Characterization of ISRgn1, a Novel Insertion Sequence of the IS3 Family Isolated from a Bacteriocin-Negative Mutant of *Ruminococcus gnavus* E1

Ana Gomez,[1]* Monique Ladire,[2] Francoise Marcille,[2] Michelle Nardi,[3] and Michel Fons[4]

Mixis France, Faculte de Medicine Necker-Enfants Malades, 75730 Paris Cedex 15,[1] France, Unité d'Ecologie et Physiologie du Système Digestif, Institut National de la Recherche Agronomique, 78352 Jouy en Josas Cedex[2], Unité de Biochimie et Structure des Protéines, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas Cedex,[3] and Institut Méditerranéen de Recherche en Nutrition, Faculté des Sciences et Techniques de Marseille St Jérôme, Av. Escadrille Normandie Niemen, 13397 Marseille Cedex 20,[4] France

ISRgn1, an insertion sequence of the IS3 family, has been identified in the genome of a bacteriocin-negative mutant of *Ruminococcus gnavus* E1. The copy number of ISRgn1 in *R. gnavus* E1, as well as its distribution among phylogenetically E1-related strains, has been determined. Results obtained suggest that ISRgn1 is not indigenous to the *R. gnavus* phylogenetic group but that it can transpose in this bacterium.

Insertion sequences (ISs) are mobile DNA elements capable of mediating various types of genome rearrangements. They also play an important role in evolution by facilitating horizontal gene transfer (for a review on insertions sequences, see reference 11). One of the largest families of ISs is the IS3 family, which is widely distributed in both gram-negative and gram-positive bacteria. Members of this group have sizes from 1.2 to 1.5 kb. They often possess imperfect terminal inverted repeats and generate a target site duplication of 3 to 5 bp. The majority of the IS3 insertion sequences have at their ends the dinucleotides 5′-TG...CA-3′. They contain two consecutive and partially overlapping open reading frames (ORFs), with the second or downstream ORF (*orfB*) at phase −1 relative to the first or upstream ORF (*orfA*). The functional transposase is proposed to be a fusion protein, OrfAB, generated by programmed ribosomal frameshifting between both *orf* genes.

Ruminococcus gnavus E1 is a gram-positive strictly anaerobic strain isolated from a human intestinal microbiota that produces ruminococcin A (RumA), a trypsin-dependent lantibiotic (5). Lantibiotics are posttranslationally modified peptides characterized by the presence of lanthionines, β-methyl-lanthionines, and didehydrated amino acid residues (15). Recently, a 12.8-kb region from the genome of *R. gnavus* E1 containing the genetic determinants for RumA production has been characterized (9). The *rumA* gene cluster is organized in three operons with predicted functions in lantibiotic biosynthesis and secretion (*rumA₁A₂A₃MTX*), signal-transduction regulation (*rumRK*), and producer self-immunity (*rumFEGHR2*) (see Fig. 1A). Transcription of *rum* operons has been shown to be induced in the presence of trypsin, probably through the two-component system RumRK (9).

In this work, we report on the identification and characterization of a novel insertion sequence present in the genome of a spontaneous E1 bacteriocin-negative mutant (E1-Bac⁻). This element, designated ISRgn1, showed the typical features of the IS3 family and to our knowledge is the first insertion sequence isolated from *R. gnavus*.

Identification of ISRgn1 in spontaneous bacteriocin-negative mutant of *R. gnavus* E1. A mutant of the RumA producer *R. gnavus* E1 unable to produce bacteriocin activity in a trypsin-supplemented medium was isolated under standard laboratory culture conditions. To test the presence of the genes (*rumA₁A₂A₃*) encoding the RumA precursor, total DNA of E1-Bac⁻ was digested with different restriction enzymes and subsequently subjected to Southern hybridization analysis using the *rumA*-specific primer ol30 (5) as a probe. Restriction DNA fragments hybridizing to the probe were obtained, but the hybridization pattern differed from that of the original producer strain used as a control (data not shown).

Based on these data, a genomic library of E1 Bac⁻ was constructed in *E. coli* DH5α cells by using established protocols for molecular biology techniques (16). Total DNA was isolated using the NucleoSpin C+T kit (Machery-Nagel GmbH & Co., Düren, Germany), digested with *Pst*I and *Sst*I, and subsequently cloned into the corresponding sites of pBluescript SK(+). After electroporation of *E. coli* DH5α cells, clones harboring the plasmid pLEM501 (Fig. 1) were identified by colony blot hybridization with ol30. The DNA sequence of the fragment cloned in pLEM501 comprised 5,451 bp, with an overall G+C content of 32%.

