J. Mol. Microbiol. Biotechnol. (2001) 3(4): 545-555.

JMMB Research Article

Studies on the Genomic Organization of Recombinant *Streptococcus gordonii* and the Development of a Novel Intergenic Integration Site for Foreign Gene Expression

Christine A. Franke[1, 2], Tové C. Bolken[1] and Dennis E. Hruby[1,2]*

[1] SIGA Research Laboratories, Inc., 4575 SW Research Way, Suite 230, Corvallis, Oregon, 97333 USA
[2] Department of Microbiology and Center for Gene Research and Biotechnology, Oregon State University, Corvallis, Oregon, USA

Abstract

The methods currently employed to produce recombinant *Streptococcus gordonii* strains for use as vaccines and/or protein expression vectors result in the insertion of foreign genes into an unknown integration site with no information on the transcriptional context or potential phenotypic consequences. Therefore, the genomic organization surrounding the insertion site of a recombinant strain of *S. gordonii* (GP1223) containing a portion of the *emm6* gene of *Streptococcus pyogenes* was determined. The nucleotide sequence of chromosomal walks in both directions from the insertion site revealed that the insert was flanked by a duplicated 3061-bp *Cla*I fragment. A consensus gram-positive promoter and a factor-independent RNA polymerase terminator sequence could be deduced in the fragment immediately upstream of the insertion site. The *Cla*I fragment also encoded open reading frames (ORFs) with high homology and parallel structural organization to the leucine biosynthesis operon of *Lactococcus lactis subsp. lactis*. Chromosomal walks downstream of the identified promoter region on the non-recombinant parental strain, GP204, yielded the sequence of two ORFs which would be normal targets of the transcription derived from this promoter. Northern analyses detected a highly expressed M6-specific transcript in recombinant strain GP1223 consistent in size with the proposed transcription unit. Transcripts analogous in length to those observed for the leucine biosynthesis operon of *L. lactis subsp. lactis* were also detected encompassing the homologous ORFs of *S. gordonii*. This information has enabled the construction of a recombinant *S. gordonii* strain in which the *emm6* gene from *S. pyogenes* was targeted to a distinct intergenic locus within the *S. gordonii* genome. This new recombination site allows for expression of foreign gene products with minimal perturbation of the genomic organization of the wild-type *S. gordonii* strain and has provided information essential for further optimization of foreign gene expression levels.

*For correspondence. Email dhruby@sgph.com; Tel. (541) 753-2000; Fax. (541) 753-9999.

© 2001 Horizon Scientific Press

Introduction

Streptococcus gordonii is a commensal bacterium of the human oral cavity. Recently, there has been a great deal of interest in engineering *S. gordonii* for use as a vaccine delivery vector. To that end, a large number of heterologous antigens have been expressed on the surface of *S. gordonii* (Pozzi *et al*., 1988; Pozzi *et al*., 1992a; Medaglini *et al*., 1995) and these live recombinant bacteria have been shown to colonize the oral mucosa of recipient animals, inducing both a local and a systemic immune response (Medaglini *et al*., 1995).

Pozzi and coworkers constructed the prototypical *S. gordonii* chromosomal recombinant by inserting the *emm6* gene (encoding the M6 protein of *S. pyogenes*) as follows: i) Chromosomal DNA from *S. gordonii* strain GP204 was digested by *Cla*I and then re-ligated in combination with a promoterless *emm6/ermC* (M6 protein of S. pyogenes/ erythromycin resistance gene) cassette contained on a *Cla*I fragment. ii) The ligated DNA was transformed into a highly related *S. gordonii* strain (V288) iii) and a recombinant strain that displayed erythromycin resistance and expressed M6 protein was isolated. This strain was designated GP230 (Oggioni and Pozzi, 1996). In order to develop a more universal recipient *S. gordonii* strain for targeted insertion and expression of heterologous genes by homologous recombination into the bacterial chromosome Pozzi *et al* replaced the majority of emm6/ ermC gene cassette with the chloramphenicol acetytransferase (*cat*) gene via homologous recombination and this became the universal recipient *S. gordonii* strain, GP251 (Oggioni and Pozzi, 1996). This genetic system for creating *S. gordonii* recombinant expression strains has been very successful. However, the genetic methods used to create the recipient *S. gordonii* strain insert genes into an unknown locus and could potentially rearrange the chromosome of wild-type *S. gordonii*.

The initial aim of this study was to characterize the contextual nature of the genomic locus, as well as the transcriptional aspects of the insertion site that Pozzi developed to express foreign genes on the surface of *S. gordonii* and to use this information to optimize heterologous gene expression. Nucleotide sequence analyses of genomic walks in both directions revealed that

Table 1. Bacterial strains, plasmids, and oligonucleotides.

Strain, plasmid, or oligonucleotide Strains	Relevant markers and characteristics	Reference or source
E. coli		
INVαF'	F' *end*A1 *rec*A1 *hsd*R17(r_k^-, m_k^+) *sup*E44 *thi*-1 *gyr*A96 *rel*A1φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169	Invitrogen
XL1 Blue	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F'*pro*AB *lac*I^{q}ZΔM15 Tn10 (Tet^r)]^c	Stratagene
S. gordonii Challis		
V288	Wild-type (ATCC 35105)	American Type Culture Collection
GP204	Spontaneous Sm^r mutant of V288	Pozzi *et al.,* 1988
GP230	Recombinant strain contains the *emm6* gene (*S. pyogenes*) and an *erm*C gene, Em^r, parent strain (V288)	Pozzi *et al.,* 1992
GP251	Recombinant recipient strain contains the *cat* gene flanked by 145 bp of *emm6* gene and 202 bp of *erm*C gene, Cm^r, parent strain (GP230)	Oggioni and Pozzi, 1996
GP1214	Recombinant strain that expresses M6 protein (*S. pyogenes*) residues 1 to 16 fused to residues 222-441 and contains an *erm*C gene, Em^r, parent strain (GP251)	Oggioni *et al.,* 1994
GP1218	Recombinant strain that expresses M6 protein (*S. pyogenes*) residues 1 to 16 fused to residues 222-441 and contains an *aphIII* gene, Km^r, parent strain (GP1214)	Oggioni *et al.,* 1994
GP1223	Recombinant strain that expresses M6 protein (*S. pyogenes*) residues 1 to 16 fused to residues 222-441 and contains an *aphIII* gene, Km^r, and has been converted to Sm^r , parent strain (GP1218)	Oggioni *et al.,* 1994
SP-02	M protein recombinant strain, p635:M/*aph*III in V288, Km^r	This study
635/ermC	M protein recombinant strain, p635/ermC in V288, Em^r	This study
Plasmids		
pCR2.1	Km^r, Amp^r	Invitrogen
pSMB104	Contains the sequences encoding the CRR of M6 protein (*S. pyogenes*) residues 1 to 16 fused to residues 222-441 in tandem with an MspI/ClaI fragment of pE194 () encoding ermC cloned into pBluescript SK-.	Oggioni *et al.,* 1994
pCR2.1:635	1.1-kb PCR-amplified 6-35 walk from V288 cloned into pCR2.1 at EcoRI, Amp^r	This study
p635(NdeI)	NdeI site incorporated in between orf 1 and orf 2 in pCR2.1:635, Amp^r	This study
p635/ermC	1.2-kb *ermC* fragment from pSMB104 cloned into NdeI site in p635(NdeI), Amp^r	This study
p635:M/aphIII	2.7-kb M/*aphIII* fragment from GP1223 cloned into NdeI site in p635(NdeI), Amp^r	This study
Oligonucleotides		
CF4	5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCC GGGCAGGT-3'; Adaptor	Siebert *et al.,* 1995
CF5	5'-GGATCCTAATACGACTCACTATAGGGC-3'; AP1	Siebert *et al.,* 1995
CF6	5'-AATAGGGCTCGAGCGGC-3'; AP2, SEQ	Siebert *et al.,* 1995
CF7	5'-ACCTGCCC-(C3-Icaa-CPG spacer); AP1	This study
CF8	5'-TCTAGAGGTACCTTCTCGTGCTTTGTCCGG-3'; PCR (GP1223)	This study
CF9	5'-TACCGTCCCCCTAGGAAACACTCTTGCAC-3'; SEQ, PCR (GP1223)	This study
CF10	5'-TGACTTACTGGGGATCAAGCCTGATTGGGAG-3'; PCR (GP1223)	This study
CF11	5'-AAGTACATCCGCAACTGTCCATACTCTGATG-3'; PCR (GP1223)	This study
CF14	5'-GTTTTTCGTGTGCCTATTTTTTTGTG-3', SEQ 1223	This study
CF15	5'-GAGCGCATCGAAAATGCTGTTG-3'; SEQ, PCR (GP204	This study
CF16	5'-CTCAGTGTAAAGAGGAAATCC-3'; SEQ	This study
CF17	5'-GAGTTTCAATGGTCTTGTCTGG-3'; SEQ, PCR (GP204,GP1223)	This study
CF18	5'-CTTGAAAAGCCTGAGGGCTGGTTAC-3'; SEQ, PCR (GP204)	This study
CF19	5'-CTTGACCTTTGGTACCTTTGAC-3'; SEQ	This study
CF20	5'-GATAGTCACACGGCTACTCACG-3'; SEQ	This study
CF21	5'-CGTGAGTAGCCGTGTGACTATC-3'; SEQ	This study
CF22	5'-GTCCATAGAGTTTGGATCCAAG-3'; SEQ	This study
CF23	5'-GTCAAAGGTACCAAAGGTCAAG-3'; SEQ	This study
CF24	5'-CCAGAAATTCGCGATATGAAC-3'; SEQ	This study
CF25	5'-GAATGAATCCAGATAAGGTGC-3'; SEQ	This study
CF26	5'-GATATCTTCAACTCATGGGATTAC-3'; SEQ,	This study

