Site-specific integration of corynephage ϕ 16: construction of an integration vector

Sylvie Moreau,[†] Carlos Blanco and Annie Trautwetter

Author for correspondence: Annie Trautwetter. Tel/Fax: +33 2 99 28 61 40.
e-mail: annie.trautwetter@univ-rennes1.fr

Membranes et
Osmorégulation, UPRES-A
6026, Université de Rennes I,
Avenue du Général Leclerc,
35042 Rennes Cedex, France

ϕ 16, a temperate phage induced from *Corynebacterium glutamicum* ATCC 21792, lysogenizes its host via site-specific recombination. The phage attachment site, *attP*, was located to a 6.5 kb *Bam*HI fragment of the ϕ 16 genome. This fragment also contained ϕ 16 integrative functions. The minimal phage DNA fragment required for integration was defined. This 1630 bp region contained a large open reading frame, *int*, encoding a protein of 416 amino acids with similarity in its carboxyl-terminal domain to tyrosine recombinases and particularly to the Xer recombinases. The comparison of the nucleotide sequences of *attB*, *attL*, *attR*, and *attP* identified a common 29 bp sequence, the core sequence. It lies 11 bp downstream of the 3' end of the integrase gene. ϕ 16 integrase was shown to catalyse site-specific integration *in trans* to *attP* with an efficiency of 5×10^3 integrants per μ g DNA. The integrating fragment catalysed integration in several *Corynebacterium* strains that are not infected by ϕ 16, thus enlarging the host spectrum of integrating vectors derived from ϕ 16. In these strains, the ϕ 16 *attB* site was located in a conserved intergenic region and lies downstream of a *clp* gene.

Keywords: temperate phage, corynebacteria, site-specific integration, integrative vector

INTRODUCTION

Only a few temperate corynephages have been described to date (Kato *et al.*, 1984; Le Marrec *et al.*, 1994; Sonnen *et al.*, 1990). Studies on site-specific integration functions, the phage attachment site and the integrase gene were initiated by studies on phage β of *Corynebacterium diphtheriae* that defined the bacterial (*attB*) and phage (*attP*) attachment sites (Cianciotto *et al.*, 1990; Ratti *et al.*, 1997). The wide distribution of the *attB* ^{β} -related sequence among *Corynebacterium* species was also demonstrated (Cianciotto *et al.*, 1986). Recently, the characterization of the integrative functions of ϕ AAU2 infecting '*Arthrobacter aureus*' C70 was reported (Le Marrec *et al.*, 1996). Surprisingly, ϕ AAU2 integrase revealed only a low similarity to other members of the λ integrase family of site-specific recombinases. Moreover, a 800 bp non-coding region immediately upstream of *int* was shown to be absolutely

required for site-specific integration. As these findings differ from what is known about the well-documented lambdoid phage family (Campbell, 1962; Weisberg & Landy, 1983) and in order to get a better understanding of site-specific recombination mediated by corynephages, we examined the integrative functions carried on other temperate corynephages. We previously isolated, by UV induction of 20 coryneform strains, four temperate corynephages and reported their characterization (Moreau *et al.*, 1995). Three of them are present as integrated prophages in their original host(s): ϕ 16, lysogenizing *Corynebacterium glutamicum* ATCC 21792, and ϕ 304L and ϕ 304S, both lysogenizing *C. glutamicum* ATCC 13058, 21650, 21488 and 21649. These phages were not detected in any other of the 20 strains, as shown by Southern hybridization analysis (Moreau *et al.*, 1995). This paper deals with ϕ 16. We have located and characterized the ϕ 16 integrative functions, i.e. *attP* and the putative integrase gene, *int*. We were able to separate the *int* gene from the *attP* site and subsequently demonstrated that ϕ 16 integrase, when produced *in trans* to the *attP* site, was still able to catalyse site-specific integration. Finally, we derived from ϕ 16 an integrative vector which recombines into the chromosome of several corynebacteria species, which are not susceptible to infection by ϕ 16.

[†]Present address: Department of Microbiology, Columbia University College of Physicians and Surgeons, 701 West 168th St, New York, NY 10032, USA.

The DDBJ/EMBL/GenBank accession numbers for the sequences determined in this work are Y12471 and Y12472, for the region encoding ϕ 16 integrative functions and the *attB* sites, respectively.

