VIROLOGY 237, 148–158 (1997)
ARTICLE NO. VY978769

The Site-Specific Integration System of the Temperate

Anne Bruttin,[1] Sophie Foley,[1] and Harald Brüssow[2]

Nestlé Research Centre, Nestec Ltd., Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland

Received June 13, 1997; returned to author for revision July 1, 1997; accepted August 7, 1997

The temperate bacteriophage ϕSfi21 integrates its DNA into the chromosome of *Streptococcus thermophilus* strains via site-specific recombination. Nucleotide sequencing of the attachment sites identified a 40-bp identity region which surprisingly overlaps both the 18-terminal bp of the phage integrase gene and the 11-terminal bp of a host $tRNA^{Arg}$ gene. A 2.4-kb phage DNA segment, covering *attP,* the phage integrase, and a likely immunity gene contained all the genetic information for faithful integration of a nonreplicative plasmid into the *attB* site. A deletion within the *int* gene led to the loss of integration proficiency. A number of spontaneous deletions were observed in plasmids containing the 2.4-kb phage DNA segment. The deletion sites were localized to the tRNA side of the identity region and to phage or vector DNA with 3- to 6-bp-long repeats from the border region. A similar type of deletion was previously observed in a spontaneous phage mutant.

INTRODUCTION

Streptococcus thermophilus, a gram-positive thermophilic lactic acid bacterium, is used as a bacterial starter in the industrial fermentation of milk. Phage attack has always been a major problem in the dairy industry leading to delays in the acidification of milk and even loss of the product (Bruttin *et al.,* 1997a; Peitersen, 1991). Efficient phage control measures depend on a detailed knowledge of the properties of *S. thermophilus* phages. A number of groups have described their *S. thermophilus* phage collections (reviewed by Brüssow *et al.,* 1997) and we are currently characterizing the temperate *S. thermophilus* phage ϕSfi21 in molecular detail (Brüssow *et al.,* 1994a,b; Brüssow and Bruttin, 1995; Bruttin and Brüssow, 1996; Bruttin *et al.,* 1997b; Desiere *et al.,* 1997). Based on DNA cross-hybridization experiments, between the temperate ϕSfi21 and numerous lytic phages of our collection, and on sequence analysis we have identified a region of 6.6 kb containing the lysogeny module (Bruttin *et al.,* 1997b). Spontaneous mutants of ϕSfi21, in which the putative integrase, immunity, and an unattributed gene were deleted, demonstrated the importance of these genes for the establishment of lysogeny since the deletion mutants were unable to lysogenize their host (Bruttin and Brüssow, 1996). In this paper we describe the integration of a plasmid, containing a ϕSfi21DNA fragment corresponding to the deleted region, into the bacterial chromosome by a mechanism which is apparently analogous to the prophage integration mechanism.

Most temperate bacteriophages integrate their DNA into the host chromosome by a site-specific recombination process following the Campbell model (Campbell, 1962). This mechanism involves two specific attachment sites, one on the bacterial chromosome (*attB*) and the other on the phage genome (*attP*). The recombination process is catalyzed by a phage-encoded integrase. There are many well-characterized examples of site-specific recombination in gram-negative bacteria (Campbell, 1992), the best-studied being that of bacteriophage λ (Landy, 1989). The integration system of phages of gram-positive bacteria has been documented for *Staphylococcus aureus* (Carroll *et al.,* 1995), *Arthrobacter aureus* (Le Marrec *et al.,* 1996), *Mycobacterium* ssp. (Lee *et al.,* 1991), *Lactobacillus gasseri* (Raya *et al.,* 1992), *Lactobacillus delbrueckii* (Dupont *et al.,* 1995), *Lactococcus lactis* (Boyce *et al.,* 1995; Christiansen *et al.,* 1994, 1996; Lillehaug and Birkeland, 1993; van de Guchte *et al.,* 1994), *Streptococcus pyogenes* (McShan *et al.,* 1997), and *Streptomyces rimosus* (Gabriel *et al.,* 1995). However, until now no data are available for *S. thermophilus.* Earlier studies (Brüssow and Bruttin, 1995) showed that one integrated copy of *S. thermophilus* phage ϕSfi21 was present in the chromosome of the lysogenic strain Sfi19. Identical signals were obtained for three other lysogens in Southern hybridization experiments (Brüssow and Bruttin, 1995), indicating the presence of a single phage integration site on the bacterial chromosome. The phage attachment site *attP* was localized on the 7-kb *Pvu*II/*Xba*I restriction fragment 3, as previously described (Brüssow and Bruttin, 1995).

In order to improve our understanding of this industri-

[1] Both authors contributed equally to this work.

[2] To whom correspondence and reprint requests should be addressed. Fax: 0041 21 785 8925. E-mail: Harald.Bruessow@CHLSNR.NESTRD.CH.

0042-6822/97 $25.00

ally important host–phage relationship, we have studied the site-specific integration of the φSfi21 genome into the *S. thermophilus* chromosome. To our knowledge the identity region of the attachment sites is unique in duplicating both a phage and a bacterial gene sequence. A nonreplicative vector based on phage integration elements was able to integrate in a specific *attB* site in the *S. thermophilus* chromosome. Spontaneous deletion events originating in the *att* site of these plasmids were observed and compared to spontaneous deletions obtained from phages undergoing serial passages.

MATERIALS AND METHODS

Bacterial strains and plasmids

S. thermophilus strains Sfi1, 3, 8, 10, 11, 16, 21, W3, YS3, and ST44 were from the Nestec strain collection. All *S. thermophilus* strains were propagated in lactose M17 broth as described previously (Bruttin and Brüssow, 1996). *Escherichia coli* JM101 (Stratagene) was grown in LB broth or on LB broth solidified with 1.5% agar. *Lactococcus lactis* MG1363 and 135 were propagated in glucose M17 broth. *Lactobacillus helveticus* N2 was propagated in MRS broth. Transformation of *E. coli* strain JM101 was done according to the manufacturer's instructions (Stratagene). *Lactococcus* and *S. thermophilus* were electrotransformed using the procedures of Holo and Nes (1989) and Slos *et al.* (1991), respectively. Chloramphenicol was used when required at a final concentration of 20, 5, and 2.5 μg/ml for *E. coli, L. lactis,* and *S. thermophilus,* respectively.

Three cloning vectors were used: pUC19, the *E. coli*/lactococcal/streptococcal shuttle vector pNZ124 (Platteeuw *et al.,* 1994), and the *E. coli*/lactococcal shuttle vector pCI372 (Hayes *et al.,* 1990). General DNA techniques were as described in previous publications (Bruttin and Brüssow, 1996; Brüssow and Bruttin, 1995).

Cloning

The integration vector pSF28 was constructed in *L. lactis* MG1363 by cloning the 2.4-kb *Xba*I/*Bgl*II phage Sfi21 DNA fragment containing the integrase gene in the *Xba*I and *Bam*HI sites of the shuttle vector pCI372. In order to clone *attL,* the 0.8-kb *Bgl*II/*Eco*RI fragments from chromosomal DNA of the lysogen Sfi19 were size-selected on preparative agarose gels and inserted into *Bgl*II/*Eco*RI-digested pUC19 vector. Ligation was done overnight at 12° in ligation buffer (20 m*M* Tris/HCl, pH 7.6, 5 m*M* $MgCl_2$, 5 m*M* DTT, 1 m*M* ATP, 50 μg/ml BSA, and 5 U T4 DNA ligase (Boehringer)).