	PCR (GP204)	
CF27	5'-CAAGATTCTCACCAGTTTTATG-3'; SEQ	This study
CF28	5'-GCTGCGATGCTTATGATTACC-3'; SEQ	This study
CF29	5'-GCTACCAATGCTGACAATAG-3'; SEQ	This study
CF31	5'-CCTAAGCAGTTTCTCAAGTTG-3'; SEQ	This study
CF32	5'-CATGTTGCCTATCGTCCAGC-3'; SEQ PCR (GP204, GP1223)	This study
CF35	5'-CGATTCGACATAGAAATAAATTGGAG-3'; SEQ, PCR (GP204)	This study
CF36	5'-CTATAGTCAGTGTGGTTTAGACAAGC-3'; SEQ	This study
CF39	5'-GATTATGCTGAATCAAATAGTC-3', SEQ	This study
CF40	5'-GAGCACGATAGTAGTCAATCAC-3'; SEQ	This study
CF41	5'-CAATTTTTGACTGATACGATGGC-3'; SEQ	This study
CF42	5'-CTGTTCTTCCAACTTTTTCAGC-3'; SEQ	This study
CF43	5'-GTTTGGTGACCTATAGTCAGTG-3'; SEQ	This study
CF44	5'-ATCTATACATTCATGCCATCCA-3'; SEQ	This study
CF45	5'-TGGATGGCATGAATGTATAGAT-3'; SEQ	This study
TB59	5' -AAAGAAGCATAACATATGTCAAAACAAG- 3'	This study
TB103	5' -GGAATTCCATATGCGGATAATAAATATATATAAACG- 3'	This study
TB104	5' -GGAATTCCATATGCGATTCACAAAAAATAGGCACACG- 3'	This study
TB117	5' -GGATCCCATATGTAAGGAGCATAAAAATGGC- 3'	This study

the site of integration is flanked by a duplicated 3-Kb *Cla*I fragment containing several genes homologous in amino acid sequence and organization to the metabolic leucine biosynthesis operon, as well as putative promoter and terminator elements. This structural arrangement differed from the chromosomal organization of the wild-type parental strain GP204, which contained only a single 3-Kb *Cla*I fragment and revealed the sequence of two ORFs (ORF A and B), the normal targets of transcription of the promoter within this fragment. Since these analyses have demonstrated one chromosomal rearrangement (a fragment duplication) and the genetic methods used to produce Pozzi's GP251 recipient strain conceptually predict the possibility of additional alterations, the information gained as a result of this study has been used to develop an alternative approach in which a foreign gene is targeted to a stable insertion site at a distinct locus within the *S. gordonii* chromosome perhaps more suitable to vaccine applications. The site is intergenic between the two open reading frames, ORF A and ORF B, downstream of the promoter that Pozzi *et al.* have used previously for protein expression (Oggioni *et al.*, 1994).

Results

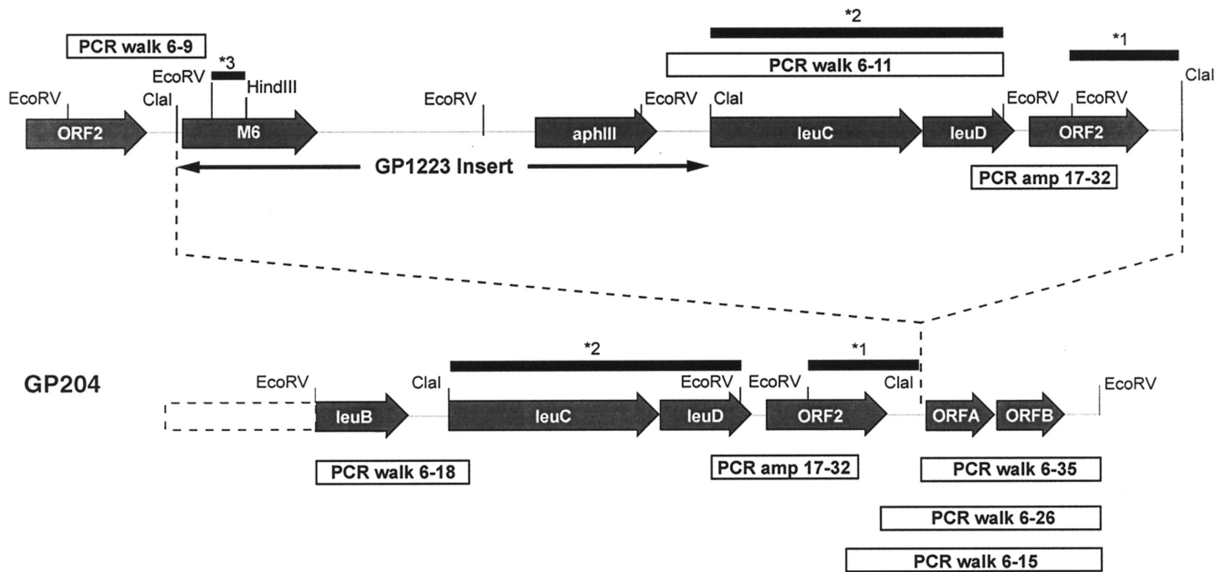
Chromosomal Walks Upstream and Downstream of the GP1223 Insert

