Interruption of the *Streptococcus gordonii* M5 *sspA/sspB* intergenic region by an insertion sequence related to IS1167 of *Streptococcus pneumoniae*

Donald R. Demuth,¹ Yang Duan,¹ Howard F. Jenkinson,² Rod McNab,² Sucheol Gil³ and Richard J. Lamont³

Author for correspondence: Donald R. Demuth. Tel: +1 215 898 2125. Fax: +1 215 898 3695.
e-mail: demuth@biochem.dental.upenn.edu

¹ Department of Biochemistry, University of Pennsylvania, School of Dental Medicine, Philadelphia, Pennsylvania 19104-6002, USA

² Department of Oral Biology and Oral Pathology, University of Otago, Dunedin, New Zealand

³ Department of Oral Biology, University of Washington, Seattle, Washington 98195, USA

Streptococcus gordonii M5 and DL1 each express two related adhesin polypeptides, SspA and SspB, which are members of the antigen I/II family of streptococcal surface proteins. The *sspA* and *sspB* genes are tandemly arranged in both strains, with *sspA* residing upstream of *sspB*. The genes are separated by approximately 400 nucleotides in *S. gordonii* DL1 and 1300 nucleotides in *S. gordonii* M5. The nucleotide sequence of the *sspA/sspB* intergenic region of strain M5 is reported and the difference in length compared to *S. gordonii* DL1 shown to arise from the presence of an insertion sequence, designated ISSg1, consisting of 1197 bp. The nucleotide sequence of ISSg1 is highly homologous to IS1167 of *Streptococcus pneumoniae* and is related to a lesser extent to other members of the IS1096 family of bacterial insertion sequences. It contains a single ORF of 1026 bp, encoding a putative transposase polypeptide of 342 amino acids. The deduced transposase sequence exhibits 93% identity with the transposase polypeptides encoded by IS1167. However, the *S. gordonii* protein lacks a 90 residue central domain that is present in the IS1167 transposase and in the transposase polypeptides encoded by the related IS elements. In addition, the organization of the inverted repeats flanking the transposase gene in *S. gordonii* differs from IS1167. Extension products generated from a *sspB*-specific primer indicated that transcription initiates within the intergenic region in both *S. gordonii* strains, suggesting that *sspA* and *sspB* are independently transcribed. Transcription appears to initiate 42 bases upstream of *sspB* in *S. gordonii* DL1. In contrast, *sspB* transcription in M5 initiates at least 125 bases upstream of *sspB*, in close proximity to the terminal inverted repeat of ISSg1. These results indicate that the *sspB* promoters of *S. gordonii* M5 and DL1 are not conserved and suggest that ISSg1 sequences may play a role in directing the expression of *sspB* in *S. gordonii* M5.

Keywords: *Streptococcus gordonii*, *sspB*, insertion sequence

INTRODUCTION

The human oral cavity presents a complex environment comprising a unique range of habitats that are colonized by specific bacterial populations. For example, organisms such as the oral streptococci colonize both hard tissue (tooth enamel) and soft tissue (mucosa) surfaces in the oral cavity (Frandsen *et al.*, 1991; Nyvad & Kilian, 1990) and must adapt to the constant flow of saliva that bathes these surfaces. Indeed, the survival of

the cells may be dependent on their ability to adhere tightly to these host tissue surfaces and to evade the host defences in this open flow system. This is particularly true for strains of *Streptococcus gordonii*, which are considered to be primary colonizers of oral tissues (Nyvad & Kilian, 1990). This organism expresses a vast repertoire of adherence properties that enable avid attachment to the mineralized and/or mucosal surfaces, and to other oral bacteria and various salivary constituents (Demuth *et al.*, 1990a; Gibbons *et al.*, 1991;

Jenkinson *et al.*, 1993; Kolenbrander & London, 1993; Lamont *et al.*, 1994; Scannapieco *et al.*, 1989). Once established, these streptococcal cells then become part of the foundation upon which dental plaque develops.

