Characterization of *Streptococcus gordonii* prophage PH15: complete genome sequence and functional analysis of phage-encoded integrase and endolysin

Jan R. van der Ploeg

Institute of Oral Biology, University of Zürich, Plattenstrasse 11, 8032 Zürich, Switzerland

**INTRODUCTION**

In healthy individuals, mitis streptococci make up a large fraction of the total bacterial microbiota of the oral cavity. One of the members of the mitis streptococci, *Streptococcus gordonii*, is among the first colonizers of teeth. Compared to *Streptococcus mutans*, the principal aetiological agent of dental caries, *S. gordonii* has low cariogenic potential (Tanzer et al., 2008). On the contrary, it has been suggested that *S. gordonii* can modulate the virulence properties of *S. mutans* by interfering with the *comCDE* quorum-sensing system, which regulates biofilm formation, bacteriocin production and genetic competence (Wang & Kuramitsu, 2005). Because of its pioneer colonizing properties and because of its potential to elicit immune responses in infants and children, it has been proposed to use engineered strains of *S. gordonii* for live vaccine delivery after mucosal colonization (Fischetti et al., 1996; Lee, 2003). Ideally, such antigens would be exported or displayed on the cell surface of *S. gordonii* (Lee, 2003). Although *S. gordonii* belongs to the GRAS (generally regarded as safe) organisms, it has been implicated in causing infective endocarditis (Douglas et al., 1993). This colonization of damaged heart valves can occur when *S. gordonii* or other viridans streptococci enter the bloodstream after dental procedures.

It is becoming increasingly apparent that temperate bacteriophages play an important role in bacterial genome diversification, in particular in the dissemination of virulence factors. The availability of bacterial genome sequences has contributed considerably to this understanding. Prophages have been detected in the majority of the sequenced genomes (Canchaya et al., 2003). Some *Streptococcus pyogenes* and *Streptococcus agalactiae* strains have been found to contain multiple prophages or prophage remnants (Canchaya et al., 2003). It has been suggested that epidemic outbreaks caused by *S. pyogenes* are related to the acquisition of prophages, and hence phage-encoded virulence factors (Banks et al., 2002). The function of prophages in contributing to the virulence of *Streptococcus pneumoniae* is less evident, but it has recently been reported that adherence of *S. pneumoniae* to inert surfaces and pharyngeal cells is increased after lysogenization with phage MM1 (Loeffler & Fischetti, 2006).

*Streptococcus gordonii* OMZ1039, isolated from supragingival dental plaque, was found to harbour a prophage, PH15, whose excision could be induced by mitomycin treatment. Phage PH15 belongs to the *Siphoviridae*. The complete genome sequence of PH15 was determined. The genome was 39 136 bp in size and contained 61 ORFs. The genome of PH15 was most similar in the structural module to the temperate bacteriophages MM1 and φNIH1.1 from *Streptococcus pneumoniae* and *Streptococcus pyogenes*, respectively. In strain OMZ1039, PH15 was found to reside as a prophage in the cysteinyl-tRNA gene. A plasmid, harbouring the attP site and the integrase gene downstream of a constitutive promoter, was capable of site-specific integration into the genomes of different oral streptococcal species. The phage endolysin was purified after expression in *Escherichia coli* and found to inhibit growth of all *S. gordonii* strains tested and several different streptococcal species, including the pathogens *Streptococcus mutans*, *S. pyogenes* and *Streptococcus agalactiae*.

Abbreviation: CRISPR, clustered regularly interspaced palindromic repeats.

The GenBank/EMBL/DDBJ accession number for the complete genome sequence of *Streptococcus gordonii* prophage PH15 is FM163528.