To determine the chromosomal site of insertion of recombinant strains isolated from recipient *S. gordonii* strain GP251, a recombinant *S. gordonii* strain, GP1223 was used as template for directed chromosomal walks upstream and downstream of the GP1223 insert. GP1223 was isolated by the targeted insertion of the coding sequence of the CRR of M6 protein of *S. pyogenes* into the insertion site of GP251. Chromosomal DNA from *S. gordonii* strain GP1223 was purified and a special adaptor, CF4 and CF7 (Table 1) was ligated to the ends of the DNA fragments generated by digestion of the chromosomal DNA with *Eco*RV. The adaptor-ligated DNA was used as template for primary and secondary PCR reactions using nested pairs of adaptor primers (CF5, CF6) and a nested pair of specific gene primers (CF8, CF9) to walk upstream of the *Cla*I M6/aphIII insert or (CF10, CF11) to walk downstream of the insert (Table 1 and Figure 1A). The walk upstream of the insert yielded an 881-bp product designated "PCR walk 6-9" and the walk downstream, a 2175-bp product designated "PCR walk 6-11" as depicted in Figure 1A. The PCR walk products were sequenced directly using sequencing primers containing the sequences indicated in Table 1.

The Region Upstream of the GP1223 Insert Contains Regulatory Signals

Immediately upstream of the GP1223 insert, nucleotide sequences that conform to the consensus for promoters from gram-positive bacteria (Graves and Rabinowitz, 1986; De Vos, 1987) were located (designated P2 in Figure 2B). The alignment of the gram-positive promoter consensus with the sequence determined from "PCR walk 6-9" is shown in Figure 2A. This sequence shows the following features in common with the gram-positive promoter consensus: (i) the canonical –35 and –10 sequences; (ii) a spacing between those hexanucleotides of 16 to 18 nucleotides; (iii) the conserved dinucleotide sequence TG, immediately preceding the –10 sequence; and (iv) the AT-rich regions upstream of the –35 sequence (AT-box). Approximately 150 nucleotides upstream of the P2 promoter region, a sequence containing dyad symmetry followed by a stretch of thymidine residues conforms to a prokaryotic factor-independent RNA polymerase terminator sequence (designated T2, Figure 2B). Also, a region containing five direct repeats, 4 perfect and 1 imperfect, of 18 nucleotides (AGTTTAAAATCTTTATTC) was observed between the terminator and the promoter sequences (Figure 2B). The significance of this repeat region is not clear at this time. Upstream of the T2 terminator sequence, the nucleotide sequence of the 881-bp "PCR walk 6-9" also contained a partial ORF (designated ORF2, see Figure 1) encoding 169 residues. The sequence of ORF 2 had not terminated when the walk fragment ended at an *Eco*RV site to which the walking adaptor was ligated.

A. GP1223



B.

Predictions for EcoRV digestion:

Probe	GP204	GP1223
*1	1875 bp	1875 bp 936 bp
*2	2753 bp	2753 bp 2345 bp
*3	-	1755 bp

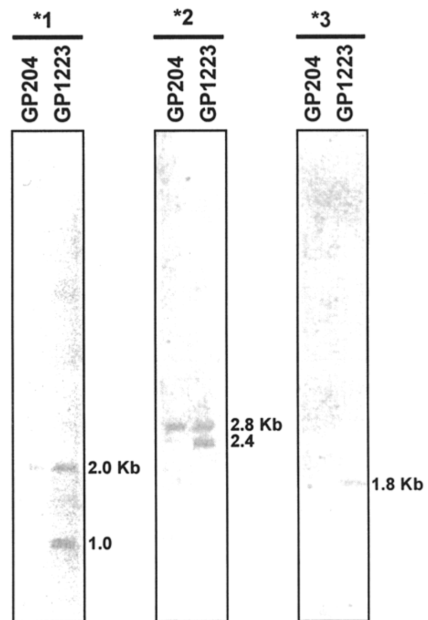


Figure 1. Analysis of the genomic organization of *S. gordonii* parental (GP204) and recombinant (GP1223) strains. A. Schematic representation of data from sequence determination of PCR walks on chromosomal DNA from *S. gordonii* strains. The organization and identity of predicted open reading frames are depicted on the genetic map as gray arrows. Restriction endonuclease sites are indicated above the genetic map. The *Cla*I fragment containing the M6 and *aphIII* gene insertion in the recombinant strain are underscored by solid arrows (GP1223 Insert). PCR walks are depicted by open boxes. The site of insertion of the recombinant strain (GP1223) into the parental strain (GP204) is shown by the dashed lines. Probes (P1, P2, P3) utilized in Southern blot analyses are shown as solid black boxes above the genetic map. B. Southern blot analyses of *Eco*RV-digested parental (GP204) and recombinant (GP1223) *S. gordonii* genomic DNA. Predicted fragment sizes based upon sequence determined above with each probe (P1, P2, P3, see above) are listed in tabular form alongside of the blots.

