Volume 17 Number 5 1989 Nucleic Acids Research

Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements

Wolf-Dieter Reiter, Peter Palm and Siobhan Yeats

Max-Planck-Institut für Biochemie, 8033 Martinsried, FRG

Received December 2, 1988; Revised and Accepted January 7, 1989

ABSTRACT

The DNA sequences were determined at the boundaries of the integrated copy of the archaebacterial genetic element SSV1. A 44 bp sequence present as a single copy on the 15.5 kb circular SSV1 DNA flanked the integrated copy as a direct DNA sequence repeat, suggesting that SSV1 integration occurred by recombination between this 44 bp SSV1 sequence and an identical sequence on the bacterial chromosome. At the left attachment site, a region encompassing the 44 bp attachment core sequence and the 31 nucleotides upstream of it displayed all characteristics expected for an arginine tRNA gene. An analysis of published attachment site sequences of other systems revealed that tRNA genes also constitute the bacterial attachment site in the case of three temperate phages and two transmissible plasmids in eubacteria, indicating a widespread occurrence of tRNA genes as integration target sites. This finding may be important for the understanding of mechanisms and evolution of site-specific recombination.

INTRODUCTION

Site-specific recombination is a common mechanism by which temperate bacteriophages and other autonomous genetic elements integrate into the bacterial chromosome. Recombination occurs within a short region of perfect homology between host chromosome and phage DNA, and in those cases studied in detail, sequence-specific recombination enzymes have been implicated in the recombination process (for reviews see refs. 1 and 2). The DNA sequences of several eubacterial integration sites are known (3-11) and it is generally assumed that these sequences do not fit into a general pattern. To investigate site-specific recombination in archaebacteria, we studied the UV-inducible virus-like particle SSV1 of the extremely thermophilic archaebacterium *Sulfolobus* strain B12 (12,13). SSV1 contains a

15.5 kb plasmid of known nucleotide sequence as its genome (P.P., manuscript in preparation). Within the cells of the carrier strain, SSV1 DNA is present both in its circular form and as a site-specifically integrated copy (12).

Here we report the nucleotide sequences of the attachment sites of the integrated SSV1 DNA and we show that the left attachment site displays all characteristics of an arginine tRNA gene. An analysis of published attachment site sequences of eubacterial genetic elements revealed that there are five other instances where tRNA genes serve as integration target sites. In three of these cases this had not been noted before.

MATERIALS AND METHODS

Materials.

The Klenow fragment of *Escherichia coli* DNA polymerase I was obtained from Pharmacia and [^{35}S]dATPαS was from Amersham. Restriction endonucleases were obtained from Pharmacia or from Boehringer Mannheim.

Cloning of the SSV1 attachment sites.

Total chromosomal DNA of *Sulfolobus* strain B12 purified as described previously (12) was digested with EcoRI and ligated into the EcoRI site of pBR325 (14). This ligation mixture was used to transform *Escherichia coli* HB 101 (15). A total number of 4000 colonies were screened by colony hybridization (16) for the presence of SSV1 sequences using nick-translated (16) SSV1 DNA as a probe. Plasmid was prepared from hybridizing clones, cleaved with EcoRI and analyzed by Southern blotting using EcoRI-cleaved *Sulfolobus* B12 DNA as a reference. Clones containing the attachment site fragments were identified by their size and by their hybridization to the largest EcoRI-fragment of SSV1 DNA which contains the integration site (12).

DNA sequence determination.

DNA sequences were determined by the enzymatic method (17) using the M13 cloning and sequencing technique (18,19). The sequencing strategy is shown in Fig. 1.

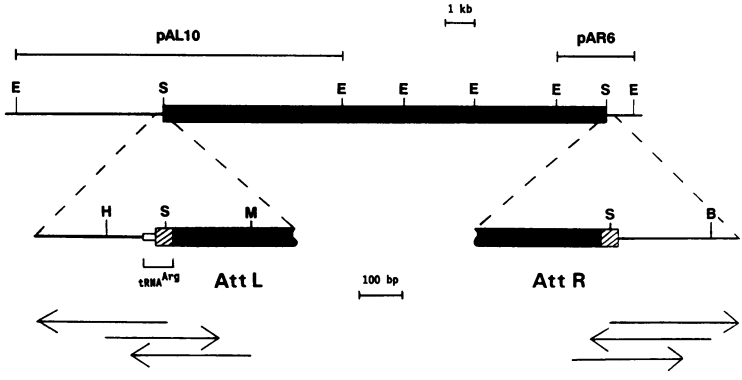

Figure 1. Physical map of the integrated SSV1 genome and sequencing strategy for the SSV1-chromosome junctions.
Filled bars represent SSV1 sequences and hatched bars represent the 44 bp attachment core sequence. The cloned EcoRI fragments containing the SSV1-chromosome junctions (pAL10 and pAR6) are indicated. Arrows represent DNA sequences determined from M13 clones. Abbreviations used for restriction sites: E: EcoRI, B: BglII, H: HpaI, M: MluI, S: SmaI.

