genomes is intriguing. In the case of *Bacillus subtilis* phage SP82, the intron-encoded endonuclease has been shown to confer a selective advantage on intron-flanking markers during mixed infections with SP01 (Goodrich-Blair & Shub, 1996).

gpl3 presented both the signature of the HNH endonuclease family and a potential zinc-binding domain ($CX_2CX_{36}CX_2C$). A number of proteins reported by different authors (Gorbalenya, 1994; Dalgaard *et al.*, 1997; Smith *et al.*, 1999; Aravind *et al.*, 2000) exhibit the same structural feature as gpl3, that is two couples of putative zinc ligands bracketing the ‘HN part’ of the HNH motif, defining a potentially zinc-binding subfamily of HNH-related proteins. Some of them are part of introns (probably as homing endonucleases) whereas others are free-standing phage- or bacterial-encoded ORFs. Their exact role in the host organism is likely to depend on the target of the endonuclease activity, if any. This subfamily comprises many sequence-specific endonucleases (like the homing endonucleases and some restriction enzymes such as the methylcytosine-specific McrA of *E. coli*) but also, as identified recently, a structure-specific endonuclease, endonuclease VII, of phage T4, a well-studied Holliday-junction resolvase (Aravind *et al.*, 2000) (see Fig. 4b). HNH proteins are found both in phages infecting Gram-positive and Gram-negative bacteria and are potentially involved in different stages of phage development; the role of most of them is unknown, with the notable exception of endonuclease VII of phage T4, which is involved in phage DNA packaging (Golz & Kemper, 1999). In our view, the lambda *rap* gene product (Fig. 4b) could also belong to this family. *rap* encodes a structure-specific endonuclease that is involved in phage recombination (Sharples *et al.*, 1998). Located in the *nin* non-essential region of lambdoid phages, it seems to be a (non-homologous) alternative to *rusA*-like genes, also encoding structure-specific endonucleases (Hendrix *et al.*, 1999). In the zinc-binding HNH family, we point out phage gene products (including gpl3 of bIL170), whose genes are adjacent to sites involved in phage DNA packaging and encode rather small proteins (Fig. 4b and 4c). They could be included in the same functional module as terminases and are possibly involved in phage DNA packaging. If so, they could theoretically act either as site-specific endonucleases, playing a role analogous to the cutting (*a priori* large) subunit of the terminase, or (our preferred hypothesis) as structure-specific endonucleases clearing branched replicative DNA prior to packaging, as the endonuclease VII from phage T4 associated with the so-called packasome (Golz & Kemper, 1999). In the latter case, phage bIL170 would possess two structure-specific endonucleases, one of the RuvC family (gpm3) and the other of the HNH family (gpl3).

Morphogenesis

Structural proteins. gpl6, gpl7 and gpl13 are highly conserved ($>85\%$ aa identity) among phages of the 936 group [F4-1, sk1 and bIL41 (only partial sequence for *l*13 from bIL41)] and would correspond to structural proteins characterized for phage F4-1. gpl6 and gpl7 corresponds to the P43 and P35 minor structural proteins of F4-1 (Kim & Batt, 1991b) and gpl13 corresponds to the major structural protein (Chung *et al.*, 1991). Synteny relationships observed with lambda and among dairy phages have led to the proposition that gpl13-like proteins are major tail proteins (Chandry *et al.*, 1997; Brussow & Desiere, 2001). gpl4 is homologous to putative portal proteins, a function also confirmed by synteny relationships (Brussow & Desiere, 2001). Other identified structural proteins, gpl12, gpl20 and gpl16, are discussed in separate paragraphs.

Packaging. gpl1 and gpl2 are very likely to be involved in maturation and packaging of phage DNA as terminase subunits. In double-stranded DNA bacteriophages, terminase is usually composed of a so-called small subunit, responsible for specific DNA binding, and a large subunit responsible for cutting the phage DNA into genome units and for prohead binding.

gpl1 and gpl2 share 95 % identity with the first two gene products of the late cluster of phages bIL41 (Parreira *et al.*, 1996b) and sk1 (Chandry *et al.*, 1997), which were proposed to encode the small and large subunits of the