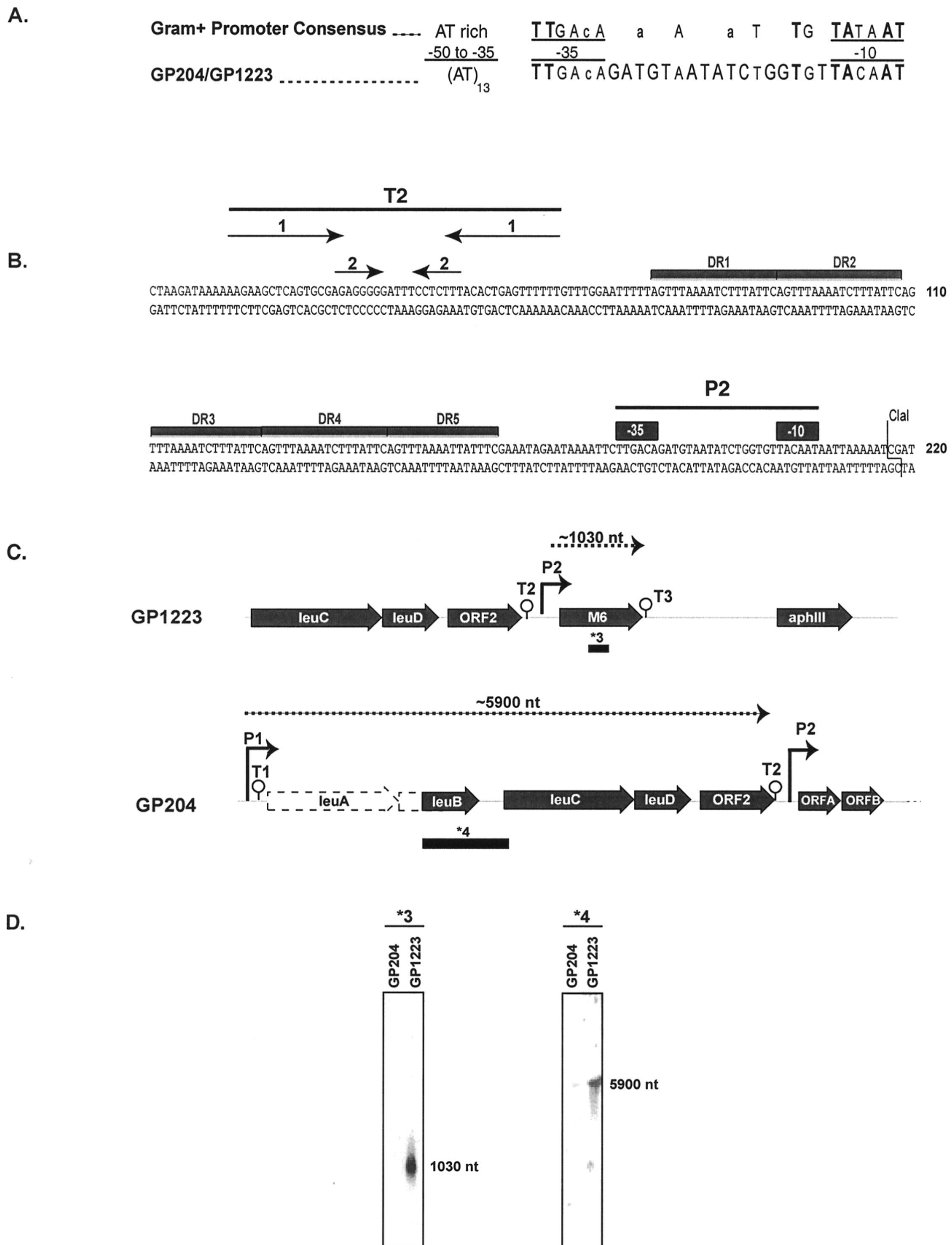


Figure 2. Analysis of the sequence of regulatory regions and transcription from *S. gordonii* parental and recombinant strains. A. Alignment of putative promoter sequences of GP204 and GP1223 with the consensus for promoters in gram-positive organisms. The alignment is illustrated with the most highly conserved bases in tall, bold capital letters. B. Sequence of the region upstream of the *Cla*I insertion site of recombinant strain GP1223. The –10 and –35 regions of the P2 promoter and the five direct repeats (DR1-DR5) are delineated above the sequence. Arrows numerated 1 and 2 overlie the regions of dyad symmetry predicted to function as factor-independent transcription termination sequences (T2). C. Schematic representation of the transcription units predicted from the parental (GP204) and recombinant (GP1223) *S. gordonii* strains. Location of promoters (P1, P2) and terminators (T1, T2, T3) are indicated relative to gene order. Predicted transcripts and sizes are indicated as dashed arrows above the maps. Probes (*3, *4) utilized in Northern analyses are localized by solid bars below the maps. D. Northern blot analyses of total RNA purified from *S. gordonii* strains GP204 and GP1223. Probes utilized are indicated above the blots and size (nt) of transcripts detected are indicated to the right of the blots.

The Region Downstream of the GP1223 Insert Contains *leuC* and *leuD* Homologues

Analysis of the nucleotide sequence of "PCR walk 6-11" (Figure 1A) revealed the presence of two partial ORFs encoding predicted proteins with significant homologies to the large subunit (*leu*C, pir S35134) and small subunit (*leu*D, pir E36889) of alpha-isopropylmalate isomerase (EC 4.2.1.33) of *Lactococcus lactis subsp. lactis*, respectively. These gene products are involved in the biosynthesis of the branched-chain amino acids leucine, isoleucine and valine (Godon *et al.*, 1992). The ORF encoding the *leu*C homologue was partial in that it did not contain the initiation codon for the reading frame, but was open from the start of the sequence at the *Cla*I site. The partial *leu*C ORF of *S. gordonii* encoded 456 amino acid residues, whereas the complete *leu*C ORF of *L. lactis subsp.lactis* is 460 residues in size. Nine nucleotides separate the termination codon of the *leu*C ORF and the initiation codon of the next ORF encoding the *leu*D homologue. The sequence of the *leu*D ORF was also partial because the ORF had not terminated when the walk fragment ended at an *Eco*RV site to which the walking adaptor was ligated. The partial *leu*D ORF of *S. gordonii* consisted of 172 residues as compared to the complete *leu*D ORF (191 residues) of *L. lactis subsp. lactis*.

The *ClaI* Fragment Flanking the GP1223 Insert is Duplicated

In order to corroborate and extend the structural organization deduced from the genomic walks described above, Southern blot analyses were carried out on chromosomal DNA from *S. gordonii* strains, GP204 and GP1223. Initially, chromosomal DNA was digested with restriction endonuclease *Cla*I, electrophoretically separated fragments blotted to membranes and probed with radiolabeled DNA fragments obtained from "PCR walk 6-9" and "PCR walk 6-1 1" digested with *Cla*I. Interestingly, both the probe specific for the upstream "PCR walk 6-9" and the probe specific for the downstream "PCR walk 6-11" hybridized to fragments that were indistinguishable in size (~ 3,000-bp) from both GP204 and GP1223 DNA (data not shown). This result suggested that "PCR walk 6-9" and "PCR walk 6-11" might be contained within the same or a related DNA fragment. In order to determine if an internal *Eco*RV fragment linked the upstream "walk 6-9" and downstream "walk 6-11" PCR fragments on a single *Cla*I fragment, PCR amplification with primers CF17 and CF32 was performed utilizing either GP204 or GP1223 chromosomal DNA as template. As predicted, a ~ 930-bp PCR amplification product was produced from both GP204 and GP1223 template DNA (Figure 1A and 1B) and the nucleotide sequence of these products was determined. The nucleotide sequence of the "PCR amp 17-32" product from both templates was identical and analysis revealed that they encoded the remaining nine residues of the previously determined *leu*D ORF (for a total *leu*D ORF of 181 residues). After a gap of 144 nt, a predicted ORF contiguous with the partial ORF2 determined on the sequence of "PCR walk 6-9" added 90 amino acid to the previously determined partial ORF2 of 169 residues yielding a total size for ORF2 of 259 residues. ORF2 encodes a protein with no homology with the branched chain amino acids biosynthesis enzymes or other proteins in the databases, but is similar in size (259 versus 254 residues) and contains 42 percent similarity when aligned with ORF2 in *L. lactis subsp. lactis* (Godon *et al.*, 1992).

Adding the internal 441-bp *Eco*RV fragment, revealed from the sequence analysis of "PCR amp 17-32", the proposed structure of the 3057-bp *Cla*I fragment present as a single copy in GP204 and duplicated on either side of the M6 insertion site of GP1223 is depicted in Figure 1A. Corroboration of this proposed genomic structure is demonstrated by the Southern blot analyses shown in Figure 1B.

The Region Upstream of the *leuC* ORF Contains *leuB* ORF, *leuA* and Regulatory Regions

Once the duplication of the *Cla*I fragment containing the *leu*C and *leu*D ORFs was comfirmed, it was of interest to determine the nucleotide sequence of the region of the chromosome immediately upstream of the *leu*C ORF from the parental strain, GP204. This would determine if the analogous gene order of the leucine biosynthetic operon from *L. lactis subsp. lactis* (*leuABCD*) (Godon *et al.*, 1992) was conserved in *S. gordonii*. Genomic walks upstream of the *leu*C ORF on parental strain GP204 were performed producing "PCR walk 6-18" (976-bp) (Figure 1A). Analysis of this sequence for ORFs predicted it to encode a partial rightward reading ORF of 193 amino acid residues (Figure 1A). This partial ORF encodes a predicted protein with significant homologies to 3-isopropylmalate dehydrogenase (EC 1.1.1.85) of *Lactococcus lactis subsp. lactis* (*leu*B, pir S35133).