Database similarity searches allowed the identification of eight ORFs, each of them preceded by potential ribosome binding sites (RBSs). Comparison of the sequenced fragment of the E1-Bac⁻ DNA (Fig. 1B) with the RumA locus of the E1 wild-type strain (Fig. 1A) showed that pLEM501 contained *rumA₁A₂A₃* as well as a 3′-truncated *rumM* gene, the gene presumably encoding the pre-RumA modifying enzyme (9).

However, upstream of *rumA₁*, the DNA sequence of the pLEM501 insert differed from that of the corresponding region

* Corresponding author. Mailing address: Mixis France. Faculte de Medicine Necker-Enfants Malades, 6eme etage. 156, rue de Vaugirard, 75730 Paris Cedex 15, France. Phone: (33) 1 40 61 56 65. Fax: (33) 1 40 61 56 61. E-mail: ana.gomez@mixis.fr.
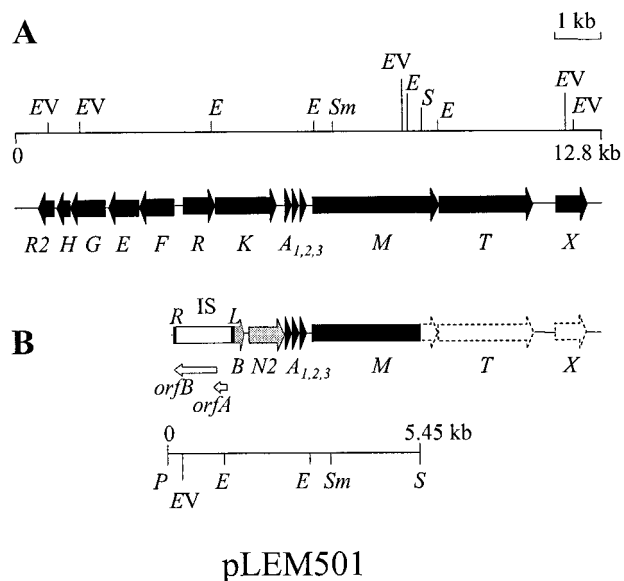
FIG. 1. Organization of RumA loci in *R. gnavus* E1 (A) and *R. gnavus* E1 Bac⁻ (B). Arrows represent the ORFs, and their orientation shows the transcriptional direction. (B) The region cloned into pLEM501 is shown below the diagram; dashed arrows represent genes that were amplified by PCR. ISRgn*1* (IS) is represented by a white box and delimited by the left (L) and right (R) inverted repeats. Both *orfA* and *orfB* within the insertion sequence are indicated below as open arrows. Restriction enzyme abbreviations: P, *Pst*I; EV, *Eco*RV; E, *Eco*RI; Sm, *Sma*I; S, *Sst*I. Not all restriction sites are shown.

FIG. 2. Putative frameshift region in the ISRgn*1* RNA, including a predicted stem-loop structure with a ΔG of -12.9 kcal/mol (19). Only one of several possible secondary structures is shown for simplicity. The putative RBS of OrfB is underlined twice, and the proposed frameshift heptanucleotide is in bold. Codons indicated above the RNA sequence correspond to the product of direct translational readout from the ORFA start codon. Those below correspond to the product of a -1 translational frameshift. The ORFA stop codon is circled.

in the wild-type RumA locus, suggesting that a genetic rearrangement occurred in the E1-Bac⁻ mutant. Starting at 1,147 bp upstream of the start codon of *rumA1*, the gene *rumB* encoded a 66-amino-acid peptide with a theoretical molecular mass of 7.39 kDa, showing similarity (33% identity, 53% similarity) with the precursor of mersacidin, a globular type B lantibiotic produced by *Bacillus* sp. strain HIL Y-85,54728 (1). The following ORF, *rumN*, encoded a putative protein of 248 amino acids (29.2 kDa) exhibiting significant sequence similarity (25% identity, 43% similarity) with the N-terminal region of MrsM, the modifying enzyme of the mersacidin precursor (1), and other LanM proteins supposed to catalyze dehydration and subsequent thioether ring formation in lantibiotic prepeptides (17, 21). Both *rumB* and *rumN* were named on the basis of database similarity searches and were in accordance with conventional lantibiotic clusters nomenclature (6).