The specific surface components, or adhesins, of *S. gordonii* that mediate the adherence reactions described above have been the focus of considerable interest. A variety of polypeptides that may be involved in adhesion of *S. gordonii* to host cell surface molecules, salivary glycoproteins and other oral bacteria have been described (Andersen *et al.*, 1993; Hasty *et al.*, 1992; Jenkinson & Easingwood, 1990; Lamont *et al.*, 1994; McNab & Jenkinson, 1992). Two of the best characterized adhesin polypeptides of *S. gordonii* are SspA and SspB (Demuth *et al.*, 1996). SspA and SspB (formerly Ssp5) are members of a large family of related polypeptides, the antigen I/II proteins, and are present in virtually all species of streptococci that are indigenous to the human oral cavity (Ma *et al.*, 1991). The genes encoding at least ten separate members of the antigen I/II family of proteins, including *sspA* (4740 bp) and *sspB* (4500 bp) have been cloned and sequenced. These polypeptides are comprised of two repetitive domains, an N-terminal alanine-rich region and a proline-rich repetitive region, both of which are conserved in all members of the antigen I/II family (Demuth *et al.*, 1990a; Kelly *et al.*, 1989; LaPolla *et al.*, 1991). The amino acid sequence of the C-terminal regions of these proteins is also highly conserved. In contrast, a central divergent domain consisting of approximately 400 amino acid residues is poorly conserved between SspB and the *Streptococcus mutans* antigen I/II proteins (Demuth *et al.*, 1990a, 1996). *In vitro* functional studies of SspA, SspB and antigen I/II polypeptides from *S. mutans* showed that these proteins bind to a specific salivary glycoprotein, salivary agglutinin (SAG), in a lectin-like reaction (Demuth *et al.*, 1990a, b). However, the interaction of SspB with SAG requires sialic acid residues of SAG, whereas the *S. mutans* I/II protein interacts with SAG in a sialic-acid-independent manner (Demuth *et al.*, 1990a, b). The role of sialic acid in the interaction of SspA and SAG has not yet been established. In addition, isogenic mutants which lack SspA and/or SspB or which express substantially reduced levels of these polypeptides exhibit reduced levels of adhesion to saliva-coated or SAG-coated hydroxyapatite and reduced binding to actinomycetes (Demuth *et al.*, 1996; Jenkinson *et al.*, 1993). Furthermore, SspA and SspB may mediate the binding of *Porphyromonas gingivalis* by *S. gordonii*. For example, the expression of *sspB* in a non-adherent strain of *Enterococcus faecalis* results in a recombinant organism that binds *P. gingivalis* (Lamont *et al.*, 1994). In addition, the purified SspA and SspB proteins bind to *P. gingivalis* cells (W. Brooks, D. R. Demuth, S. Gil & R. J. Lamont, unpublished). Collectively, these data suggest that the Ssp polypeptide proteins may be important for streptococcal adherence to the acquired salivary pellicle and to other oral bacteria *in vivo*.

S. gordonii strains M5 and DL1 are unique in that they

are the only oral streptococci known thus far to possess two genes, *sspA* and *sspB*, encoding antigen I/II polypeptides (Demuth *et al.*, 1996). Interestingly, the divergent central domain of SspA is more related to the *S. mutans* antigen I/II than it is to the corresponding domain of SspB, suggesting that SspA may represent a naturally occurring chimera of *S. gordonii* and *S. mutans* antigen I/II polypeptides. The *S. gordonii* DL1 *sspA* and *sspB* genes are tandemly arranged and separated by an intergenic sequence of approximately 400 bp (Demuth *et al.*, 1996). However, while the tandem arrangement of *sspA* and *sspB* in strain M5 is similar to DL1, approximately 1.3 kbp separates the *sspA* and *sspB* genes. Here we report the nucleotide sequence of the *sspA/sspB* intergenic region of *S. gordonii* M5 and show that this region is highly related to IS1167 of *Streptococcus pneumoniae*. The *S. gordonii* insertion sequence contains a single ORF as well as several other features that distinguish it from IS1167. In addition, primer extensions mapped the DL1 and M5 *sspB* promoters within the *sspA/sspB* intergenic regions. Transcriptional initiation of the *S. gordonii* DL1 *sspB* gene occurs 42 bp upstream of the coding region, and putative -10 and -35 promoter elements exist upstream of this site. In contrast, transcriptional initiation of M5 *sspB* occurs 125 bases upstream of the coding region, near the junction of the IS element. These results suggest that the *sspB* promoters of *S. gordonii* DL1 and M5 are not conserved and raise the possibility that integration of the IS element altered the *sspB* promoter of strain M5.

METHODS

Bacterial strains and growth conditions. *Streptococcus gordonii* strains M5 and DL1 were grown overnight in brain heart infusion (BHI) at 37 °C. For RNA isolation, 10 ml of an overnight streptococcal culture was inoculated into 200 ml BHI and cells were harvested after 6 h at 37 °C. *Escherichia coli* DH5 α and XL1-Blue were grown at 37 °C with aeration in Luria-Bertani (LB) medium [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl] supplemented with 100 μ g ampicillin ml⁻¹ (Sigma). Cultures of *E. coli* DH5 α F' were used to support propagation of M13 phage for sequencing experiments. These cells were plated onto LB containing 1.5% (w/v) agar, 0.1 mM IPTG (Bethesda Research Laboratories) and 50 μ g X-Gal ml⁻¹. Broth cultures were grown in LB without ampicillin, IPTG or X-Gal. *E. coli* DH5 α (pEV2) was grown in LB with ampicillin as described above and carries plasmid pEV2. This plasmid contains a 2.7 kbp insert spanning the region from nucleotide 4191 of *sspA* to nucleotide 1926 of *sspB* of *S. gordonii* DL1 that encompasses the entire *sspA/sspB* intergenic region (see below).

Cloning of the *sspA/sspB* intergenic regions of *S. gordonii* strains M5 and DL1. Our previous work suggested that clone SSP-5 contained the entire *S. gordonii* M5 *sspB* gene and approximately 5.5 kbp of sequence upstream of *sspB* (Demuth *et al.*, 1988). Subsequent sequencing of portions of this upstream region showed that it encoded approximately 4 kbp of *sspA* but lacked the 5' end of the gene (Demuth *et al.*, 1996). The region situated between *sspA* and *sspB* was not previously sequenced. From the restriction map of the *ssp* locus (Demuth *et al.*, 1988), it was determined that the intergenic region is largely contained within a 1.1 kbp *Bam*HI-*Eco*RI restriction fragment. This fragment was generated by digestion of SSP-5

DNA with *Bam*HI and *Eco*RI, purified from 0.8% low-melting-point agarose (Bethesda Research Laboratories) and ligated into *Bam*HI/*Eco*RI-cleaved M13mp19 replicative form DNA. Transformation into *E. coli* DH5 α F' was carried out as described by Hanahan (1983). The presence of the correct insert in the recombinant plaques was verified by restriction analysis.