Inverse PCR

The inverse PCR strategy was used to obtain the *attR* site which was located on a 4.4-kb *Hin*dIII fragment. Chromosomal DNA from the lysogen Sfi19 was digested with *Hin*dIII, diluted, and self-ligated. Primers 1 and 5 were used for PCR. The amplification program consisted of an initial denaturation step at 95° for 2 min, followed by 30 cycles of 30 sec at 94°, 30 sec at 55°, and 2 min at 72°. The PCR product was purified using the QIAquick PCR purification kit (Qiagen) and sequenced.

PCR

PCR products were prepared using pairs of synthetic oligonucleotides, the relevant template DNA, and Taq polymerase Fermentas. The primers used were as follows: Primer 1 (nt position 3626): 5′-GCTATTTAATTGCCCAC-3′; primer 2 (nt position 3278): 5′-GTCTATTAGTATGTCGG-3′; primer 3 (bacterial): 5′-CACAATCTTCATCAAGC-3′; primer 4 (bacterial): GCAAGGTAAAGCTGCAC; primer 5 (nt position 1600): 5′-GACGACATCGTTCTTGG-3′. The nt positions given for primers 1 and 2 refer to the sequence given in Bruttin *et al.* (1997b, Accession No. X95646), while that of primer 5 refers to the sequence given in Desiere *et al.* (1997, Accession No. AF004379). The *attB* site was amplified by PCR using primers 3 and 4 from chromosomal DNA of the prophage-free strain Sfi1. The *attL* and *attR* site were amplified from chromosomal DNA of lysogens or cells containing integrated plasmids by PCR using the primer pairs 2/3 and 1/4, respectively. The inverse PCR was done using the primer pair 1 and 5. The *attP* site was PCR-amplified using the primer pair 1 and 2.

DNA sequencing

DNA was sequenced on both strands by the Sanger method of dideoxy-mediated chain termination using the fmol DNA Sequencing System of Promega (Madison, WI). The sequencing primers indicated in the text were end-labeled using [γ-^{33}P]ATP according to the manufacturer's protocol. The thermal cycler was programmed as described previously (Bruttin *et al.,* 1997).

Database searches

The sequences were analyzed as described previously (Bruttin *et al.,* 1997b), with the following updates: GenBank release 100, EMBL release 50, PIR-Protein release 52.

RESULTS

Localization and characterization of the *attP* site

Previously we had mapped the *attP* site of the temperate *S. thermophilus* phage φSfi21 to the 7-kb *Pvu*II/*Xba*I restriction fragment 3, PX3 (Brüssow and Bruttin, 1995). To improve the resolution of this attribution, we used a panel of PCR probes generated from the PX3 fragment in Southern hybridization experiments (Fig. 1A). Probes 1 and 2, respectively, were used on Southern blots of *Bgl*II/*Xba*I- and *Bgl*II/*Pvu*II-digested DNA from phage φSfi21 and its lysogen Sfi19 (Brüssow and Bruttin, 1995).

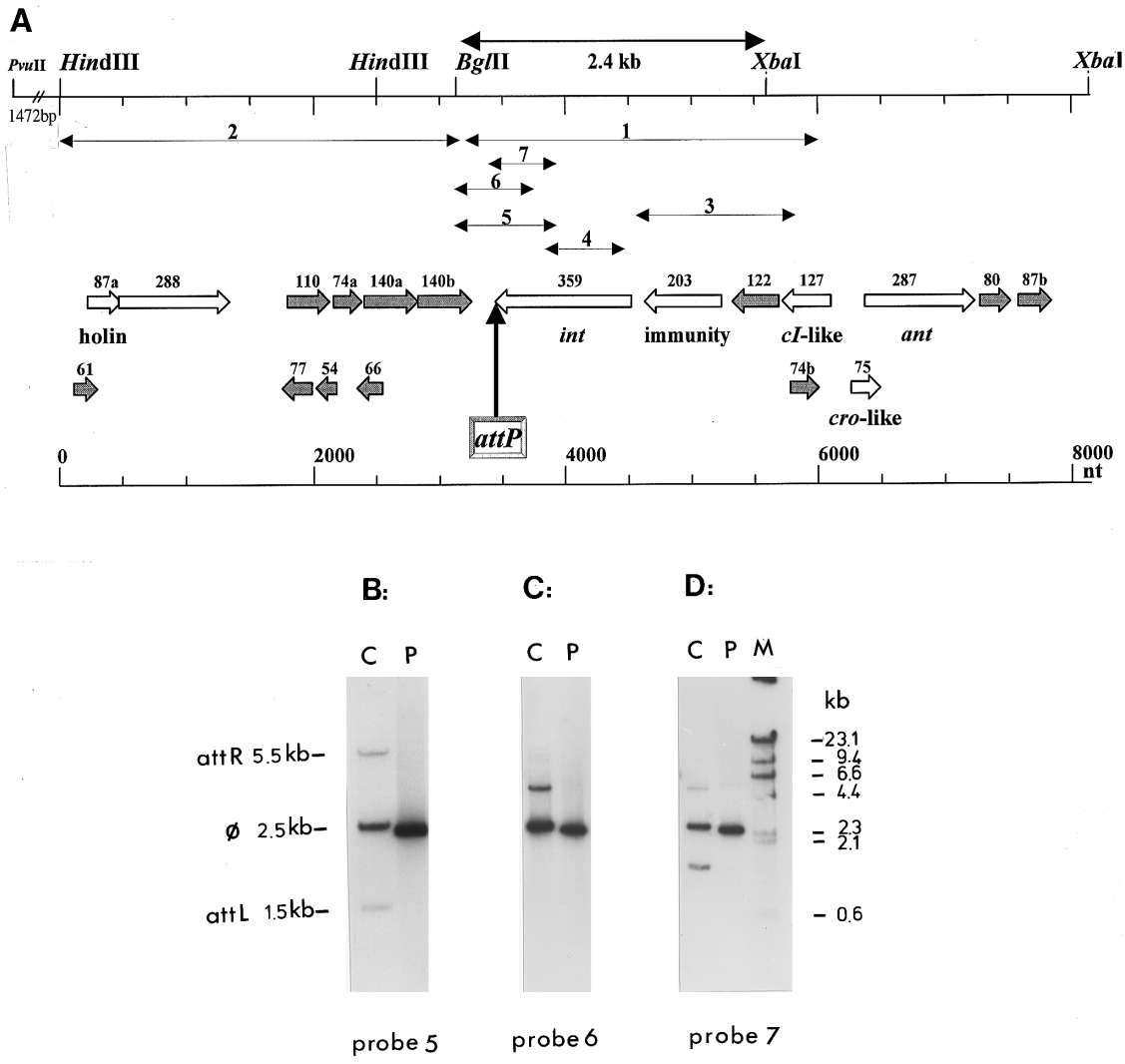

FIG. 1. Localization of the *attP* site by Southern hybridization. (A) Orientation of the PCR probes used in the Southern hybridization experiments on the map of an 8-kb region of the phage ϕSfi21. The probes are indicated by horizontal arrows and are numbered 1 to 7. The position of the 2.4-kb *Bgl*II/*Xba*I restriction fragment of ϕSfi21 is given by a bold arrow. The vertical arrow indicates the position of the *attP* core sequence. The map provides the predicted orfs named according to the number of aa. Orfs showing matches to the database are indicated by white arrows and annotated with genetic symbols (Bruttin *et al.,* 1997a). (B to D) Hybridizations using the PCR probes 5 (B), 7 (C) and 6 (D) against *Bgl*II/*Xba*I-digested chromosomal DNA from lysogen Sfi19 (C) and phage ϕSfi21 DNA (P). M, molecular weight marker in kb.