Two supplementary tables, showing the oligonucleotides used in this study and the ORFs encoded by *S. gordonii* prophage PH15, and three supplementary figures, showing phylogenetic analysis of part of sodA from strains OMZ1038, OMZ1039 and OMZ1081, electron microscopy of phage PH15 adsorption to *S. gordonii* and *S. oralis*, and a comparison of PH15 ORFs with proteins encoded by streptococcal phages, are available with the online version of this paper.
Although the role played by oral streptococci in the ecology of the oral cavity has been studied to some extent (Kuramitsu et al., 2007), little is known about the presence and impact of bacteriophages from these organisms. The genome sequence of the virulent phage M102, which infects S. mutans, has recently been published (van der Ploeg, 2007). The sequenced genomes of three oral streptococcal strains, including S. gordonii CH1, do not appear to contain prophages (Ajdic et al., 2002; Tanzer et al., 2008; Vickerman et al., 2007; Xu et al., 2007). Temperate phages from Streptococcus mitis, an abundant inhabitant of the oral cavity, have been described (Romero et al., 2004a; Siboo et al., 2003). S. mitis phage SM1 has been found to encode two proteins, PblA and PblB, which mediate binding to platelets and affect virulence in a rat model of infective endocarditis (Mitchell et al., 2007; Siboo et al., 2003).

Here, the isolation and the genome sequence of what is believed to be the first temperate bacteriophage from a fresh isolate of S. gordonii and the functional analysis of the phage-encoded endolysin and integrase are reported.

**METHODS**

**Bacterial strains and growth conditions.** Streptococcus, Lactococcus lactis and Enterococcus faecalis strains are listed in Table 1 and were routinely grown without agitation in Todd–Hewitt broth supplemented with 0.3 % (w/v) yeast extract (THY) at 37 °C under aerobic conditions or in a mixture of 10 % CO₂ and 90 % air. When required, antibiotics were used at the following final concentrations: erythromycin, 15 μg ml⁻¹; kanamycin, 750 μg ml⁻¹. Streptococci from human supragingival plaque were isolated on mitis salivarius agar (Difco). Escherichia coli strains JM109 and BL21(DE3) were used for routine plasmid propagation and protein production, respectively, and grown in Luria–Bertani (LB) medium. When required, antibiotics were added at the following final concentrations: erythromycin, 200 μg ml⁻¹; kanamycin, 50 μg ml⁻¹.

**DNA manipulations.** Oligonucleotides used in this study are listed in Supplementary Table S1. Plasmid purification from E. coli, restriction enzyme digestion, agarose gel electrophoresis, DNA ligation and transformation of E. coli were carried out using standard methods (Ausubel et al., 1987). Chromosomal DNA was isolated by using the GenElute Bacterial Genomic DNA kit (Sigma). DNA sequencing employed dye terminators and was performed on an ABI 3730xl DNA analyser (Applied Biosystems).

For strain identification, part of the 16S rRNA gene was amplified by using PCR with the universal primers 27f and 1492r. The PCR product was sequenced using primer 519r. For sequencing of sodA, PCR was performed with primers sodAf and sodAr, and the product was sequenced with both primers.

Transformation of streptococci was done as follows. An overnight culture of the strain of interest was diluted 40-fold in fresh THY medium and incubated for 3–4 h at 37 °C. In some cases, heat-inactivated horse serum was added to a final concentration of 5 % (v/v). A suitable amount of plasmid (0.1–1.0 μg) was added to 0.5 ml of the culture, which was further incubated for 1 h at 37 °C. Appropriate dilutions were plated on THY medium containing erythromycin. Plates were incubated for 48 h in 10 % CO₂/90 % air at 37 °C.

**Phage propagation, isolation of phage DNA and sequencing strategy.** For isolation of phage DNA from OMZ1039, an overnight culture was diluted 100-fold in fresh THY medium. After 3 h of incubation at 37 °C, mitomycin was added to a final concentration of 0.1 μg ml⁻¹. The culture was incubated for another 20 h at 37 °C. To remove debris, the lysate (designated as phage lysate below) was centrifuged for 10 min at 7000 g and passed through a 0.4 μm pore-size filter. Phage DNA was isolated from the lysate by using the LambdaPrep kit (Promega). Digestion of phage DNA with EcoRV yielded 11 fragments with a total size of about 38 kb. Nine of these EcoRV restriction fragments (with a total size of about 28 kb) were cloned in the vector pSMART-Lc-Kan (Lucigen). The inserts were end-sequenced. Sequences were assembled with the program CAP3 (Huang & Madan, 1999). Direct sequencing of PCR fragments obtained with custom primers was utilized to confirm junction sequences across the ends of cloned fragments and to fill gaps.