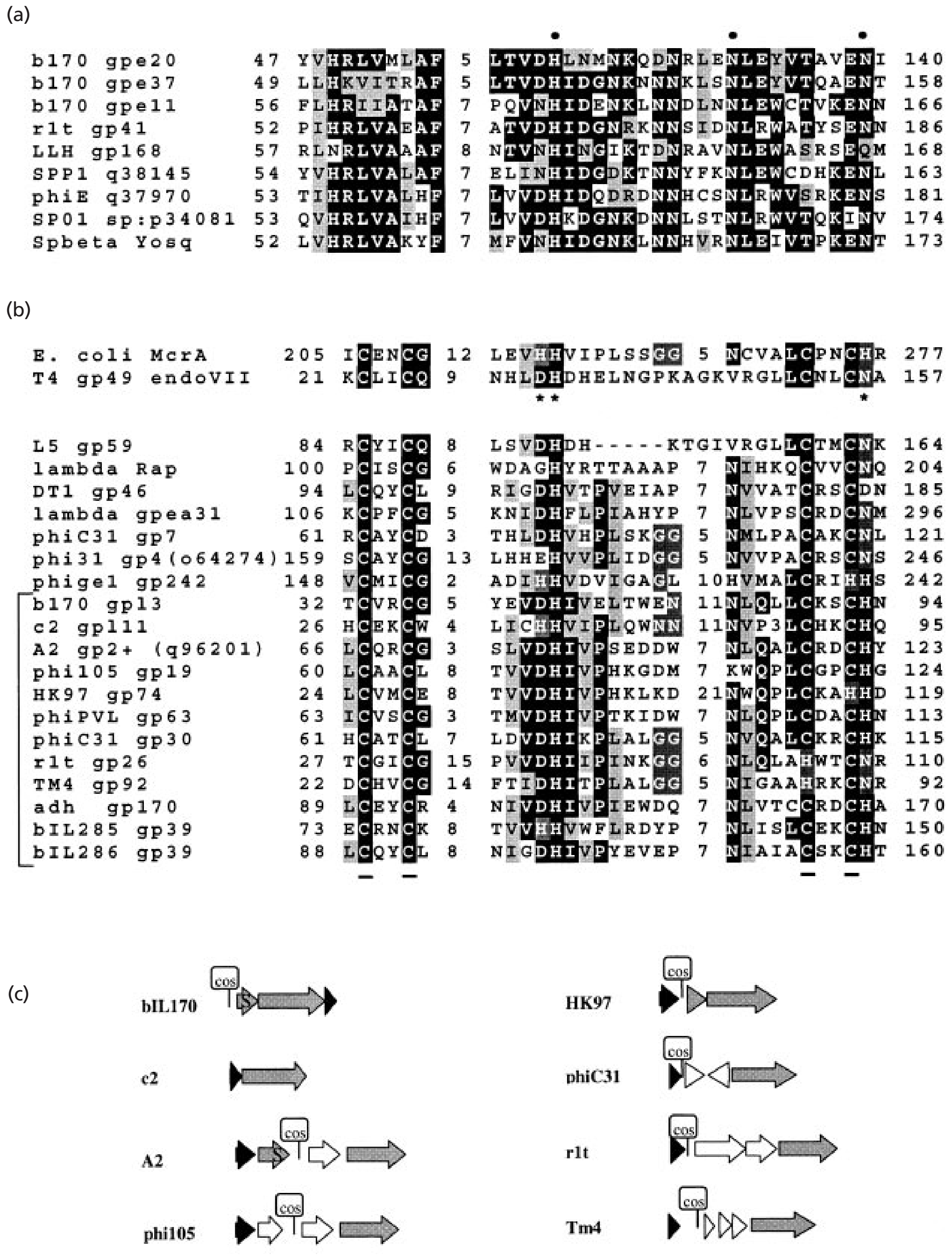

Fig. 4. Subfamilies of HNH proteins and related phage proteins. (a) HNN motif in phage-encoded homing endonucleases. Amino acids that are identical or similar in at least one third of the sequences are shaded black or grey, respectively. Numbers within the alignments indicate the numbers of aa residues not shown at that position. Numbers at the left and at the right of the alignments indicate numbers of residues preceding the aligned sequences and total number of residues in the proteins respectively. (b) HNH proteins with zinc-binding motif. First block: conserved motif between McrA, a restriction endonuclease, and the His-acid cluster of endonuclease VII (gp49) of phage T4, a Holliday-junction resolvase. Residues involved in the cleavage activity of endonuclease VII are indicated by asterisks under the sequence. Second block: phage proteins. Those whose gene maps close to packaging sites are bracketed (see panel c). The shading of amino acids is the same as in (a) with an additional dark-grey shading for amino acids commonly found at that position in other HNH proteins. Columns of residues potentially involved in zinc liganding are underlined. Note that proteins highly homologous to gp46 of DT1 (gp3 of *L. lactis* phage phi 31, gp170 of Sfi19, gp175 of Sfi21, gp20 of phi 7201) or gpl3-like proteins in other phages of the 936 group are not shown here. gp2+ of phage A2 is the product of a modified *orf2*, corrected arbitrarily for two frameshifts to maximize homology in the C terminal part. gp19 of phi 105 was probably misannotated as holin in GenBank reference L35561. (c) Potential functional modules including an HNH endonuclease (in black) and subunits of terminase (in grey; S, small subunit).

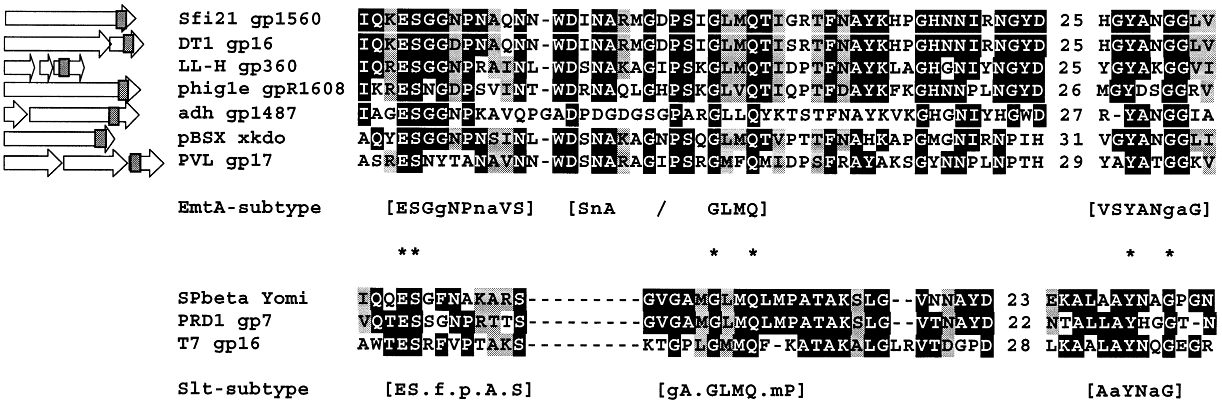

Fig. 5. Motifs associated with putative lytic transglycosylase activity in phage proteins. Top block: EmtA subfamily (Blackburn & Clarke, 2001). ORFs related to gp1560 of Sfi21 are represented by arrows for each indicated phage. The location of transglycosylase motifs are indicated by the grey boxes. gpR1608 in phi g1e, a minor structural protein (Kodaira *et al.*, 1997), is probably a tail tape measure protein (Desiere *et al.*, 2000) and misannotated as a capsid protein. Bottom block: Slt subfamily (Blackburn & Clarke, 2001). Residues conserved among both subfamilies, belonging to the same superfamily, are indicated by asterisks.

terminase respectively. gpl2 exhibits expected Walker's A and B motifs for ATP binding, and fits well in the multiple sequence alignment of terminases (not shown) (Smith *et al.*, 1999).