Only probe 1 yielded hybridization signals in the lysogen which were not detected in the phage DNA (data not shown) localizing the *attP* site to the right 2.4 kb of the PX3 fragment, PX3R. In the next step probes 3, 4, and 5 (Fig. 1A) were used against *Bgl*II/*Xba*I-digested DNA. In addition to a 2.4-kb signal possibly corresponding to unintegrated phage DNA (Brüssow and Bruttin, 1995), only probe 5 yielded two further hybridization signals of about 5.5 and 1.5 kb (Fig. 1B), the putative *attR* and *attL* fragments created by a Campbell-like integration event. To further reduce the size of the phage DNA fragment containing the *attP* site, probes 6 and 7 were used. Probe 6, which covered the 3′ part of the integrase gene (*int*) plus 300 bp downstream of the *int* gene, yielded the two supplementary 5.5- and 1.5-kb hybridization signals previously observed (Fig. 1D). In contrast, probe 7, which covered the 3′-half of the *int* gene, excluding the terminal 18 bp, gave only one supplementary signal of 5.5 kb (Fig. 1C). Therefore, we had to conclude that the recombination site must be located in a region covering the last 18 bp of the *int* gene and a further 300 bp downstream. This location of the *attP* site was further substantiated by sequence analysis in that a conspicuous array of direct and inverted repeats, typical for *attP* sites, was observed at the end of the *int* gene (Fig. 2, first line).

The *attL* fragment

When *Bgl*II/*Eco*RI-digested chromosomal DNA from the lysogenic *S. thermophilus* strain Sfi19 was hybridized with probe 7 (Fig. 1A), a 0.8-kb signal was obtained which

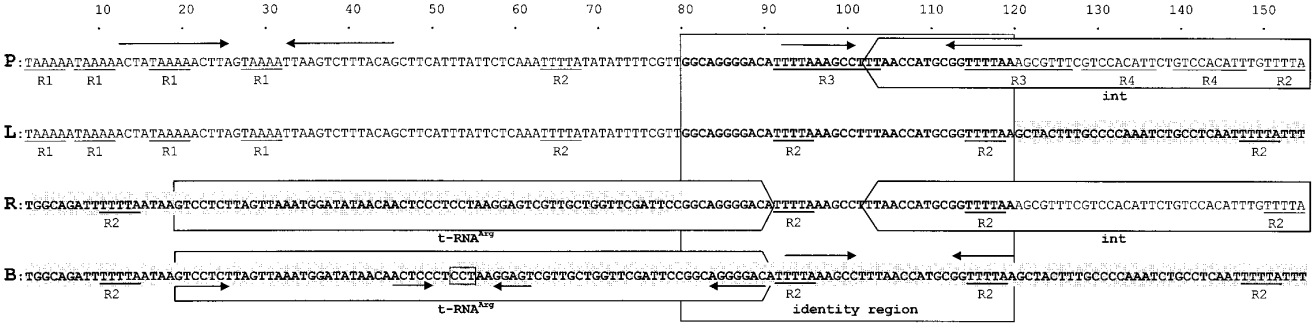

FIG. 2. Comparison of the DNA sequences of the phage *attP* (P), the lysogen *attL* (L) and *attR* (R), and bacterial *attB* (B) attachment sites. The box indicates the 40-bp segment of identity between the four sequences. The bacterial DNA sequence is shaded. The position of the tRNA gene is represented by an arrow on the *attB* and *attR* sequence. The 3′-end of the phage integrase gene is represented by an arrow in the *attP* and *attR* sequence. The arrows point to the 3′-end of the genes. R1, R2, R3, and R4 type direct repeats are underlined. Two regions of dyad symmetry were marked by opposing arrows in *attP.* The aminoacyl-stem and anticodon-stem repeats in the tRNA gene were marked by opposing arrows in *attB* and the anticodon was boxed. The DNA sequences shown in this figure were deposited under Access Nos. AF013584 (*attP*), AF013585 (*attB*), AF013586 (*attL*), and AF013587 (*attR*) in GenBank.

was not observed for *Bgl*II/*Eco*RI-digested phage DNA (data not shown). In order to clone this fragment, chromosomal *Bgl*II/*Eco*RI fragments of this size were recovered from preparative agarose gels and ligated to *Bgl*II/*Eco*RI-digested pUC19 vector. Transformation of *E. coli* yielded several positive clones, all of which showed an identical sequence (Fig. 2). The chimeric origin of this sequence, which we have designated as *attL,* was demonstrated by the fact that the clone hybridized with chromosomal DNA from the nonlysogenic *S. thermophilus* strain Sfi1 and with phage DNA (data not shown). No significant DNA homology was found between the cloned DNA of bacterial origin and entries in the databases. The transition point from the known phage DNA sequence into bacterial DNA sequence was within the *int* gene, 15 bp upstream of the stop codon (Fig. 2, second line).

PCR using oligonucleotides 2 and 3 designed according to the bacterial and phage DNA sequences bracketing the *attL* region resulted in an amplification product from the lysogenic strains Sfi19 and Sfi1L, a ϕSfi21-lysogenized Sfi1 (Brüssow and Bruttin, 1995), but not from the nonlysogenic strain Sfi1 or ϕSfi21 DNA (Fig. 3A). The amplification products comigrated with the PCR product obtained from the *attL* clone.

The *attR* fragment

A number of attempts at cloning the *attR* fragment were unsuccessful. To circumvent this, we applied the technique of inverse PCR. Lysogen Sfi19 DNA was digested with the restriction enzyme *Hin*dIII, diluted, and religated. Using primer 1 which anneals within the *int*

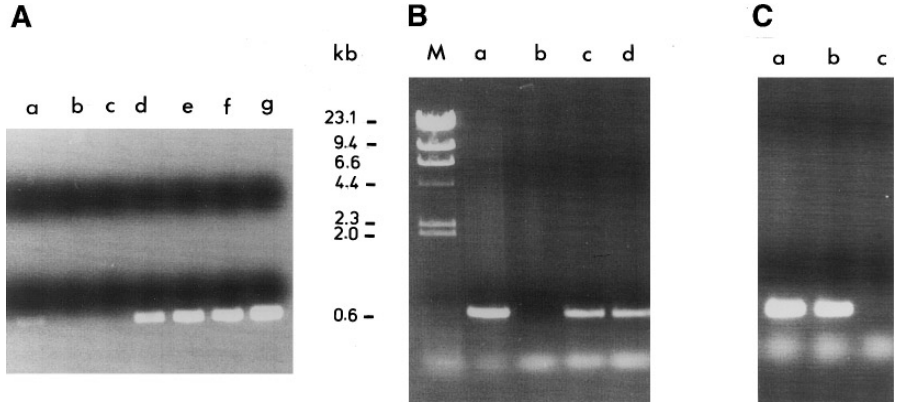

FIG. 3. PCR amplification of the *attL*/*attR* segments. (A) PCR amplification of *attL* using primers 2 and 3 and the following DNA templates: crude phage ϕSfi21 (a), plasmid pPX3R (b), chromosomal DNA from the nonlysogen Sfi1 (c), the lysogens Sfi19 (d) and Sfi1L (e), the nonlysogen Sfi1 transformed with plasmid pPX3R containing the phage *int* gene (f) and *attL* cloned from the lysogen Sfi19 (g). The weak reaction in lane a is likely to represent cellular DNA from the infected cell as it was not observed with cloned phage DNA shown in lane b. (B) PCR amplification of *attL* using primers 2 and 3 and as DNA template chromosomal DNA from the lysogen Sfi19 (lane a), the nonlysogen Sfi1 (lane b), and survivors of phage ϕSfi21 infections of the strains Sfi1 (lane c) and Sfi8 (lane d). M, molecular weight marker; lambda DNA digested with *Hin*dIII. (C) PCR amplification of the *attR* segment of chromosomal DNA from the lysogen Sfi19 (lane a), the nonlysogenic Sfi1 transformed with plasmid pPX3R (b) and with plasmid pSF101 (c), a derivative of pPX3R with an internal deletion in the *int* gene (Bruttin *et al.,* 1997b). The primer pair 1 and 4 was used.