Table 1. Bacterial strains, bacteriophage, plasmids and primers

	Relevant characteristics	Reference/source
Bacteria		
<i>Corynebacterium glutamicum</i> ATCC 21792	φ16 natural lysogen	Moreau <i>et al.</i> (1995)
<i>C. glutamicum</i> 21792c	ATCC 21792 cured of φ16 prophage	Moreau <i>et al.</i> (1995)
<i>C. glutamicum</i> ATCC 13032 RM3	Restriction and modification defective	Schäfer <i>et al.</i> (1990)
<i>C. glutamicum</i> ATCC 21086 BL15	Restriction and modification defective	Bonnassie <i>et al.</i> (1990b)
' <i>Arthrobacter aureus</i> ' C70	φAAU2 indicator strain	Le Marrec <i>et al.</i> (1994)
<i>Escherichia coli</i> NM522	<i>supE thi Δ(lac-proAB) Δ(hsdMS-mcrB)5 F'(proAB lacI^a ΔlacZM15)</i>	Mead <i>et al.</i> (1985)
<i>E. coli</i> MC4100 (Mucts)(MudIIPR13)	<i>araD139 Δ(argF-lac)U169 rspL150 (Sm^r) relA1 flbB5301 pstF25 deoC1 (Mucts)</i> , bearing MudIIPR13 (<i>lacZ'</i> YA Cm ^r) integrated in the chromosome	Ratet <i>et al.</i> (1988)
Phage		
φ16	Temperate phage isolated by UV induction from <i>C. glutamicum</i> ATCC 21792	Moreau <i>et al.</i> (1995)
Plasmids		
pJRD16905	<i>aphIII</i> inserted between <i>MluI</i> - <i>PvuII</i> sites of pJRD184; Amp ^r Tet ^r Kan ^r	Trautwetter & Blanco (1991)
pACYC184	Cm ^r Tet ^r	Chang & Cohen (1978)
pBluescriptIISK(+)	Amp ^r <i>lacZα</i>	Stratagene
pBK	Kan ^r derivative of pBluescript; <i>lacZα</i>	Le Marrec <i>et al.</i> (1996)
p25435	<i>E. coli</i> / <i>Corynebacterium</i> shuttle vector, Kan ^r <i>lacZα</i>	Trautwetter & Blanco (1991)
pBBR1MCS-5	Vector replicative in <i>E. coli</i> but not in <i>corynebacteria</i> , Gm ^r <i>lacZα</i>	Kovach <i>et al.</i> (1995)
p11915	pBluescriptIISK(+) ΔCfr10I; Amp ^r <i>lacZα</i>	This study
p16046	6.5 kb <i>Bam</i> HI fragment from φ16 DNA cloned in pBK; integrating	This study
p13348	2.5 kb <i>Bam</i> HI fragment from the p16046::MudIIPR13 E mutant cloned in pBK; integrating	This study
p16150	Exo III digestion of p13348 (2.2 kb <i>attP</i> -bearing fragment); integrating	This study
pA3781	p13348 ΔEcoRI (1.79 kb <i>Bam</i> HI- <i>Eco</i> RI insert); integrating	This study
pA6114	p13348 ΔPstI (1.67 kb <i>Bam</i> HI- <i>Pst</i> I insert); integrating	This study
p16130	Exo III digestion of p13348 (1.63 kb <i>attP</i> -bearing fragment); integrating	This study
p16136	Exo III digestion of p13348 (1.53 kb <i>attP</i> -bearing fragment); non-integrating	This study
p11199	<i>Kpn</i> I <i>attL</i> rescue from RM3::p16046, Kan ^r	This study
p11257	<i>Xba</i> I <i>attR</i> rescue from RM3::p16046, Kan ^r	This study
p12399	0.3 kb <i>Bam</i> HI- <i>Sal</i> I fragment from p11199 cloned in pBK, <i>attL</i>	This study
p12945	0.6 kb <i>Bam</i> HI- <i>Sal</i> I fragment from p11257 cloned in pBK, <i>attR</i>	This study
pA7039	<i>Sal</i> I- <i>Bam</i> HI fragment from the 501 bp PCR product amplified from INT1 to INT2 primers, carrying the 3' end of the φ16 <i>int</i> gene cloned in pACYC184	This study
pA7078	pA7039 inserted in the unique <i>Sal</i> I site (bordering <i>attP</i>) of p16150; <i>int</i> unlinked from the <i>attP</i> site	This study
pB780	1.67 kb <i>Bam</i> HI- <i>Pst</i> I fragment from pA7078 cloned in p25435; <i>int</i> ; no <i>attP</i> site	This study

Table 1. (cont.)

	Relevant characteristics	Reference/source
pB594	0.3 kb <i>Sall</i> – <i>Bam</i> HI <i>attP</i> containing fragment from p16150 in pBBR1MCS-5	This study
pA4350	<i>Hind</i> III rescue of RM3 ^{attB} from a p12945::RM3 recombinant	This study
p14409	<i>Bst</i> EII– <i>Eco</i> RV ϕ AAU2 integrative fragment in pJRD184, Δ <i>Apa</i> I, Kan ^r cassette inserted in <i>Hind</i> III site	Le Marrec <i>et al.</i> (1994); unpublished results
pB1714	<i>Pst</i> I– <i>Bam</i> HI <i>int</i> -carrying fragment from pA7078 in the <i>Nsi</i> I– <i>Bgl</i> II sites of p14409	This study
pA3250	<i>Bam</i> HI– <i>Pst</i> I integrative fragment from pA6114 cloned in pJRD16905	This study
pA3253	<i>Bam</i> HI– <i>Bal</i> I integrative Kan ^r fragment from pA3250 cloned between the <i>Dra</i> I sites of p11915	This study
Primers		
INT1	5' CGCTGAAAACCGAGAGA 3' (position 931 to 947 of <i>int</i>)	This study
INT2	5' AGCCGCACTGAGGATCC 3' (position 1415 to 1431 of <i>int</i>)	This study

METHODS

Bacteriophage, bacterial strains and plasmids. These are listed in Table 1, together with the primers used.

Growth conditions. *C. glutamicum* BL15 was grown in Brain Heart Infusion (BioMérieux, France) at 30 °C; other corynebacteria and *E. coli* strains were grown in L-broth (Miller, 1972), at 30 °C for the former and 37 °C for the latter. Bacterial growth was monitored as OD₅₇₀. When needed, antibiotics were used at the following final concentrations: kanamycin at 90 µg ml⁻¹ for '*A. aureus*' and 50 µg ml⁻¹ for all other strains; chloramphenicol at 50 µg ml⁻¹ for *E. coli*, 10 µg ml⁻¹ for *C. glutamicum* RM3 and 5 µg ml⁻¹ for *C. glutamicum* BL15; ampicillin at 50 µg ml⁻¹; and gentamicin at 10 µg ml⁻¹ for *E. coli* and 1 µg ml⁻¹ for *C. glutamicum* RM3 and BL15.