**Sequence analysis.** The assembled sequence of PH15 was analysed for the presence of ORFs with the program Genemark.hmm (Lukashin & Borodovsky, 1998), with the codon usage of S. pneumoniae strain R6 and a minimum size of 42 bases. Further sequence analysis used the programs from the GCG package (Accelrys). BLAST sequence similarity searches were done at http://www.ncbi.nlm.nih.gov/blast/. For multiple sequence alignments, CLUSTAL W (http://www.ebi.ac.uk/clustalw/) was used. Analysis of CRISPR (clustered regularly interspaced palindromic repeats) sequences in the genome of S. gordonii was done at http://crispr.u-psud.fr/. Phylogenetic analysis of sodA sequences was performed at http://www.phylogeny.fr/, using the one-click mode (Dereeper et al., 2008).

**Construction of streptococcal integration vector.** The integrase gene and attP from PH15 were amplified by PCR using primers PH15-int1 and PH15-int2. The product was digested with SadI and EcoRI and cloned into plasmid pNE1 (Bartillon et al., 2001) to give pPH15int. A fragment harbouring a constitutive promoter (cp32) (Jensen & Hammer, 1998) was generated by annealing of two complementary primers (cp32f and cp32r) and cloned upstream of the int gene in the BamHI/Smal sites of PH15int to give pcPH15int. Subsequently, the BamHI/EcoRI fragments from pPH15int and pcPH15int were cloned into similarly digested plasmid pVA891 (Macrina et al., 1983) to make plasmids pOMZ292 and pOMZ291, respectively. A plasmid with a truncated integrase gene was constructed by digestion of pOMZ291 with BglII, filling in the ends with Klenow polymerase and religation, resulting in plasmid pOMZ308.

**Cloning, expression and activity of the lytic enzyme of PH15.** The lysA gene from PH15 was amplified by PCR using primers PH15lysF and PH15lysR and DNA from strain OMZ1039 as template. The product was digested with Ndel and Xhol and cloned into similarly digested pET28b (Novagen) to give plasmid pOMZ276. For expression of lysA, a culture of E. coli BL21(DE3) (Novagen) harbouring plasmid pOMZ276 was grown at 37 °C to OD₆₀₀ 0.8, upon which IPTG was added to 0.1 mM. Incubation of the culture at 37 °C was continued for 5 h. Cells were harvested, resuspended in 1 × Bugbuster (Novagen) and incubated with shaking for 1 h at room temperature. The extract was centrifuged for 5 min at 16,000 g and the supernatant purified by Ni-NTA agarose affinity chromatography as recommended by the manufacturer (Qiagen), using 250 mM imidazole to elute the protein. The elution buffer was exchanged with 20 mM sodium phosphate buffer, pH 7.1, by using an Amicon Ultra-4 centrifugal filter device.
Table 1. Micro-organisms used in this study and their lysis by purified endolysin from PH15