RESULTS AND DISCUSSION

Cloning and characterization of the chromosomal attachment sites of SSV1.

The junctions between the integrated SSV1 DNA and the host chromosome were cloned into pBR325 using a shotgun approach (see the Materials and methods section). A recombinant plasmid containing the large EcoRI junction fragment (about 11.3 kb) was designated pAL10 and a recombinant plasmid containing the small EcoRI junction fragment (about 2.7 kb) was designated pAR6 (Fig. 1). Both plasmids were further characterized by restriction mapping to determine the transition sites between SSV1 DNA and host sequences. In the case of pAL10, a 6.2 kb SmaI/EcoRI restriction fragment was identical in its length and in its restriction map to a segment of circular SSV1 DNA ranging from map position 1756 (SmaI) to map position 7932 (EcoRI), numbering referring to the standard SSV1 map (EMBL database code SASSV1XX). In the case of pAR6, a 1.75 kb EcoRI/SmaI fragment was identical in its length and its restriction map to a segment of the circular SSV1 genome ranging from map position 1 (EcoRI) to map position 1756

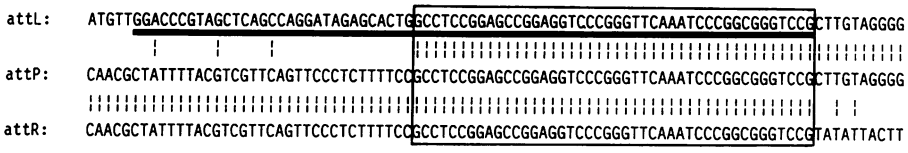

Figure 2. DNA sequence comparison between the boundaries of the integrated SSV1 DNA (*attL* and *attR*) and the corresponding region on the circular SSV1 genome (*attP*).
The 44 bp attachment core sequence is boxed and the putative arginine tRNA gene is marked by a bar.

(SmaI). The restriction maps of both pAL10 and pAR6 diverged from that of SSV1 beyond the respective SmaI sites. Based on these mapping data, the DNA sequences of the junction sites were determined by the strategy shown in Fig. 1. This sequence analysis indicated that the SSV1 genome was flanked by a 44 bp direct sequence-repeat present as a single copy on the circular SSV1 DNA (Fig. 2). This finding suggests that SSV1 integration occurred by recombination between a 44 bp sequence present on the circular SSV1 genome and an identical sequence present on the host chromosome. Since *Sulfolobus* B12 variants lacking an integrated SSV1 genome are not available, the sequence of an

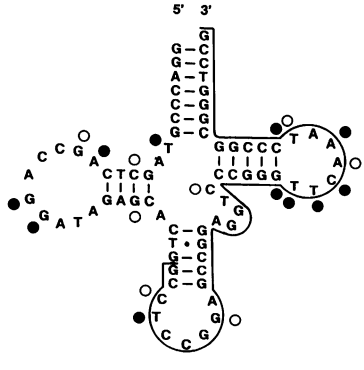

Figure 3. Cloverleaf representation of the putative tRNA gene within SSV1 *attL*.
Residues which are invariant or semi-invariant in tRNAs (20,21) are marked by closed and open circles respectively. The position of the 44 bp attachment core sequence is indicated by a line surrounding the DNA sequence.

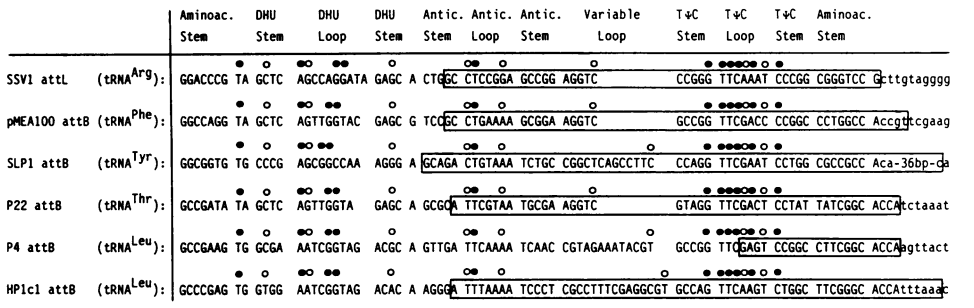

Figure 4. Comparison of bacterial attachment sites corresponding to tRNA genes.