Plasmid electrotransformation in corynebacteria. Corynebacteria were electrotransformed with a gene pulser (Bio-Rad) apparatus. *C. glutamicum* BL15 and *Arthrobacter* strains were electroporated as reported previously (Bonnassie *et al.*, 1990a; Le Marrec *et al.*, 1994). *C. glutamicum* 21792c and RM3 were transformed as follows. Cells were grown in L-broth with added sodium succinate (10%, w/v) and glycine (2%, w/v) until the cultures reached an OD₅₇₀ of 0.4. Then penicillin was added to a final concentration of 0.4 U ml⁻¹ and cells were grown for a further doubling. Cells were harvested, washed twice with 20 ml cold 0.5 M sucrose, and concentrated 60-fold in the same medium. Then 200 µl of concentrated cells were mixed with plasmid DNA and immediately electroporated at 25 µF, 10 kV cm⁻¹ and 400 Ω, using a pre-cooled 2 mm cuvette. After electroporation, cells were immediately diluted fivefold in L-broth with 0.5 M added sucrose, incubated at 30 °C for 1 h and plated.

DNA amplification. The PCR amplification of the 3' end of ϕ 16 *int* gene was done on p16150 DNA with INT1 and INT2 primers (Table 1) and Goldstar DNA polymerase (Eurogentec). Reaction was carried out for 35 cycles as

follows: denaturation at 94 °C for 3 min, primer annealing at 43 °C for 1 min and primer extension at 72 °C for 2 min.

Plasmid integration assay. Non-replicative plasmids were introduced into corynebacteria by electrotransformation; p25435 was used as a positive transformation control. Antibiotic-resistant clones were selected on the appropriate solid medium. Following the extraction of total DNA, the cellular fate of the plasmid was analysed in two ways. First, DNA was digested with restriction enzymes, and then analysed by Southern hybridization to search for plasmid–chromosome junctions. Secondly, the junction fragments were rescued independently of each other by using an enzyme able to liberate a DNA fragment including one junction and both the antibiotic-resistance marker and the replication origin of the integrated plasmid. This fragment was the only one able to transform *E. coli* to kanamycin resistance after self-ligation of the restriction digest.

Rescue of *attL* and *attR* sites from p16046 integrated in RM3. The *attL* and *attR* junctions resulting from the integration of p16046 into the *attB* site of the RM3 chromosome (see Fig. 1c) were rescued by *Kpn*I and *Xba*I cleavage respectively. The digestion mixture was self-ligated, introduced into *E. coli*, and replicating fragments were selected for as Kan^r clones, giving p11199 (*attL*) and p11257 (*attR*). Next, the 0.3 and 0.6 kb *Bam*HI–*Sall* fragments, carrying *attL* and *attR* sites respectively, were introduced into pBK, giving p12399 (*attL*) and p12945 (*attR*) and selected for in *E. coli* NM522 as Kan^r Lac⁻ clones.

Rescue of *attB* from RM3. For this purpose we used p12945, which carries the *attR* junction from a RM3::p16046 integrant (Fig. 1c), cloned in pBK. This plasmid, which did not replicate in corynebacteria, was electrotransformed into RM3. Kan^r clones harbouring p12945 integrated in the chromosome via a single crossover occurring in the *attR* region were obtained. Hence, their chromosome contained the intact *attB* region and the *attR* junction spaced by the pBK sequence. Thus it became possible to rescue the *attB* region using the *Hind*III site present

Fig. 1. Characterization of the minimal $\phi 16$ integrating fragment. (a) Site-specific integration mediated by the 6.5 kb *Bam*HI phage DNA fragment containing the *attP* site and carried by p16046. All DNAs were cleaved with *Bam*HI and analysed by Southern blotting with $\phi 16$ or p16046 DNAs as probes. The lysogen was *C. glutamicum* ATCC 21792. Lanes M1 and M2, *Hind*III and *Eco*RI fragments, and fragments produced by an *Eco*RI/*Hind*III double digest of λ DNA respectively, probed with λ . The sizes of relevant fragments, in kb, are indicated on the right. (b) Map of the integrating plasmid p16046 resulting from the insertion of a 6.5 kb *Bam*HI $\phi 16$ DNA fragment into the *Bam*HI site of pBK. (c) Site-specific integration of p16046 into the *attB* site of 21792c, generating *attL* and *attR* junctions (the map is the same for strains BL15 and RM3). Chromosome and vector are drawn as thick and thin lines respectively, and phage sequences as boxes. (d) Localization of $\phi 16$ integrative functions within the 6.5 kb *Bam*HI fragment. The top line shows a detailed restriction map of the 6.5 kb fragment contained in p16046. The vertical lines below the map indicate the position of the insertion points of MudIIPR13 in mutants A to H: —, Loss of integration; +, 5×10^4 transformants per μg DNA in *C. glutamicum* RM3 and BL15. The hatched box represents the *attP* site and the black box the *int* gene. The lines below show the recovery of $\phi 16$ integrating functions from the E mutant. p13348, pA3781 and pA6114 were obtained by

on the left side of *attB* (see Fig. 1c) and the unique pBK *Hind*III site located downstream of *ori* and *aphIII* sequences in the recombinant. This replicating fragment was recovered in *E. coli* after transformation of the self-ligated *Hind*III digest of the RM3::p12945 chromosome. All Kan^r clones harboured the same plasmid, pA4350.

Insertional mutagenesis. Mutagenesis of p16046 by the insertion of MudIIPR13 was performed in *E. coli* as reported previously (Ratet *et al.*, 1988). Insertion points were determined by *Bam*HI and *Hind*III restriction analysis.

DNA manipulations. Chromosome, phage and plasmid DNAs from coryneform strains were prepared as previously described (Le Marrec *et al.*, 1994; Moreau *et al.*, 1995). Southern hybridizations and DNA sequencing were described elsewhere (Sambrook *et al.*, 1989; Sanger *et al.*, 1977). Database searches were performed by using FASTA (Pearson & Lipman, 1988) and BLAST (Altschul *et al.*, 1990) programs.