Upstream of *rumB*, two ORFs, *orfA* and *orfB*, and their neighboring sequences fulfilled typical features of an IS of the IS*3* family. This element was named ISRgn*1*, and it is described below.

To complete the comparison between the RumA loci of the E1 wild-type strain and the E1- Bac⁻ mutant, PCR amplifications were performed using primer sets based on the nucleotide sequences of genes surrounding the *rumA* copies on the RumA wild-type locus. The primers utilized and their corresponding target sites within the *R. gnavus* E1 RumA locus (EMBL accession no. AJ276653) were rtR2 (positions 663 to 692), rtF (positions 2862 to 2833), NrumF (positions 3034 to 3063), NrumK (positions 4533 to 4499), rtM (positions 9256 to 9290), rtT1 (positions 10807 to 10778), rtT2 (positions 10778 to
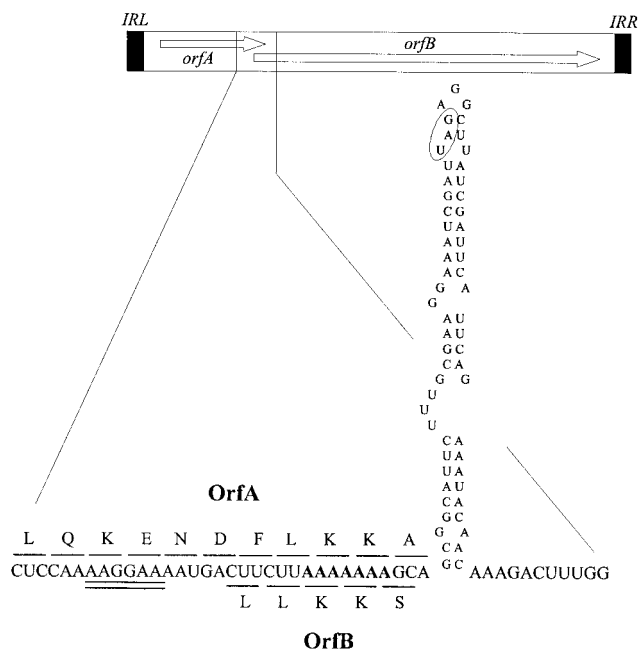
10812), and rtX (positions 12176 to 12142). The results obtained demonstrated that E1-Bac⁻ conserved the genes *rumMTX* belonging to the hypothetical biosynthetic operon, whereas the operons presumably involved in regulation (*rumRK*) and immunity (*rumFEGHR2*) were absent (Fig. 1B).

Taken together, the data obtained by DNA sequencing and PCR amplification showed that the genetic rearrangement that occurred in E1-Bac⁻ resulted in the deletion of at least 5 kb at the 5′ end of the RumA locus. On the other hand, the rearrangement could also affect the hypothetical *rumB* gene cluster. First, sequence comparisons suggested that RumN could be truncated. It is 248 amino acids long, whereas typical LanM proteins consist of over 900 residues. Furthermore, RumN lacks the C-terminal region, which is thought to be involved in the formation of thioether bonds from previously didehydrated residues in lantibiotic prepeptides (17, 21). Second, no other genes typically associated with bacteriocin production, as that encoding the specific transporter, have been found in the vicinity of *rumN*. Further studies are now necessary to analyze the chromosomal region surrounding *rumB* in the wild-type strain and to determine if a type B lantibiotic is really produced.

