Cloning and DNA sequence analysis of the region containing attP of the temperate phage φAR29 of Prevotella ruminicola AR29

Keith Gregg, Brett G. Kennedy† and Athol V. Klieve‡

Author for correspondence: Athol V. Klieve. Tel: +61 7 362 9483. Fax: +61 7 892 5374.

Institute of Biotechnology, University of New England, Armidale, NSW, Australia

Phage φAR29 was shown to exist as a prophage integrated into the chromosome of Prevotella ruminicola AR29. By DNA hybridization studies, the point of integrative recombination on the phage genome (attP) was located on a 4.5 kb EcoRV fragment. After preliminary mapping with restriction endonucleases, a 2.8 kb EcoRV/HindIII fragment was isolated, cloned in Escherichia coli and sequenced. DNA hybridization localized the attP site to the vicinity of an internal Dral site. Sequence analysis showed the presence of several direct and inverted repeats around the attP site, with consensus core sequences similar to the integrase binding sites of phage λ. Two open reading frames are present adjacent to attP (ORF1 and ORF2). The predicted polypeptide product of ORF1 has a region of structural similarity to known integrases. Although the predicted product of ORF2 shows at best weak homology with known excisionases, no other ORFs occur in the sequence upstream from ORF1, leaving ORF2 as the most likely candidate for this role. However, if ORF2 does represent an xis gene, then this putative integration module would possess a notable difference from that of other temperate phages in the inversion of the positions of int and xis relative to attP. The proposed φAR29 integration module is being used to develop phage-based integrative vector systems for the genetic manipulation of rumen bacteria.

Keywords: Prevotella ruminicola, temperate bacteriophage, integration, DNA sequence, rumen

INTRODUCTION

The potential for manipulating rumen fermentation, especially using recombinant DNA technology, has received considerable attention in recent years (Smith & Hespell, 1983; Teather, 1985; Gregg et al., 1987; Mackie & White, 1990). To ensure that newly introduced DNA remains stable within the altered bacterium, and to reduce the likelihood of transfer to other organisms, it must be incorporated into the bacterial chromosome. For reasons of ecological safety, it is possible that the release of genetically altered bacteria into the rumen will only be permitted if such precautions are taken.

Possible means of incorporating DNA permanently into the chromosome include homologous recombination, the use of transposons, or the integration system of temperate bacteriophage. Bacteriophage integration mechanisms have recently been used to develop integrative vectors for Mycobacterium spp., Staphylococcus aureus, and Streptomyces spp. (Kuhstoss et al., 1991; Lee et al., 1991a, b) and it may be possible to develop equivalent systems for rumen bacteria.

The temperate bacteriophage φAR29 was identified after mitomycin C induction from Prevotella ruminicola AR29 (Klieve et al., 1989), and has been selected as a likely candidate for use in the integration of novel DNA into the P. ruminicola genome. P. ruminicola is an important fibrolytic rumen bacterium that is capable of degrading hemicellulose, and has been nominated as a species that might be modified to improve ruminal digestion, by the incorporation of genes for additional or enhanced fibrolytic enzymes (Russell & Wilson, 1988; Woods et al., 1989; Klieve et al., 1991). In this paper we describe the
identification, cloning and DNA sequence analysis of the site at which recombination occurs during integration of phage $\phi$AR29.

**METHODS**

**Bacteria and bacteriophages.** Rumen bacterial isolates, their culture conditions and media, have been described previously (Klieve et al., 1989). Phage $\phi$AR29 was obtained from $P$. ruminicola subsp. brevis strain AR29 by induction with mitomycin C (Klieve et al., 1989). Plasmid clones of $\phi$AR29 DNA were grown in *Escherichia coli* strain K803 (Raleigh et al., 1988), using media and culture conditions described elsewhere (Woods et al., 1989).

**DNA manipulation.** Methods for the isolation of phage and bacterial chromosomal DNA, electrophoresis, restriction endonuclease digestion, Southern blotting and hybridization, have been reported previously (Klieve et al., 1991).