Unlinking of the integrase gene from the *attP* site and cloning of *attP*. p16150, carrying the ϕ 16 integrating functions (see Fig. 1d), was used to amplify (using PCR), a fragment of 501 bp containing the 3' end of the *int* gene followed by the few bases separating it from the core of *attP*. The primers used were INT1 (Table 1) and INT2 (Table 1, Fig. 2b), with INT2 designed to create a *Bam*HI site flanking the 5' end of the core region. The amplified DNA was cleaved with *Bam*HI and *Sal*I and introduced into pACYC184 cut with the same enzymes. The resulting pA7039 was selected for in *E. coli* as Cm^r Tet^s clones. The next step aimed to replace the natural 3' end of *int* in p16150 by the mutated one. Thus, pA7039 was cut with *Sal*I and inserted in *Sal*I-cut p16150 in the correct orientation, which reconstituted an entire *int* gene while separating it from *attP*, yielding pA7078. This latter construction enabled the recovery of the entire *int* gene deprived of *attP*, by a *Pst*I/*Bam*HI double digestion. The resulting fragment was introduced between the corresponding sites present in the multiple cloning site of p25435 to obtain pB780, which was selected for as Kan^r Lac⁻ clones in *E. coli* NM522. This plasmid replicates in corynebacteria. The *attP* site was cloned as the 0.3 kb *Sal*I-*Bam*HI fragment from p16150 between the corresponding sites of pBBR1MCS-5, giving pB594, which was obtained as Gm^r Lac⁻ clones in *E. coli* NM522. The ϕ 16 *int* gene was also integrated into the chromosome of RM3 using the ϕ AAU2^{attB} site (Le Marrec *et al.*, 1996). We used the ϕ AAU2 integrative plasmid p14409 (Amp^r Kan^r). Briefly it corresponds to the *Bst*EII-*Eco*RV integrative fragment of ϕ AAU2 (Le Marrec *et al.*, 1996) inserted in the corresponding sites of pJRD184 and further deleted by *Apal*. Finally, a Kan^r gene was introduced in the *Hind*III site of the recombinant pACYC184 (unpublished data). The *Pst*I-*Bam*HI *int*-carrying fragment from pB780 was introduced between the *Nsi*I and *Bgl*II sites of p14409, yielding pB1714. After electrotransformation into RM3, pB1714 integrated into the ϕ AAU2^{attB} site.

RESULTS

Localization of ϕ 16 integrative functions

The ϕ 16 *attP* site was previously assigned to 18 kb *Mlu*I and 2.8 kb *Nsi*I phage DNA fragments (Moreau *et al.*,

1995). In the same way, when the ATCC 21792 chromosome was cleaved with *Bam*HI and probed with radiolabelled ϕ 16 DNA, all the vegetative phage restriction fragments but one, the 6.5 kb fragment, were detected in normal amounts and a new fragment, 7.4 kb in size, was revealed (Fig. 1a). This result is consistent with a recombination event occurring between the phage attachment site, *attP*, borne by the 6.5 kb fragment and the chromosomal attachment site, *attB* (Campbell, 1962). This event results in the splitting of the *attP* fragment, accounting for the appearance of the 7.4 kb prophage-chromosome junction fragment ('junction'). The second junction fragment is presumably obscured by the other restriction fragments (Fig. 1a). The small amount of 6.5 kb fragment detected in the pattern of the lysogen probably derived from the free intracellular ϕ 16 genomes which result from spontaneous prophage induction events (Moreau *et al.*, 1995). If the restriction maps of the 18 kb *Mlu*I, 2.8 kb *Nsi*I and 6.5 kb *Bam*HI *attP*-containing fragments are aligned, a 0.5 kb *Bam*HI-*Nsi*I overlap, that must bear the *attP* site, is observed (data not shown). The 6.5 kb *Bam*HI fragment was cloned into pBK, a plasmid vector which cannot replicate in corynebacteria, giving p16046 (Fig. 1b). When p16046 was electrotransformed into *C. glutamicum* 21792c, 10 Kan^r transformants per μ g DNA were obtained. The total cellular DNA content of 20 independent clones was extracted, cleaved with *Bam*HI and analysed by Southern blotting with p16046 as a probe. The hybridization patterns were identical for all clones (data not shown), and the two junctions that were similar to those seen in *C. glutamicum* ATCC 21792 were now detectable (Fig. 1a). The 1.4 kb junction was barely visible, suggesting that it had only a short sequence that was homologous to the p16046 probe. Arbitrarily, the 7.4 kb junction was attributed to *attR*, and the 1.4 kb fragment to *attL* (Fig. 1c). When introduced into RM3, BL15 and '*A. aureus*', strains already shown to be nonlysogenic for ϕ 16 (Moreau *et al.*, 1995) and deficient in restriction systems, p16046 gave 5×10^4 integrants per μ g of DNA in the first two strains and no integrants in the last strain. Chromosomal DNA from twenty or so p16046 integrants in RM3 and BL15 was analysed by *Bam*HI digestion followed by Southern hybridization. Hybridization was detected to 7.4 kb and 1.4 kb *attL* and *attR* fragments respectively, suggesting that integration occurs at a unique *attB* site present in all three strains.

Since the 6.5 kb *Bam*HI fragment contained all the functions required for integration, it was subjected to insertional mutagenesis using MudIIPR13 (Fig. 1d). MudIIPR13 insertion points were determined by *Hind*III and *Bam*HI restriction analysis, and the mutant plasmids were electrotransformed into BL15 or RM3 strains to

restriction-enzyme-mediated deletions and p16150, p16130 and p16136 by exonuclease III deletion. The name of each plasmid is followed by the size of the phage DNA insert in parentheses, and the efficiency of transformation expressed as the number of transformants per μ g DNA. B, *Bam*HI; E, *Eco*RV; El, *Eco*RI; H, *Hind*III; K, *Kpn*I; N, *Nsi*I; P, *Pst*I; S, *Sal*I; X, *Xba*I.