Additional genomic walks and analyses of *S. gordonii* databases has confirmed the presence of the *leu*A ORF upstream of the *leu*B ORF and the presence of homologous promoter (P1) and terminator (T1) sequences prior to *leu*A (data not shown). These data confirm the gene order as *leuABCD* in *S. gordonii*.

Comparison of the Leucine Operon of *S. gordonii* to Other Organisms

Analysis of the complete nucleotide sequence of the compiled sequence of strain GP204 revealed significant amino acid and structural organization homology to the leucine operon of *Lactococcus lactis subsp. lactis* and numerous other gram-positive and gram-negative bacteria. Specifically, the predicted products of translation of two of the three reading frames sequenced in their entirety display significant homologies with the large subunit (*leu*C, pir S35134) and small subunit (*leu*D, pir E36889) of alpha-isopropylmalate isomerase (EC 4.2.1.33) of *Lactococcus lactis subsp.lactis*. The *S. gordonii* reading frame with homology to leuC of *L. lactis subsp.lactis* was 67% identical (207 identities over 307 residues) and 81% positive (250 positives over 307 residues). As mentioned in the previous section, partial nucleotide sequence of regulatory regions (P1, T1) and *leu*A and *leu*B have confirmed their presence and arrangement (data not shown). However, since these are only partial sequences further analysis of the extent of their homology will not be analyzed presently.

Nucleotide Sequence of the Region Downstream of the Promoter of Parental Strain, GP204

In order to identify the gene(s) endogenously expressed by the promoter directing expression of the CRR insert of GP1223, genomic walks downstream of the promoter region on parental strain GP204 were performed. This produced PCR walks 6-15 (1637-bp), 6-26 (1409-bp) and 6-35 (1152-bp) (Figure 1A). The nucleotide sequence of all three PCR walks was identical from the regions in which the PCR products overlapped and downstream of the putative promoter were sequences that encoded two complete ORFs (designated ORF A and ORF B). ORF A and ORF B were predicted to encode polypeptides of 145 and 156 amino acid residues respectively. Homology searches of ORF A or ORF B sequences versus the GenBank non-redundant database and the TIGR microbial genomes database using the BLAST program were performed. ORF A bore homology over the C-terminal half of the predicted protein to the regulatory protein, *Sly*A, found in *Escherischia coli*, *Salmonella typhimurium* and other Enterobacteriaceae. The alignment of ORF A with *Sly*A(*E. coli*) contained 31% identities (23 identities over 74 residues) and 49% positives (37 positives over 74 residues). *Sly*A is a member of the *Mar*R family of transcriptional regulators and a BLOCKS search (Henikoff and Henikoff, 1994) revealed ORF A to be a member of the *Mar*R family as well. The search of the unfinished microbial genomes database revealed only one highly homologous predicted protein in the TIGR-1313 (sp12 contig) *Streptococcus pneumoniae* database that contained 85% residue identities (122 identities over 143 residues) and 90% positives (129 positives over 143 residues).

Similar searches with the predicted peptide sequence of ORF B revealed no known functional homologies or patterns. However, the search of the unfinished microbial genomes database also revealed only one highly homologous predicted protein in the TIGR-1313 *Streptococcus pneumoniae* database that contained 90% residue identities (140 identities over 155 residues) and 93% positives (145 positives over 155 residues). This ORF in the *S. pneumoniae* database was located in the same contig as the ORF A homologue described above (sp12) and the ORF B homologue was located immediately downstream of ORF A homologue revealing conservation in structural arrangement between the two subspecies as well.

Northern Analyses

A schematic diagram of the predicted transcription units of parental (GP204) and recombinant (GP1223) *S. gordonii* strains is shown in Figure 2C. This representation is based upon the gene organization (solid filled) and promoter (P2) and terminator (T2) homologies determined by analyses of the DNA sequenced. Promoter (P1), terminator (T1), the *leu*A gene and remainder of *leu*B gene (illustrated in dashed lines) have not been sequenced in their entirety. However, as mentioned above, partial nucleotide sequence of these regions has confirmed the presence and arrangement of these genes and regulatory elements (data not shown). Northern analyses (Figure 2D) utilizing probe *3 (specific for M6 sequences) detected a unique (not present in GP204 RNA), highly expressed M6-specific transcript of ~1,030 nt. This transcript size is in agreement with transcription originating at the P2 promoter element and terminating at the T3 factor-independent transcription terminator sequence predicted to normally terminate the M6 protein. Analyses with probe *4 (specific for a portion of the leuB/leuC ORFs) detected a major transcript of ~ 5,900 nt in RNA from both GP204 and GP1223. The predicted transcriptional unit for this polycistronic RNA species is from promoter P1 to terminator T2 and is in agreement with size predictions for this transcript (5,947 nt) from *L. lactis subsp. lactis* as well (Godon *et al*., 1992).

Construction of an Intergenic Mutant

Genomic analysis of recombinant *S. gordonii* GP1223 and the parent GP204 strain revealed that the inserted foreign sequences (recombined into GP1223 chromosome) were being driven by a promoter (P2, Figure 2C) normally located in front of two unknown open reading frames, ORF A and ORF B, which are just downstream of the leucine operon. A *Nde*I site was introduced between ORF A and ORFB in p635 (Figure 3A) to serve as an insertion site between the two genes. Initially, an erythromycin gene was inserted and the resulting plasmid p635:ermC was transformed into strain V288. Since p635:ermC has no gram-positive origin of replication it can not replicate in *S. gordonii* and was forced to integrate into the chromosome via homologous recombination and yielded erythromycin resistant colonies. PCR analysis of this double crossover mutant 635/ermC with the primer pair CF43-CF45 produced a product that was 1.2 kb larger than wild type V288 due to insertion of the erythromycin gene (data not shown).

The *emm6* (Oggioni *et al*., 1994) gene fused to the *aphIII* gene (M/aphIII) was then inserted into p635 at the engineered NdeI site and the resulting plasmid p635:M/aphIII was transformed into V288 yielding kanamycin resistant colonies. This recombinant, SP-02 (Figure 3A), containing the M/aphIII fusion between ORF A and ORF B was verified by PCR using the primer pair CF43-CF45. SP-02 produced a product that was 2.7 kb larger than wild type V288 due to insertion of the M/aphIII cassette (data not shown). Southern blot analyses on SP-02 genomic DNA, restricted with *Cla*I and *Sma*I, using a portion of the 635 sequence and the M/aphIII sequence as labeled probes showed an intergenic insertion event had occurred. The 635 probe reacted with a 4.7 kb band in V288 (also restricted with *Cla*I and *Sma*I) and a 7.5 kb band in SP-02 which is a difference of 2.7 kb, the size of the insert (Figure 3B), suggesting that it is a double cross-over mutant. The M/aphIII probe did not react with V288 DNA, which does not have the M/aphIII gene, and reacted with a 7.5 kb band in SP-02 as expected (Figure 3C).