<table>
<thead>
<tr>
<th>Organism</th>
<th>Strain number</th>
<th>Alternative name, characteristics</th>
<th>Source or reference</th>
<th>Lysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. gordonii</td>
<td>OMZ1039</td>
<td>Harbours prophage PH15</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>S. gordonii</td>
<td>OMZ504</td>
<td>ATCC 12396 (SK6)</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>S. gordonii</td>
<td>OMZ505</td>
<td>ATCC 10558T (SK3)</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>S. gordonii</td>
<td>OMZ606</td>
<td>NCTC 3165 (SK51)</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>S. gordonii</td>
<td>OMZ938</td>
<td>Challis DL1*</td>
<td>H. Jenkinson†</td>
<td></td>
</tr>
<tr>
<td>S. oralis</td>
<td>OMZ8</td>
<td>ATCC 9811</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. oralis</td>
<td>OMZ307</td>
<td>SK2 (NCTC 7864)</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>S. oralis</td>
<td>OMZ607</td>
<td>SK248</td>
<td>M. Kilian§</td>
<td></td>
</tr>
<tr>
<td>S. oralis</td>
<td>OMZ1038</td>
<td>Harbours prophage PH10</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>S. oralis</td>
<td>OMZ1081</td>
<td>Harbours prophage PH5</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>S. mitis</td>
<td>OMZ508</td>
<td>SK24 (NCTC 8029)</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>S. mitis</td>
<td>OMZ991</td>
<td>DSMZ12643 (NCTC 12261)</td>
<td>DSMZ</td>
<td></td>
</tr>
<tr>
<td>S. mitis</td>
<td>OMZ1043</td>
<td>BMS</td>
<td>H. Busscher‖</td>
<td></td>
</tr>
<tr>
<td>S. mitis</td>
<td>OMZ1049</td>
<td>BA</td>
<td>H. Busscher‖</td>
<td></td>
</tr>
<tr>
<td>S. pneumoniae</td>
<td>OMZ810</td>
<td>Wu2</td>
<td>B. Diamond¶</td>
<td></td>
</tr>
<tr>
<td>S. sanguinis</td>
<td>OMZ499</td>
<td>ATCC 10556</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>S. sanguinis</td>
<td>OMZ500</td>
<td>SK36</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>S. mutans</td>
<td>OMZ381</td>
<td>P42</td>
<td>G. Tiraby#</td>
<td></td>
</tr>
<tr>
<td>S. mutans</td>
<td>OMZ918</td>
<td>UA159 (ATCC 700610)</td>
<td>ATCC</td>
<td></td>
</tr>
<tr>
<td>S. mutans</td>
<td>OMZ1001</td>
<td>Derivative of UA159</td>
<td>van der Ploeg (2005)</td>
<td></td>
</tr>
<tr>
<td>S. mutans</td>
<td>OMZ62</td>
<td>AHT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. mutans</td>
<td>OMZ125</td>
<td>LM7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. mutans</td>
<td>OMZ175</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptococcus</td>
<td>OMZ165</td>
<td>HS6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptococcus</td>
<td>OMZ176</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptococcus</td>
<td>OMZ510</td>
<td>NCTC 10713</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>S. agalactiae</td>
<td>OMZ1092</td>
<td></td>
<td>R. Zbinden**</td>
<td></td>
</tr>
<tr>
<td>S. pyogenes</td>
<td>OMZ1091</td>
<td></td>
<td>R. Zbinden**</td>
<td></td>
</tr>
<tr>
<td>Streptococcus</td>
<td>OMZ511</td>
<td>ATCC 27823</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>Streptococcus</td>
<td>OMZ512</td>
<td>ATCC 27335</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>Streptococcus</td>
<td>OMZ513</td>
<td>NCTC 8618</td>
<td>Kilian et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>S. thermophilus</td>
<td>OMZ1082</td>
<td>DSM 20617</td>
<td>M. Loesner†</td>
<td></td>
</tr>
<tr>
<td>Ent. faecalis</td>
<td>OMZ940</td>
<td>GC110</td>
<td>D. Clewell‡</td>
<td></td>
</tr>
<tr>
<td>L. lactis</td>
<td>OMZ1041</td>
<td>DSM 20729</td>
<td>DSMZ</td>
<td></td>
</tr>
<tr>
<td>L. lactis</td>
<td>OMZ1045</td>
<td>NZ9800</td>
<td>M. Kleerebezem§§</td>
<td></td>
</tr>
</tbody>
</table>

* S. gordonii Challis DL1 originates from the same ancestor as strain CH1, whose genome has been sequenced, but the strains have apparently diverged during maintenance in different laboratories (Howard Jenkinson, personal communication).
†University of Bristol.
§Århus University.
ND, Not done.
‖University of Groningen.
¶Albert Einstein College of Medicine of Yeshiva University, New York.
#Université Paul Sabatier, Toulouse.
**University of Zurich.
§§Swiss Federal Institute of Technology Zurich.
¶¶University of Michigan.
§§§NIZO Food Research.

(cut-off 10 000 Da). Cell wall lytic activity of the purified protein with different bacterial strains was evaluated as follows. One hundred microlitres of an overnight culture was mixed with 4 ml THY soft agar and poured on a THY plate. After the soft agar had hardened, 5 μl enzyme was spotted on the plate. The plate was incubated for 16 h at 37 °C in 10 % CO₂ and 90 % air and examined for clear zones indicative of growth inhibition.

Electron microscopy. Grids were prepared by applying drops of phage lysates (prepared as described above) to Formvar-coated, carbon-reinforced copper grids for about 45 s, then excess solution was removed with filter paper. Subsequently, a drop of 1 % phosphotungstic acid (PTA), pH 4.2, was applied for ~15 s and similarly removed with filter paper. The grids were allowed to dry and were examined under a Philips EM 400T electron microscope at
60 kV. Images were taken at magnifications of ×33 000, ×42 000 and ×54 500.