Sequence analysis and features of ISRgn*1*. Similarities found after database comparisons of nucleotide and amino acid sequences, suggested that ISRgn*1* belonged to the IS*3* subgroup of the IS*3* family (11). ISRgn*1* was 1,278 bp long, with a GC content of 36%. It carried 38-bp imperfect inverted repeats at the ends and was flanked by identical 3-bp direct
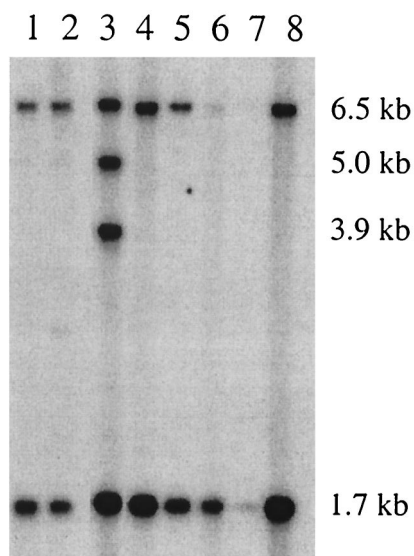
FIG. 3. Southern blot hybridization showing the copy number of ISRgn1 in the genomes of different *R. gnavus* E1 isolates. Sizes of the fragments containing a copy of ISRgn1 are indicated (in kilobases) on the right. The DNA probed in lane 1 was from the same sample used for cloning the RumA locus shown in Fig. 1A. The relatively lower signal intensity displayed by the isolate in lane 7 was due to an incorrect quantification of the DNA concentration in the sample, whereas DNA probed in lane 6 was partially degraded. Lane 8, *R. gnavus* E1 Bac⁻ mutant.

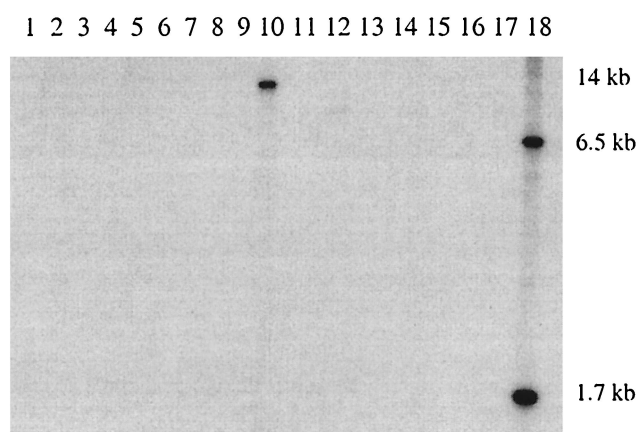FIG. 4. Southern blot hybridization of *Eco*RI-digested total DNA from a selection of *R. gnavus* E1 phylogenetically related strains using an ISRgn1-specific probe. Sizes of the fragments containing a copy of the insertion sequence are indicated (in kilobases) on the right. Lane 1, *R. gnavus* ATCC 29149ᵀ; lane 2, *R. gnavus* LEMG25; lane 3, *R. gnavus* LEMV95; lane 4, *R. gnavus* LEMV99; lane 5, *R. gnavus* LEMB86; lane 6, *R. gnavus* LEMV62; lane 7, *R. gnavus* LEMV50; lane 8, *R. gnavus* LEMV58; lane 9, *R. gnavus* LEMV66; lane 10, *R. gnavus* LEMB53; lane 11, *R. hansenii* DSMZ 20583ᵀ; lane 12, *R. hansenii* LEMV98; lane 13, *C. nexile* DSMZ 1787ᵀ; lane 14, *C. nexile* LEMB04; lane 15, New operational taxonomic unit (OTU) LEMV61; lane 16, New OTU LEMV78; lane 17, New OTU LEMV63; lane 18, *R. gnavus* E1 Bac⁻, which was included as a positive control. ATCC and DSMZ strains were laboratory reference strains from American Type Culture Collection (Manassas, Va.) and Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Braunschweig, Germany), respectively. The other strains were from our laboratory collection and are available upon request.

repeats, probably generated by duplication of the target insertion site. Another IS3-conserved feature was the presence of the terminal dinucleotides TG (5′) and CA (3′).

ISRgn1 contained two overlapping ORFs, *orfA* and *orfB*, in the reading phases 0 and −1, respectively. The putative proteins deduced from these *orf* genes exhibited a high degree of similarity to transposases of the IS3 family (between 30% and 56% of identity). The predicted product of *orfA* was a protein of 97 amino acids possibly starting at an ATG initiation codon. ISRgn1-OrfA contained a putative DNA-binding helix-turn-helix motif, which could be involved in the sequence-specific recognition of the terminal inverted repeats by the transposase (20). *orfB* encoded a hypothetical protein of 303 amino acids containing a highly conserved DDE motif, which has been proposed to be the catalytic site of IS3 transposases and retroviral integrases (3, 13).