Surface Expression

M6 surface protein expression was demonstrated on SP-02 by streak blot (data not shown), and competition ELISA (Figure 4). The competition ELISA was carried out by utilizing purified ColiM6 protein and anti-M6 monoclonal antibody 10F5. *S. gordonii* strain V288 was used as a negative control and GP1223 was used as a positive control. The results of the ELISA showed that surface expressed M6 protein on SP-02 bound 10F5 and

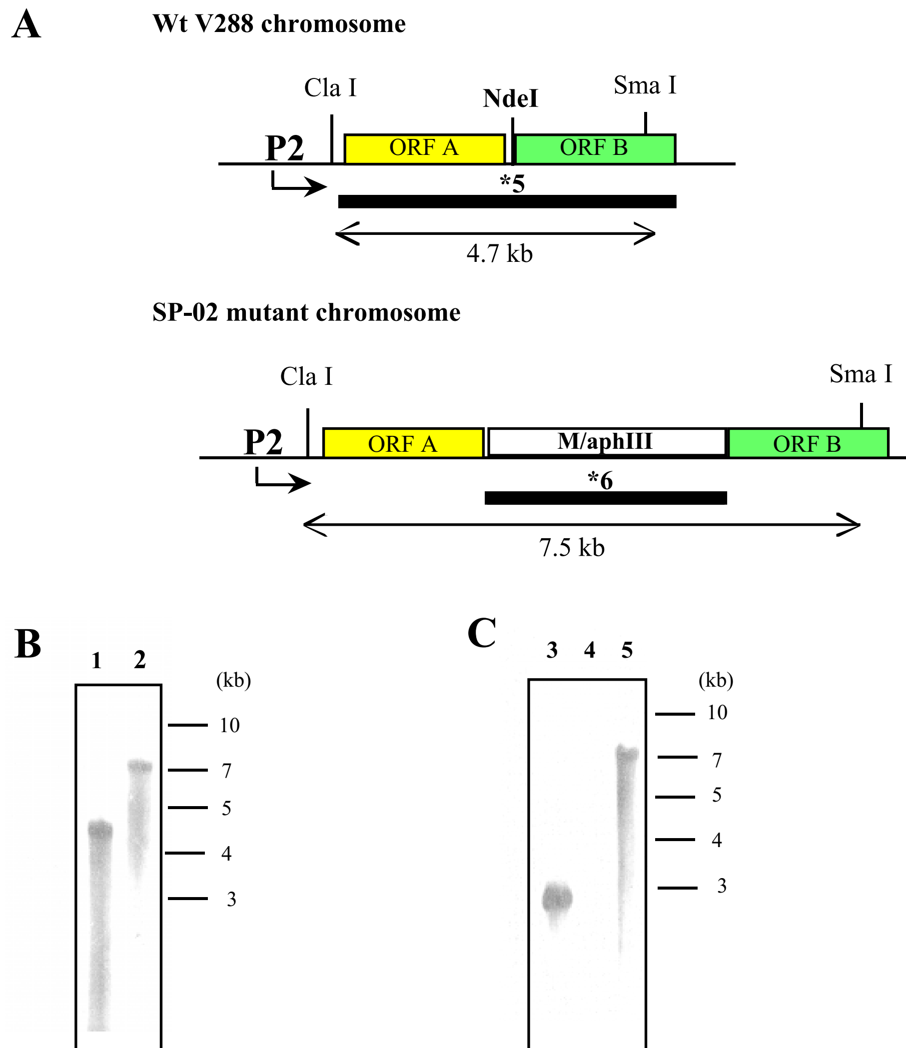


Figure 3. Southern blot analysis depicting insertion of the M/aphIII cassette into 6-35 site. (A) Schematic representation of the wild type V288 and mutant SP-02 chromosomes. The 6-35 (*5) and M/aphIII (*6) DNA probes are shown by solid lines. The genomic DNA was digested with *Cla*I and *Sma*I. (B) Southern blot of genomic DNA from V288 (lane 1) and SP-02 (lane 2) probed with 6-35 probe. (C) Southern blot of M/aphIII DNA fragment (lane 3), V288 genomic DNA (lane 4) and SP-02 genomic DNA (lane 5) probed with M/aphIII (*6) probe.

competitively inhibited it from binding ColiM6. The level of inhibition of SP-02 was approximately 3-fold lower than that of GP1223, which has the M6 protein gene inserted directly behind the promoter. This shows that the level of expression of heterologous proteins is dependent on its position after the P2 promoter. The expected size of the M6 protein (28 KDa) was verified by western blot (data not shown).

Discussion

Several genetic systems have been developed for the stable expression of heterologous proteins from nonpathogenic oral streptococci, such as *S. gordonii* (Pozzi *et al.*, 1992a; Medaglini *et al.*, 1995). Pozzi and coworkers established a chromosomal insertion site downstream of a promoter in *S. gordonii* for protein expression (Pozzi *et al.*,1992b, Oggioni *et al.*, 1994). However, the consequences of the genetic methods utilized to develop the streptococcal recipient strain, GP251, have not been determined. The nature of the integration site, the regulatory elements (promoters, terminators, etc.) acting on the expression cassette, and possible phenotypic consequences on the recombinant strains have not been investigated. Conceivably, the chromosome of the recipient strain, and consequently the recombinant strain, may have undergone some genetic rearrangement as a result of the chromosomal digestion with *Cla*I followed by re-ligation used to create the recipient strain. The inceptive goal of this work was to characterize the nature of the genomic locus, as well as the transcriptional aspects of the insertion site that Pozzi developed to express foreign genes on the surface of *S. gordonii* and to use this information to optimize heterologous gene expression.

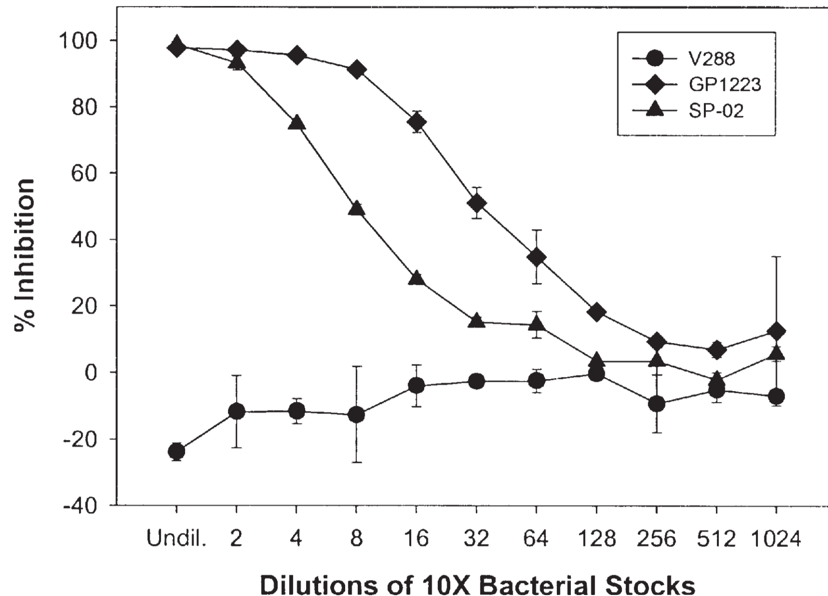


Figure 4. Competition ELISA assay with M protein surface expression strains versus *coli* M6 protein. Graph shows percent inhibition of binding of mAb 10F5 to *coli* M6 protein by decreasing concentrations of cells. Strains are as shown in the legend.