Fig. 2. Characterization of the $\phi 16$ *att* sites. (a) DNA sequence of $\phi 16$ *attB*. The amino acid sequence of the end of the ORF overlapping the 5' end which shows homology to CLP proteases is given above the nucleotide sequence. The core site is boxed. Arrows indicate the positions of the direct repeats. (b) Sequence organization of the $\phi 16$ *attP* region lying from position 1260 to position 1630 of the minimal integrating fragment carried by p16130. The deduced amino acid sequence of the 3' end of *int* is shown above the nucleotide sequence. The core site is boxed. Beneath is shown the sequence of the primer INT2 used to separate *int* from *attP*, with an asterisk indicating the mismatch introduced to create a *Bam*HI site. Arrows indicate the positions of the direct repeats d1 to d4. The direct repeat d1' is a part of the repeat d1.

assay their integration ability (Fig. 1d). All insertions located upstream of the *Sal*I₍₂₎ site abolished integration. Therefore we made use of restriction sites created in the *E* mutant (Fig. 1d) to create a pBK subclone carrying a 2.5 kb *Bam*HI fragment, p13348, which should contain the integrating functions. This plasmid was subsequently subjected to restriction digest and progressive exonuclease III deletions (Fig. 1d). p16130, carrying a 1.63 kb DNA fragment, was the shortest construction giving an electrotransformation efficiency equal to that of p16046 (5×10^4 transformants per μ g DNA). Deletion of a further 96 bp segment yielded p16136, which failed to integrate.

Finally we located the crossover point within the *attP* region more precisely. Chromosomal DNA from a p16046 integrant was cleaved with *Sal*I and analysed by Southern hybridization with p16046. This analysis demonstrated that the crossover occurs within the 0.3 kb *Bam*HI–*Sal*I₍₁₎ fragment of p16046 (data not shown and Fig. 1d).

Isolation of *attL*, *attR* and *attB* sites, and definition of the $\phi 16$ core

A physical map of the *attB* region was established in each host starting from p16046 integrants obtained in *C. glutamicum* 21792c, RM3 and BL15. This map was identical in all the strains for the region delineated by *Kpn*I and *Xba*I sites (Fig. 1c).

When the sequences obtained from $\phi 16$ *attP*, *attL* and *attR* from 21792c, RM3 and BL15 (five clones each), and *attB* from RM3 were compared, an identical 29 bp

sequence (5' TCAGTGCGGCTGCGCTTGATTTCA-AAGAA 3') was found in each. This region was designated as the core sequence.

The nucleotide sequence of the RM3 *attB*-containing region, determined for about 900 bp, contains one ORF ending 252 bp upstream of the core sequence (Fig. 2a). The inferred amino acid sequence is 50–60% identical to several prokaryotic and eukaryotic CLP proteases. The homology is centred on the enzyme active site. The remainder of the *attB* region contains numerous stop codons in all six possible frames. Moreover, *attB* neither shows similarity to tRNA genes nor displays secondary structure. Thus, the *attB* ^{$\phi 16$} sites are identical and are located in a conserved intergenic region in RM3, BL15 and ATCC 21792.

Nucleotide sequence of the $\phi 16$ *int*–*attP* region

The shortest $\phi 16$ DNA fragment sustaining a site-specific integration into the *attB* site, contained in p16130, was entirely sequenced. It is 1630 nt in size and its overall G + C content is 52.4 mol%. The core site extends from position 1421 to 1449 and has a G + C content of 48.2 mol%. The region neighbouring the core contains three direct repeats of 8–9 bp (numbered d2 to d4), each present twice, and one sequence of 11 bp, placed in tandem downstream of the core site, whereas 8 contiguous base pairs of this sequence constitute the d1' repeated sequence located upstream of the core. Five repeats of the 5' AAATT 3' sequence, generally in association with a repeat, are also found in the surroundings of the core (Fig. 2b). None of these repeated sequences is identified in the *attB* region (Fig. 2a, b).

Fig. 3. Alignment of the carboxyl-terminal region of φ16 Int and its closer homologues. The conserved regions are framed. Box I and box II were defined according to Esposito & Scocca (1997) and Nunes-Düby *et al.* (1998), patches I to III according to Nunes-Düby *et al.* (1998). Consensus residues are indicated at the top of each region. The three residues perfectly conserved among tyrosine recombinases are boxed. h indicates the presence of hydrophobic amino acids. In the consensus of the patches, the hydrophobic character is more conserved than the consensus. Residues identical to the consensus sequence are indicated in bold, similar residues are indicated by shadowed characters. Similarity was defined by the Dayhoff PAM250 matrix. Presented sequences are: EC XERC (XerC of *E. coli*); ST XERC (*Salmonella typhimurium* XerC); PM XERC (*Proteus mirabilis* XerC); ST XERD (*Salmonella typhimurium* XerD); MS6 (phage MS6 Int), P22 (phage P22 Int); φ16 (φ16 Int); BC YDCL (*Bacillus subtilis* YDCL); and SA INT (*Staphylococcus aureus* Int).

The *int* ORF was the only one preceded by a potential ribosome-binding site. The deduced polypeptide contained 416 amino acids. To identify related sequences, BLAST searches were performed using the whole protein. Thirty proteins with high scores and a significant probability ($<10^{-5}$) were identified, including phage, plasmid, transposon and chromosome site-specific recombinases belonging to the λ integrase family (Argos *et al.*, 1986). The homology was mainly centred in the C-extremity residues 180–375 of φ16 Int. A second BLAST search using this region identified 100 related sequences showing 18–28% identity with φ16 Int. These regions were aligned using CLUSTAL W, which allowed the identification of highly conserved domains. Only the conserved region of the seven proteins displaying the highest scores of similarity to φ16 Int were used to construct the CLUSTAL alignment presented in Fig. 3. The highly conserved boxes described by Esposito & Scocca (1997) (boxes A and B) and refined by Nunes-Düby *et al.* (1998) (boxes I and II), are found. Box I contains the Arg residue previously identified (Abremski & Hoess, 1992). Box II contains the His-X-X-Arg motif commonly considered a hallmark of the recombinases of the λ Int family (Argos *et al.*, 1986), and the tyrosine residue implicated in the formation of a covalent intermediate between the recombinase and the DNA. The alignment