Functionally equivalent segments of the tRNA structures have been aligned. The tRNA sequences are represented by uppercase letters; lowercase letters denote 3′ flanking regions. Nucleotides which are invariant or semi-invariant in tRNAs (20,21) are marked by closed and open circles respectively. The attachment core sequences are boxed. For references see text. The eleven 5′-terminal nucleotides of the HP1c1 *att*B sequence were taken from recent work of M. Hauser and J. Scocca (personal communications).

"empty" bacterial attachment site could not be determined. In the following, the SSV1-chromosome junction contained within pAL10 will be referred to as *att*L and the junction contained within pAR6 will be referred to as *att*R.

A putative tRNA gene is found within SSV1 *att*L.

The DNA sequence at the left border of the integrated SSV1 genome encompassing the 44 bp attachment core sequence and the 31 nucleotides upstream of it could be folded into a cloverleaf structure typical for tRNAs (Fig. 3). Within this structure all positions known to contain invariant or semi-invariant nucleotides in tRNAs conformed to the established rules (20,21), indicating that the SSV1 DNA was integrated into an arginine tRNA gene (Fig. 3).

tRNA genes as integration sites for other genetic elements.

Site-specific integration into tRNA genes has recently been described for the bacteriophages P4 and P22 which infect the closely related enteric bacteria *Escherichia coli* and *Salmonella typhimurium* respectively (8). Despite different target site specificities, the attachment core sequences of both phages extend precisely to the 3′-termini of the

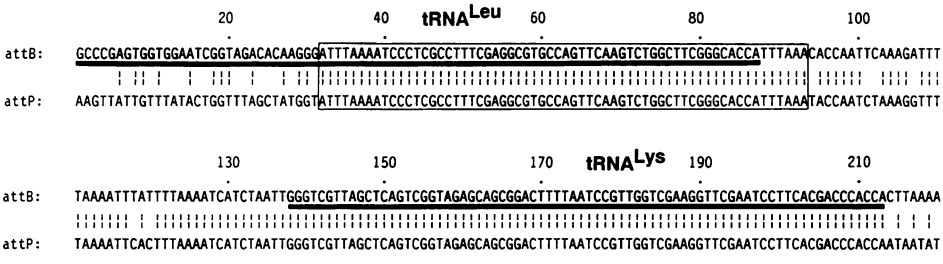

Figure 5. Comparison of the *att*P sequence of HP1c1 and the *att*B sequence of its host.

The numbering of nucleotides is arbitrary and serves reference purposes only. Putative tRNA genes are indicated by bars. The nucleotide sequences shown represent the reverse complement of the data published by Waldman *et al.* (ref. 9) with exception of the eleven 5'-terminal nucleotides of the *att*B sequence, which were taken from recent work by M. Hauser and J. Scocca (personal communications). The 62 bp sequence identity containing the recombination site is boxed. No role in site-specific recombination has been established for the 93 bp sequence identity (9) containing a gene for $tRNA^{Lys}$.

respective tRNA genes, a situation identical to that found in the case of SSV1 (Fig. 4).

DNA sequences of attachment sites have also been published for the eubacterial phages lambda (3) φ80 (4), HP1c1 (9), L54a (5) and φ11 (11) as well as for the transmissible plasmids SLP1 (6) and pMEA100 (7) and the *Escherichia coli* genetic element e14 (10), but in these cases no involvement of tRNA genes was reported. Our analysis of these sequences revealed, however, that putative tRNA genes constitute the target sites for site-specific integration in the case of HP1c1, SLP1 and pMEA100 (Fig. 4), whereas no tRNA-like sequences were found in the other five cases.

As far as the *Haemophilus influenzae* phage HP1c1 is concerned, its attachment site consists of two neighbouring regions of 62 bp and 93 bp which are identical in phage and host bacterium (ref. 9; see Fig. 5). These two regions are separated by a 27 bp sequence containing six mismatches between *att*P and *att*B (Fig. 5). We discovered that the 62 bp sequence identity which harbours the recombination site (9), represents the 56 3'-terminal nucleotides of a putative leucine tRNA gene plus six nucleotides immediately 3' to this gene (Fig. 4 and

Fig. 5). Within the 93 bp sequence identity, we detected a complete gene for a $tRNA^{Lys}$ which ends precisely at the position up to which the bacterial and the phage sequences are identical (Fig. 5). The complete conservation of this gene between bacterium and phage suggests that it may also play a role in the site-specific integration of HP1c1.