**DNA cloning and sequencing.** $\phi$AR29 DNA was ligated to the phagemid pTZ19U (T4 DNA ligase, Promega), using methods described by Vercoe & Gregg (1992), and cloned in *E. coli* strain K803 by electroporation (Dower et al., 1988). The sequence of cloned $\phi$AR29 DNA was determined by the dideoxynucleotide chain-termination method, using both single- and double-stranded sequencing techniques (Vercoe & Gregg, 1992).

**Computer analysis of DNA and protein sequences.** DNA sequence analysis used computer software prepared by Dr R. W. Bottomly (formerly of CSIRO, Division of Plant Industry) and sequence comparisons were performed using the computer sequence analysis package of the Wisconsin Genetics Computer group (WGGC version 6.1; Devereux et al., 1984). The WGGC programs COMPARE and DOTPLOT were used to compare polypeptide sequences, as described elsewhere (Brown et al., 1990).

**RESULTS**

**Identification of the integration site**

*EcoRV* digests of DNA from $\phi$AR29 particles, and *P. ruminicola* AR29 genomic DNA, were electrophoresed on a 1% (w/v) agarose gel and blotted onto nylon membrane (Hybond-N; Amersham). DNA from $\phi$AR29 was labelled with digoxigenin dUTP (Boehringer-Mannheim) and hybridized to the Southern blots. The hybridization pattern from the *P. ruminicola* genome was essentially similar to that of DNA from phage particles, indicating the presence of phage DNA within the genomic DNA preparation. However, important differences between band patterns of phage particle DNA and bacterial genomic DNA suggested that $\phi$AR29 was integrated into the bacterial genome and not simply present as contaminating phage (Fig. 1a, b). The major differences in hybridization patterns were: (i) a 4.5 kb band present in the $\phi$AR29 digest was absent from digests of the bacterial genome; and (ii) a number of bands present in the *P. ruminicola* chromosomal digest were not observed in the $\phi$AR29 digest, including one very prominent band of approximately 7 kb.

The 4.5 kb fragment from $\phi$AR29 was excised from a gel, extracted and labelled as a hybridization probe. This fragment was postulated to contain the integration site, and was predicted to split into two separate fragments upon integration. Fig. 1(c) shows that the 4.5 kb fragment hybridized strongly to fragments of 7 kb and 2 kb in an *EcoRV* digest of *P. ruminicola* AR29 genomic DNA. It was concluded that the 4.5 kb fragment contained the $attP$ region of phage $\phi$AR29. This DNA fragment also hybridized weakly to a number of other bands within the AR29 genomic DNA digest.

An 8.6 kb length of DNA centred about the $attP$ region was partially mapped with restriction endonucleases, revealing a number of DraI sites in the vicinity of $attP$. To localize the position of $attP$ more accurately, adjacent restriction fragments were isolated and used as hybridization probes to *EcoRV* digests of *P. ruminicola* genomic DNA. These were a 0.8 kb DraI/DraI fragment and a 0.6 kb *EcoRV*/DraI fragment (bases 697–1509 and 1510–2139, respectively in Fig. 2). The 0.6 kb fragment hybridized most strongly to the 2 kb genomic band, and the 0.8 kb fragment bound most strongly to the 7 kb band (data not shown). However, both probes also hybridized significantly to the alternative band, indicating that the homology region containing the $attP$ site encompassed the DraI site between the 0.6 kb and 0.8 kb fragments.

Cleavage of the original 4.5 kb *EcoRV* fragment with *HindIII* produced a 2.8 kb fragment encompassing the fragments described above.

