Characterisation of IS1126 from Porphyromonas gingivalis W83: a new member of the IS4 family of insertion sequence elements

Janette Maley and Ian S Roberts *

Department of Microbiology and Immunology, University of Leicester, P.O. Box 138, Medical Sciences Building, University Road, Leicester LE1 9HN, UK

(Received 13 June 1994, revision received 15 August 1994, accepted 23 August 1994)

Abstract The nucleotide sequence of IS1126, the only insertion sequence so far isolated from the oral pathogen Porphyromonas gingivalis, has been determined. It had a nucleotide sequence of 1338 base pair (bp) flanked by 12 bp perfect inverted repeats and generated a 5 bp target site duplication. The single major open reading frame encoded a predicted protein of 361 amino acids and molecular mass of 41 kDa. The gene encoding the transposase was subcloned into pUC18 and the transposase expressed in Escherichia coli minicells. The predicted amino acid sequence of the transposase had homology to putative transposases of IS1106 and IS1186 both of which belong to the IS5 group within the IS4 super-family of insertion elements. On the basis of this homology we propose that IS1126 should also be included in the IS5 group. Southern-blot analysis of a number of P. gingivalis strains using IS1126 as a probe revealed a unique pattern of hybridisation for each strain and the absence of IS1126 from other closely related Porphyromonas species. This should allow IS1126 to be used as a rapid epidemiological tool in studying oral infections by P. gingivalis.

Key words Porphyromonas gingivalis, Insertion sequence, IS4 family

Introduction

Bacterial insertion sequence (IS) elements are segments of DNA capable of altering gene expression and of mediating a variety of DNA rearrangements through their insertion into the host chromosome [1]. These elements represent the simplest form of transposable elements, encoding only the proteins required for the transposition process and such elements have been described for a wide range of bacterial species and are generally species-specific [1]. Bacterial transposable elements have become extremely valuable tools in molecular biology where they have been exploited to provide systems which permit insertional mutagenesis of genes and also analysis of gene regulation [2].

Porphyromonas gingivalis is a Gram-negative black-pigmented anaerobe which is believed to be one of the main causative organisms in a variety of human periodontal diseases [3–5]. At present, our ability to analyse the virulence factors of P. gingivalis has been limited by the lack of naturally
occurring plasmids and transposons in *P. gingivalis* [6,7] One approach to overcome this problem has been to exploit plasmid vectors and transposons developed for use in colonic *Bacteroides* species [8–10] During experiments involving the transfer of *Bacteroides-Escherichia coli* shuttle-vector pNJR12 into *P. gingivalis* W83, a putative IS-element, designated IS1126, was found to have transposed into pNJR12 to generate plasmid pNJR12-1 [10] IS1126 is particularly interesting since it represents the only transposable element so far identified in either *P. gingivalis* or any pathogenic Gram-negative oral anerobe In this paper we report the analysis of IS1126 and show

---

**Fig 1** Complete nucleotide sequence of the 1338 bp insertion element IS1126 and deduced amino acid sequence of ORF1 The 12 bp terminal inverted repeats are shown in bold and unique restriction endonuclease cleavage sites are indicated The helix-turn-helix motif in ORF1 is shown in bold
that it belongs to the IS4 family of IS elements and demonstrate the suitability of IS1126 as an epidemiological tool for studying P gingivalis.

Materials and Methods

Bacterial strains, plasmids and bacteriophage
The bacterial strains used were P gingivalis W83, HG241, wpH 35, 23A-3, and 11834, P asaccharolytica NCTC 9337, and P endodontalis HG370. All were grown in BM broth [11] supplemented with 5 μg ml⁻¹ of haemin and 0.5 μg ml⁻¹ of menadione and cultured in an atmosphere of 80% nitrogen, 10% hydrogen and 10% carbon dioxide.

E. coli strain JM101 was used as a bacterial host for bacteriophages M13mp18/19 [12].

Plasmid pNJR12-1, which carries a copy of IS1126, has been described previously [10] and was used to provide DNA for cloning.

Preparation and manipulation of DNA
Total DNA was prepared from Porphyromonas strains by the method of Sato and Muira [13]. Plasmid DNA was extracted according to the method of Birnboim and Doly [14]. Restriction endonucleases were used as recommended by the manufacturers. Single-stranded phage DNA was prepared by standard methods [12]. Southern blotting procedures were conducted as previously described [10] using IS1126 as a radiolabelled probe [15].

DNA sequencing and sequence analysis
Nucleotide sequences were determined using single-stranded template DNA [16] and carried out using the Sequenase system (United States Biochemical Corp.). Nucleic acid and protein sequences were analysed using the GCG package [17].