To visualize phages adhering to bacteria, 5 ml of exponentially grown cultures of S. gordonii OMZ1039 and Streptococcus oralis OMZ607 were incubated with 5 ml of phage lysate PH15 for 30 min at 37 °C. Glutaraldehyde was added to a final concentration of 2.5% for 30 min at room temperature. Fixed cells were harvested by centrifugation, washed three times with distilled water and resuspended to a final volume of 1 ml. Electron microscopy was as described above.

RESULTS AND DISCUSSION

Presence of temperate phages in fresh streptococcal supragingival plaque isolates

It was investigated whether cultures of fresh isolates from human supragingival plaque harboured prophages. Out of 15 isolates tested, 10 lysed following addition of mitomycin. The lysates from these 10 isolates were examined for the presence of bacteriophages by electron microscopy. In five lysates, bacteriophages were found to be present, suggesting that these isolates harboured a prophage. Analysis of the sequence of the 16S rRNA gene showed that four isolates belonged to the mitis group of streptococci; the fifth was identified as Staphylococcus epidermidis. Two of the sequences were identical and might represent the same isolate. Therefore, investigations were continued with the three streptococcal isolates, which were designated OMZ1038, OMZ1039 and OMZ1081. Since identification of mitis streptococci on the basis of 16S rRNA sequences at the species level is not reliable (Kawamura et al., 1999), part of the sodA gene was sequenced (Poyart et al., 1998). Strains OMZ1038 and OMZ1081 were both assigned as S. oralis, whereas OMZ1039 was identified as S. gordonii (Supplementary Fig. S1). In this study, the temperate phage (designated PH15) from OMZ1039 was characterized.

Phage PH15 did not form plaques with S. gordonii strains OMZ505 and OMZ938, which suggests either that the host range is narrow or that the phage is defective. Electron microscopic analysis of PH15 showed that it belongs to the Siphoviridae, having a long, presumably non-contractile tail of 231 ± 39 nm and an icosahedral head with a diameter of 59.7 ± 5.4 nm (Fig. 1). Incubation of PH15 with bacteria indicated that the phage adhered to OMZ1039 (Supplementary Fig. S2a), but not to S. oralis strain OMZ607 (Supplementary Fig. S2b), which suggests that phage adherence is specific.

Genome sequence of PH15

The complete genome sequence of PH15 was determined to be 39 136 bp in size, with a GC percentage of 40.51%, which is identical to the GC percentage of the genome sequence of S. gordonii Chassis CH1 (accession no. CP000725). A total of 61 ORFs were encoded on the phage genome. They are listed in Supplementary Table S2, and a sequence comparison of the ORFs from PH15 with proteins from other streptococcal phages is summarized in Supplementary Fig. S3. Three of the ORFs had a GTG start codon, five a TTG start codon and the rest an ATG start codon. The transcription of most ORFs was directed to the right, but six ORFs, all located near attL (Fig. 2), were oriented in the opposite direction. There were four sequences with 100% identity to the E. coli consensus promoter (ttgaca-n17-tataat). These were located just upstream of ORF5, ORF9, ORF56 and ORF59.

According to their putative functions and assuming a genetic organization similar to that in other bacteriophages from low-GC-content Gram-positive organisms (Canchaya et al., 2003), the ORFs were assigned to different functional groups, i.e. lysogeny, replication and regulation, packaging, morphogenesis and host lysis (Fig. 2).

The lysogeny module is located on the most leftward part of the genome. The module consists of ORFs, the order of which follows the conserved pattern found in Siphoviridae from low-GC-content Gram-positive bacteria (Lucchini et al., 1999). ORF1 showed similarity to numerous putative integrases from low-GC-content Gram-positive streptococci, lactococci, lactobacilli and bacilli. The highest similarity (78% identity) was found with the putative integrase from bacteriophage LambdaSa2, which has been identified as a prophage in S. agalactiae strain 2603 V/R by genome sequencing (Tettelin et al., 2002). One of the few similar proteins with demonstrated activity is the integrase
from *L. lactis* phage TPW22 (Petersen *et al.*, 1999). ORF1 appears to be preceded by its own promoter and separated from ORF2 by an inverted repeat (free energy of 14.6 kcal mol\(^{-1}\); 61.1 kJ mol\(^{-1}\)). This suggests that ORF1 is transcribed independently of the genes upstream.