In the case of the *Streptomyces coelicolor* transmissible plasmid SLP1 (6), site-specific integration occurs within a putative tyrosine tRNA gene, with the attachment core sequence extending from the 5'-terminus of the anticodon stem to a position 39 nucleotides downstream of the gene (Fig. 4).

The integration site of the *Nocardia mediterranei* transmissible plasmid pMEA100 (7) is a putative gene for $tRNA^{Phe}$ and the position of the attachment core sequence within this gene is very similar to that found for the attachment sites of SSV1, P22 and HP1c1 (Fig. 4).

Evolutionary implications.

The frequent involvement of tRNA genes in site-specific integration and the surprising similarity in the position of the attachment core sequence within these tRNA genes could be due to convergent evolution, to a high degree of evolutionary conservation, or to horizontal gene transfer. As far as the latter possibility is concerned, genetic elements capable of integration into tRNA genes may easily spread among heterogeneous bacterial populations due to the high conservation of tRNA structure. The regions of dyad symmetry characteristic for all tRNA genes could serve as binding sites for enzymes involved in recombination, as in the case of lambda integrase (22) and the Cre (23) and Flp (24) resolvases. Intriguingly, the integrases of P4 and P22, which catalyse phage insertion into tRNA genes, appear to be evolutionarily related to several recombinases with other target site specificities (25), and it is therefore tempting to speculate that tRNA genes may have served as recognition sequences for a primordial recombination enzyme.

ACKNOWLEDGEMENTS

We thank W. Zillig for support of this work and J. Scocca for the communication of unpublished data. Thanks are also due

to B. Grampp and G. Simon for excellent technical assistance and to our colleagues for critical comments on the manuscript.

REFERENCES

1. Campbell, A. (1983) In Shapiro, J. (ed), Mobile Genetic Elements, Academic Press, New York, pp. 66-103.
2. Weisberg, R.A. and Landy, A. (1983) In Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (eds), Lambda II, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 211-250.
3. Landy, A. and Ross, W. (1977) Science **197**, 1147-1160.
4. Leong, J.M., Nunes-Düby, S., Lesser, C.F., Youderian, P., Susskind, M.M. and Landy, A. (1985) J. Biol. Chem. **260**, 4468-4477.
5. Lee, C.Y. and Iandolo, J.J. (1986) Proc. Natl. Acad. Sci. USA **83**, 5474-5478.
6. Omer, C.A. and Cohen, S.N. (1986) J. Bacteriol. **166**, 999-1006.
7. Madon, J., Moretti, P. and Hütter, R. (1987) Mol. Gen. Genet. **209**, 257-264.
8. Pierson III, L.S. and Kahn, M.L. (1987) J. Mol. Biol. **196**, 487-496.
9. Waldman, A.S., Goodman, S.D. and Scocca, J.J. (1987) J. Bacteriol. **169**, 238-246.
10. Brody, H. and Hill, C.W. (1988) J. Bacteriol. **170**, 2040-2044.
11. Lee, C.Y. and Iandolo, J.J. (1988) J. Bact. **170**, 2409-2411.
12. Yeats, S., McWilliam, P. and Zillig, W. (1982) EMBO J. **1**, 1035-1038.
13. Martin, A., Yeats, S., Janekovic, D., Reiter, W.-D., Aicher, W. and Zillig, W. (1984) EMBO J. **3**, 2165-2168.
14. Bolivar, F. (1978) Gene **4**, 121-136.
15. Boyer, H.W. and Roulland-Dussoix, D. (1969) J. Mol. Biol. **41**, 459-472.
16. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York.
17. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
18. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. **143**, 161-178.
19. Messing, J. and Vieira, J. (1982) Gene **19**, 269-276.
20. Rich, A. and RajBhandary, U.L. (1976) Ann. Rev. Biochem. **45**, 805-860.
21. Gauss, D.H. and Sprinzl, M. (1983) Nucl. Acids Res. **11**, r1-r53.
22. Ross, W. and Landy, A. (1983) Cell **33**, 261-272.
23. Hoess, R.H., Ziese, M. and Sternberg, N. (1982) Proc. Natl. Acad. Sci. USA **79**, 3398-3402.
24. Senecoff, J.F., Bruckner, R.C. and Cox, M.M. (1985) Proc. Natl. Acad. Sci. USA **82**, 7270-7274.
25. Argos, P., Landy, A., Abremski, K., Egan, J.B., Haggard-Ljungquist, E., Hoess, R.H., Kahn, M.L., Kalionis, B., Narayana, S.V.L., Pierson III, L.S., Sternberg, N. and Leong, J.M. (1986) EMBO J. **5**, 433-440.