[SVI-2]

A Group I Self-splicing Intron in the *recA* Gene of *Bacillus anthracis*

Minsu Ko[1], Hyang Choi[2] and Chankyu Park

Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Yusong-Ku, Taejon 305-701, Republic of Korea

[1]Present address: Institute of Molecular Genetics, Solgent Inc.

[2]Present address: Institute of Science & Technology, Cheil-Jedang Co.

The self-splicing introns are rarely found in bacteria and bacteriophages. They are classified into group I and II according to their structural features and splicing mechanisms. While the group I introns are occasionally found in protein coding regions of phage genomes and in several tRNA genes of cyanobacteria and proteobacteria, they were never found in protein-coding regions of bacterial genomes. Here we report a group I intron in the *recA* gene of *Bacillus anthracis*, which was initially found by DNA sequencing as an intervening sequence (IVS). Using reverse transcriptase PCR, the IVS was shown to be removable from the *recA* precursor mRNA for RecA being translated in *E. coli*. The splicing was visualized *in vitro* with labeled free GTP, indicating that it is a group I intron, which is also implied by its predicted secondary structure. The RecA protein of *B. anthracis* expressed in *E. coli* was functional in its ability to complement a *recA* defect. When *recA*-negative *E. coli* cells were irradiated with UV, the *Bacillus* RecA reduced the UV susceptibility of the *recA* mutant, regardless of the presence of intron.

While most of the introns found in eukaryotes and archaebacteria require functions of several proteins for their processing, introns found in eubacterial species are removed by self-splicing. The bacterial group I introns are found in several tRNA genes of cyanobacteria and purple proteobacteria (13, 20, 25). Some protein-coding genes, associated with DNA metabolism, of phages from *Escherichia coli*, *Bacillus subtilis*, *Lactococcus lactis*, *Lactococcus delbruckii*, and *Staphylococcus aureus*, contain group I introns (1, 3, 11, 15, 24). However, no group I intron has been reported so far from a protein-coding gene of an eubacterial genome. The group II introns of bacteria have been identified in cyanobacteria (*Calothrix* spp.), proteobacteria (*Azotobacter vinelandii* and *E. coli*, 9), and gram-positive bacteria (*L. lactis* and *Chlostridium difficile*, 19, 23). Most of the eubacterial group II introns are found in genes on plasmids or associated with conjugal transfer (7).

The bacterial RecA protein is ubiquitous and highly conserved (12). It is involved in homologous recombination, DNA repair, and in promoting proteolysis (18, 21) In *E. coli*, the transcription of the *recA* gene is negatively regulated by the LexA protein, which binds to the upstream regions of the SOS genes including *recA* (17).

Several *recA* genes of mycobacteria are interrupted by an intein that is removed from its precursor by a

protein splicing mechanism to generate a functional RecA protein (5, 6). The inteins in six species including *M. leprae* and *M. chitae* were found in the same insertion sites (RecA-b) immediately downstream of glycine 205 in *M. leprae*, while the one (RecA-a) for *M. tuberculosis* is after the lysine 251 (22).

During the study of the *recA* genes from several gram-positive bacteria, we discovered that a putative *recA* gene of *Bacillus anthracis* (BA) is interrupted by an intervening sequence (IVS). Here we show that the IVS is a group I intron and that the *recA* gene codes for a functional RecA protein in *E. coli*. This is the first demonstration of the presence of a group I intron in a protein-coding gene of a bacterial genome.

Results

A putative *recA* gene from *B. anthracis*

A partial sequence (292 bp) of a putative *recA* gene of *B. anthracis* was obtained by PCR amplification using a set of degenerate primers recognizing conserved regions of the *recA* gene in Gram-positive bacteria (Y. Chae, unpublished). This sequence is homologous to the region encoding amino acid residues 104-202 of the *B. subtilis* RecA with 96 % identity. Using inverse PCR technique with a pair of outwardly directed primers, we were able to determine the full sequence of the *B. anthracis recA* gene (Materials and Methods for details). As shown in Fig.1, the putative *recA* gene is interrupted by a 327 bp fragment after amino acid residue 203. This intervening sequence (IVS) contains eleven stop codons when translated continuously. There is no inverted repeat at the ends of the IVS, indicating that the IVS is neither an IS element nor a transposon.