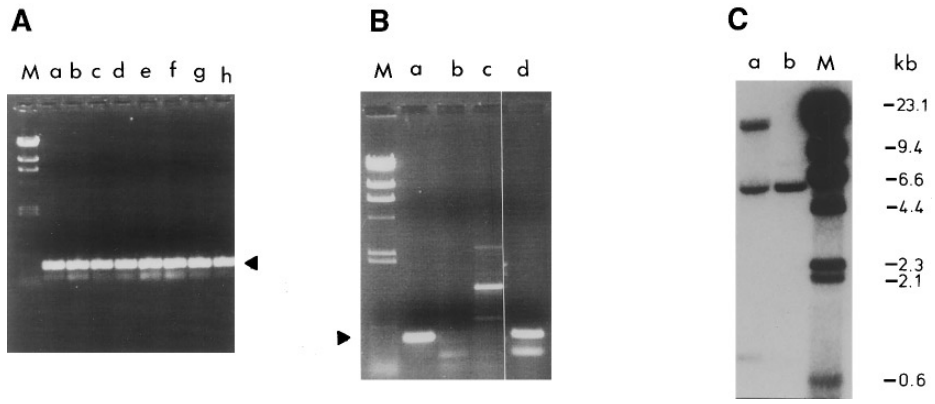

FIG. 4. Analysis of the *attB* site. (A) PCR amplification of the *attB* region from chromosomal DNA of *S. thermophilus* strains Sfi2 (a), 3 (b), 6 (c), 8 (d), 9 (e), 10 (f), 11 (g), and 13 (h). (B) PCR amplification of *attB* region from chromosomal DNA of *S. thermophilus* Sfi1 (a) and Sfi19 (d), *Lactobacillus helveticus* N2 (b), and *Lactococcus lactis* 135 (c). (C) Southern blot hybridization of *Pvu*II-digested chromosomal DNA from the lysogenic Sfi1L (lane a) and nonlysogenic Sfi1 (lane b) probed with the labeled *attB* PCR amplification product. M, molecular weight marker in kb; lambda DNA digested with *Hin*dIII.

gene 200 bp from the *attP* region and primer 5, located next to the *Hin*dIII site in orf 443, 5.3 kb upstream of *int* (Desiere *et al.,* 1997), a 4.3-kb PCR product was obtained. The inverse PCR product hybridized with DNA from both ϕSfi21 and the nonlysogen strain (data not shown), confirming its hybrid character. Sequence analysis located the transition point from phage to bacterial DNA 22 bp downstream of the phage *int* gene stop codon (Fig. 2, third line) and differed thus from the transition point determined for the *attL* clone. This discrepancy can only be explained by a core region common to both bacterial and phage DNA which covers 22 bp downstream and 15 bp upstream of the *int* gene stop codon, thereby yielding a 40-bp identity region. This was confirmed by analysis of the *attB* region and is described below.

attB

In order to amplify the bacterial attachment site *attB,* oligonucleotides 3 and 4 were designed according to the bacterial DNA sequences obtained for the phage–host junctions *attL* and *attR.* A PCR product of the expected size of 0.7 kb was synthesized when DNA from 15 nonlysogenic *S. thermophilus* strains, including SFi1, were used as the template (Fig. 4A). Sequence analysis of this amplification product from six strains revealed a sequence of 40 bp which was identical to DNA from *attP* (Fig. 2, fourth line). This identity region overlapped the 3′-end of the ϕSfi21 *int* gene by 18 bp.

Seven of the 15 nonlysogenic strains could be lytically infected with ϕSfi21 and lysogenized cells could be recovered from all infections confirmed by the amplification of a PCR signal indicative of the phage–host junction *attL* (Fig. 3B).

tRNA gene

The DNA sequence of the *attB* region was investigated for similarities with database entries. The sequence showed strong similarity with the *tRNA*^{arg5} gene from *E. coli* with 63 identical bp of a total of 75 (Fig. 5). The attribution of a tRNA gene to this *S. thermophilus* sequence was strengthened by the observation that the differences compared to the *E. coli* sequence represented complementary bp changes that reconstituted a perfect anticodon–stem structure. The anticodon recognized the AGG codon. The 11-terminal bp of the putative bacterial tRNA gene were located in the identity region of *attB* and *attP* (Fig. 2). In the *attR* junction the tRNA and the *int* gene were in a tail-to-tail position and separated by 11 bp.

Approximately 50 bp upstream of the tRNA gene we identified a host gene putatively coding for the *S. thermophilus* homologue of the large subunit ribosomal protein L19 (>85 and >50% aa identity with the *Bacillus stearothermophilus* homologue, $P = 10^{-56}$, and the *E. coli rplS* gene product, $P = 10^{-36}$, respectively). It should be noted that the *tRNA*^{arg5} gene of *E. coli* is not adjacent to the *rplS*

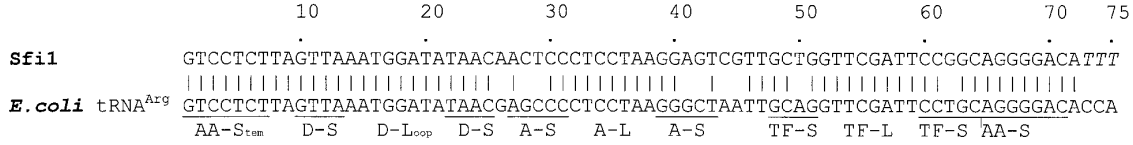

FIG. 5. Alignment of a 75-bp DNA sequence from the *attB* site with the tRNA^{arg5} gene from *E. coli.* The stem (S) and loop (L) structures of the tRNA are indicated. Note the complementary changes in the Ant (anticodon) stem structures. Abbreviations: AA-S/L, aminoacyl-stem/loop; D-S/L, D-stem/loop; TF-S/L, TF-stem/loop; A-S/L, anticodon-stem/loop.

gene (positions 62.4 and 59 min, respectively). The AGG codon recognized by the *tRNA^{arg5}* gene is one of the rarest codons used in *E. coli* (1.5/1000 codons). It is also a rarely used codon in *S. thermophilus* (1.7/1000; 8000 codons evaluated). Since some phages use this rare AGG codon for regulation of protein expression, we investigated the AGG codon usage in ϕSfi21 genes. The AGG codon was used with a higher frequency in phage (6/1000; 15 kb phage DNA evaluated) than in host genes. The frequency was higher in genes from the phage DNA replication module (8/1000; Desiere *et al.,* 1997) than in genes from the lysogeny module (4/1000; Bruttin *et al.,* 1997b), e.g., 2 AGG codons were seen in orf 359 coding for the phage integrase, while 5 AGG codons were seen in orf 443 coding for a distant member of the helicase family.

attB in lysogen

Surprisingly, PCR using the primer pair 3 and 4 yielded an amplification of the *attB* site from the lysogen Sfi19 (Fig. 4 B). The specificity of the reaction was demonstrated by the absence of a PCR signal from chromosomal DNA of *Lactobacillus helveticus* N2 and the presence of a PCR signal from chromosomal DNA of *Lactococcus lactis* ssp. *lactis* 135 at a different molecular weight position (Fig. 4B).

This observation could indicate a spontaneous excision of the phage from the lysogen without inducing a lytic infection cycle with concomitant degradation of DNA. We searched for independent evidence for intact *attB* sites in lysogenic strains. In fact, Southern blots of lysogen DNA probed with the *attB* PCR product yielded in addition to the *attL* and *attR* junctions an *attB* signal which comigrated with the single hybridization signal obtained for the nonlysogenic strain Sfi1 (Fig. 4C).