The *sspA*/*sspB* intergenic region of strain DL1 was amplified by PCR from genomic DNA using the oligonucleotide primers 5'-ACTGAAGCGTTGACAGCTGGA-3' (a) and 5'-CGCTTCTGCGAGTGCTGCACT-3' (b). Oligonucleotide (a) anneals at the 3' end of the *S. gordonii* DL1 *sspA* gene, spanning the region corresponding to nucleotides 3058 to 3078 of the *S. gordonii* M5 *sspA* gene (GenBank accession number U40025). The complete sequence of the 3' end of *sspA* from DL1 has not yet been determined. Oligonucleotide (b) anneals to nucleotides 1920 to 1940 of the *S. gordonii* DL1 *sspB* gene (GenBank accession number U40027). Amplification reactions were carried out using 5 ng *S. gordonii* DL1 genomic DNA as the template and 3 U *Pfu* polymerase (Stratagene) with the following cycle parameters: 95 °C for 5 min (1 cycle); 94 °C for 25 s, 53 °C for 40 s, 72 °C for 6 min (32 cycles); 72 °C for 6 min (1 cycle). The 3.2 kbp amplification product was isolated from low-melting-point agarose and cleaved with *Pvu*II and *Pst*I. The resulting 2.7 kbp fragment was ligated into pBluescript, which was then cleaved with *Eco*RV and *Pst*I and transformed into *E. coli* XL-1 Blue to generate strain pEV2.

Nucleotide sequencing and analysis. Nucleotide sequencing was carried out by the dideoxy chain-termination method using Sequenase version 2 (United States Biochemical). The DNA fragments containing the entire *sspA*/*sspB* intergenic regions of *S. gordonii* M5 and DL1 were sequenced in a stepwise manner by employing oligonucleotide primers (16–20 nucleotides) that were synthesized from the nucleotide sequence determined with the previous primer. Thus, the same template was used for all sequencing reactions. The initial primers utilized were the commercially available M13 –40 primer for the M5-derived fragment and the T7 primer for pEV2. Both strands were sequenced in their entirety. Comparisons of the nucleotide sequences with the GenBank database and analysis of the nucleotide and deduced amino acid sequences were carried out using the Wisconsin Genetics Group series of sequence analysis programs (Devereux *et al.*, 1984).

Primer extension of *S. gordonii* M5 and DL1 RNA. Total RNA was isolated from 200 ml exponential-phase cultures of *S. gordonii* M5 and DL1 as described by Reddy *et al.* (1990). Initially, cultures were put on ice and 10 ml 200 mM Tris pH 8.0, 20 mM EDTA, 20 mM sodium azide, 20 mM aurintricarboxylic acid (Sigma) was added. The cultures were centrifuged at 8000 *g* for 15 min and the cell pellet was suspended in 5 ml STET buffer [8% sucrose, 5% Triton X-100, 5 mM EDTA, 50 mM Tris pH 7.0, 10 mM vanadyl ribonucleoside complex, (VRC)]. Cell suspensions were extracted with equal volumes of phenol and chloroform, and nucleic acids were precipitated with ethanol. The nucleic acid pellet was suspended in 2 ml distilled water (dH₂O) containing 10 mM VRC and extracted and precipitated as above. The resulting pellet was suspended in 2 ml dH₂O containing 1 g CsCl, layered onto a 2.5 ml cushion of 5.7 M CsCl, 0.1 M EDTA and centrifuged at 105 000 *g* for 18 h. The RNA pellets were then suspended in 1 ml dH₂O and precipitated several times with ethanol. Annealing reactions were carried out with 20–40 μ g RNA and 0.1 μ g end-labelled primer in 10 mM Tris pH 7.5, 30 mM NaCl, 80% (v/v) formamide as described by Sambrook *et al.* (1989). The oligonucleotide primer 5'-

CTACCACAACCGTACT-3' (c) was used for annealing reactions with strain M5 RNA; the oligonucleotide primer 5'-TAACTGCTACATTTGC-3' (d) was used for strain DL1 RNA samples. Primer (c) specifically anneals to a region situated 110–126 bp downstream of the M5 *sspB* gene start codon and does not hybridize with *sspA* since this sequence differs in 11 of 16 nucleotide positions between the M5 *sspA* and *sspB* genes. Primer (d) anneals to a region situated 139 to 154 bp downstream of the DL1 *sspB* gene start codon. After precipitation of the annealing mixture with ethanol, the dried RNA pellet was suspended in 100 μ l 50 mM Tris pH 8.3, 40 mM KCl, 6 mM MgCl₂, 1 mM DTT, 0.1 mg bovine serum albumin ml⁻¹, 0.5 mM each of dATP, dGTP, dCTP and dTTP, and 200 U of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories). After 2 h incubation at 37 °C, RNaseA was added to 50 μ g ml⁻¹ and maintained at 37 °C for an additional 30 min. Samples were extracted with phenol/chloroform, precipitated with ethanol and suspended in 10 μ l dH₂O containing 2 μ l dye mix (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanole). Reaction mixtures were heated at 75 °C for 15 min and electrophoresed on a 6% (w/v) sequencing gel.