In the promoter region of the *recA*-like coding sequence are putative –10 and –35 promoter sequences between 32 and 63 bp upstream of the predicted start codon. The putative promoter sequence (TTGGCA-18 bp-TATAAT) is very similar to the *E. coli* consensus (TTGACA-16 to 18 bp-TATAAT). The putative LexA binding site, located 72-83 bp upstream of the start codon, is exactly the same as that of *B. subtilis recA*. The nucleotide and deduced amino acid sequ ences of the *B. anthracis recA* gene have been submitted to GenBank under accession no. of AF229167.

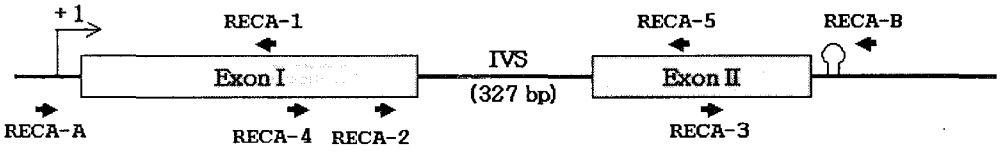

Fig. 1. Schematic diagram of *B. anthracis recA* gene. The gene includes an intervening sequence (IVS) with a size of 327 bp, located after the residue 203. The thick arrows indicate locations of oligonucleotides used. The predicted transcriptional start site and the termination loop are also indicated. The exons are shadowed.

In vivo and *in vitro* splicings of the intervening sequence

We examined whether the IVS of putative *recA* gene is excisable from the primary transcript. Using a pair of primers, RECA-4 and RECA-5 as in Fig.1, RT-PCR was carried out for total RNA from *B. anthracis*,

resulting in a product smaller than that from the genomic DNA (Fig. 2, Left). Essentially the same result was obtained for total RNA extracted from *E. coli* containing *B. anthracis recA* (DH10β/pUC*recA*[int$^+$], data not shown). The size difference matches to the size of the IVS, indicating that the IVS was excised, which was also confirmed by DNA sequencing. The spliced form of the mRNA encodes an open reading frame of 343 amino acids, showing 85 % identity with the RecA protein of *B. subtilis*.

Using the clone of *recA* whose transcription was under the control of T7 promoter, we obtained RNA *in vitro* using T7 RNA polymerase from a pT7*rec*[int$^+$] plasmid linearized with *Xba*I (Materials and Methods). When RT-PCR was performed using RECA-4 and RECA-5 primers, a similar result was obtained as in the *in vivo* splicing experiment with total RNA from *B. anthracis* (data not shown), indicating that the splicing of $recA_{BA}$ (the *recA* gene of *B. anthracis*) occurs by a self-splicing mechanism.