As other IS3 members, the nucleotide sequence of ISRgn1 contained a conserved frameshift motif (2, 4), 5′-AAAAAAA-3′, in the 46-bp region of overlap between *orfA* and *orfB*. For some IS3 elements, it has been demonstrated that the functional transposase is an OrfAB fusion protein generated by a −1 translational frameshift as a consequence of ribosome slippage in this motif (14, 22). The efficiency of the frameshifting depends also on the presence of an RBS and a secondary structure upstream and downstream, respectively, of the frameshift window (4). As is shown in Fig. 2, these features were present in the *orfA-B* overlapping region of ISRgn1.

Genomic copy number and distribution of ISRgn1. To determine the genomic copy number of ISRgn1, total DNA was obtained from the E1 Bac⁻ mutant and seven independent isolates of *R. gnavus* E1 exhibiting a wild-type phenotype.

DNAs were digested with *Eco*RI and subsequently hybridized with an internal *Eco*RV-*Eco*RI fragment of 0.9 kb from ISRgn1. Since one cleavage site for *Eco*RI is present in the cloned ISRgn1 but not in the probe (Fig. 1B), the number of hybridization bands was assumed to be the number of the IS copies.

As is shown in Fig. 3, two hybridizing fragments, with molecular sizes of 6.5 and 1.7 kb, were detected in the genome of the E1-Bac⁻ mutant (lane 8) as well as in six other E1 isolates. These results suggested that chromosomal copies of the IS have the same localization in the E1 wild-type strain and in the nonproducer mutant and, consequently, that ISRgn1 has not transposed in E1-Bac⁻. However, the presence of two additional copies of ISRgn1 in the genome of one E1 isolate (Fig. 3, lane 3) strongly suggested that ISRgn1 is an active insertion sequence that could transpose in *R. gnavus* E1. Nevertheless, the mechanism by which ISRgn1 transposes is unclear. The fact that the copy number of the IS increases suggests a replicative mechanism of transposition, but these data are also consistent with a nonreplicative process followed by segregation of the donor molecule (11).

The presence of ISRgn1 in the genome of 17 *R. gnavus* E1-phylogenetically related strains was also investigated. All belonged to the *Clostridium coccoides* phylogenetically defined group (RDP registration no. 2.30.4.1) (12), which includes some of the predominant bacterial genera found in the human large bowel (18). Fourteen of these strains had been isolated from human fecal samples according to their capability to

produce a trypsin-dependent antimicrobial substance active against *Clostridium perfringens* (12a).

Strains were cultivated as described by Dabard et al. (5), and *Eco*RI-digested total DNAs were hybridized with the *Eco*RV-*Eco*RI fragment of ISRgn*1* (Fig. 4). The results obtained demonstrated that ISRgn*1* was present in a single copy in the genome of only one tested strain, *R. gnavus* LEMB53 (Fig. 4, lane 10).

Restricted distribution of ISRgn*1* among closely E1-related strains, along with the discrepancy between the GC content of the sequenced copy of IS (36%) and that of the *R. gnavus* E1 genome (43%; 12a), suggest that ISRgn*1* is not indigenous to *R. gnavus* but could be acquired by horizontal transfer from a microorganism with a lower GC%.

In fact, results reported by Marcille et al. strongly suggest that genes encoding RumA-like bacteriocins might be located in a mobile genetic element able to transfer among bacteria belonging to the *C. coccoides* phylogenetic group (12a). Several lantibiotics, such as nisin, lacticin 3147, and lacticin 481, are encoded by gene clusters located in transposons (10) or composite transposons (7, 8).

The characterization of such a mobile structure and its putative participation in the dissemination of adaptive functions among bacteria inhabiting the digestive ecosystem will be the subject of future research.

Nucleotide sequence accession number. The sequence reported here has been deposited in the GenBank Sequence Database from the National Center for Biotechnology Information (Bethesda, Md.) under accession number AF320327.