RESULTS

Comparison of the *S. gordonii* DL1 and M5 *sspA*/*sspB* intergenic region

S. gordonii M5 and DL1 have recently been shown to contain two highly related but distinct copies of the *ssp* gene, designated *sspA* and *sspB* (Demuth *et al.*, 1996). During the course of these studies, restriction mapping of the *ssp* locus in these strains suggested that *sspA* and *sspB* of M5 were separated by approximately 1.3 kbp whereas the two genes appeared to be closely organized in strain DL1. In addition, Zhou *et al.* (1995) recently reported that the *S. pneumoniae* insertion element, IS1167, was similar in sequence to a short non-coding region in our previously published *sspB* sequence (Demuth *et al.*, 1990a). Therefore, to investigate the difference between the DL1 and M5 *ssp* loci and the relationship of the M5 locus to *S. pneumoniae* sequences, the complete nucleotide sequence of the *sspA*/*sspB* intergenic region of *S. gordonii* M5 was determined (Fig. 1). This sequence was also compared to a fragment from the *S. gordonii* DL1 *ssp* locus that contained the 3' end of *sspA* (900 bp), the entire intergenic region (420 bp) and 1100 bp of *sspB* (Demuth *et al.*, 1996). The nucleotide sequences of the 3' ends of the M5 and DL1 *sspA* genes were 95% identical and this level of similarity extended 120 bp downstream of the *sspA* stop codon. This highly conserved region of the *sspA*/*sspB* intergenic sequence is indicated by the bold italicized sequence in Fig. 1. Within this 120 bp non-coding region is an inverted repeat (IR-C) that resembles a rho-independent transcriptional terminator, suggesting that transcription of *sspA* terminates within the *sspA*/*sspB* intergenic region. Further downstream, the intergenic sequences of *S. gordonii* M5 and DL1 diverge and exhibit little similarity until the coding region of *sspB*. The nucleotide sequences of the 5' ends of the M5 and DL1 *sspB* genes exhibit 85% identity. Thus, the *sspA* and *sspB* genes of strains DL1 and M5 are highly related, whereas the intergenic sequences differ significantly

Fig. 1. Nucleotide sequence of the *S. gordonii* *sspA/sspB* intergenic region. The sequence extends from the first nucleotide upstream of the *sspB* start codon (nucleotide 1) to the stop codon of *sspA* (nucleotides 1332–1334). The putative *ISSg1* transposase is transcribed in the opposite direction relative to the *sspA* and *sspB* genes. The portion of the sequence presented in bold italics (nucleotides 1215–1334) is conserved within the intergenic regions of both *S. gordonii* M5 and DL1. The remainder of the sequence is unique to strain M5. The four inverted repeat sequences, IR-A, IR-B, IR-C and IR-D are indicated by opposing arrows. The deduced amino acid sequence of the putative *ISSg1* transposase is indicated in bold. A ribosome-binding site (RBS) is shown in bold and double-underlined, and potential –10 and –35 promoter elements are shown in bold and single-underlined.

that approximately 1200 bp of the intergenic region, including the entire ORF and the inverted repeat, IR-A/A', exhibited similarity to the *S. pneumoniae* insertion element IS1167 (Zhou *et al.*, 1995) and several other members of the IS1096 family of bacterial IS elements (Zhou & Morrison, 1995; see below). This result suggests that the *ssp* locus of *S. gordonii* M5 has been interrupted by a sequence that has features resembling those of known streptococcal insertion elements, although it is not known if the putative IS

Fig. 2. Comparison of ISSg1 (a), IS1167 (b) and the *sspA/sspB* intergenic region (c) of *S. gordonii* DL1. The transcriptional orientations of the *sspA*, *sspB* and ORF (transposase) genes are shown by arrows. The inverted repeat sequences (IR-A) and direct repeat sequences that flank the insertion elements are indicated. Inverted repeats IR-B, IR-C and IR-D that are present in the *S. gordonii* sequences are not shown. The structure of IS1167 is based on that presented by Zhou *et al.* (1995). The diagrams are not drawn to scale.

element is active. We have designated this sequence ISSg1.

Structure of ISSg1 and comparison with IS1167

The overall structures of the putative *S. gordonii* IS sequence and IS1167 are shown in Fig. 2(a) and (b), respectively. The ISSg1 element consists of 1197 bp whereas IS1167 is 1435 bp in length and the arrangement of the inverted repeats differs in the IS1167 and ISSg1 sequences. IS1167 is flanked by a 24 bp terminal imperfect inverted repeat that is similar to IR-A/A' and a direct repeat of 8 bp that has no homologue in the ISSg1 sequence. In contrast, ISSg1 has a second copy of IR-A' that is present upstream of the ORF. No corresponding repeat sequence exists in IS1167 at this position. In addition, the ISSg1 element lacks the direct repeat sequences that flank IR-A and IR-A' in IS1167. The *sspA/sspB* intergenic region of *S. gordonii* DL1 is presented in Fig. 2(c). Interestingly, a sequence that resembles IR-A' is present within the conserved sequence that resides downstream of *sspA* in *S. gordonii* DL1. The remainder of the DL1 intergenic region exhibits no significant similarity to either IS1167, ISSg1 or other prokaryotic insertion sequences.