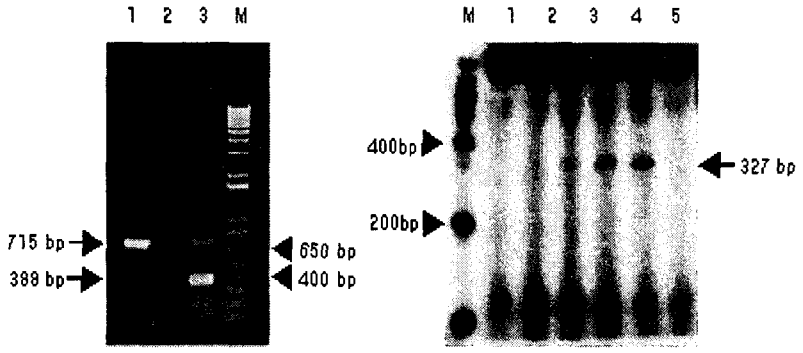

Fig. 2. *In vivo* and *in vitro* splicing of *recA* transcript. [Left] *In vivo* splicing of *recA* transcript. The lane 1 shows a PCR product obtained from *B. anthracis* chromosomal DNA using the RECA-4 and RECA-5 primers. Results of RT-PCR from total RNA of *B. anthracis* using RECA-4 and RECA-5 are also shown, in the absence (lane 2) or presence (lane 3) of reverse transcriptase. M represents the DNA size markers. [Right] Labeling of intron RNA with [α-^{32}P]GTP. Total RNAs were extracted from strains with various plasmids as described in Materials and Methods, and incubated with [α-^{32}P]GTP at 37℃ in a splicing buffer : lane 1, BL21/pBluescript SK(+) incubated for 60 min; lane 2-5, BL21/pT7*recA* [int$^+$] incubated for 0 (2), 10 (3), 30 (4), and 60 min (5); lane 6, BL21/pT7*recA* [int$^-$] incubated for 60 min. After the reaction, samples were analyzed by 5% acrylamide /8 M urea gel electrophoresis and exposed to an X-ray film. The size of detected band was approximately the same as that of the intron (327 bp). The ^{32}P-labeled DNA ladder was used as a marker (M).

Secondary structure of the *recA* intron

Prediction of secondary structure for the *recA* intron revealed the distinctive features of a group I intron (Fig. 3). Group I introns contain regions of base-pairing (P1-P10), as well as the conserved sequence elements (P, Q, R, S), which are necessary for proper folding and excision (2). The *B. anthracis recA* intron is spliced after the 'u' residue in the P1 stem that forms a pair with G, and the intron ends with G as in most group I introns. In addition, the 3' terminal guanine of the intron is followed by three residues (CCA) capable of

forming a base-pair (P10) with residues (UGG) in the 3' side of P1. Thus, the nucleotides on the 3'side of P1 may serve as an 'internal guide sequence (IGS)', promoting an alignment between the 3' and 5' splice sites for ligation of the two exons (4). The predicted structure of this intron resembles that of the *bnrdE*-I1 group I intron inserted in the ribonucleotide reductase gene of the SPβ-related prophage from *Bacillus* spp. BSG40 (16). Most of the nucleotide sequences in the catalytic core, stem-loops, and single-stranded regions are identical (see Fig. 3). Structural differences lie in the extra pairings of P1a and P1b, besides of an absence of an open reading frame in the *recA* intron after P6a.

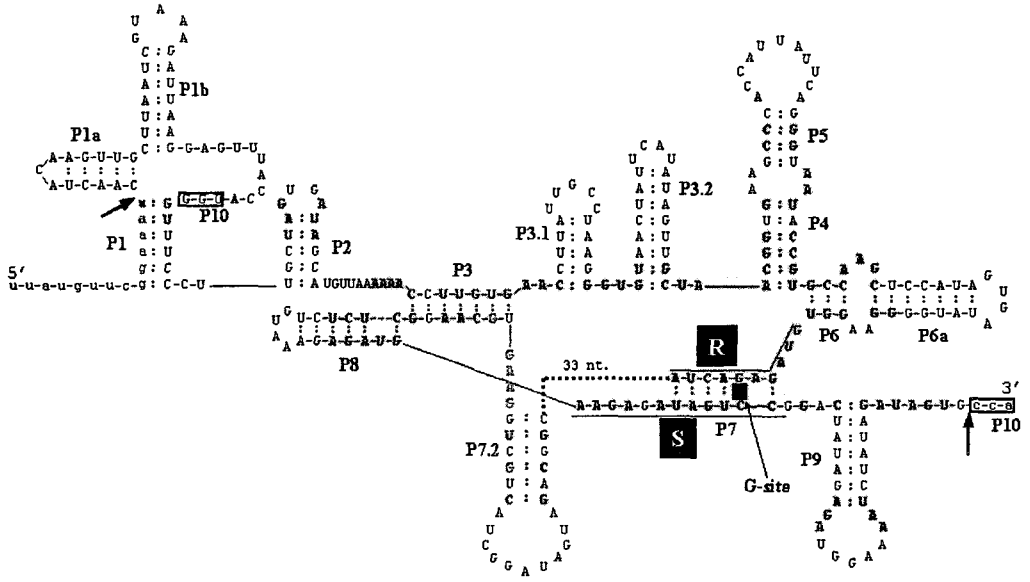

Fig. 3. Predicted secondary structure of the *recA* intron. The arrows indicate the boundaries between exons (lowercase) and intron (uppercase). The conserved base-paired regions (P1-P9), extra pairings (P1a, P1b, P3.1, P3.2, P6a, P7.2), and conserved sequence elements (R and S) are shown. The boxed residues, UGG near P1 and CCA near P9, can make a pair (P10) to promote an alignment between the 3' and 5' splice sites required for the ligation of exons. The G-site (black box) is responsible for the binding of the guanosine cofactor. The shaded boxes represent residues identical with the *bnrdE*-I1 intron from the Bacillus strain BSG40 (18).