Cell lysis

Cell lysis most probably relies on gpl21 and gpl22, as holin and endolysin respectively. gpl21 is highly homologous to the sk1 holin (Chandry *et al.*, 1997) and exhibits structural characteristics such as two predicted transmembrane helices and a highly charged carboxy terminus. gpl22 is composed of two domains, as expected for lytic enzymes; an N-terminal domain possessing catalytic activity and a C-terminal one involved in substrate binding (Garcia *et al.*, 1990). The N-terminal moiety (first 160 aa) is highly similar to the N moiety of endolysins in the closely related phages sk1 (Chandry *et al.*, 1997) and phi US3 (Platteeuw & de Vos, 1992). It has a catalytic activity of the amidase type since it shares 50% identity with a central region of two different amidases from staphylococcal phages, LytA (Wang *et al.*, 1992) and PlyTW (Loessner *et al.*, 1998). The C-terminal part of gpl22 shared 50% identity with that of endolysins encoded by lactococcal phages of the c2 and P335 groups. The bIL170 endolysin looks as if it is a chimera between the catalytic N-terminal domain shared by phages of the 936 group and a C-terminal domain found in other lactococcal phages. Such natural chimeras have been found in different bacteriophages (Fastrez, 1996; Sheehan *et al.*, 1997) and are likely to play a role in evolution of host range.

Tail tape measure protein

The longest gene product encoded by bIL170 is gpl16. It is related to a number of phage proteins (Table 1), generally also among the longest encoded by the phages, most of them assumed to be functional analogues of the tail tape measure protein (gpH) of phage lambda. This protein is involved in the determination of tail length (Katsura & Hendrix, 1984), being used as a template for tail polymerization and remaining inside the tail tube. Although the amino acid conservation is poor between proteins of distantly related phages, analysis of the synteny among dairy and other phages (Brussow & Desiere, 2001; Chandry *et al.*, 1997) emphasized the function assignment. The first experimental evidence among lactococcal phages was recently obtained for gp45 (TMP) of phage TP901-1 (Pedersen *et al.*, 2000). Surprisingly, gpl16 of bIL170 also shows low aa identity (Table 1) with a large gene product from lactococcal phages of the c2 group (gpl10 of c2, gp31 of bIL67) located by immunomicroscopy at the tip of the tail and proposed to encode the tail adsorption protein (Lubbers *et al.*, 1995; Schouler *et al.*, 1994). The latter should thus be an analogue of the tail tip protein gpJ of lambda, which is involved in the adsorption of the phage to its host. Significant BLAST E values and low amino acid identity may simply reflect a common structure of the fibrous type, expected for a tail fibre or a tail length template protein. Other hypotheses are that the function may have been wrongly attributed in the case of the c2 gene product, as suggested by Desiere *et al.* (2001), or that the same structural protein can serve several functions depending on the phage. Immuno-detection of the tail tape measure in preparations of phages c2 may have resulted from a deterioration of the baseplate (rendering the tape measure protein accessible) or to the real accessibility of this protein beyond the baseplate in these phages.

gpl16 in bIL170 and its gp14 homologue in sk1 show some divergence (mainly deletion but also lower aa conservation) in their central region. It is a characteristic shared with other pairs of highly conserved tail tape measure proteins in dairy phages, as gp1560, gp1626 and gp15 in Sfi21, Sfi19 and DT1 respectively (Lucchini *et al.*, 1999). Interestingly, it is also true for the gpl10/gp31

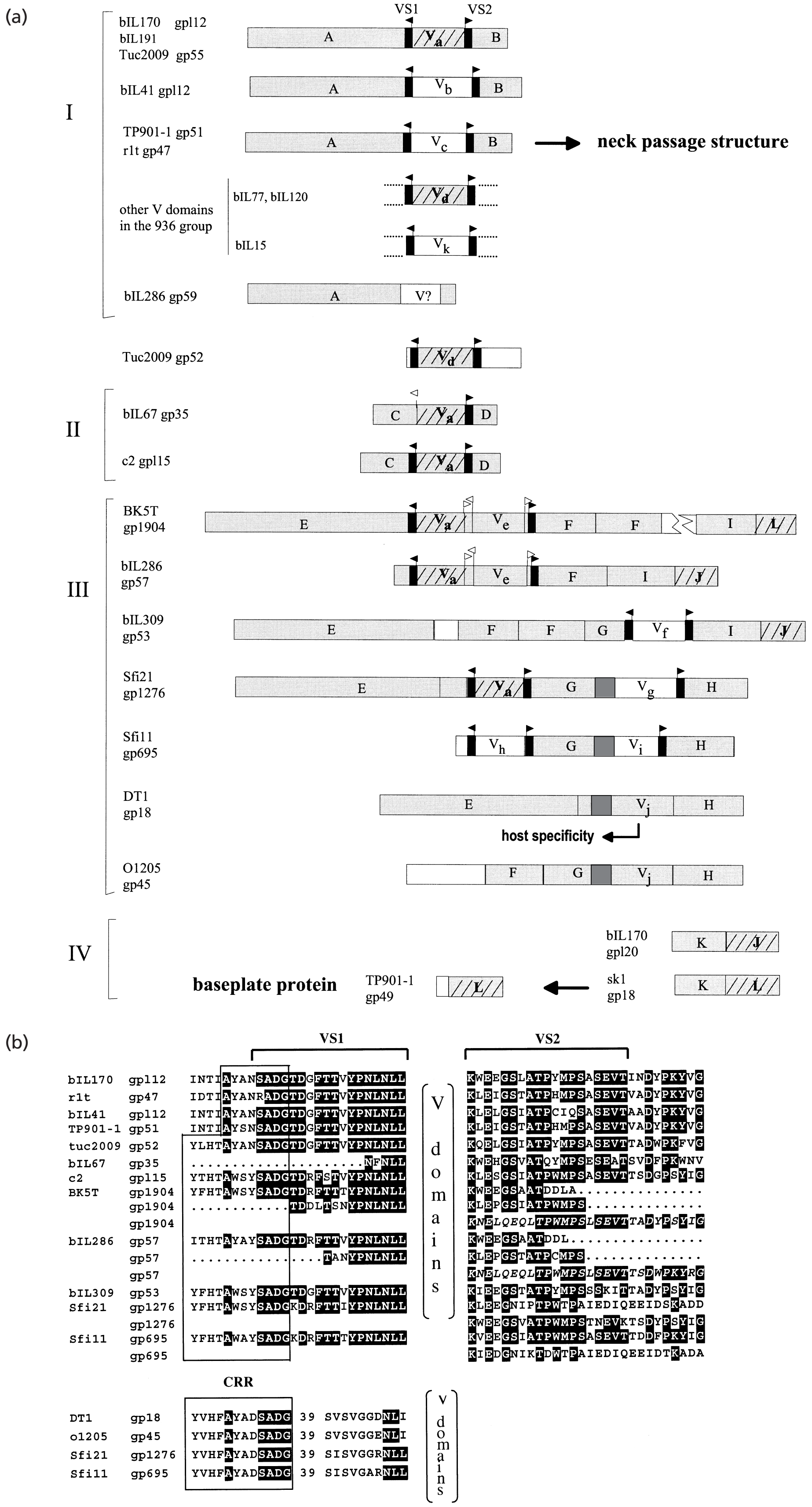

Fig. 6. For legend see facing page.