The structure of IS1167 differs from other members of the IS1096 family of IS elements since it possesses two overlapping ORFs (ORF1 and ORF2) encoding peptides of 172 and 253 amino acids, respectively (Zhou *et al.*, 1995). These genes presumably encode the transposase activity required for the insertion and/or excision of the IS element (Galas & Chandler, 1989). The ISSg1 sequence encodes a single putative transposase poly-

peptide, and thus is more similar to the structure of IS1096 and its related sequences. However, it lacks a region of approximately 100 amino acid residues that is present in both the transposase sequences of the IS1096-like IS elements and in ORF2 of IS1167 (see below). Although the organization of the transposase ORF resembles IS1096, the deduced amino acid sequence of the ISSg1 transposase is most similar to the IS1167 ORFs (Fig. 3). Indeed, the ISSg1 ORF is almost identical to portions of the IS1167 ORFs. Significantly lower levels of sequence identity (from 30 to 45 %) exist between the ISSg1 transposase and the transposase polypeptides encoded by IS1181, IS1165, IS1251 and IS1001 (not shown). The N-terminal 156 residues of the ISSg1 ORF sequence are 92 % identical to ORF1 of IS1167 (Fig. 3a) and residues 188–342 of the putative ISSg1 transposase are 92 % identical to residues 103–254 of ORF2 (Fig. 3b). However, the C-terminus of ORF1 and 99 residues at the N-terminus of ORF2 have no counterpart within the ISSg1 deduced polypeptide sequence (see Fig. 3). This suggests that a portion of the IS1167 transposase sequences have been deleted in ISSg1.

Mapping of the *S. gordonii* M5 and DL1 *sspB* promoters

The existence of a sequence that resembles a rho-independent transcriptional terminator downstream of *sspA* suggested that the *S. gordonii* M5 and DL1 *sspB* genes are not transcribed as part of an operon with *sspA* but are transcribed from their own promoters. Presumably, the *sspB* promoter resides within the *sspA/sspB* intergenic region. Therefore, we were interested in determining if the presence of ISSg1 affected

Fig. 3. Comparison of the deduced amino acid sequence of the putative ISSg1 transposase with ORF1 (a) and ORF2 (b) of IS1167. The ISSg1 transposase is the top sequence in each panel. Sequences were aligned using the BESTFIT program of Devereux *et al.* (1984). Identical residues are indicated with colons. Numbering is based on the deduced amino acid sequence of the ISSg1 polypeptide.

the structure of the putative *sspB* promoter in this strain. To accomplish this, primer extension reactions were carried out utilizing primers that were specific for the *sspB* gene of *S. gordonii* M5 or DL1. As shown in Fig. 4, a single primer extension product was obtained for each strain, suggesting that a single promoter directs the expression of *sspB*. The size of the extension product obtained from DL1 RNA predicts that transcription initiates 42 bases upstream of the *sspB* translational initiation codon (Fig. 5a). Sequences that may represent potential -10 and -35 promoter elements exist upstream of the putative initiation site. In contrast, the

Fig. 4. Primer extensions of *S. gordonii* DL1 and M5 *sspB* genes. Extension reactions are shown in lanes designated P. Two concentrations of extension product are shown for strain M5. The predicted transcriptional initiation sites are indicated by the arrows in the expanded sequences shown on the left. Sequences (lanes A, C, G and T) were derived from the same primer utilized for the primer extensions. The numbering of the expanded sequences is consistent with those shown in Fig. 5, but the sequences shown are the reverse complements of those in Fig. 5. Smaller bands in the primer extension lanes presumably represent products of premature termination by the reverse transcriptase.

extension product obtained with strain M5 suggests that transcription initiates 125 bases upstream of *sspB* (Fig. 5b). However, no -10, -35 or other promoter consensus sequences were identified at the appropriate distances upstream from this site. Instead, two inverted repeat sequences exist upstream of the putative transcriptional initiation site: IR-B, which is of unknown origin and function, and IR-A, which represents an ISSg1 terminal inverted repeat sequence and defines the

Fig. 5. Comparison of the *S. gordonii* DL1 and M5 *sspB* promoter regions. Transcriptional initiation sites predicted from primer extension reactions are indicated by the bold arrows. Inverted repeat sequences are labelled IR-A, IR-B and IR-C and indicated by opposing arrows. The putative *sspB* ribosome-binding sites (RBS) are shown in bold and double-underlined, and -10 and -35 promoter elements are shown in bold and single-underlined. IR-C is present in the intergenic regions of *S. gordonii* DL1 and M5 and resembles a rho-independent transcriptional terminator.

junction with the ISSg1 element. Since it is possible that these inverted repeats form secondary structures that may cause the reverse transcriptase to pause, primer extension reactions were also conducted at 50 °C to destabilize the potential stem-loop that may form at IR-B. The extension product obtained under elevated reaction temperatures was identical to that observed in Fig. 4 (not shown). Taken together, these results suggest that the sequences that direct *sspB* expression in M5 reside near or within the ISSg1 element and are clearly different from the *sspB* promoter sequences of strain DL1.