Minicell analysis
Minicells were prepared from E. coli strain DS410 and the proteins radiolabelled with 35S-methionine as described previously [18].

Nucleotide sequence accession number
The nucleotide sequence reported in this paper has the Genbank accession number X77924.

Results and Discussion
The site of IS1126 insertion in pNJR12-1 was mapped to 171 bp upstream of the start codon [10]. Transposition of IS-elements into their target DNA often cause a duplication of the target site sequence [1]. In this instance, 5 bp direct repeats of the sequence GTAGG were found to flank the element and represent duplication of pNJR12 DNA.

The sequence of IS1126 was 1338 bp in length with 12 bp perfect inverted repeats (IR) at either end (Fig 1) with the G + C content being 47.45%, which is comparable with the expected ratio within this species (46–48%) [19]. A computer search of DNA databases revealed 58.2% identity in a 306 bp overlap with the sequence of IS1106 from Neisseria meningitidis [20] which stretched from position 698 bp to 1002 bp of IS1126 (data not shown). IS1126 possessed a single open reading frame (ORF1) (Fig 1) which spanned most of IS1126 from position 106 bp to 1191 bp and encoded a basic protein of 361 amino acids with a predicted molecular mass of 41 kDa and a theoretical isoelectric point of 10.26 (Fig 1). This arrangement is typical of most IS elements which usually contain a single gene encoding for a transposase enzyme [1]. Protein sequence database searches revealed 25.4% and 29.4% identity over 327 and 200 amino acids respectively between the protein encoded by ORF1 and the transposases of IS1186 from Bacteroides fragilis [21] and IS1106 [20] (data not shown). A common feature of a number of DNA-binding proteins is the presence of a conserved helix-turn-helix motif which is found in a number of proteins involved in DNA transposition such as Tn3 resolvase and the InsA protein of IS1 [1]. A helix-turn-helix motif was identified between residues 206–225 in the predicted amino acid sequence of the protein encoded by ORF1 (Fig 1). These results would suggest that the protein encoded by ORF1 is likely to be the transposase enzyme of IS1126.

An EcoRI-SstI fragment from pNJR12-1 containing ORF1 [10] was subcloned into pUC18 downstream of the lac promoter to generate plas-
Fig 2 SDS-PAGE analysis of proteins encoded by pJM1 and pUC18. Minicells were isolated from strain DS410 carrying pJM1 (A) or pUC18 (B) labelled with [35S]methionine and the labelled polypeptides were separated by SDS-PAGE followed by autoradiography. The numbers on the left refer to molecular mass markers in kilodaltons.

Mid pJM1 Minicell analysis of the pJM1 encoded proteins showed the expression of a non-vector encoded protein of 36 kDa (Fig 2). This is smaller than the predicted molecular mass of the transposase. However, this apparent discrepancy between predicted molecular mass and migration of the protein on SDS-PAGE has been seen with other transposases and may be a reflection of the highly basic nature of these enzymes [22].

IS1106 and IS1186 belong to the recently identified IS4 super-family of insertion sequences [23]. This family comprises more than 40 unrelated IS elements which originate from a wide variety of prokaryotic hosts including Gram-negative, Gram-positive, cyanobacteria and even archaeabacteria [23]. Members of this family all have two highly conserved domains within the transposase suggesting a common step in the transposition mechanism of these otherwise unrelated IS elements. One domain, consists of a 60 amino acid C-terminal region termed C1, whilst the second is a 29 amino acid N-terminal region termed N3 [23]. Within these two regions are two core signatures. The C1 core being Y-(2)-R-(3)-E-(6)-K of which the R and E residues are invariant and the N3 core being D-(1)-(G/A)-(Y/F) [23]. Depending upon the relative proximity of the N3 and C1 domains within the transposase it has been possible to further divide the IS4 family into the IS4 and IS5 groups. In the IS4 group the N3 and C1 domains are separated in the transposase by between 20 to 110 amino acids, whilst in the IS5 group the N3 and C1 domains are adjacent [23]. IS1106 and IS1186 have been previously assigned to the IS5 group with the N3 and C1 domains being located adjacent in the transposase [23]. The putative transposase of IS1126 exhibits homology with both of these proteins within this region (Fig 3). It has a perfect match with the N3 motif (D-(1)-G-Y) (Fig 3) and has the invariant R and E residues within the C1 signature (Fig 3). The close proximity of these domains within the putative transposase of IS1126 would support the inclusion of IS1126 into the IS5 group of the IS4 family of IS elements. In all three transposases the Y residue at the start of the C1 signature has been replaced by an S residue (Fig 3). It has been reported that chang-