Cloning of the *attP/int* region in the vector pNZ124

The 2.4-kb *Bgl*II/*Xba*I phage DNA fragment (Fig. 1A) was cloned in the *E. coli*/*Lactococcus lactis*/*S. thermophilus* shuttle vector pNZ124 (Platteeuw *et al.,* 1994), yielding plasmid pPX3R (Bruttin *et al.,* 1997b). When *S. thermophilus* strain Sfi1 was transformed with this plasmid, we observed integration of the plasmid into the host chromosome as demonstrated by PCR, using the primer pair 2 and 3 which amplify the cloned *attL* site (Fig. 3A, lane f). The PCR product was sequenced and found to be identical to the *attL* sequence previously obtained, suggesting faithful integration of the plasmid pPX3R into the *attB* site. The specificity of the PCR and the implication of the phage *int* gene for plasmid integration was demonstrated by the absence of a PCR amplification signal for *attL* and *attR* from chromosomal DNA of Sfi1 transformed with plasmid pSF101 (Fig. 3C). pSF101 differs from pPX3R only by an internal deletion in the *int* gene (Bruttin *et al.,* 1997b).

Construction of a nonreplicative integration vector

pCI372 is a shuttle vector capable of replicating in *E. coli* and *Lactococcus lactis,* but not in *S. thermophilus,* and contains the *cat* gene as a selectable marker (Hayes *et al.,* 1990). In order for *S. thermophilus* cells to express chloramphenicol resistance when transformed with this vector, chromosomal integration of the vector is required. We cloned the 2.4-kb *Bgl*II/*Xba*I phage DNA fragment in this vector, yielding plasmid pSF28, to test whether the *int* gene could mediate chromosomal integration of this vector (Fig. 6A). pSF28 was electroporated into *S. thermophilus* Sfi1. Seven randomly selected chloramphenicol-resistant colonies yielded a PCR product corresponding to the *attL* junction, while the plasmid pSF28 and the untransformed Sfi1 cell gave no amplification product (data not shown). The PCR product was sequenced and found to be identical to the *attL* region shown in Fig. 2. The *attP* site could not be amplified by PCR using primers 1 and 2, thus proving the absence of the original plasmid (data not shown). To independently demonstrate chromosomal integration of pSF28, cellular DNA was isolated from the integrant Sfi1::pSF28 and digested with *Eco*RI, *Sac*I, and *Kpn*I, all of which cut once in pSF28. Southern blots were probed with P^{32}-labeled pCI372. A single hybridization signal of variable size was observed indicative of chromosomal integration of the plasmid (Fig. 6B).

Deletions

During the course of this study we have observed a considerable number of deletions in the cloned 2.4-kb *Bgl*II/*Xba*I phage DNA fragment, some of which have been sequenced.

One such deletion was observed during our attempt to clone the *attR* region by the method of plasmid rescuing from the integrant Sfi1::pSF28. In order to achieve this, *Sac*I-digested chromosomal DNA was self-ligated and electroporated into *E. coli.* Since *Sac*I cuts once within the integrated plasmid DNA and at a second site within the chromosome (Fig. 6A), the colonies obtained should possess a single plasmid made up of pCI372 vector DNA plus the phage–host junction *attR.* However, analysis of the recovered plasmid DNA indicated a mixed population which included a deletion derivative. Sequence analysis of this deletion revealed phage DNA including the *int* gene and a further 14 or 15 bp belonging to the identity region. This was immediately followed by a sequence corresponding to the vector located 34 bp from the *Bgl*II site in the direction of the *ori 305* (Fig. 6A). Examination of the pCI372 vector showed a 6-bp sequence (GGCAGG) repeated in the left side of the identity region (Fig. 7A). The 6-bp repeat on the plasmid was surrounded by two R2 (TTTTA) and two R1 (TAAAA) repeats (see Fig. 2 for the definition of the repeats).

A second interesting deletion was observed in the plasmid pPX3R. The deletion sites could be mapped to

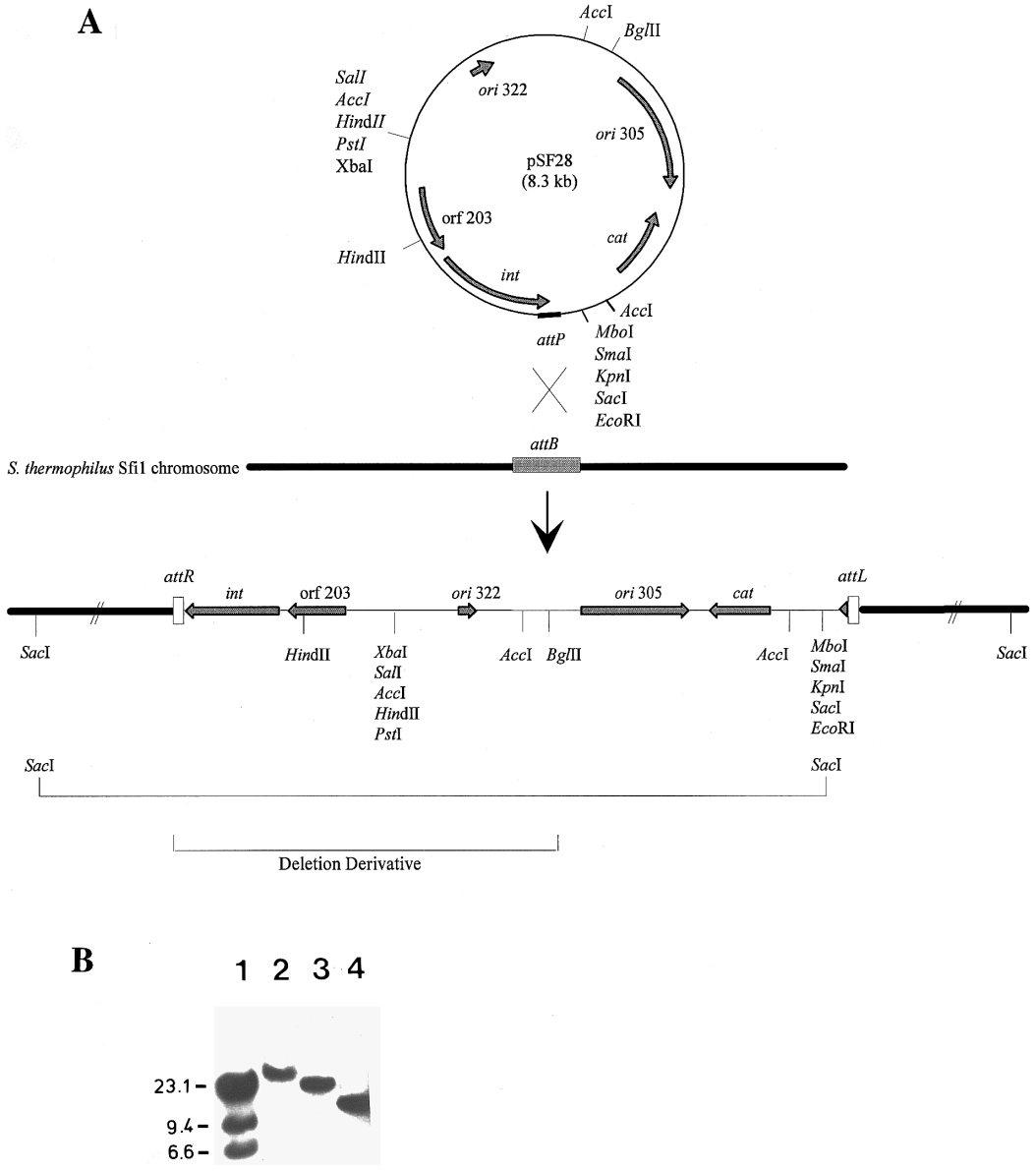

FIG. 6. Construction of an integrative plasmid. (A) Schematic representation of pSF28 site-specific integration into the $tRNA^{Arg}$ gene of the *S. thermophilus* Sfi1 chromosome. The deletion plasmid described in the legend to Fig. 7A is shown. (B) Southern blot hybridization of *Kpn*I (lane 2), *Sac*I (lane 3), and *Eco*RI (lane 4) digested chromosomal DNA from Sfi1 transformed with plasmid pSF28 and probed with labeled vector pCI372 DNA. Lane 1 gives the positions of the molecular weight markers: lambda DNA digested with *Hin*dIII.