of 105 recombinases of the λ Int family (Nunes-Düby *et al.*, 1998) showed that the two arginines and the tyrosine are indeed invariant. The high sequence conservation of boxes I and II (Fig. 3) is reflected in the conserved secondary structure of the λ Int c170, HPC, XerD and Cre (Guo *et al.*, 1997; Hickman *et al.*, 1997; Kwon *et al.*, 1997; Subramanya *et al.*, 1997). In each of these proteins the R-H-R residues form a cluster on the protein surface, located at the centre of the DNA interaction surface in the Cre–DNA complex. Moreover a high degree of conservation and clustering of hydrophobic residues was evident from the alignment (Nunes-Düby *et al.*, 1998). Boxes I and II are well conserved in φ16 Int, the latter being the most conserved between φ16 and its closer homologues depicted in Fig. 3 (39–82% identity and 68–96% similarity). In addition to these highly conserved boxes and the pattern of core hydrophobic residues, three patches of conserved sequences were identified by Nunes-Düby *et al.* (1998). These patches, including charged amino acids and precisely spaced hydrophobic residues that are highly conserved, are also found in Int φ16 (Fig. 3). Considering the five conserved structural motifs, the greatest scores of homology to φ16 Int were obtained with Xer proteins, which function in circular chromosome separation. Moreover the spacing between the perfectly conserved arginine and tyrosine

Fig. 4. The ϕ 16-based integrative cloning vector, pA3253. The restriction sites present in the multiple cloning site (MCS), clockwise are: *KpnI*, *Apal*, *Dral*, *XhoI*, *HincII*, *AccI*, *Sall*, *ClaI*, *HindIII*, *EcoRV*, *EcoRI*, *PstI*, *SmaI*, *BamHI*, *SpeI*, *XbaI*, *NotI*, *SadI*, *SacI*. Only sites in bold type are unique.

residues in box II of ϕ 16 integrase is close to that of Xer proteins, suggesting that ϕ 16 Int is a member of the Xer family of site-specific recombinases.

This, added to the localization of the *int* ORF close to sites of insertional inactivation in mutants C and D, made this ORF a good candidate for ϕ 16 integrase. The deletion of the 96 bp at the 5' end of the integrative fragment (p16136) abolished integration, while leaving 48 bp in front of this ORF. So the transcriptional promoting region or a regulatory sequence might be altered in p16136.

ϕ 16 integrase catalyses recombination when produced *in trans* to *attP*

To establish whether the integrase of ϕ 16 might catalyse recombination in corynebacteria, when it is produced at a distance from *attP*, we (i) separated these two elements by about 4 kb within the same construct, pA7078 (see Methods and Table 1), and (ii) isolated the *int* gene on a *Corynebacterium* replicative plasmid, pB780 (Kan^r), and the *attP* site on a non-replicative plasmid, pB594 (Gm^r) (see Methods and Table 1). Both experiments yielded 5×10^3 transformants per μ g DNA while p16130 gave 5×10^4 transformants per μ g DNA. To ascertain whether this *trans* activity resulted from the overproduction of ϕ 16 integrase by the replicative pB780, we integrated ϕ 16 *int* into the chromosome of the recipient strain (see Methods) making use of the *attB* ϕ AAU² site (Le Marrec *et al.*, 1996): 1×10^3 transformants per μ g DNA were obtained. The integrase production resulting from the expression of a single copy of ϕ 16 *int* was therefore sufficient to catalyse recombination. When pB594 was introduced into '*A. aureus*' C70 harbouring pB780 no Kan^r clones were recovered, suggesting that host-specific factors are important for the integration event.

Construction of an integrative vector

Finally, we constructed an integrative cloning vector based on ϕ 16 and pBluescript, pA3253 (Fig. 4). The *Bam*HI–*Pst*I fragment of pA6114 which carries ϕ 16 integrating functions was first inserted between the corresponding sites of pJRD16905 (Table 1), yielding pA3250. This construct enabled the recovery of the integrative functions linked to the *aph*III selection marker through a *Bam*HI/*Bal*I double digest. This 2·870 kb fragment, whose *Bam*HI end was filled in with Klenow polymerase, was introduced between the *Dra*I sites of p11915 (Table 1), yielding pA3253. pA3253 carries unique restriction sites for *Kpn*I, *Apal*, *Xho*I, *Acc*I, *Cla*I, *Pst*I, *Sma*I, *Bam*HI, *Spe*I, *Xba*I, *Not*I, *Sac*II and *Sac*I, located within the multiple cloning site of p11915 (Fig. 4), therefore enabling direct recombinant screening in *E. coli* NM522 by the Lac[–] phenotype. Since it does not replicate in corynebacteria, this vector can only confer kanamycin resistance upon transformants in which site-specific recombination occurred in *attB* ^{ϕ 16}.

DISCUSSION

In this study we present the nucleotide sequence and the genetic organization of the *int*–*attP* region of ϕ 16 corynephage. The *attP* site was assigned to the 0·3 kb *Sall*–*Bam*HI fragment overlapping the 3' end of the *int* ORF encoding ϕ 16 integrase. The 29 bp core site, corresponding to the nucleotide sequence where the integrase catalyses strand exchange between *attP* and *attB* (Burgin & Nash, 1992), lies within the *attP* site but is 11 bp downstream of the 3' end of *int*. Therefore, the establishment of ϕ 16 prophage does not disrupt *int*. This is generally the case, the unique exception to date being phage Mx8 of *Myxococcus xanthus* (Tojo *et al.*, 1996). Several direct repeated sequences are located on both sides of the core site. This is a common feature among *attP* regions. In the case of phage λ , these sequences are bound by factors involved in the recombination event: the phage integrase and host factors such as IHF (Craig & Nash, 1984) and FIS (Thompson *et al.*, 1987) that play a role in the bending of the double helix. To date, no functions have been assigned to the direct repeats found in ϕ 16 *attP*. Repeats d3 and d4 (Fig. 2b), located upstream of the *Sall* site, are likely to be nonessential to the recombination process since integration efficiency is unaltered by their removal. One possibility is that they are involved in the excision process. Although a high A + T content is frequently a characteristic of *attP* sites (Hsu *et al.*, 1980; Landy, 1989) or at least of core sites, the A + T content of the ϕ 16 *attP* site (51·7 mol %) is not significantly different from that of the remaining nucleotide sequence of the integrating fragment (47·6 mol %).