<table>
<thead>
<tr>
<th>IS1106</th>
<th>184-TVYADGYSQHKLKHEHQLQDGIMRKA-213</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS1186</td>
<td>250-DLYLDAGYAGQEST--VEEKNWPILCEK-277</td>
</tr>
<tr>
<td>IS1126</td>
<td>258-AVLADGYSQHKLKHEHQLQDGIMRKA-287</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IS1106</th>
<th>221-ETQTERNYLSKTRYVQSFQTLHKPRYARAAYFGLIKVSAQSHKAMCLN-274</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS1186</td>
<td>286-EQKLENRKSKRCLVEHVFPEEQSMHGLIERTIGIVRAKANMTNLNYN-337</td>
</tr>
<tr>
<td>IS1126</td>
<td>294-EQQGQIHQSRPISTLIERTFPGDIWPHGCRYGRGLAKTHQINLENIFAN-347</td>
</tr>
</tbody>
</table>

Fig 3 The N3 (A) and C1 (B) conserved regions of IS1186, IS1106 and IS1126. The underlined amino acids represent the core regions. Asterisks correspond to perfect matches, whilst dots denote conservative amino acid changes. The positions of the first and last residues within each protein are shown by the numbers on the left and right.
ing this Y residue to an S by site-directed mutagenesis in transposase of IS10R, an archetypal member of the IS4 family, reduces transposition to undetectable levels and that a Y residue at this point in the transposase may be essential for the function of the IS4 family of transposases [24]. Clearly, on the basis of the results presented here this would appear not to be the case for IS1106, IS1126 or IS1186 and casts some doubt on applicability of this prediction to all members of the IS4 family.

In addition, it is possible to discriminate between the IS4 and IS5 groups on the basis of the nucleotide sequence of the IR. Typically the external trinucleotides of the IR of IS5 group are GAG or GGC, in contrast to that of the IS4 group which are typically CAT [23]. Since the external trinucleotide of IS1126 is GAG (Fig 1) this would support the assignment of IS1126 to the IS5 group of the IS4 family.

IS-elements are usually species-specific and, because of their mobile nature, can be located at different sites in the chromosome with variable copy number [1,2,25]. Southern blot analysis of chromosomal DNA from five P. gingivalis strains using IS1126 as a radiolabelled probe revealed a unique pattern of hybridisation for each strain (Fig 4), reflecting multiple copies of IS1126 on the chromosome. Since there are no BamHI sites within IS1126 (Fig 1) the number of fragments of BamHI digested chromosomal DNA which hybridised to the radiolabelled probe is the minimum number of copies of IS1126. The probe hybridised to at least 6 fragments in strain HG421 and 12 fragments in strain W83 (Fig 4). This number may be an underestimate of the number of copies of IS1126 since some large BamHI fragments may contain more than one copy of IS1126. This range of copy number of IS1126 of 6 to 12 is comparable to the number of copies of IS1 in E. coli which range from 6 to 17 depending on the strain [1]. No detectable hybridisation was seen between IS1126 and the chromosomal DNA of P. asaccharolytica NCTC 9337 and P. endodontalis HG370 (Fig 4). This would suggest that IS1126 is species-specific and should allow IS1126 to be used as a rapid epidemiological tool for the detection of P. gingivalis from the remainder of the genus and the differentiation of P. gingivalis strains. This should circumvent problems associated with isolation and lengthy subculture of these fastidious anaerobic bacteria [11].

Strains of P. gingivalis exhibit extensive genetic heterogeneity. Analysis of 33 strains by restriction enzyme fragment length polymorphism showed 29 different groups [26] whilst multi-locus enzyme electrophoresis of 100 strains identified 78 different groups [27]. It is possible that IS1126 could have contributed to this diversity by its ability to insert at a number of sites and thereby bring about DNA rearrangements of the host chromosome.

Genetic manipulation of P. gingivalis has relied on the use of plasmid vectors and transposons constructed for colonic Bacteroides species [8–10]. Unfortunately, due to the multiple copies of IS1126 on the chromosome of all of the P. gingivalis strains examined it is unlikely that IS1126 could be used for random insertional mutagenesis in P. gingivalis. However, its absence in the two other members of the genus may allow...
it to be exploited as a starting point for the development of transposons for insertional mutagenesis in the oral pathogens \textit{P. asaccharolytica} and \textit{P. endodontalis}. These studies are currently underway in our laboratory.

Acknowledgements

This work was supported by a grant from the Medical Research Council of the United Kingdom to I S R I S R is a Lister Institute Research Fellow.

References