ORF2 and ORF3 did not show sequence similarity to genes of known function. ORF4 was similar to cl-like repressors from lactococcal phages LC3 (Blatny *et al.*, 2003) and Tuc2009 (Kenny *et al.*, 2006). As in most streptococcal and lactococcal phage genomes, the genetic switch for lysogeny/lysis is most likely located upstream of the gene encoding the cl-like repressor. This region contained two putative operators for the cl-like repressor which overlap with the two putative promoters for cl and ORF5. ORF5 was preceded by an *E. coli* 70 promoter sequence with an extended \(-10\) region. The imperfect inverted repeat cacgaagctg, which overlaps with this promoter, might serve as a binding site for the cl-like repressor. The gene encoding the latter protein was also preceded by a putative operator sequence (cgcacgagaa). The sequences of the operators resemble those of the operator from lactococcal phage Tuc2009 (Kenny *et al.*, 2006). An additional copy of the operator (ccacaagctg) with perfect dyad symmetry was present in the 3' region of ORF8, encoding a putative Cro-like repressor. Such an additional copy of the operator has also been found in the Cro-like repressor of lactococcal phages LC3 and Tuc2009 (Blatny *et al.*, 2003; Kenny *et al.*, 2006). In contrast to most other streptococcal and lactococcal phages (Lucchini *et al.*, 1999), the Cro-like repressor was separated from the cl repressor by three ORFs. Of these, ORF5 was very small and without similarity to other genes, whereas ORF6 and ORF7 showed similarity to ORFs from *S. mitis* phage SM1 (Siboo *et al.*, 2003).

In the module that putatively encodes proteins involved in regulation and transcription, only a few of the ORFs could be assigned a function based on sequence similarity with known proteins or on the presence of conserved motifs. The packaging module consists of ORF33 and ORF34, which most likely encode the small and large subunits, respectively, of a terminase. The structural module of PH15 encodes the structural proteins of the phage in the order capsid and tail, which corresponds to the usual order in streptococcal and lactococcal phages. Overall, the ORFs from the structural module showed a similar synteny and sequence similarity to the structural proteins from *S. pyogenes* φNIH1.1 (Ikebe *et al.*, 2002) and *S. pneumoniae* MM1 (Obregón *et al.*, 2003).

The lysis modules of *Siphoviridae* generally consist of two genes, which encode a holin and an endolysin. This appears to be also the case for PH15. ORF54 is predicted to contain two transmembrane helices and was most similar to the holins from *S. mitis* phage Sk137 (Llull *et al.*, 2006) and *S. pneumoniae* phage EJ-1 (Díaz *et al.*, 1996). ORF55 has an amidase-5 domain, suggesting that it cleaves the peptidoglycan at the amide bond between N-acetylmuramic acid and L-alanine of the stem peptide. ORF55 was similar to a number of putative endolysins from *Streptococcus thermophilus* bacteriophages (Sheehan *et al.*, 1999). In about half
of these *S. thermophilus* bacteriophages, an intron has been found to be present in the endolysin gene (Foley et al., 2000), but there was no indication that this is the case for the endolysin gene of PH15 (see below). Following ORF55, a putative termination signal was present with a free energy of $-13.5$ kcal mol$^{-1}$ ($56.5$ kJ mol$^{-1}$).

The region downstream of the lysis module and flanking the *attR* site frequently encodes lysogenic conversion genes, whose expression changes the phenotype of the host strain (Canchaya et al., 2003). In PH15, this region contains six ORFs (ORF56–ORF61), whose amino acid sequences did not show sequence similarity to those of proteins of known function. However, some of the ORFs (ORF56, ORF57 and ORF60) showed similarity to ORFs encoded in the lysogenic conversion region of *Streptococcus dysgalactiae* phage 3396 (Davies et al., 2007). ORF56 and ORF59 were preceded by a consensus $\sigma^{70}$ *E. coli* promoter sequence, suggesting that these genes are expressed in the lysogenic state.