To date, ϕ 16 has only been shown to infect its cured original lysogenic host (Moreau *et al.*, 1995). However, the ϕ 16 integrative fragment carried by p16130, first

identified through transformation in *C. glutamicum* 21792c, was also able to produce plasmid integration in two strains of industrial importance, *C. glutamicum* ATCC 13032 (RM3) and *C. glutamicum* ATCC 21086 (BL15), restriction-deficient derivatives that are apparently unrelated to *C. glutamicum* ATCC 21792, the original host of ϕ 16. This difference in behaviour might be explained by the lack of phage receptors on the recipient cell wall or possibly the inability to replicate ϕ 16. These two parameters were not important for p16130 integration since plasmids were introduced by electroporation. The p16130 integration spectrum may possibly be even larger, since the low efficiency of plasmid transformation in coryneform strains is caused by the presence of restriction systems (Bonamy *et al.*, 1990; Bonnassie *et al.*, 1990b; Vertès *et al.*, 1993). It could also be limited by the absence of an *attB* site or possibly the lack of host factors. In contrast to ϕ AAU2, whose integration functions work in *E. coli* (Le Marrec *et al.*, 1996) we were unable to obtain any recombination between the ϕ 16 minimum integrating fragment and an *attB*-bearing plasmid in *E. coli* (data not shown).

The ϕ 16 *attB* site was characterized from hosts allowing plasmid integration, either directly (by a cloning strategy) as for RM3, or indirectly (by nucleotide sequence comparison of *attL* and *attR*) for BL15 and 21792c. The core of the *attB* site was identical in the three strains, as was the surrounding sequence. The ϕ 16 *attB* site lies in an intergenic region, yet is conserved. Such a location is rather infrequent for phages of Gram-positive bacteria (Pattee *et al.*, 1977).

ϕ 16 integrase contains in its carboxyl-terminal region the five structural motifs identifying tyrosine recombinases. Therefore it seems not to be related to ϕ AAU2 integrase, which is a very atypical recombinase having the three conserved residues, R-R-Y, but showing poor homology to the consensus sequence of recombinases elsewhere in its C-extremity (Le Marrec *et al.*, 1996; Nunes-Düby *et al.*, 1998).

We were able to uncouple *int* and *attP* by introducing a point mutation by PCR, generating a *Bam*HI site in the short sequence spacing the integrase gene from the *attP* site. In such conditions ϕ 16 integrase was still able to catalyse integration. Nevertheless, the integration efficiency decreased by a factor of 10 compared to that produced by the native fragment. This did not result from a mutation introduced in the *attP* site during the PCR amplification since the nucleotide sequence of this fragment was verified. The size of the *attP* fragment used was likewise irrelevant since the same experiment performed with the *Nsi*I-*Bam*HI *attP*-containing fragment (456 bp) gave the same result, unless a binding site for a protein factor involved in recombination was present upstream of the *Nsi*I site. Finally, it was unlikely to be due to insufficient production of integrase by pB780, since the amount of integrase produced by only one copy of the gene was enough to catalyse the integration as observed for the non-replicating p16130 carrying the *int-attP* fragment (yielding 5×10^4 integrants per μ g DNA). However, it is possible that a

decrease in the integrase concentration in the vicinity of *attP*, resulting from the unlinking of these two sequences, causes a reduction in the integration efficiency.

Starting from the fact that lysogenic strains behave like non-lysogenic ones, and that integration of genetic material in the *attB* sites does not alter any vital function, in contrast to random recombination, phage *attB* sites might be considered for the introduction of heterologous genetic material into corynebacteria of industrial importance without requiring a selection pressure.

ACKNOWLEDGEMENTS

We thank L. Communier for photographic work.

This work was supported by grants from the European Economic Community (BRIDGE grant BIOT-CT91-0264), The Direction de la Recherche et des Etudes Doctorales, and The Centre National de la Recherche Scientifique.

REFERENCES

- Abremski, K. E. & Hoess, R. H. (1992). Evidence for the second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng* 5, 87-91.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.
- Argos, P., Landy, A., Abremski, K. & 9 other authors (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J* 5, 433-440.
- Bonamy, C., Guyonvarch, A., Reyes, O., David, F. & Leblon, G. (1990). Interspecies electro-transformation in corynebacteria. *FEMS Microbiol Lett* 66, 263-270.
- Bonnassie, S., Burini, J. F., Oreglia, J., Trautwetter, A., Patte, J. C. & Sicard, A. M. (1990a). Transfer of plasmid DNA to *Brevibacterium lactofermentum* by electroporation. *J Gen Microbiol* 36, 2119-2126.
- Bonnassie, S., Oreglia, J., Trautwetter, A. & Sicard, A. M. (1990b). Isolation and characterization of a restriction and modification deficient mutant of *Brevibacterium lactofermentum*. *FEMS Microbiol Lett* 72, 143-146.
- Burgin, A. B. & Nash, H. A. (1992). Symmetry in the mechanism of bacteriophage λ integrative recombination. *Proc Natl Acad Sci USA* 89, 9642-9646.
- Campbell, A. M. (1962). Episomes. *Adv Genet* 11, 101-145.
- Chang, A. C. Y. & Cohen, S. N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. *J Bacteriol* 134, 1141-1156.
- Cianciotto, N., Rappuoli, R. & Groman, N. (1986). Detection of homology to the beta bacteriophage integration site in a wide variety of *Corynebacterium* spp. *J Bacteriol* 168, 103-108.
- Cianciotto, N., Serwold-Davis, T., Groman, N., Ratti, G. & Rappuoli, R. (1990). DNA sequence homology between *attB*-related sites of *Corynebacterium diphtheriae*, *Corynebacterium ulcerans*, *Corynebacterium glutamicum*, and the *attP* site of γ -corynephage. *FEMS Microbiol Lett* 66, 299-302.
- Craig, N. L. & Nash, H. A. (1984). *E. coli* integration host factor binds to specific sites in DNA. *Cell* 39, 707-716.
- Esposito, D. & Socca, J. J. (1997). The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res* 25, 3605-3614.