pair in c2/bIL67 (Lubbers *et al*., 1995). In contrast, some of these proteins share a conserved domain (Desiere *et al*., 1998) that is not detected in gpl16/gp14 in bIL170/sk1, which appears to have the features of lytic transglycosylases (Lehnherr *et al*., 1998), as do other phage proteins (Fig. 5). It has been proposed, and very recently demonstrated for phages T7 and PRD1 (Moak & Molineux, 2000; Rydman & Bamford, 2000), that phage-encoded transglycosylase activity may help the process of infection probably by partial lysis of the bacterial peptidoglycan. Provided it comes into contact with the bacterial peptidoglycan in the infection process, transglycosylase activity can be borne by various proteins in virions. For example, in the T7 virion, the protein involved is an internal protein. It seems that for some of the dairy phages (Sfi21, phi adh and phi g1e), this function could be encoded by the same gene as the predicted tape measure protein. In other dairy phages, the activity is encoded within an adjacent gene, as in DT1 (see Fig. 5). In the P335-type lactococcal phage TP901, a function of endopeptidase has been predicted for a protein different from the tape measure protein (Brondsted *et al*., 2001). Whether a cell-wall hydrolase activity, fulfilling a function different to the endolysin involved in cell lysis at the end of the phage cycle, is also present in bIL170 and other virulent lactococcal phages, but not yet detected by sequence analysis, remains to be determined.

Modular structural proteins

gpl12 and gpl20 are each related to a number of late products found in lactococcal or streptococcal phages with different proposed functions and/or localization in the virion. To analyse more accurately the homology relationships, we split up the proteins into homologous domains. The complex pattern of similarity reflects the modular structure of these proteins, schematized and classified into four families in Fig. 6(a).

gpl12, a putative fibre. gpl12 shares at least one homologous domain with three families of proteins (I to III, Fig. 6a). It is highly similar to a 'neck passage structure' protein (NPS, gp51) of TP901-1 (Brondsted *et al*., 2001). This protein, localized by immuno-microscopy, could be associated with thin fibres at the collar (Johnsen *et al*., 1995), reminiscent of the whiskers of phage T4. However the exact role of the protein in the life cycle of TP901-1 remains to be determined. Both proteins seem to be composed of three domains, as are other similar proteins constituting family I (Fig. 6a). They all exhibit, between two constant parts (A and B in Fig. 6a), an internal variable (V) domain.

gpl12 is also related, but only by its V domain, to other putative fibres (family III) from streptococcal and P335-type lactococcal phages, as well as to minor structural proteins (not localized) from lactococcal phages of the c2 group (Lubbers *et al*., 1995). Family III is the most heterogeneous and comprises large multi-domain proteins with collagen-type motifs (repetitions of GXX), reminiscent of the long tail fibres of coliphages by their mosaic organization and assumed to be involved in host specificity (Boyce *et al*., 1995; Chopin *et al*., 2001; Lucchini *et al*., 1999; Tremblay & Moineau, 1999). Whether they are indeed fibres remains to be determined experimentally but their direct involvement in host-range determination was recently demonstrated for gp18 of phage DT1 (Duplessis & Moineau, 2001). The high level of conservation of their primary structure suggests that gene products of family I, including gpl12 of bIL170, are likely to adopt the same structure as gp51 of TP901-1, even though they may have a different location in the virion.

Variable domains. Variable domains, about 110–150 aa long, flanked by two different stretches of about 20 well conserved amino acids each, designated VS (for V-domain Signature), were observed in proteins of families I to III. In some cases, variable regions may have lost one or both of their VS signatures (eg, last four of family III in Fig. 6a). The sequence polymorphism of the central region in gpl12-like proteins from lactococcal phages of the 936 group was assessed by sequencing that part in other phages of this group. We defined a new type of V domain when it shows no significant homology, that is less than 25 % aa identity on its whole length, with the others. Four types of V domains were found in gpl12-like proteins from phages of the 936 group and 11 throughout all groups of structural proteins (Fig. 6a).

Fig. 6. (a) Schematic representation of the modular structure of gpl12 and gpl20 in phage bIL170 and other phage related proteins. Proteins were divided into four families (I, II, III and IV) according to their pattern of homology. Large segments homologous to each other are labelled with the same letter (>30 % aa identity) or the same subscript for V domains (if >25 % aa identity on the whole length). Observed aa identity among each labelled segment were as follows: Va domains, from 35 % to 97 %; A, B, C, K, more than 80 %; D, 45 %; E, 35–80 %; F, 55–75 %; G, 60–90 %; H, 45–75 %; I, 40–90 %; J, 40–95 %; L, 30–35 %. Domains found in proteins of different groups are indicated by hatched segments. Conserved motifs VS1 (N-proximal) and VS2 (C-proximal), delimiting V domains, are shown as black boxes topped by leftward- and rightward-oriented triangles respectively (white triangles without black box for truncated motifs). A putative conserved region alternative to VS1 is indicated by a dark grey box. Representation of the proteins has been simplified: in particular the existence of some repeated regions in proteins of family III (including collagen-type repeats) is not reported. Note that Duplessis & Moineau (2001) designated VR1 the Va domain of gp1276 in Sfi21 and VR2 the Vg, Vi and Vj domains in streptococcal phage proteins. (b) VS1 and VS2 conserved motifs delimiting V domains. Top panel: identical aa in more than two thirds of the sequences are shaded in black. In italics, a complete VS2 motif observed just after a truncated motif. Part of the conserved motif reported by Duplessis & Moineau (2001) at the end of collagen-type repeats in the streptococcal phage proteins is boxed (named here CRR). Bottom panel: aa residues preceding some variable domains in streptococcal phage proteins. Those conserved with VS1 are shaded in black. Numbers within the alignments indicate the numbers of aa residues not shown at that position.