the *attP* core and orf 203. Sequence analysis revealed the presence of a 4-bp sequence $^{5'}$AGGG$^{3'}$ at both sites: this sequence is part of the identity region of the attachment sites and overlaps with the 6-bp repeat observed in the deletion described above (Fig. 7B). As in the identity region the $^{5'}$AGGG$^{3'}$ sequence in orf 203 was situated in an A-T rich DNA sequence. An identical deletion was obtained in an independent experiment, while no such deletions were observed with plasmid pSF101 which differed from pPX3R by a deletion within the *int* gene.

A third deletion originating in the attachment site was observed during rescue of *attR* from Sfi1::pSF28. The deletion sites could be mapped to the *attP* core and a plasmid sequence. Sequence analysis revealed the presence of a 3-bp sequence $^{5'}$GGC$^{3'}$ at both sites, one of which overlaps with the tRNA side of the identity region. This 3-bp sequence also overlaps with the 6-bp repeat of the first deletion described above (Fig. 7D). Note that the 6 bp 5′ of the GGC sequence of the plasmid are also present in the *attP* site (TTCGTT-GGC, compare Fig. 7D with 7B).

Finally, we have previously reported on a spontaneous phage deletion mutant obtained during serial passages of ϕSfi21 in the laboratory (Bruttin and Brüssow, 1996). In this mutant the *int* gene and the upstream orf 203, orf 122, and orf 127 were deleted (Bruttin *et al.,* 1997b).

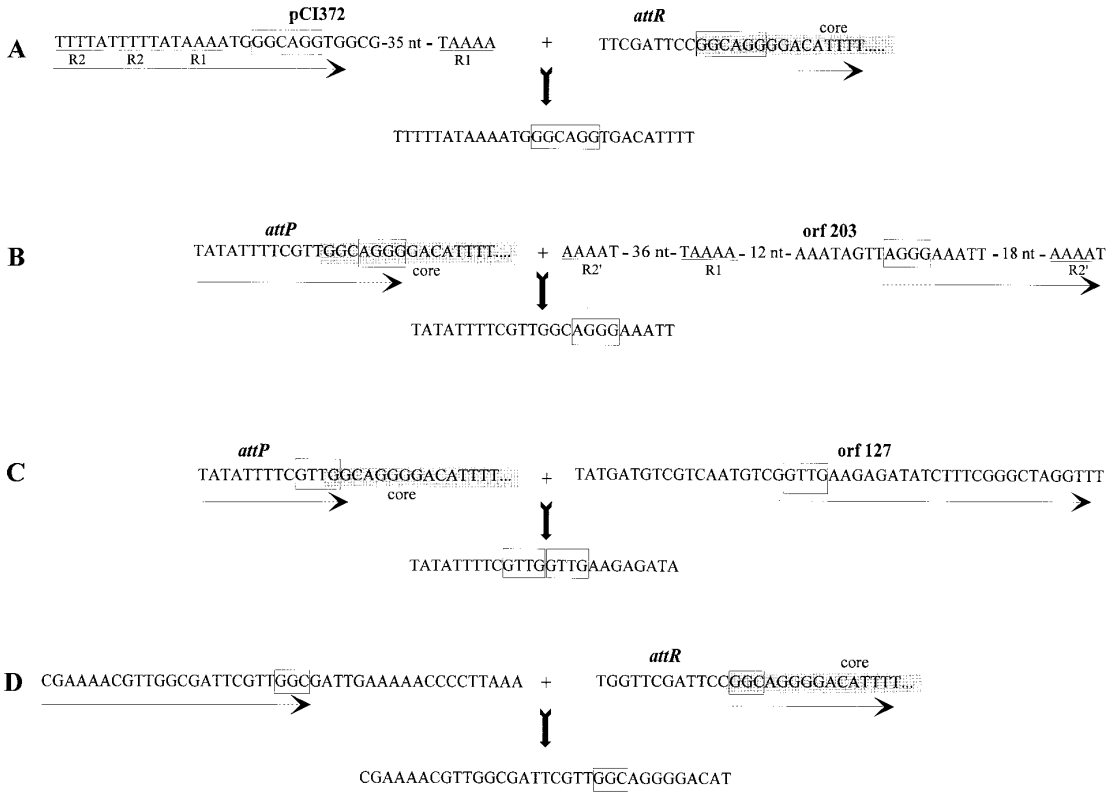

FIG. 7. Molecular analysis of plasmid and phage deletions. (A) Alignment of the DNA regions leading to the deletion observed in the rescued integration vector pSF28. The first line shows the sequence surrounding the 6-bp repeat GGCAGG (boxed) in the plasmid pCI372 sequence (left) and in the *attR* region (right). The identity region in *attR* is shaded. R1 and R2 repeats (see Fig. 2) are underlined. The contribution of bp from the vector pCI372 and the *attR* region to the deleted plasmid (shown in the second line) is indicated by arrows (dotted line: area of ambiguous attribution). −35nt-, 35-bp intervening sequence not depicted. (B) Alignment of the DNA regions leading to the deletions observed in the plasmid pPX3R. The first line shows the sequence surrounding the 4-bp repeat AGGG (boxed) in the *attP* region (left) and phage orf 203 sequence (right) in the plasmid pPX3R. The contribution of bp from both regions to the deleted plasmid (shown in the second line) is indicated by arrows; dotted line, area of ambiguous attribution. R2′, complementary sequence to R2. The identity region is shaded. (C) Alignment of the DNA regions leading to the phage deletion. The first line shows the sequence surrounding the 4-bp repeat GTTG (boxed) in the *attP* region (left) and phage orf 127 sequence (right). The identity region within *attP* is shaded. (D) Alignment of the DNA regions leading to the deletion obtained during plasmid rescuing from the integrant Sfi1::pSF28. The first line shows the sequence surrounding the 3-bp repeat GGC (boxed), the region of the two deletion points.

Sequence analysis revealed that a deletion had occurred between two 4-bp repeats (GTTG), one located at the tRNA side of the identity region and the other in the upstream orf 127 (Bruttin *et al.,* 1997b). In contrast to the plasmid deletions where one repeat was always lost, the two GTTG repeats were conserved in the phage deletion mutants (Fig. 7C). This deletion was observed in a number of independent phage isolates.

DISCUSSION

The temperate bacteriophage ϕSfi21 integrates its DNA into the chromosome of *Streptococcus thermophilus* strains via site-specific recombination. Nucleotide sequencing of *attP*-containing phage DNA, the bacterial attachment site *attB,* and the host–phage junctions *attL* and *attR* identified a 40-bp identity region which is substantially larger than those reported for the other phages from lactic acid bacteria (5 to 17 bp: Boyce *et al.,* 1995; Christiansen *et al.,* 1994, 1996; Dupont *et al.,* 1995; Lillehaug and Birkeland, 1993; Raya *et al.,* 1992; van de Guchte *et al.,* 1994). Longer core sites are, however, not unusual as demonstrated by the phage T12 of *Streptococcus pyogenes* which showed a 96-bp region of identity with the chromosomal DNA (McShan *et al.,* 1997).