To this end, the genomic context surrounding the insertion site of recombinant *S. gordonii* strain (GP1223) was determined. The nucleotide sequence of chromosomal walks in both directions from the insertion site revealed that the insert was flanked by a duplicated 3061-bp *Cla*I fragment. The presence of the identical duplicated flanking *Cla*I fragments in the recipient strain GP251 (containing the *cat* gene) was validated as well (data not shown). The *Cla*I fragment was present in the parental *S. gordonii* strain GP204, but was not duplicated as in the recipient and recombinant strains. This finding confirms that genetic rearrangement has occurred due to the manner in which the recipient strain was constructed. The likelihood of additional perturbations at other genomic locations is high. The potential consequences of such changes on the virulence of this normally nonpathogenic organism are worthy of careful consideration.

With regard to transcription, a consensus gram-positive promoter and a factor-independent RNA polymerase terminator sequence were detected immediately upstream of the insertion site. Chromosomal walks downstream of the identified promoter region on the non-recombinant parental strain, GP204, yielded the sequence of two ORFs (ORF A and ORF B) which would be the conventional subjects of the transcription derived from this promoter. ORF A (residues 72 to 145) shares homology with salmolysin [cytolysin *SlyA*(EC), P55740], the transcriptional activator for the cryptic hemolysin gene *sheA*(E. coli). New homologues of *SlyA* form a subgroup of a growing superfamily of bacterial regulatory proteins controlling diverse physiological processes (Buchmeier *et al.*, 1997; Thompson *et al.*, 1997), as well as virulence in *Salmonella typhimurium* (Libby *et al.*, 1994).

The *Cla*I fragment, as well as the fully and partially sequenced upstream region of the wild-type strain GP204 encoded open reading frames (ORFs) with high homology and parallel structural organization to the leucine biosynthesis operon of *Lactococcus lactis subsp. lactis*. The organization of the putative promoters and terminators (P1, T1) as well as the gene order of the leucine biosynthesis operon (*leuABCD)* was highly similar between *L. lactis* and *S. gordonii*. Possible regulation of the *S. gordonii* leucine operon by the attenuation mechanism, mediated by a leucine-rich leader peptide that has been well-characterized in *L. lactis subsp. lactis* (Godon *et al.*, 1992) is currently under additional investigation.

Based upon the information developed from the sequence analysis of the chromosomal walks we sought to develop a stable chromosomal insertion site in the wild type *S. gordonii* genetic background for use without significant disruption of the chromosome that might have deleterious effects on the phenotype of derived recombinants which could compromise their eventual use as vaccines. The work done in this study provides some insight into the promoter driving transcription of the recombinant genes and the surrounding region in the parental strain. We have taken advantage of this new genomic information to design plasmids that allow insertion of heterologous genes between ORFA and ORFB (Figure 3A). This allows for a clean and stable chromosomal insertion site that does not disrupt any other loci. Protein expression from this locus was achieved, albeit at about a 3-fold lower level than insertion directly behind the promoter (Figure 4). Homology searches to identify this operon and promoter have not identified any known functions, but with further studies this operon may provide several other intergenic insertion sites that provide high levels of expression or that are possibly inducible.

With the current interest in using commensal gram-positive bacteria as vaccine delivery vectors, there is a need for additional clean, stable insertion sites that do not appreciably disrupt the bacterial chromosome. One such site has been discussed in this paper.

Experimental Procedures

Bacterial Strains, Plasmids, and Oligonucleotides

The bacterial strains, plasmids, and oligonucleotides used or relevant to this study are listed in Table 1. *Escherichia coli* strains were grown in Luria-Bertani broth or on Luria-Bertani medium containing 1.5% agar. *S. gordonii* was plated on or cultured in brain heart infusion medium (BHI; Difco Laboratories, Detroit, MI) with or without agar respectively. All bacterial cultures were incubated at 37°C. Kanamycin (500 µg/ml) erythromycin (5 µg/ml) and streptomycin (500 µg/ml) were used whenever required for *S. gordonii* strains and ampicillin (50 µg/ml) was used for the selection and growth of *E. coli* strain INVαF' containing the plasmid pCR2.1 clones. The oligonucleotides, described in Table 1 were synthesized by either the Central Services Laboratory (Oregon State University) or Gibco-BRL Laboratories (Grand Island, NY). Frozen cells of naturally competent *S. gordonii* strain V288 were prepared and transformed as previously described (Pozzi *et al*., 1990). Standard procedures were used for gene fusions and mutagenesis in *E. coli* vectors (Maniatis *et al*., 1982).

Chromosomal Walks

Chromosomal DNA was prepared from GP204 and GP1223 cells lysed with lysozyme and sodium dodecyl sulfate at pH 8.0 followed by three cycles of freezing and thawing and purified by phenol extraction. Chromosomal walks from a known region to an unknown region in uncloned genomic DNA were accomplished using an improved adaptor ligation PCR method (Siebert *et al*., 1995).

Nucleotide Sequence Methods and Analysis

PCR products of chromosomal walks were either sequenced directly or cloned into a TA-cloning vector pCR2.1 (Invitrogen, Carlsbad, CA) prior to sequence determination. Sequence determinations were performed at the Central Services Laboratory of the Center for Gene Research and Biotechnology (Oregon State University) using the dideoxy chain termination method. The M13 reverse sequencing primer and the T7 promoter primer were utilized to determine the sequence of PCR inserts cloned into pCR2.1, as well as the specifically designed primers listed in Table 1. Sequences were compiled and DNA and amino acid sequences were analyzed using programs developed by the Genetic Computer Group at the University of Wisconsin (Devereux *et al*., 1984). The BLAST programs (Altschul *et al*., 1997) were used to compare the determined nucleotide sequences to the sequences in the GenBank databases.

Southern Blot Analysis

S. gordonii chromosomal DNA (1 µg), purified as described above, was digested with either *Eco*RV or *Cla*I/*Sma*I (New England Biolabs; Beverly, MA). DNA fragments were separated in 0.8% agarose-Tris-borate-EDTA and then transferred to Nytran Plus (Schleicher and Schuell; Keene, NH) membranes. Probe *1 (comprised of the C-terminal portion of ORF 2 and the P2 promoter region) was derived by digestion of the PCR product generated by PCR amplification with primers CF4 and CF9 from GP204 chromosomal template with *Cla*I followed by the gel isolation of the 722-bp digestion product. Probe *2 (encompassing the *leu*C and *leu*D ORFs) was obtained by digestion of the PCR product generated by PCR amplification with primers CF4 and CF11 from a GP204 chromosomal template with *Cla*I followed by the gel isolation of the 1894-bp digestion product. Probe *3 (a portion of C-repeat region of M6 protein of *S. pyogenes*) was obtained by isolation of the 247-bp *Eco*RI/*Hin*dIII digestion product of pSMB104. Probes *1, *2 and *3 were labeled and hybridization products visualized using the Rad-Free Psoralin Biotin Probe Labeling and Hybridization Kit (Schleicher and Schuell; Keene, NH). Probe *5 (containing ORF A and ORF B) was derived by digestion of pCR2.1:635 with *Eco*RI followed by the gel isolation of the 1153-bp digestion product. Probe *6 (containing M/aphIII) was derived by digestion of p635:M/aphIII with *Nde*I followed by the gel isolation of the 2702-bp digestion product. Probes *5, *6 were radiolabeled with Redivue [α-^{32}P] dCTP Rediprime II random primed labeling system (Amersham, Picastaway, NJ).