DISCUSSION

S. gordonii M5 and DL1 are the only strains of oral streptococci that have been shown to possess two distinct copies of the gene encoding antigen I/II (Demuth *et al.*, 1996). However, the structure of the *ssp* locus differs in these strains. The *sspAB* genes of DL1 are arranged in tandem and separated by approximately 400 non-coding nucleotides, whereas *sspAB* of M5 are separated by over 1300 bp. The nucleotide sequence of the M5 *sspAB* intergenic region presented here shows

that this difference arises from the presence of an insertion sequence that is related to, but distinct from the pneumococcal insertion sequence IS1167. This is consistent with the findings of Zhou *et al.* (1995) who reported that IS1167 was similar to a short non-coding region in our previously published nucleotide sequence of the M5 *sspB* gene (Demuth *et al.*, 1990a). The ISSg1 element has been stable during 10 years of laboratory culture and results from Southern blotting experiments suggest that it is present in a single copy in the M5 genome. In contrast, IS1167 is present in multicopy (from 3 to 12 copies per cell) in strains of *S. pneumoniae* examined by Zhou *et al.* (1995) but, like ISSg1, appears to be very stable under laboratory culturing conditions.

ISSg1 encodes a single polypeptide of 342 amino acids with a predicted isoelectric point of 10.67, consistent with the highly basic nature of transposases encoded by other bacterial insertion sequences (Bennett, 1991). The putative transposase polypeptide is highly similar (85–95% sequence identity) to regions of both ORF1 and ORF2 of IS1167 and exhibits 30–40% sequence identity with transposases encoded by IS1165 (Johansen & Kibenich, 1992), IS1181 (Derbise *et al.*, 1994), IS1251 (Handwerger *et al.*, 1995) and IS1001 (van der Zee *et al.*,

1993). However, the ISSg1 protein lacks a central domain that is present in IS1167 (encompassing the C-terminus of ORF1 and the N-terminal residues of ORF2 in IS1167) and in all of the transposases of the IS1096 family of IS elements. Thus, ISSg1 may represent a new member of the IS1096 family (Cirillo *et al.*, 1991; Zhou & Morrison, 1995) of IS elements that may have arisen from IS1167 by deletion of this central coding domain. It is not known if IS1167 expresses separate proteins from ORF1 and ORF2 or if translational readthrough, occurring from a -1 frameshift, generates a single transposase polypeptide (Chandler & Fayet, 1993). Whether the lack of these sequences in the ISSg1 transposase influences its activity remains to be determined.

The organization of the inverted repeat around ISSg1 also differs from IS1167, suggesting that rearrangements in the *sspA/sspB* intergenic region may have occurred after the initial insertion of ISSg1. This is consistent with the ability of some IS elements (e.g. IS257) and transposons (e.g. Tn5404) to actively promote deletions, rearrangements and integration of foreign sequences (Byrne *et al.*, 1991; Derbise *et al.*, 1995; Matthews & Stewart, 1988; Wada *et al.*, 1991). Indeed, rearrangements within this region might also explain the lack of similarity within the *sspB* promoters of M5 and DL1 (see below).

S. gordonii DL1 does not possess the entire ISSg1 sequence. However, a 120 bp portion of the M5 intergenic region situated immediately downstream of *sspA* is conserved in DL1. This conserved region extends into the coding region of the putative ISSg1 transposase gene and contains an inverted repeat similar to IR-A', raising the possibility that *S. gordonii* DL1 might have previously carried a copy of the ISSg1 element. The second inverted repeat (IR-C) downstream of *sspA* in both M5 and DL1 resembles a rho-independent transcriptional terminator (Platt, 1986), suggesting that *sspA* and *sspB* are independently transcribed. Several other lines of evidence support this observation: (a) the primer extension results show that *sspB* transcription initiates within the *sspA/sspB* intergenic region in both strains, (b) expression of the M5 *sspB* gene in *E. coli* and *Enterococcus faecalis* does not require sequences that reside within *sspA* or the ISSg1 transposase gene (Demuth *et al.*, 1989), and (c) Northern blots of *S. gordonii* DL1 RNA showed that *sspB* is encoded by a message of approximately 5 kb (H. F. Jenkinson, unpublished data). Therefore, we conclude that *sspA* and *sspB* are independently transcribed in *S. gordonii*.

The *sspA* and *sspB* coding sequences exhibit substantial sequence identity extending into the intergenic region downstream of *sspA*, suggesting that the sequence and overall organization of the *ssp* locus was conserved in DL1 and M5 until the integration of the ISSg1 element into the *S. gordonii* M5 genome. Thus, it was surprising to find that the M5 and DL1 *sspB* promoters differ. Transcription initiates at least 125 nucleotides upstream of *sspB* in M5. However, it is possible that the inverted repeat at -125 (IR-B) may block extension by the

reverse transcriptase. Thus, the true initiation site and promoter of *sspB* may reside even farther upstream, within the IS sequence. Outwardly facing promoters within IS elements have been shown to direct expression of *katA* in *Bordetella pertussis* (DeShazer *et al.*, 1994) and *cfiA* in *Bacteroides fragilis* (Podglajen *et al.*, 1994). In addition, sequences within the terminal inverted repeat of IS905 (from *Lactococcus lactis*) are potentially promoter-active (Dodd *et al.*, 1994). In contrast, transcription of DL1 *sspB* initiates at nucleotide -42 . Putative -10 and -35 sequences reside upstream of the initiation site, but the -35 sequence does not closely match the *E. coli* consensus sequence. One possible explanation is that *sspB* expression in DL1 does not require a -35 sequence but is directed from an extended -10 site. Indeed, Sabelnikov *et al.* (1995) suggested that extended -10 sequences frequently direct gene expression in *S. pneumoniae*. However, further studies will be required to precisely map the *sspB* promoter sequences and to evaluate the role of ISSg1 in directing and/or regulating the expression of *sspB* in *S. gordonii* M5.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health grants DE08239, RR00040 and DE09439. H. F. J. gratefully acknowledges support from the Health Research Council of New Zealand.