Interestingly, there are two features which set the identity region of ϕSfi21 apart. First, in contrast to most phage systems where the identity region is localized downstream of the integrase gene, the identity region of phage ϕSfi21 includes the terminal 18 bp of the integrase gene. The only other bacteriophage system for which the identity region is located within the integrase gene is the *Myxococcus xanthus* phage Mx8 (Tojo *et al.,* 1996). Integration of this phage into the chromosome of its host results in the disruption of the *int* gene and thereby the production of an altered Int protein. In contrast, *S. thermophilus* lysogens possess an intact *int* gene since the bacterial *attB* site provides 18 bp which complement exactly the end of the *int* gene. Since the phage DNA sequence duplicated in the bacterial DNA is too long to be

explained by chance, we postulate that a faithful extension of the phage integrase gene provides a selective advantage for the streptococcal host. One apparent advantage of lysogeny is resistance to superinfection with lytic phages (Bruttin *et al.,* 1997b). Advantages of lysogeny might be even more subtle as demonstrated in *E. coli* where a nonessential phage λ gene confers serum resistance to its host (Barondess and Beckwith, 1995).

The second interesting feature is that the identity region complements both a phage and a bacterial gene since it also duplicates the 11 bp of a putative $tRNA^{arg}$ gene of *S. thermophilus.* Integration of the prophage into tRNA genes is common (Campbell, 1992; Reiter *et al.,* 1989). Several rules have been established that seem to be important for integration into tRNA genes (Gabriel *et al.,* 1995): (i) The core sequence comprises the 3′ terminal part of the tRNA gene. (ii) Integration restores an intact tRNA gene (with unchanged upstream region) at the host–phage junction. (iii) The tRNA used as target site is a functional gene. (iv) The anticodon is part of the core region. The first two rules are fulfilled in the *S. thermophilus* integration system. Whether the third rule is fulfilled is uncertain. The putative $tRNA^{arg}$ gene appears to be a class II-type tRNA gene requiring a post-transcriptional modification by nucleotidyltransferase to acquire 5′-CCA-3′ (Deutscher, 1990), while the only other tRNA gene sequenced in *S. thermophilus* includes the 3′ terminal CCA sequence (Tilsala-Timisjarvi and Alatossava, 1997). This difference does not, however, exclude that the $tRNA^{arg}$ gene is functional as the *B. subtilis* genome has a mixture of the two types of tRNA genes, with one quarter coding for type II tRNAs (Vold, 1985). The fourth rule is definitively not fulfilled by the identity region of φSfi21, since it does not include the CTT anticodon of the tRNA gene (Fig. 2). A similar observation was made for the *Lactobacillus delbrueckii* phage mv4 (Dupont *et al.,* 1995).

Like the actinophage RP3 of *Streptomyces rimosus* (Gabriel *et al.,* 1995), φSfi21 integrates into a putative $tRNA^{Arg}$ (AGG) gene. Since the 5′-AGG-3′ codon is used at a relatively low frequency in *Streptomyces* genes, Gabriel *et al.* (1995) postulated that the $tRNA^{Arg}$ (AGG) gene is not only important for the recombination system, but also for the regulation of the *int* mRNA translation which contains 2 AGG codons (Gabriel *et al.,* 1995). In the *Streptomyces* phage φC31 14 orfs are free of AGG codons (Hartley *et al.,* 1994), while the *int* gene contains 5 AGG codons (Kuhstoss and Rao, 1991). It should be noted that the φC31 *int* gene is not a member of the integrase/recombinase family and prophage is not integrated into a tRNA gene (Kuhstoss and Rao, 1991). Modulation of Int synthesis by the rare $tRNA^{Arg}$ codons AGA and AGG was also reported for phage λ which also does not integrate into a tRNA gene (Zahn and Landy, 1996). We examined the codon usage of the available φSfi21 genome sequence (Bruttin and Brüssow, 1996; Bruttin *et al.,* 1997b; Desiere *et al.,* 1997) and found that the AGG codon was used more frequently in phage genes than in the host genes where it was as rare as in *E. coli.* However, the rare AGG codon was used more frequently in a protein involved in lytic function (helicase) than in lysogeny (integrase).

Within *attP* a complex array of repeats was observed. Two inverted repeats (IR) of 13 bp (two mismatches) and 9 bp (one mismatch), each with an interval of an odd number of bp, were observed. The 13-bp IR could represent a weak transcriptional terminator of the *int* gene, while the 9-bp IR could represent a recognition site for the integrase as proposed by Campbell (1992). In addition, two types of 5-bp (R1 and R2), one 9-bp (R4), and one 13-bp (R3, one mismatch) direct repeats were observed. By analogy with better defined phage systems the repeats are likely to represent binding or recognition sites for Int or IHF (integrative host factor) proteins, but attribution of these sites in φSfi21 must await the purification of these proteins.

In the present communication we report on a number of deletion events that originated at the tRNA side of the identity region resulting in the fusion of phage to plasmid DNA sequences sharing 4- to 6-bp repeats. It is difficult to explain this deletion "hot spot" by slippage of the DNA polymerase during replication (Mollet and Delley, 1990; Peeters *et al.,* 1988) as more than 100 4-bp repeats exist in the 2.4-kb phage DNA fragment. This selectivity could, however, be explained by implicating the phage integrase. Similar deletion mutants have been described for phage λ where an Int-mediated cross-over event between the phage *attP* site and secondary *att* sites in nonessential regions of the λ genome led to deletions (Davis and Parkinson, 1971). All deletion types described in this report would identify the left border (tRNA side) of the identity region as the crossover site. Interestingly, Auvray *et al.* (1997) observed that for integrative plasmids, based on *Lactobacillus* φmv4 DNA, a 3- to 4-bp homology with the tRNA side of the identity region was sufficient for plasmid integration into heterologous hosts.

We have demonstrated that a 2.4-kb phage DNA fragment contains all of the genetic information required for faithful integration of a plasmid in the bacterial attachment site. Loss of integration activity in plasmid pSF101 identified orf 359 as the gene coding for the integrase. The phage excisionase is commonly found upstream of the *int* gene. This is not the case for φSfi21 since the upstream orf 203 was implicated in superinfection immunity (Bruttin *et al.,* 1997b). Furthermore, the integrated plasmid is maintained in contrast to the less stable prophage, as demonstrated by the release of infectious phage from the lysogen (Brüssow and Bruttin, 1995) and the presence of free *attB* sites. Mobilization of the integrated plasmid by cell transformation with different cloned phage segments should lead to the identification of the excisionase.

ACKNOWLEDGMENTS

We thank Frank Desiere for help in the final part of the work and the Swiss National Science Foundation for financial support of Sophie Foley in the framework of its Biotechnology Module (Grant 5002-044545/1).

REFERENCES

Auvray, F., Coddeville, M., Ritzenthaler, P., and Dupont, L. (1997). Plasmid integration in a wide range of bacteria mediated by the integrase of *Lactobacillus delbrueckii* bacteriophage mv4. *J. Bacteriol.* 179, 1837–1845.

Barondess, J. J., and Beckwith, J. (1995). *bor* gene of phage λ, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. *J. Bacteriol.* 177, 1247–1253.

Boyce, J. D., Davidson, B. E., and 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.

Brüssow, H., and Bruttin, A. (1995). Characterization of a temperate *Streptococcus thermophilus* bacteriophage and its genetic relationship with lytic phages. *Virology* 212, 632–640.

Brüssow, H., Bruttin, A., Desiere, F., Lucchini, S., and Foley, S. (1997). Molecular ecology and evolution of *Streptococcus thermophilus* bacteriophages-a review. *Virus Genes,* in press.