The Va-type of V domain seems to be the most widespread. V domains are involved in host specificity of streptococcal phages: Lucchini *et al.* (1999) pointed out a correlation between host range and variability in gp695 of Sfi11 and homologues in Sfi21/Sfi19, O1205 and DT1, which fall in the domain designated Vi in gp695 and positional equivalents in other proteins (see Fig. 6a). As elegantly demonstrated by Duplessis & Moineau (2001), the V domain of gp18 (which they designated VR2 and we name Vj in Fig. 6a) is responsible for the host-specificity of phage DT1 as anti-receptor. V domains are clearly reminiscent of the receptor-recognizing domains of long tail fibres of T-even phages. For instance, receptor specificity of T4-type phages was shown to be associated with an area of 70–100 aa in gp37 and homologues, variable in sequence and often flanked by conserved direct repeats of 14 aa (Montag *et al.*, 1990). In streptococcal and P335-type lactococcal phages (family III), two types of motifs can be observed in the vicinity of V-domains: repetition of collagen-type motifs followed by a stretch of conserved amino acid residues (Duplessis & Moineau, 2001) – conserved amino acids that we will designate here CCR – and VS signatures described in this paper (Fig. 6b). Actually, VS1 and CCR overlap (Fig. 6b). Interestingly, gpl15 in phage c2 exhibits both CCR and VS1, but no collagen-type repeats, which were only found in proteins of family III. The role of these conserved motifs is unknown. As observed for CCR in streptococcal phages (Duplessis & Moineau, 2001), we observed that VS motifs are less conserved at the nucleotide level: in particular VS1 motifs identical in aa for bIL41 and bIL170 diverge at the nt level.

gpl20, a putative baseplate protein. gpl20 in bIL170 and the probable orthologue in sk1 (gp18) are made up of two domains. They could be two different versions of baseplate proteins, related to the baseplate protein (BBP, gp49) experimentally identified in the lactococcal phage TP901-1 (Pedersen *et al.*, 2000). The latter polypeptide, essentially composed of a single domain, is homologous to the C-terminal part of gp18 of sk1 with which it shares 35 % aa identity. They constitute family IV (Fig. 6a). They appeared to be related to family III by one of their domains (L and J, Fig. 6a).

A multi-component system potentially involved in host-range determination in dairy phages. The exact nature of V-domain-containing proteins in dairy phages awaits further determination. None of them have had both their localization and function experimentally determined; V domains are part of proteins with different primary structures and maybe with different locations in the virion. Interestingly, large gene products of family III (putative tail fibres) seem to gather domains found in baseplate proteins (eg. gp49 of TP901-1) and variable domains found in families I and II. Although the V domain of gp18 has been characterized as an essential host-range determinant for DT1 (Duplessis & Moineau, 2001), the authors are aware of other phage factors also involved in host specificity, either inside (for example other variable domains) or outside gp18-like proteins in streptococcal phages. Our bioinformatic analysis of families of multi-domain structural proteins in dairy phages also suggests that specific recognition of the cell by dairy phages may involve different partners. The constellation is not completely described here since other proteins, not found in bIL170, are already known to share homologous domains with proteins of family III (e.g. gp1000 of Sfi11) (Lucchini *et al.*, 1998). The presence of multiple V domains seems to be a common characteristic in streptococcal and P335-type lactococcal phages, either as domains of a single gene product, as in proteins of group III, or in separate gene products, as in phage Tuc2009 with gp55 (group I) and gp52 (unclassified). This multiplicity, possibly allowing diversity and extension of host range, seems to be worth investigating. Extension of host range was shown to result from duplication within the adhesin of T4 (Tetart *et al.*, 1998).

What could be host-range determinants in bIL170? In comparing the genetic maps of lactococcal phages of the 936 group and lambda, it is noteworthy that no prediction was made for proteins involved in specific adsorption to the host, except for gpl12 depicted as a putative 'ectopic' tail fibre (Chandry *et al.*, 1997;Desiere *et al.*, 2001). gpl12 is dispensable for infection of the laboratory host strain, since no homologue was found in the closely related phage sk1, nor possibly in F4-1 (for which only partial genome sequence is available), and since its deletion did not impair phage bIL170 propagation (A.-M. Crutz-Le Coq, unpublished results). Although gpl12 of bIL170 contains a V domain, it clearly does not encode an essential determinant of adsorption, at least on its propagation strain, which is still to be identified. Taking into account the little knowledge we have on gpl12, two simple roles may be hypothesized: it could be an accessory gene product participating directly in a possible extension of host range by recognizing additional cell receptors, such as the Stf fibre of lambda (Hendrix & Duda, 1992), or it could interact with another partner involved in bacterial cell recognition. For instance, the whiskers of phage T4 that are involved in the assembly and retraction of the long tail fibres, have been proposed to prevent adsorption under certain environmental conditions (Conley & Wood, 1975). At present, we have no other obvious candidate for a tail fibre with which, gpl12, if really a component of whiskers, could interact. It is worth noting that BBP and NPS (gpl12-like) in TP901-1 may be part of the same functional module not found in lambda (Brondsted *et al.*, 2001). The exact function of these proteins in the life cycles of the phages is still to be determined and will require more than simply knowing their location in the virion.

Conclusions

bIL170 and sk1, belonging to the 936 group of lactococcal phages, were expected to be closely related at the nucleotide level. Knowledge of both complete genomic sequences reveals the existence of segments of high genetic diversity between the two phages, mainly represented by additional genes (within indels), protein domain shuffling and exchange of genomic segments. This is roughly similar to what was observed for the other group of virulent lactococcal phages (c2) (Lubbers *et al.*, 1995). The early regions of sk1 and bIL170 are rather divergent, with a total of 15 ORFs (excluding the three putative homing endonucleases) with no detectable homologous counterpart in one or the other phage and 26 homologous ORFs (>30% aa identity) in the two phages. In the late cluster, the divergent regions could point out modules specifically involved in host specificity.