REFERENCES

- Andersen, R. N., Ganeshkumar, N. & Kolenbrander, P. E. (1993). Cloning of the *Streptococcus gordonii* PK488 gene, encoding an adhesin which mediates coaggregation with *Actinomyces naeslundii* PK606. *Infect Immun* **61**, 981–987.
- Bennett, P. M. (1991). Transposable elements and transposition in bacteria. In *Modern Microbial Genetics*, pp. 324–364. Edited by U. N. Streips & R. E. Yasbin. New York: Wiley-Liss.
- Byrne, M. E., Gillespie, M. T. & Skurray, R. A. (1991). 4',4" adenylyltransferase activity on conjugative plasmids isolated from *Staphylococcus aureus* is encoded on an integrated copy of pUB110. *Plasmid* **25**, 70–75.
- Chandler, M. & Fayet, O. (1993). Translational frameshifting in the control of transposition in bacteria. *Mol Microbiol* **7**, 497–503.
- Cirillo, J. D., Baretta, R., Bloom, B. R. & Jacobs, W. R. (1991). A novel transposon trap for mycobacteria: isolation and characterization of IS1096. *J Bacteriol* **173**, 7772–7780.
- Demuth, D. R., Davis, C. A., Corner, A. M., Lamont, R. J., Leboy, P. S. & Malamud, D. (1988). Cloning and expression of a *Streptococcus sanguis* surface antigen that interacts with a human salivary agglutinin. *Infect Immun* **56**, 2484–2490.
- Demuth, D. R., Berthold, P., Leboy, P. S., Golub, E. E., Davis, C. A. & Malamud, D. (1989). Saliva-mediated aggregation of *Enterococcus* transformed with a *Streptococcus sanguis* gene encoding the Ssp-5 surface antigen. *Infect Immun* **57**, 1470–1475.
- Demuth, D. R., Golub, E. E. & Malamud, D. (1990a). Streptococcal-host interactions: structural and functional analysis of a *Streptococcus sanguis* receptor for a human salivary glycoprotein. *J Biol Chem* **265**, 7120–7126.
- Demuth, D. R., Lammey, M. S., Huck, M., Lally, E. T. & Malamud, D. (1990b). Comparison of *Streptococcus mutans* and *Strepto-*