Brüssow, H., Frémont, M., Bruttin, A., Sidoti, J., Constable, A., and Fryder, V. (1994a). Detection and classification of *Streptococcus thermophilus* bacteriophages isolated from industrial milk fermentation. *Appl. Environ. Microbiol.* 60, 4537–4543.

Brüssow, H., Probst, A., Frémont, M., and Sidoti, J. (1994b). Distinct *Streptococcus thermophilus* bacteriophages share an extremely conserved DNA fragment. *Virology* 200, 854–857.

Bruttin, A., and Brüssow, H. (1996). Site-specific spontaneous deletions in three genome regions of a temperate *Streptococcus thermophilus* phage. *Virology* 219, 96–104.

Bruttin, A., Desiere, F., d'Amico, N., Guérin, J.-P., Sidoti, J., Huni, B., Lucchini, S., and Brüssow, H. (1997a). Molecular ecology of *Streptococcus thermophilus* bacteriophage infections in a cheese factory. *Appl. Environ. Microbiol.* 63, 3144–3150.

Bruttin, A., Desiere, F., Lucchini, S., Foley, S., and Brüssow, H. (1997b). Characterization of the lysogeny DNA module from the temperate *Streptococcus thermophilus* bacteriophage ϕSfi21. *Virology* 232, 136–148.

Campbell, A. (1962). Episomes. *Adv. Genet.* 11, 101–145.

Campbell, A. (1992). Chromosomal insertion sites for phages and plasmids. *J. Bacteriol.* 174, 7495–7499.

Carroll, D., Kehoe, M. A., Cavanagh, D., and Coleman, D. C. (1995). Novel organization of the site-specific integration and excision recombination functions of the *Staphylococcus aureus* serotype F virulence-converting phages ϕ13 and ϕ42. *Mol. Microbiol.* 16, 877–893.

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

Christiansen, B., Johnsen, M. G., Stenby, E., Vogensen, F. K., and Hammer, K. (1994). Characterization of the lactococcal temperate phage TP901-1 and its site-specific integration. *J. Bacteriol.* 176, 1069–1076.

Davis, R. W., and Parkinson, J. S. (1971). Deletion mutants of bacteriophage lambda. III. Physical structure of attϕ. *J. Mol. Biol.* 56, 403–429.

Desiere, F., Lucchini, S., Bruttin, A., Zwahlen, M.-C., and Brüssow, H. (1997). A highly conserved DNA replication module from *Streptococcus thermophilus* phages is similar in sequence and topology to a module from *Lactococcus lactis* phages. *Virology* 234, 372–382.

Deutscher, M. P. (1990). Ribonucleases, tRNA nucleotidyltransferase, and the 3′ processing of tRNA. *Prog. Nucleic Acid Res. Mol. Biol.* 39, 209–240.

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.

Gabriel, K., Schmid, H., Schmidt, U., and Rausch, H. (1995). The actinophage RP3 DNA integrates site-specifically into the putative $tRNA^{Arg}$ (AGG) gene of *Streptomyces rimosus. Nucleic Acids Res.* 23, 58–63.

Hartley, N. M., Murphy, G. O., Bruton, C. J., and Chater, K. F. (1994). Sequence of the essential early region of ϕC31, a temperate phage of *Streptomyces* ssp. with unusal features in its lytic development. *Gene* 147, 29–40.

Hayes, F., Daly, C., and Fitzgerald, G. F. (1990). Identification of the minimal replicon of *Lactococcus lactis* subsp. *lactis* UC317 plasmid pCI305. *Appl. Environ. Microbiol.* 56, 202–209.

Holo, H., and Nes, I. F. (1989). High frequency transformation by electroporation of *Lactococcus lactis* ssp. *cremoris* grown in glycine in osmotically stabilised media. *Appl. Environ. Microbiol.* 55, 3119–3123.

Kuhstoss, S., and Rao, R. N. (1991). Analysis of the integration function of the streptomycete bacteriophage ϕC31. *J. Mol. Biol.* 222, 897–908.

Landy, A. (1989). Dynamic, structural, and regulatory aspects of λ site-specific recombination. *Annu. Rev. Biochem.* 58, 913–949.

Lee, M., Pascopella, L., Jacobs, W., and Hatfull, G. (1991). Site-specific integration of mycobacteriophage L5: Integration-proficient vectors for *Mycobacterium smegmatis, Mycobacterium tuberculosis,* and bacille Calmette-Guérin. *Proc. Natl. Acad. Sci. USA* 88, 3111–3115.

Le Marrec, C., Moreau, S., Loury, S., Blanco, C., and Trautwetter, A. (1996). Genetic characterization of site-specific integration functions of ϕAAU2 infecting "*Arthrobacter aureus*" C70. *J. Bacteriol.* 178, 1996–2004.

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.

McShan, W. M., Tang, Y.-F., and Feretti, J. J. (1997). Bacteriophage T12 of *Streptococcus pyogenes* integrates into the gene encoding a serine tRNA. *Mol. Microbiol.* 23, 719–728.

Mollet, B., and Delley, M. (1990). Spontaneous deletion formation within the β-galactosidase gene of *Lactobacillus bulgaricus. J. Bacteriol.* 172, 5670–5676.

Peitersen, N. (1991). Practical phage control. *Bull. Int. Dairy Fed.* 263, 1–43.

Peeters, B. P. H., de Boer, J. H., Bron, S., and Venema, G. (1988). Structural plasmid instability in *Bacillus subtilis:* Effect of direct and inverted repeats. *Mol. Gen. Genet.* 212, 450–458.

Platteeuw, C., Simons, G., and de Vos, W. M. (1994). Use of the *Escherichia coli* β-glucuronidase (gusA) gene as a reporter gene for analyzing promoters in lactic acid bacteria. *Appl. Environ. Microbiol.* 60, 587–593.

Raya, R., Fremaux, C., De Antonioni, G., and Klaenhammer, T. (1992). Site-specific integration of the temperate bacteriophage ϕadh into the *Lactobacillus gasseri* chromosome and molecular characterization of the phage (*attP*) and bacterial (*attB*) attachment sites. *J. Bacteriol.* 174, 5584–5592.

Reiter, W.-D., Palm, P., and Yeats, S. (1989). Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res.* 17, 1907–1914.

Slos, P., Bourquin, J.-C., Lemoine, Y., and Mercenier, A. (1991). Isolation and characterization of chromosomal promoters of *Streptococcus salivarius* subsp. *thermophilus. Appl. Environ. Microbiol.* 57, 1333–1339.

Tilsala-Timisjarvi, A., and Alatossava, T. (1997). Development of oligo-

nucleotide primers from the 16S-23S rRNA intergenicx sequences for identifying different dairy and probiotic lactic acid bacteria by PCR. *Int. J. Food Microbiol.* 35, 49–56.

Tojo, N., Sanmiya, K., Sugawara, H., Inouye, S., and Komano, T. (1996). Integration of bacteriophage Mx8 into *Myxococcus xanthus* chromosome causes a structural alteration at the C-terminal region of the IntP protein. *J. Bacteriol.* 178, 4004–4011.

Van de Guchte, M., Daly, C., Fitzgerald, G. F., and Arendt, E. K. (1994). Identification of *int* and *attP* on the genome of lactococcal bacteriophage Tuc2009 and their use for site-specific plasmid integration in the chromosome of Tuc2009-resistant *Lactococcus lactis* MG1363. *Appl. Environ. Microbiol.* 60, 2324–2329.

Vold, B. S. (1985). Structure and organization of genes for transfer ribonucleic acid in *Bacillus subtilis. Microbiol. Rev.* 49, 71–80.

Zahn, K., and Landy, A. (1996). Modulation of lambda integrase synthesis by rare arginine tRNA. *Mol. Microbiol.* 21, 69–76.