**Presence of CRISPR in *S. gordonii* Challis**

It has recently been shown that one of the defence mechanisms against bacteriophage infection involves the CRISPR system (Barrangou et al., 2007). CRISPR arrays consist of identical repeats that are separated by spacer sequences. Bacteria can acquire resistance to bacteriophage infection by the incorporation of short phage-derived sequences between the repeats (Barrangou et al., 2007). The precise mechanism of phage resistance mediated by CRISPR is unknown, but is thought to be the counterpart of RNA interference in eukaryotes. The genome from *S. gordonii* Challis CH1 (accession no. CP000725; Vickerman et al., 2009) contains a CRISPR array with repeats of 36 bp and a total of 26 spacers of 30 bp in size. The sequences of these repeats were identical to the repeats of CRISPR1 from *S. thermophilus* (Romero et al., 2007). ORF56 and ORF59 were preceded by a consensus $\sigma^{70}$ *E. coli* promoter sequence, suggesting that these genes are expressed in the lysogenic state.

**Characterization of the phage attachment site in *S. gordonii* OMZ1039**

To determine the phage and bacterial attachment sites *attP* and *attB*, genomic DNA from OMZ1039 was directly sequenced. The choice of primers was based on the assumption that the attachment sequences are in proximity to the putative phage integrase at the 5' end and of the putative lyisin at the 3' end. Thus, one primer (15att2 and 15att3) was located in the integrase gene (ORF1), the other (15att1 and 15att4) was located in ORF61. The genomic sequencing resulted in distinct sequence traces, up to a point where overlapping peaks were observed, probably due to a mixture of phage and bacterial sequences. By subtracting the known phage sequence from the mixed sequence, the bacterial sequence was deduced. Subsequently, the bacterial sequence that bordered the phage sequence was confirmed by PCR, combining primers located in the phage sequence and the bacterial sequence. The PCR products had the expected sizes and sequencing confirmed the correctness of the attachment sites.

The *attP* site could be deduced from these sequences and was 18 bp with the sequence $5'\text{-TAAGGCGGTAGAC GGATT-3'}$. The *attB* site was found in the genome of *S. gordonii* Challis NCTC 7868. Further analysis showed that the first 16 bp of the *attP* site were identical to the last 16 bp of the cysteinyl-tRNA, indicating that in OMZ1039 phage PH15 has integrated just downstream of this gene (Fig. 3a). By PCR amplification of chromosomal DNA with primers 15att2 and 15att3, it was determined whether other *S. gordonii* strains possess the same *attB* sequence. Template DNA from *S. gordonii* strains OMZ504, OMZ505 and OMZ606 gave a PCR product of the expected size, whereas template DNA from OMZ938 did not yield a product. Sequencing of the products from OMZ504 and OMZ505 confirmed that the *attB* site was identical to that in OMZ1039.

The *att* sequence from PH15 was similar at the 5' end to the *attP* sequence from *L. lactis* phage TPW22. This phage has been shown to integrate in *L. lactis* at the identical position of the cysteinyl-tRNA (Petersen et al., 1999; Williams, 2002). Due to sequence differences in the cysteinyl-tRNA gene, the attachment sites are also different. The integrase from TPW22 has 49% identity to the integrase of PH15. It thus appears that the integrase has coevolved with the cysteinyl-tRNA gene. As described above, the putative integrase from *S. agalactiae* phage LambdaSa2 shows very high sequence similarity to the PH15 integrase. Analysis of the sequence downstream of the LambdaSa2 integrase gene indicated the presence of an identical attachment site (not shown).