REFERENCES

1. **Altena, K., A. Guder, C. Cramer, and G. Bierbaum.** 2000. Biosynthesis of the lantibiotic mersacidin: organization of a type B lantibiotic gene cluster. Appl. Environ. Microbiol. **66:**2565–2571.
2. **Atkins, J. F., R. B. Weiss, and R. F. Gesteland.** 1990. Ribosome gymnastics—degree of difficulty 9.5, style 10.0. Cell **62:**413–423.
3. **Baker, T. A., and L. Luo.** 1994. Identification of residues in the Mu transposase essential for catalysis. Proc. Natl. Acad. Sci. USA **91:**6654–6658.
4. **Chandler, M., and O. Fayet.** 1993. Translational frameshifting in the control of transposition in bacteria. Mol. Microbiol. **7:**497–503.
5. **Dabard, J., C. Bridonneau, C. Phillipe, P. Anglade, D. Molle, M. Nardi, M. Ladire, A. Gomez, and M. Fons.** 2001. Ruminococcin A, a new lantibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces. Appl. Environ. Microbiol. **67:**4111–4118.
6. **de Vos, W. M., G. Jung, and H.-G. Sahl.** 1991. Appendix: definitions and nomenclature of lantibiotics, p. 457–464. *In* G. Jung and H.-G. Sahl (ed.), Nisin and novel lantibiotics. Escom Publishers, Leiden, The Netherlands.
7. **Dougherty, B. A., C. Hill, J. F. Weidman, D. R. Richardson, J. C. Venter, and R. P. Ross.** 1998. Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasmid pMRC01 from Lactococcus lactis DPC3147. Mol. Microbiol. **29:**1029–1038.
8. **Dufour, A., A. Rince, P. Uguen, and J. P. Le Pennec.** 2000. IS1675, a novel lactococcal insertion element, forms a transposon-like structure including the lacticin 481 lantibiotic operon. J. Bacteriol. **182:**5600–5605.
9. **Gomez, A., M. Ladire, F. Marcille, and M. Fons.** 2002. Trypsin mediates growth phase-dependent transcriptional regulation of genes involved in biosynthesis of ruminococcin A, a lantibiotic produced by a *Ruminococcus gnavus* strain from a human intestinal microbiota. J. Bacteriol. **184:**18–28.
10. **Kuipers, O. P., M. M. Beerthuyzen, R. J. Siezen, and W. M. De Vos.** 1993. Characterization of the nisin gene cluster nisABTCIPR of Lactococcus lactis. Requirement of expression of the nisA and nisI genes for development of immunity. Eur. J. Biochem. **216:**281–291.
11. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
12. **Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt, and J. M. Tiedje.** 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. **29:**173–174.
12a. **Marcille, F., A. Gomez, P. Joubert, G. Veau, M. Ladire, A. Clara, F. Gavini, A. Willems, and M. Fons.** 2002. Distribution of genes encoding the trypsin-dependent lantibiotic ruminococcin A among bacteria isolated from human fecal microbiota. Appl. Environ. Microbiol. **68:**3424–3431.
13. **Polard, P., and M. Chandler.** 1995. Bacterial transposases and retroviral integrases. Mol. Microbiol. **15:**13–23.
14. **Polard, P., M. F. Prere, M. Chandler, and O. Fayet.** 1991. Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS911. J. Mol. Biol. **222:**465–477.
15. **Sahl, H. G., and G. Bierbaum.** 1998. Lantibiotics: biosynthesis and biological activities of uniquely modified peptides from gram-positive bacteria. Annu. Rev. Microbiol. **52:**41–79.
16. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
17. **Siezen, R. J., O. P. Kuipers, and W. M. de Vos.** 1996. Comparison of lantibiotic gene clusters and encoded proteins. Antonie Leeuwenhoek **69:**171–184.
18. **Suau, A., R. Bonnet, M. Sutren, J. J. Godon, G. R. Gibson, M. D. Collins, and J. Dore.** 1999. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Appl. Environ. Microbiol. **65:**4799–4807.
19. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levin, O. C. Uhlenbeck, D. M. Crothers, and J. Bralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nat. New Biol. **246:**40–41.
20. **Ton-Hoang, B., P. Polard, and M. Chandler.** 1998. Efficient transposition of IS911 circles in vitro. EMBO J. **17:**1169–1181.
21. **Uguen, P., J. P. Le Pennec, and A. Dufour.** 2000. Lantibiotic biosynthesis: interactions between prelacticin 481 and its putative modification enzyme, LctM. J. Bacteriol. **182:**5262–5266.
22. **Vogele, K., E. Schwartz, C. Welz, E. Schiltz, and B. Rak.** 1991. High-level ribosomal frameshifting directs the synthesis of IS150 gene products. Nucleic Acids Res. **19:**4377–4385.