Isolation of Total RNA and Northern Blot Analysis

S. gordonii total RNA was purified as previously described (Shaw and Clewell, 1985). RNAs (10 µg) were separated in 1% (wt/vol) agarose-2.2M formaldehyde gels and then transferred to NYTRAN MaxStrength (Schleicher and Schuell; Keene, NH) membranes. Probe *3 (a portion of C-repeat region of M6 protein of *S. pyogenes*) was obtained from by isolation of the 247-bp *Eco*RI/*Hin*dIII digestion product of pSMB104. Probe *4 (a portion of the *leu*B/*leu*C region of *S. gordonii*) by isolation of the 976-bp PCR product generated by PCR amplification with primers CF6 and CF18 from a GP204 chromosomal template. The probes were radiolabeled with [α-^{32}P]dCTP, using a random primers DNA labeling kit, Rediprime (Amersham; Picastaway, NJ) according to the manufactures instructions.

Nucleotide Sequence Accession Numbers

The sequences of the regions depicted in Figure 1A have been assigned GenBank Accession Nos. AF251027, AF251028 and AF251029.

Streak Blot Analysis

S. gordonii transformants were streaked on the surface of BHI plates by toothpick transfer of colonies from the selection plates. Each plate contained the transformants, an M6$^+$ strain (GP1223) and an M6$^-$ strain (V288) for controls. Streak blot was performed as previously described (Pozzi *et al*., 1992b), using monoclonal antibody (mAb) 10F5 (Fischetti *et al*., 1985), raised against the recombinant M6 protein purified from *E. coli*.

Western Blot Analysis

The streptococcal strains were grown to late stationary phase of growth in BHI broth. 300 µl of culture was pelleted by centrifugation in 1.5 ml microfuge tubes. The culture supernatant was acetone- precipitated and the pellet was resuspended in SDS sample buffer. The samples were run on a 4-12% Bis-Tris gel and transferred to a Millipore

Immobilon-P transfer membrane. Western blotting was performed as previously described (Fischetti *et al.*, 1985) using mAb 10F5.

Competition ELISA

Streptococcal overnight cultures were back diluted 1:100 in BHI containing the appropriate antibiotics and grown to late log (OD_{650nm} = 0.6-0.7). Fifty ml of culture was harvested by centrifugation (10,000 X *g*) for 10 min and the cell pellets were resuspended in 25 ml PBS/azide (PBS + 0.02% sodium azide). The bacterial suspensions were placed in a 56°C water bath for 60 minutes to kill the cells. The cells were centrifuged and washed with 25 ml PBS/azide. The cell pellets were resuspended in 10 ml PBS/azide and the OD_{650nm} was adjusted to 1.0 with PBS/azide. 10 ml of adjusted suspension was centrifuged and 9 ml of supernatant was removed by pipet. The pellet was resuspended with the remaining supernatant. Strain preparations were stored at 4°C for up to 1 week. The resulting cell suspensions were used to compete for the binding of mAb 10F5 to recombinant M6 protein in competition ELISA's, as described by Jones *et al.* (Jones *et al.*, 1986; Jones *et al.*, 1988). Wild type strain V288 was used as a negative control in the competition ELISA's.

Acknowledgements

This work was supported in part by grants from NIH (AI46176-01A1) and SIGA Pharmaceuticals (J0262A).

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Buchmeier, N., Bossie, S., Chen, C.Y., Fang, F.C., Guiney, D.G. and Libby, S.J. 1997. SlyA, a transcriptional regulator of Salmonella typhimurium, is required for resistance to oxidative stress and is expressed in the intracellular environment of macrophages. Infect. Immun. 65:3725-3730.

Devereux, J., Haeberli, P. and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.

De Vos, W.M. 1987. Gene cloning and expression in lactic streptococci. FEMS Microbiology Reviews 46:281-295.

Fischetti, V.A., Jones, K.F. and Scott, J.R. 1985. Size variation of the M protein in group A streptococci. J. Exp. Med. 161:1384-1401.

Godon, J.J., Chopin, M.C. and Ehrlich, S.D. 1992. Branched-chain amino acid biosynthesis genes in Lactococcus lactis subsp. lactis. J. Bacteriol. 174:6580-6589.

Graves, M.C., and Rabinowitz, J.C. 1986. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferrodoxin gene. J. Biol. Chem. 261:11409-11415.

Henikoff, S., and Henikoff, J.G. 1994. Protein family classification based on searching a database of blocks. Genomics 19:97-107.

Jones, K.F., Khan, S.A., Erickson, B.W., Hollingshead, S.K., Scott, J.R. and Fischetti, V.A. 1986. Immunochemical localization and amino acid sequences of cross reactive epitopes within a streptococcal M6 protein. J. Exp. Med. 164:1226-1238.

Jones, K.F., Holingshead, S.K., Scott, J.R. and Fischetti, V.A. 1988. Spontaneous M6 protein size mutants of group A streptococci display variation in antigenic and opsonogenic epitopes. Proc. Natl. Acad. Sci. USA. 85:8271-8275.

Libby, S.J., Goebel, W., Ludwig, A., Buchmeier, N., Bowe, F., Fang, F.C., Guiney, D. G.J., Songer, G. and Heffron, F. 1994. A cytolysin encoded by Salmonella is required for survival within macrophages. Proc. Natl. Acad.Sci USA. 91:489-493.

Maniatis, T., Fritsch, E.F. and Sambrook, J. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Medaglini, D., Pozzi, G., King, T.P., and Fischetti, V.A. 1995. Mucosal and systemic immune responses to a recombinant protein expressed on the surface of the oral commensal bacterium *Streptococcus gordonii* after oral colonization. Proc. Natl. Acad. Sci. 92:6868-6872.

Oggioni, M.R., Manganelli, R., and Pozzi, G. 1994. Construction of a recombinant *Streptococcus gordonii* expressing the C repeat region of *Streptococcus pyogenes* M protein. Internal report. M6 Pharmaceuticals.

Oggioni, M.R., and Pozzi, G. 1996. A host-vector system for heterologous gene expression in *Streptococcus gordonii*. *Gene* 169:85-90.

Pozzi, G., Musmanno, R.A., Renzoni, E.A., Oggioni, M.R., and Cusi, M.G. 1988. Host-vector system for integration of recombinant DNA into chromosomes of transformable and nontransformable streptococci. *J. Bacteriol.* 170:1969-1972.

Pozzi, G., Musmanno, R.A., Lievens, P.M.J., Oggioni, M.R., Plevani, P., and Manganelli, R. 1990. Method and parameters for genetic transformation of *Streptococcus sanguis* Challis. Res. Microbiol. 141:659-670.

Pozzi, G., Contorni, M., Oggioni, M.R., Manganelli, R., Tommasino, M., Cavalieri, F., and Fischetti, V.A. 1992. Delivery and expression of a heterologous antigen on the surface of streptococci. Infect. Immun. 60:1902-1907.

Pozzi, G., Oggioni, M.R., Manganelli, R., and Fischetti, V.A. 1992. Expression of M6 protein gene of *Streptococcus pyogenes* in *Streptococcus gordonii* after chromosomal integration and transcriptional fusion. Res. Microbiol. 143:449-457.

Shaw, J. H., and Clewell, D.B. 1985. Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn917 in *Streptococcus faecalis*. J. Bacteriol. 164:782-796.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A., and Lukyanov, S.A. 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research* 23:1087-1088.

Thompson, N.R., Cox, A., Bycroft, B.W., Stewart, G.S., Williams, P., and Salmond, G.P. 1997. The rap and hor proteins of Erwinia, Serratia and Yersinia: a novel subgroup in a growing superfamily of proteins regulating diverse physiological processes in bacterial pathogens. Mol. Microbiol. 26:531-544.