- Guo, F., Gopaul, D. N. & Van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombinase synapse. *Nature* **389**, 40–46.
- Hickman, A. B., Waninger, W., Scocca, J. & Dyda, F. (1997). Molecular organisation in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution. *Cell* **89**, 227–237.
- Hsu, P. S., Ross, W. & Landy, A. (1980). The λ phage att site: functional limits and interaction with Int protein. *Nature* **285**, 85–91.
- Kato, F., Yoshimi, M., Araki, K., Motomura, Y., Matsufune, Y., Nobunaga, H. & Murata, A. (1984). Screening of bacteriocins in amino acid or nucleic acid producing bacteria and related species. *Agric Biol Chem* **48**, 193–200.
- Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M. & Peterson, K. M. (1995). Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**, 175–176.
- Kwon, H. J., Tirumalai, R., Landy, A. & Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* **276**, 126–131.
- Landy, A. (1989). Dynamic, structural, and regulatory aspects of λ site-specific recombination. *Annu Rev Biochem* **58**, 913–949.
- Le Marrec, C., Michotey, V., Blanco, C. & Trautwetter, A. (1994). ϕ AAU2, a temperate bacteriophage specific for 'Arthrobacter aureus', whose integrative functions work in other corynebacteria. *Microbiology* **140**, 3071–3077.
- Le Marrec, C., Moreau, S., Loury, S., Blanco, C. & Trautwetter, A. (1996). Genetic characterization of site-specific functions of ϕ AAU2 infecting 'Arthrobacter aureus' C70. *J Bacteriol* **178**, 1996–2004.
- Mead, D. A., Skorupa, E. S. & Kemper, B. (1985). Single stranded DNA SP6 promoter plasmid for engineering mutant RNAs and proteins, synthesis of a 'stretched' preproparathyroid hormone. *Nucleic Acids Res* **13**, 1103–1118.
- Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, NY: Cold Spring Harbor.
- Moreau, S., Leret, V., Le Marrec, C., Varangot, H., Ayache, M., Bonnassie, S., Blanco, C. & Trautwetter, A. (1995). Prophage distribution in coryneform bacteria. *Res Microbiol* **146**, 493–505.
- Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res* **26**, 391–406.
- Pattee, P. A., Thompson, N. E., Haubrich, D. & Novick, R. P. (1977). Chromosomal map locations of integrated plasmids and related elements in *Staphylococcus aureus*. *Plasmid* **1**, 38–51.
- Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.
- Ratet, P., Schnell, J. & De Bruijn, F. J. (1988). Mini-Mulac transposons with broad-host-range origins of conjugal transfer and replication designed for gene regulation studies in *Rhizobiaceae*. *Gene* **63**, 41–52.
- Ratti, G., Covacci, A. & Rappuoli, R. (1997). A tRNA^{Arg} gene of *Corynebacterium diphtheriae* is the chromosomal integration site for toxigenic bacteriophage. *Mol Microbiol* **25**, 1179–1181.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.
- Schäfer, A., Kalinowski, J., Simon, R., Seep-Feldhaus, A. H. & Pühler, A. (1990). High frequency conjugal plasmid transfer from gram negative *Escherichia coli* to various gram positive coryneform bacteria. *J Bacteriol* **172**, 1663–1666.
- Sonnen, H., Schneider, J. & Kutzner, H. J. (1990). Characterization of ϕ GA1, an inducible phage particle from *Brevibacterium flavum*. *J Gen Microbiol* **136**, 1–5.
- Subramanya, H. S., Arciszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J. & Wigley, D. B. (1997). Crystal structure of the site-specific recombinase XerD. *EMBO J* **16**, 5178–5187.
- Thompson, J. F., Moitoso de Vargas, L., Koch, C., Kahmann, R. & Landy, A. (1987). Cellular factors couple recombination with growth phase: characterization of a new component in the λ site-specific recombination pathway. *Cell* **50**, 901–908.
- Tojo, N., Sanmiya, K., Sugawara, H., Inouye, S. & Komano, T. (1996). Integration of bacteriophage Mx8 into the *Myxococcus xanthus* chromosome causes a structural alteration at the C-terminal region of the IntP protein. *J Bacteriol* **178**, 4004–4011.
- Trautwetter, A. & Blanco, C. (1991). Structural organization of the *Corynebacterium glutamicum* plasmid pCG100. *J Gen Microbiol* **137**, 2093–2101.
- Vertès, A. A., Inui, M., Kobayashi, M., Kurusu, Y. & Yukawa, H. (1993). Presence of *mrr*- and *mcr*-restriction systems in coryneform bacteria. *Res Microbiol* **144**, 181–185.
- Weisberg, R. & Landy, A. (1983). Site specific recombination in phage lambda. In *Lambda II*, pp. 211–250. Edited by R. W. Hendrix, J. W. Roberts, F. W. Stahl & R. A. Weisberg. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Received 31 July 1998; revised 9 November 1998; accepted 23 November 1998.