**Fig. 1.** Hybridization of phage AR29 DNA to *EcoRV* digests of DNA from (a) *P. ruminicola* AR29 and (b) phage AR29. Track (c) shows hybridization of the 4.5 kb *EcoRV* fragment of phage AR29 DNA to an *EcoRV* digest of *P. ruminicola* AR29 DNA. Bands present in *P. ruminicola* bacterial DNA but not in AR29 phage DNA are indicated (▷), as is the band present in phage but not bacterial DNA (◂).
Nucleotide sequence of the φAR29 attP region

A 2.8 kb EraRV/HindIII fragment containing the attP region of φAR29 was cloned in the vector pTZ19U, which had been linearized by digestion with HindIII and Smal, and the recombinant plasmid was named pIF. Both strands of clone pIF were sequenced and the nucleotide sequence of 2.14 kb of the pIF insert is shown in Fig. 2.

Analysis of the sequence revealed an inverted repeat between bases 1617 and 1645, i.e. 108–136 bp downstream from the DraI site, near which the recombination crossover had been concluded to occur. Several repeats of this nucleotide sequence occur, three times upstream and four times downstream of the inversion point. All upstream repeats are in the same orientation, and all downstream repeats are in the opposite orientation (Fig. 3a). There are strong similarities between this set of DNA sequences and those of the integrase binding sites around the attP region of phage λ (Weisberg & Landy, 1983). The central inverted repeat corresponds to the core sites of...
phage λDNA surrounded by peripheral repeats which have been referred to as the ‘arm sites’. The nucleotide alignment of the core and arm sites, and the consensus sequence among repeats, is shown in Fig. 3(b).

Three major open reading frames (ORFs) were identified adjacent to attP, and are shown in Fig. 2. The amino acid sequences of the proteins putatively encoded by these ORFs were compared with proteins known to be involved in the integration processes of other phages and integrative plasmids.

Analysis of ORF1

The amino acid sequence encoded by ORF1 was compared to the C-terminal 163 amino acids of the int protein from the integrative plasmid pSE211 (Brown et al., 1990) using the DOTPLOT program. The most significant region of similarity between the two proteins corresponded to the region of pSE211 int that is conserved in, and defines, the integrase family of proteins (Argos et al., 1986). The corresponding amino acid sequence of ORF1 was aligned and compared with other members of the integrase family of proteins (Fig. 4) and showed similarities of structure consistent with other members of the group, i.e. it matched two of the three strongly conserved amino acids in the sequence, and in 10 other positions, amino acids were matched to four of the eight integrases examined. This was the only significant similarity observed in ORF1 and it matched the only region of each of the other proteins that is common to the family.
**Analysis of ORF2**

When analysed by the DOPLOT program, the protein encoded by ORF2 showed similarities to ORF2 protein of plasmid pSE211. A single region of similarity was found towards the N terminus of both proteins, which coincided with the region of pSE211 ORF2 that was used to relate this protein to the excisionase family of recombinase proteins (Leong et al., 1986). The corresponding amino acid sequence of φAR29 ORF2 was aligned with five other excisionases. As with integrase proteins, excisionases generally show only scant similarities, but the match of ORF2 to other excisionases was insufficient to reach a firm conclusion on its identity. The only other ORF close to attP (ORF3) was still less similar to known excisionases. It was concluded that ORF2 may encode an excisionase, but in the absence of functional evidence, this remains hypothetical.

**Comparative genetic structure of integration regions**

The spatial arrangements of integration systems from five temperate phages and one integrative plasmid (pSE211) are compared in Fig. 5. The structure of the integration systems of bacteriophages λ, φ80 and p22, were as reported by Leong et al. (1986), those of L5 and pSE211 were compiled from data presented by Lee et al. (1991b) and Brown et al. (1990), respectively. The position and existence of the xis gene of L5 has been postulated (Lee et al., 1991b) as shown in Fig. 5.

All six systems have three major elements: the attP site, and genes for an integrase and an excisionase. The integration system of φAR29 differs from the others in that the positions of the proposed int gene and the possible xis gene are reversed relative to attP.

**DISCUSSION**

A 4·5 kb restriction fragment of φAR29 DNA was concluded to encode the integration site of the phage because it was split by integration into the bacterial genome. In addition to the primary integration site, the existence of secondary sites was deduced from the presence of faintly hybridizing bands in addition to the two major new phage bands generated by the recombination process.