Adsorption of tailed phages to bacterial cells may involve different levels of complexity ranging from laboratory strains of phage lambda, possibly representing one of the simplest cases, where the major determinant of host specific recognition is a tail tip protein, to the well-studied phage T4 possessing a complex system involving long and short tail fibres as well as whiskers. Our bioinformatic analysis of multidomain proteins in dairy phages suggests that specific recognition of the cell by these phages may also involve a complex system (which should evolve easily when acquiring new domains) with different partners, some of them likely being fibres. As exemplified by morphogenesis of tail fibres from coliphages (Wood *et al.*, 1994), we believe that the basal unit of analysis should be the functional module rather than the entire gene products.

ACKNOWLEDGEMENTS

We greatly thank A. Sorokin for providing the program he created for analysing nucleotide composition of genomes, C. Gaillardin for his encouragement and critical reading of this manuscript, as well as A. Chopin and M-C. Chopin for their comments. M.-A. Petit is kindly acknowledged for her help in improving this manuscript.

REFERENCES

Altermann, E., Klein, J. R. & Henrich, B. (1999). Primary structure and features of the genome of the *Lactobacillus gasseri* temperate bacteriophage phi adh. *Gene* **236**, 333–346.

Alvarez, M. A., Herrero, M. & Suarez, J. E. (1998). The site-specific recombination system of the *Lactobacillus* species bacteriophage A2 integrates in Gram-positive and Gram-negative bacteria. *Virology* **250**, 185–193.

Aravind, L., Makarova, K. S. & Koonin, E. V. (2000). Survey and summary. Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res* **28**, 3417–3432.

Bidnenko, E., Ehrlich, S. D. & Chopin, M. C. (1995). Phage operon involved in sensitivity to the *Lactococcus lactis* abortive infection mechanism AbiD1. *J Bacteriol* **177**, 3824–3829.

Bidnenko, E., Ehrlich, S. D. & Chopin, M. C. (1998). *Lactococcus lactis* phage operon coding for an endonuclease homologous to RuvC. *Mol Microbiol* **28**, 823–834.

Blackburn, N. T. & Clarke, A. J. (2001). Identification of four families of peptidoglycan lytic transglycosylases. *J Mol Evol* **52**, 78–84.

Bolotin, A., Mauger, S., Malarme, K., Ehrlich, S. D. & Sorokin, A. (1999). Low-redundancy sequencing of the entire *Lactococcus lactis* IL1403 genome. *Antonie Leeuwenhoek* **76**, 27–76.

Borodovsky, M. & McIninch, J. D. (1993). GeneMark: parallel gene recognition for both DNA strands. *Comput Chem* **17**, 123–133.

Bouchard, J. D. & Moineau, S. (2000). Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. *Virology* **270**, 65–75.

Boyce, J. D., Davidson, B. E. & Hillier, A. J. (1995). Spontaneous deletion mutants of the *Lactococcus lactis* temperate bacteriophage BK5-T and localization of the BK5-T *attP* site. *Appl Environ Microbiol* **61**, 4105–4109.

Breuner, A., Brondsted, L. & Hammer, K. (1999). Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1. *J Bacteriol* **181**, 7291–7297.

Brondsted, L., Ostergaard, S., Pedersen, M., Hammer, K. & Vogensen, F. K. (2001). Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. *Virology* **283**, 93–109.

Brussow, H. & Desiere, F. (2001). Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. *Mol Microbiol* **39**, 213–223.

Casjens, S., Hatfull, G. & Hendrix, R. (1992). Evolution of dsDNA tailed bacteriophage genomes. *Semin Virol* **3**, 383–397.

Chandry, P. S., Moore, S. C., Boyce, J. D., Davidson, B. E. & Hillier, A. J. (1997). Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. *Mol Microbiol* **26**, 49–64.

Chevalier, B. S. & Stoddard, B. L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res* **29**, 3757–3774.

Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S. D. & Chopin, M. C. (2001). Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res* **29**, 644–651.

Christiansen, B., Brondsted, L., Vogensen, F. K. & Hammer, K. (1996). A resolvase-like protein is required for the site-specific integration of the temperate lactococcal bacteriophage TP901-1. *J Bacteriol* **178**, 5164–5173.

Chung, D. K., Kim, J. H. & Batt, C. A. (1991). Cloning and nucleotide sequence of the major capsid protein from *Lactococcus lactis* ssp. *cremoris* bacteriophage F4-1. *Gene* **101**, 121–125.

Conley, M. P. & Wood, W. B. (1975). Bacteriophage T4 whiskers: a rudimentary environment-sensing device. *Proc Natl Acad Sci U S A* **72**, 3701–3705.

Dalgaard, J. Z., Klar, A. J., Moser, M. J., Holley, W. R., Chatterjee, A. & Mian, I. S. (1997). Statistical modelling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res* **25**, 4626–4638.

Desiere, F., Lucchini, S. & Brussow, H. (1998). Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. *Virology* **241**, 345–356.

Desiere, F., Pridmore, R. D. & Brussow, H. (2000). Comparative genomics of the late gene cluster from *Lactobacillus* phages. *Virology* **275**, 294–305.

Desiere, F., Mahanivong, C., Hillier, A. J., Chandry, P. S., Davidson, B. E. & Brussow, H. (2001). Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. *Virology* **283**, 240–252.

Duplessis, M. & Moineau, S. (2001). Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. *Mol Microbiol* **41**, 325–336.

Fastrez, J. (1996). Phage lysozymes. *EXS* **75**, 35–64.

Foley, S., Bruttin, A. & Brussow, H. (2000). Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. *J Virol* **74**, 611–618.

Forde, A. & Fitzgerald, G. F. (1999). Bacteriophage defence systems in lactic acid bacteria. *Antonie Leeuwenhoek* **76**, 89–113.

Garcia, P., Garcia, J. L., Garcia, E., Sanchez-Puelles, J. M. & Lopez, R. (1990). Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Gene* **86**, 81–88.

Garcia, P., Alonso, J. C. & Suarez, J. E. (1997). Molecular analysis of the *cos* region of the *Lactobacillus casei* bacteriophage A2: gene product 3, gp3, specifically binds to its downstream *cos* region. *Mol Microbiol* **23**, 505–514.

Garcia, P., Ladero, V., Alonso, J. C. & Suarez, J. E. (1999). Co-operative interaction of CI protein regulates lysogeny of *Lactobacillus casei* by bacteriophage A2. *J Virol* **73**, 3920–3929.