**Construction of an integration vector for (oral) streptococci**

To confirm the functionality of the putative integrase, the *int* gene (ORF1) and *attP* were cloned into vector pVA891 (Macrina et al., 1983), which cannot replicate in streptococci. In plasmid pOMZ292, the *int* gene is preceded by the putative promoter present in PH15. Plasmid pOMZ291 contains an additional synthetic promoter (cp32) upstream
of this putative promoter (Fig. 3b). It was examined whether the plasmids could be integrated into the genome of S. gordonii OMZ938 after introduction by transformation of naturally competent cells. Plasmid pOMZ291 gave rise to transformants (4 \times 10^6 c.f.u. \text{mg}^{-1}), whereas transformation with plasmids pVA891 or pOMZ292 did not yield colonies. This confirms the functionality of both integrase and \texttt{attP}. It also suggests that successful integration requires the presence of the synthetic promoter. The putative promoter located upstream of the integrase may not be functional in plasmid pOMZ292, since a phage-encoded transcription factor might be lacking. In addition, a plasmid with a C-terminal-truncated integrase gene (pOMZ308) could not be introduced into S. gordonii OMZ938, which shows that the integrase is absolutely required for integration. PCR and sequence analysis confirmed that pOMZ291 had inserted at \texttt{attB} (not shown). The stability of the integrated plasmid was evaluated by cultivation in THY medium without antibiotic selection and counting the number of antibiotic-resistant colonies after plating. After about 50 generations in non-selective medium, all colonies were antibiotic-resistant. After about 100 generations, six out of 50 colonies had lost antibiotic resistance. The conservation of the \texttt{attB} sequence in streptococci should in principle allow for integration of pOMZ291 in the genomes of a wide variety of oral streptococcal species. Indeed, transformation of S. mutans strains OMZ381 and OMZ918, S. sanguinis strain OMZ500 and S. mitis strain OMZ991 with plasmid pOMZ291, but not with pVA891, yielded erythromycin-resistant colonies. Combining PCR and sequence analysis of the product, it was confirmed that integration occurred at the expected position in the genome of S. mutans OMZ918. Since the \texttt{attB} sequence is present in all sequenced streptococcal genomes, it is expected that plasmid pOMZ291 will integrate in most streptococcal genomes, provided that the organism can be transformed. It may thus be a valuable tool for delivery of foreign genes in single copy. The system may also be used to introduce single-copy transcriptional fusions at an ectopic site.

**Production of active endolysin from the lysA gene of PH15**

Phage endolysins are potentially useful as antibacterial agents to treat infections because of their specificity and high activity. To explore the specificity of the PH15 endolysin, the \texttt{lysA} gene (ORF55) was cloned into the E. coli expression vector pET28b, expressed as a fusion to a C-terminal His-tag. Upon induction with IPTG, a 31 kDa protein was produced, which is somewhat smaller than the calculated size of 32.4 kDa. Addition of the protein to a
culture of S. gordonii OMZ1039 led to a more than threefold decrease in optical density within 5 min (Fig. 4). The protein also showed weak activity towards Ent. faecalis OMZ940, but not towards L. lactis OMZ1041 (Fig. 4). The protein retained activity for 6 months when stored at 4 °C or at −20 °C in 44% glycerol. The activity of the purified protein towards different bacterial species was evaluated by spot assays (Table 1). The protein inhibited growth of all tested strains of S. gordonii, S. mutans and S. sanguinis. The protein was also active with the pathogenic streptococci S. pyogenes and S. agalactiae. However, S. mitis and S. oralis strains and one S. pneumoniae strain were not inhibited. This could be caused by a difference in murein structure, since S. mitis and S. oralis contain a direct link between the stem peptides (Lys-direct type peptidoglycan), whereas S. gordonii, S. mutans and S. sanguinis contain a Lys-Ala1,3 or Lys-Ala2,3 type peptidoglycan (Schleifer & Kilpper-Bälz, 1987). The requirement for a Lys-Ala1,3 type peptidoglycan could indicate that the enzyme acts as an endopeptidase, cleaving a bond within the peptide cross-link. Growth of S. thermophilus OMZ1082 was also inhibited by LysA, which is not surprising, since the protein was most similar in sequence to endolysins from S. thermophilus bacteriophages. The enzymes from S. thermophilus have not been characterized in detail, but the enzyme from bacteriophage O1205 has also been reported to be active towards S. mutans (Sheehan et al., 1999).

In summary, phage PH15, which is believed to be the first S. gordonii bacteriophage to be identified, appears structurally related to S. pyogenes phage φNIH1.1 and S. pneumoniae phage MM1. The phage was found to harbour a putative lysogenic conversion region, which may encode virulence factors. The phage endolysin was active towards a broad range of oral streptococci, including pathogenic streptococci. Finally, the conserved nature of the phage attachment site may be exploited for the construction of streptococcal integration vectors.

ACKNOWLEDGEMENTS

I thank Verena Osterwalder and Steve Reese for excellent technical assistance, and Rudolf Gmürr for critical reading of the manuscript. This work was supported by grant 215 from the research foundation of the Swiss Dental Association SSO.

REFERENCES


Edited by: M. Kilian