Sequence analysis of the 2·8 kb HindIII/EcoRV fragment revealed elements similar to those found in the vicinity of attP in other temperate phages and integrative plasmids. Inverted repeats of DNA sequence are commonly found near the attP region (Brown et al., 1990; Kubstoss & Rao, 1991; Lee et al., 1991a; Omer & Cohen, 1986; Waldman et al., 1987). In phage λ DNA, sites for attachment of integrase are present as short sequence repeats around attP, inverted about the point where strand exchange occurs in the integration process (core sites), with additional repeats in close proximity on both sides of the inverted repeat (arm sites; Weisberg & Landy, 1983). A conserved 14 bp sequence is repeated nine times in the attP region of phage φAR29, about a central point of inversion. Similarities between this DNA structure and the integrase binding sites of phage λ, in the size, positioning, number and distance of repeat sequences from the central inverted repeat, led us to postulate that these sequences represent the integrase binding sites of phage φAR29. The binding sites are highly conserved and have a high AT content (average 68%). It is possible that outside the central five conserved bases (GTTGT), the AT content may be more important for binding than the exact sequence. The consensus sequence has been represented here in the same orientation as the ORFs. However, the consensus sequence of the phage λ integrase binding sites is the precise complement of the first four bases, CAAC. These four nucleotides are perfectly conserved in φAR29, in all except site 7 (cTTG).

These observations suggest that DNA strand exchange during integration of φAR29 is most likely to occur between the core sites, in the sequence AACAGTATA (bases 1627-1635), which is consistent with the hybridization data.

Three ORFs were identified near the integration site, two upstream and one downstream of attP. ORF1 encoded a peptide sequence with an area of structural similarity to known integrases. Homology between int proteins is severely limited, but ORF1 possesses levels of similarity expected from members of this protein family. Importantly, the only significant similarity detected was in the region that defines the family (Argos et al., 1986), and translation of ORFs 2 and 3 showed no similarity to the int proteins.

ORF2 was compared with members of the excisionase family of recombinases. A single region of ORF2 suggested similarity to excisionases when analysed by DOPLOT, but encoded a protein structure with only very weak similarity to the xis gene of pSE211. This region aligns with peptide sequences at which excisionases are
related (Leong et al., 1986), but ORF2 could not be convincingly demonstrated to encode an xis gene. This is perhaps unsurprising because, as with the integrases, the excisionases are diverse in amino acid sequence, showing only limited homology within the group, with no strictly invariant amino acids in the conserved sequences of the excisionase family (Leong et al., 1986).

The combination of elements in the vicinity of attP in φAR29 suggests the existence of an integration module of the general type present in temperate phages and integrative plasmids of aerobic bacteria. All the modules possess an int gene, an xis gene and the crossover point in attP. Although these elements are invariably present on an uninterrupted segment of DNA, the spatial arrangement of the entities is very flexible. In φAR29 the elements are arranged in a pattern similar to phage λ, except that the positions of int and the ORF suspected to encode an xis gene are reversed in relation to attP.

In summary, DNA encoding the integration site of φAR29 was isolated, and appears to contain a module composed of an int gene, an xis gene and an attP site, similar to that of other temperate phages and integrative plasmids. The elucidation of this system may allow the development of integrative vectors for the genetic manipulation of Pa. ruminicola. Current work is directed towards development of a plasmid transformation system for Pa. ruminicola AR29, to allow in vivo studies of integration processes, and the construction of integrative vectors for the stable modification of AR29.

ACKNOWLEDGEMENTS

We thank Dr Philip Vercoe for assistance with DNA sequence analysis, and Megan Harman, Annabelle Douglas-Hill and Jennifer Druitt, for technical assistance. This study was funded by the Australian Meat Research Corporation.

REFERENCES


Received 1 March 1994; accepted 4 March 1994.