Golz, S. & Kemper, B. (1999). Association of holliday-structure resolving endonuclease VII with gp20 from the packaging machine of phage T4. *J Mol Biol* **285**, 1131–1144.

Goodrich-Blair, H. & Shub, D. A. (1994). The DNA polymerase genes of several HMU-bacteriophages have similar group I introns with highly divergent open reading frames. *Nucleic Acids Res* **22**, 3715–3721.

Goodrich-Blair, H. & Shub, D. A. (1996). Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell* **84**, 211–221.

Gorbalenya, A. E. (1994). Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. *Protein Sci* **3**, 1117–1120.

Hendrix, R. W. & Duda, R. L. (1992). Bacteriophage lambda PaPa: not the mother of all lambda phages. *Science* **258**, 1145–1148.

Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E. & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* **96**, 2192–2197.

Henrich, B., Binishofer, B. & Blasi, U. (1995). Primary structure and functional analysis of the lysis genes of *Lactobacillus gasseri* bacteriophage phi adh. *J Bacteriol* **177**, 723–732.

Jarvis, A. W., Fitzgerald, G. F., Mata, M., Mercenier, A., Neve, H., Powell, I. B., Ronda, C., Saxelin, M. & Teuber, M. (1991). Species and type phages of lactococcal bacteriophages. *Intervirology* **32**, 2–9.

Johnsen, M. G., Neve, H., Vogensen, F. K. & Hammer, K. (1995). Virion positions and relationships of lactococcal temperate bacteriophage TP901-1 proteins. *Virology* **212**, 595–606.

Johnsen, M. G., Appel, K. F., Madsen, P. L., Vogensen, F. K., Hammer, K. & Arnau, J. (1996). A genomic region of lactococcal temperate bacteriophage TP901-1 encoding major virion proteins. *Virology* **218**, 306–315.

Kakikawa, M., Oki, M., Tadokoro, H., Nakamura, S., Taketo, A. & Kodaira, K. (1996). Cloning and nucleotide sequence of the major capsid proteins of *Lactobacillus* bacteriophage phi g1e. *Gene* **175**, 157–165.

Katsura, I. & Hendrix, R. W. (1984). Length determination in bacteriophage lambda tails. *Cell* **39**, 691–698.

Kim, J. H. & Batt, C. A. (1991a). Molecular characterization of a *Lactococcus lactis* bacteriophage F4-1. *Food Microbiology* **8**, 15–26.

Kim, J. H. & Batt, C. A. (1991b). Nucleotide sequence and deletion analysis of a gene coding for a structural protein of *Lactococcus lactis* bacteriophage F4-1. *Food Microbiol* **8**, 27–36.

Kodaira, K. I., Oki, M., Kakikawa, M., Watanabe, N., Hirakawa, M., Yamada, K. & Taketo, A. (1997). Genome structure of the *Lactobacillus* temperate phage phi g1e: the whole genome sequence and the putative promoter/repressor system. *Gene* **187**, 45–53.

Lakshmidevi, G., Davidson, B. E. & Hillier, A. J. (1990). Molecular characterization of promoters of the *Lactococcus lactis* subsp. *cremoris* temperate bacteriophage BK5-T and identification of a phage gene implicated in the regulation of promoter activity. *Appl Environ Microbiol* **56**, 934–942.

Lazarevic, V., Soldo, B., Dusterhoft, A., Hilbert, H., Mauel, C. & Karamata, D. (1998). Introns and intein coding sequence in the ribonucleotide reductase genes of *Bacillus subtilis* temperate bacteriophage SPbeta. *Proc Natl Acad Sci U S A* **95**, 1692–1697.

Lehnherr, H., Hansen, A. M. & Ilyina, T. (1998). Penetration of the bacterial cell wall: a family of lytic transglycosylases in bacteriophages and conjugative plasmids. *Mol Microbiol* **30**, 454–457.

Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* **13**, 660–665.

Loessner, M. J., Gaeng, S., Wendlinger, G., Maier, S. K. & Scherer, S. (1998). The two-component lysis system of *Staphylococcus aureus* bacteriophage Twort: a large TTG-start holin and an associated amidase endolysin. *FEMS Microbiol Lett* **162**, 265–274.

Lubbers, M. W., Waterfield, N. R., Beresford, T. P., Le Page, R. W. & Jarvis, A. W. (1995). Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2 genome and identification of the structural genes. *Appl Environ Microbiol* **61**, 4348–4356.

Lucchini, S., Desiere, F. & Brussow, H. (1998). The structural gene module in *Streptococcus thermophilus* bacteriophage phi Sfi11 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts. *Virology* **246**, 63–73.

Lucchini, S., Desiere, F. & Brussow, H. (1999). Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J Virol* **73**, 8647–8656.

Ludwig, W., Seewaldt, E., Kilpper-Balz, R., Schleifer, K. H., Magrum, L., Woese, C. R., Fox, G. E. & Stackebrandt, E. (1985). The phylogenetic position of *Streptococcus* and *Enterococcus*. *J Gen Microbiol* **131**, 543–551.

Madsen, P. L. & Hammer, K. (1998). Temporal transcription of the lactococcal temperate phage TP901-1 and DNA sequence of the early promoter region. *Microbiology* **144**, 2203–2215.

Mahanivong, C., Boyce, J. D., Davidson, B. E. & Hillier, A. J. (2001). Sequence analysis and molecular characterization of the *Lactococcus lactis* temperate bacteriophage BK5-T. *Appl Environ Microbiol* **67**, 3564–3576.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

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

Mikkonen, M. & Alatossava, T. (1995). A group I intron in the terminase gene of *Lactobacillus delbrueckii* subsp. *lactis* phage LL-H. *Microbiology* **141**, 2183–2190.

Moak, M. & Molineux, I. J. (2000). Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. *Mol Microbiol* **37**, 345–355.

Montag, D., Hashemolhosseini, S. & Henning, U. (1990). Receptor-recognizing proteins of T-even type bacteriophages: the receptor-recognizing area of proteins 37 of phages T4 TuIa and TuIb. *J Mol Biol* **216**, 327–334.

Moscoso, M. & Suarez, J. E. (2000). Characterization of the DNA replication module of bacteriophage A2 and use of its origin of replication as a defence against infection during milk fermentation by *Lactobacillus casei*. *Virology* **273**, 101–111.