- coccus sanguis* receptors for human salivary agglutinin. *Microb Pathog* 9, 199–211.
- Demuth, D. R., Duan, Y., Brooks, W., Holmes, A. R., McNab, R. & Jenkinson, H. F. (1996). Tandem genes encode cell-surface polypeptides SspA and SspB which mediate adhesion of the oral bacterium *Streptococcus gordonii* to human and bacterial receptors. *Mol Microbiol* 20, 403–413.
- Derbise, A., Dyke, K. G. H. & El Solh, N. (1994). Isolation and characterization of IS1181, an insertion sequence from *Staphylococcus aureus*. *Plasmid* 31, 251–264.
- Derbise, A., Dyke, K. G. H. & El Solh, N. (1995). Rearrangements in the staphylococcal beta-lactamase-encoding plasmid, pIP1066, including a DNA inversion that generates two alternative transposons. *Mol Microbiol* 17, 769–779.
- DeShazer, D., Wood, G. E. & Friedman, R. L. (1994). Molecular characterization of catalase from *Bordetella pertussis*: identification of the *katA* promoter in an upstream insertion sequence. *Mol Microbiol* 14, 123–130.
- Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12, 387–395.
- Dodd, H. M., Horn, N. & Gasson, M. J. (1994). Characterization of IS905, a new multicopy insertion sequence identified in lactococci. *J Bacteriol* 176, 3393–3396.
- Frandsen, E. V. G., Pedrazzoli, V. & Kilian, M. (1991). Ecology of viridans streptococci in the oral cavity and pharynx. *Oral Microbiol Immunol* 6, 129–133.
- Galas, D. J. & Chandler, M. (1989). Bacterial insertion sequences. In *Mobile DNA*, pp. 109–162. Edited by D. E. Berg & M. M. Howe. Washington, DC: American Society for Microbiology.
- Gibbons, R. J., Hay, D. I. & Schlesinger, D. H. (1991). Delineation of a segment of absorbed salivary acidic proline-rich proteins which promotes adhesion of *Streptococcus gordonii* to apatitic surfaces. *Infect Immun* 59, 2948–2954.
- Hanahan, D. (1983). Studies on the transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166, 557–580.
- Handwerger, S., Skoble, J., Discotto, L. F. & Pucci, M. J. (1995). Heterogeneity of the *vanA* gene cluster in clinical isolates of enterococci from the northeastern United States. *Antimicrob Agents Chemother* 39, 362–368.
- Hasty, D. L., Ofek, I., Courtney, H. S. & Doyle, R. J. (1992). Multiple adhesins of streptococci. *Infect Immun* 60, 2147–2152.
- Jenkinson, H. F. & Easingwood, R. A. (1990). Insertional inactivation of a gene encoding a 76 kilodalton cell surface polypeptide in *Streptococcus gordonii* Challis has a pleiotropic effect on cell surface composition and properties. *Infect Immun* 58, 3689–3697.
- Jenkinson, H. F., Terry, S. D., McNab, R. & Tannock, G. W. (1993). Inactivation of the gene encoding surface protein SspA in *Streptococcus gordonii* DL1 affects cell interactions with human salivary agglutinin and oral actinomyces. *Infect Immun* 61, 3199–3208.
- Johansen, E. & Kibenich, A. (1992). Isolation and characterization of IS1165, an insertion sequence of *Leuconostoc mesenteroides* subsp. *cremoris* and other lactic acid bacteria. *Plasmid* 27, 200–206.
- Kelly, C., Evans, P., Bergmeier, L., Lee, S. F., Progulske-Fox, A., Harris, A. C., Aitken, A., Bleiweis, A. S. & Lehner, T. (1989). Sequence analysis of the cloned streptococcal cell surface antigen I/II. *FEBS Lett* 258, 127–132.
- Kolenbrander, P. E. & London, J. (1993). Adhere today, here tomorrow: oral bacterial adherence. *J Bacteriol* 175, 3247–3252.
- Lamont, R. J., Gil, S., Demuth, D. R., Malamud, D. & Rosan, B. (1994). Molecules of *Streptococcus gordonii* that bind to *Porphyromonas gingivalis*. *Microbiology* 140, 867–872.
- LaPolla, R. J., Haron, J. A., Kelly, C. G., Taylor, W. R., Bohart, C., Hendricks, M., Pyati, J., Graff, R. T., Ma, J. K.-C. & Lehner, T. (1991). Sequence and structural analysis of surface protein antigen I/II (SpA) of *Streptococcus sobrinus*. *Infect Immun* 59, 2677–2685.
- Ma, J. K.-C., Kelly, C. G., Munro, G., Whiley, R. A. & Lehner, T. (1991). Conservation of the gene encoding antigen I/II in oral streptococci. *Infect Immun* 59, 2686–2694.
- McNab, R. & Jenkinson, H. F. (1992). Gene disruption identifies a 290 kDa cell-surface polypeptide conferring hydrophobicity and coaggregation properties in *Streptococcus gordonii*. *Mol Microbiol* 6, 2939–2949.
- Matthews, P. R. & Stewart, P. R. (1988). Amplification of a section of chromosomal DNA in methicillin-resistant *Staphylococcus aureus* following growth in high concentrations of methicillin. *J Gen Microbiol* 134, 1455–1464.
- Nyvad, B. & Kilian, M. (1990). Comparison of the initial streptococcal microflora on dental enamel in caries-active and caries-inactive individuals. *Caries Res* 24, 267–272.
- Platt, T. (1986). Transcriptional termination and the regulation of gene expression. *Annu Rev Biochem* 55, 339–372.
- Podglajen, I., Brueil, J. & Collatz, E. (1994). Insertion of a novel DNA sequence, IS1186, upstream of the silent carbapenemase gene *cfiA*, promotes expression of carbapenem resistance in clinical isolates of *Bacteroides fragilis*. *Mol Microbiol* 12, 105–114.
- Reddy, K. J., Webb, R. & Sherman, L. A. (1990). Bacterial RNA isolation with a one hour centrifugation in a table top ultracentrifuge. *BioTechniques* 8, 250–251.
- Sabelnikov, A. G., Greenberg, B. & Lacks, S. A. (1995). An extended –10 promoter alone directs transcription of the DpnII operon of *Streptococcus pneumoniae*. *J Mol Biol* 250, 144–155.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Scannapieco, F. A., Bergey, E. J., Reddy, M. S. & Levine, M. J. (1989). Characterization of salivary α -amylase binding to *Streptococcus sanguis*. *Infect Immun* 57, 2853–2863.
- Wada, A., Katayama, Y., Hiramatsu, K. & Yokota, T. (1991). Southern hybridization analysis on the *mecA* deletion from methicillin resistant *Staphylococcus aureus*. *Biochem Biophys Res Commun* 176, 1319–1325.
- van der Zee, A., Agterberg, C., van Agterveld, M., Peeters, M. & Mooi, F. R. (1993). Characterization of IS1001, an insertion sequence element of *Bordetella parapertussis*. *J Bacteriol* 175, 141–147.
- Zhou, L. & Morrison, D. A. (1995). The new insertion sequence IS1167 of *Streptococcus pneumoniae* is related to IS1096 and a family of IS elements occurring widely among Gram positive bacteria. *Dev Biol Stand* 85, 77–81.
- Zhou, L., Hui, F. M. & Morrison, D. A. (1995). Characterization of IS1167, a new insertion sequence in *Streptococcus pneumoniae*. *Plasmid* 33, 127–138.

Received 19 November 1996; revised 17 February 1997; accepted 19 February 1997.