Mrazek, J. & Karlin, S. (1998). Strand compositional asymmetry in bacterial and large viral genomes. *Proc Natl Acad Sci U S A* **95**, 3720–3725.

Nauta, A., van Sinderen, D., Karsens, H., Smit, E., Venema, G. & Kok, J. (1996). Inducible gene expression mediated by a repressor-operator system isolated from *Lactococcus lactis* bacteriophage r1t. *Mol Microbiol* **19**, 1331–1341.

Neve, H., Zenz, K. I., Desiere, F., Koch, A., Heller, K. J. & Brussow, H. (1998). Comparison of the lysogeny modules from the temperate *Streptococcus thermophilus* bacteriophages TP-J34 and Sfi21: implications for the modular theory of phage evolution. *Virology* **241**, 61–72.

Oki, M., Kakikawa, M., Yamada, K., Taketo, A. & Kodaira, K. I. (1996). Cloning, sequence analysis, and expression of the genes encoding lytic functions of bacteriophage phi g1e. *Gene* **176**, 215–223.

Oki, M., Kakikawa, M., Nakamura, S., Yamamura, E. T., Watanabe, K., Sasamoto, M., Taketo, A. & Kodaira, K. (1997). Functional and structural features of the holin HOL protein of the *Lactobacillus plantarum* phage phi gle: analysis in *Escherichia coli* system. *Gene* **197**, 137–145.

Ostergaard, S., Brondsted, L. & Vogensen, F. K. (2001). Identification of a replication protein and repeats essential for DNA replication of the temperate lactococcal bacteriophage TP901-1. *Appl Environ Microbiol* **67**, 774–781.

Parreira, R. (1996). Caractérisation du mécanisme de résistance aux phages par infection abortive codé par le gène *abiB* de *Lactococcus lactis* subsp. *lactis*. PhD thesis: Université Paris XI.

Parreira, R., Ehrlich, S. D. & Chopin, M. C. (1996a). Dramatic decay of phage transcripts in lactococcal cells carrying the abortive infection determinant AbiB. *Mol Microbiol* **19**, 221–230.

Parreira, R., Valyasevi, R., Lerayer, A. L., Ehrlich, S. D. & Chopin, M. C. (1996b). Gene organization and transcription of a late-expressed region of a *Lactococcus lactis* phage. *J Bacteriol* **178**, 6158–6165.

Pedersen, M., Ostergaard, S., Bresciani, J. & Vogensen, F. K. (2000). Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. *Virology* **276**, 315–328.

Platteeuw, C. & de Vos, W. M. (1992). Location, characterization and expression of lytic enzyme-encoding gene, *lytA*, of *Lactococcus lactis* bacteriophage phi US3. *Gene* **118**, 115–120.

Rydman, P. S. & Bamford, D. H. (2000). Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. *Mol Microbiol* **37**, 356–363.

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

Sharples, G. J., Corbett, L. M. & Graham, I. R. (1998). Lambda Rap protein is a structure-specific endonuclease involved in phage recombination. *Proc Natl Acad Sci U S A* **95**, 13507–13512.

Sheehan, M. M., Garcia, J. L., Lopez, R. & Garcia, P. (1996). Analysis of the catalytic domain of the lysin of the lactococcal bacteriophage Tuc2009 by chimeric gene assembling. *FEMS Microbiol Lett* **140**, 23–28.

Sheehan, M. M., Garcia, J. L., Lopez, R. & Garcia, P. (1997). The lytic enzyme of the pneumococcal phage Dp-1: a chimeric lysin of intergeneric origin. *Mol Microbiol* **25**, 717–725.

Sheehan, M. M., Stanley, E., Fitzgerald, G. F. & van Sinderen, D. (1999). Identification and characterization of a lysis module present in a large proportion of bacteriophages infecting *Streptococcus thermophilus*. *Appl Environ Microbiol* **65**, 569–577.

Shub, D. A., Goodrich-Blair, H. & Eddy, S. R. (1994). Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem Sci* **19**, 402–404.

Smith, M. C. M., Burns, R. N., Wilson, S. E. & Gregory, M. A. (1999). The complete genome sequence of the *Streptomyces* temperate phage phi C31: evolutionary relationships to other viruses. *Nucleic Acids Res* **27**, 2145–2155.

Staden, R. (1996). The Staden sequence analysis package. *Mol Biotechnol* **5**, 233–241.

Stanley, E., Fitzgerald, G. F., Le Marrec, C., Fayard, B. & van Sinderen, D. (1997). Sequence analysis and characterization of ϕO1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiology* **143**, 3417–3429.

Tetart, F., Desplats, C. & Krisch, H. M. (1998). Genome plasticity in the distal tail fibre locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. *J Mol Biol* **282**, 543–556.

Tremblay, D. M. & Moineau, S. (1999). Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. *Virology* **255**, 63–76.

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

Vasala, A., Valkkila, M., Caldentey, J. & Alatossava, T. (1995). Genetic and biochemical characterization of the *Lactobacillus delbrueckii* subsp. *lactis* bacteriophage LL-H lysin. *Appl Environ Microbiol* **61**, 4004–4011.

Walker, S. A. & Klaenhammer, T. R. (1998). Molecular characterization of a phage-inducible middle promoter and its transcriptional activator from the lactococcal bacteriophage phi31. *J Bacteriol* **180**, 921–931.

Wang, X., Mani, N., Pattee, P. A., Wilkinson, B. J. & Jayaswal, R. K. (1992). Analysis of a peptidoglycan hydrolase gene from *Staphylococcus aureus* NCTC 8325. *J Bacteriol* **174**, 6303–6306.

Waterfield, N. R., Lubbers, M. W., Polzin, K. M., Le Page, R. W. & Jarvis, A. W. (1996). An origin of DNA replication from *Lactococcus lactis* bacteriophage c2. *Appl Environ Microbiol* **62**, 1452–1453.

Wood, W. B., Eiserling, F. A. & Crowther, R. A. (1994). Long tail fibers: genes, proteins, structure and assembly. In *Molecular Biology of Bacteriophage T4*, pp. 282–290. Edited by J. D. Karan. Washington, DC: American Society for Microbiology.

Received 7 September 2001; revised 5 December 2001; accepted 13 December 2001.