The group I-specific splicing of the *recA* gene

To test whether the splicing of the *recA* intron is mechanistically similar to group I introns, we carried out an *in vitro* splicing reaction with ^{32}P-labeled GTP. Total RNAs were extracted from IPTG induced cells of BL21[DE3] containing pBluescript SK(+), pT7*recA*[int$^+$], or pT7*recA*[int$^-$], and incubated with [α-^{32}P] GTP. After separation on urea-acrylamide gel electrophoresis, a labeled RNA product whose density increased over incubation time was detected from the sample containing the intron (Fig. 2, Right). The size of that band matched well with that of the predicted intron (328 nucleotides including GTP). The same result was obtained with the pUC*recA*[int$^+$] clone, but the band intensity was weak (data not shown) due to a relatively low copy number of

the precursor RNA. From this result, we concluded that the IVS of the $recA_{BA}$ gene is a self-splicing group I intron.

Complementation of a *recA* defect in *E. coli* by the RecA from *B. anthracis*

In order to assess the level of $RecA_{BA}$ expression in *E. coli*, we performed a western blotting with an anti-RecA antiserum raised against the *E. coli* RecA protein. As shown in Fig. 4A, the $RecA_{BA}$ proteins synthesized in *E. coli* (MG1655 Δ*recA*::Cm) from the plasmids pUC*recA*[int$^+$] and pUC*recA*[int$^-$] were identical in size and amount regardless of the presence of the intron. The size of $RecA_{BA}$ produced in *E. coli* was similar to that of $RecA_{EC}$.

To further investigate the functionality of $RecA_{BA}$, we carried out a complementation test in terms of its ability to protect against DNA-damage caused by UV irradiation. As shown in Fig. 4B, the $recA_{BA}$ gene was able to complement the *recA* defect and the degree of complementation for UV treatment was indistinguishable in the intron-positive from intron-negative *recAs*. This indicates that the $recA_{BA}$ gene produces a functional RecA protein, which is unaffected by the presence of an intron.

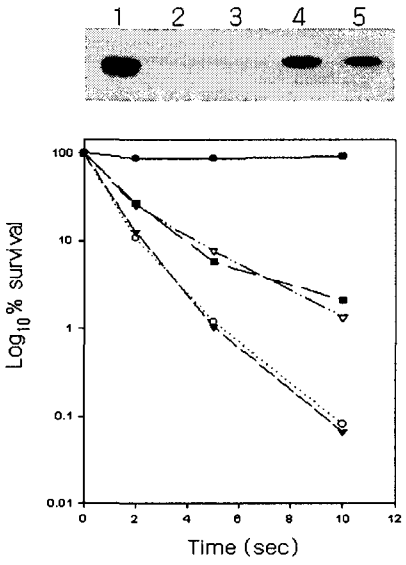

Fig. 4. Complementation of *recA* deletion in *E. coli* by expression of *B. anthracis recA*. [Top] The *Bacillus* RecA protein expressed in *E. coli*. Total proteins from the strains indicated were analyzed by immunoblotting with an anti-$RecA_{EC}$ antiserum after SDS-polyacrylamide (12%) gel electrophoresis and transfer to a nitrocellulose paper: lane 1, MG1655; 2, MG1655Δ*recA*; 3, MG1655Δ*recA*/pUC19; 4, MG1655Δ*recA* /pUC*recA* [int$^+$]; 5, MG1655Δ *recA*/pUCrecA [int$^-$]. The size of the *B. anthracis* RecA protein was about 37 kD, which is similar to that of *E. coli* RecA. [Bottom] Complementation of *E. coli recA* defect by a cloned *B. anthracis recA*. Strains used are MG1655 (●), MG1655Δ *recA* (O), MG1655Δ *recA*/pUC19 (▼), MG1655Δ *recA*/ pUCRecA [int$^+$] (▽), and MGΔ *recA*/pUCrecA [int$^-$] (■). The numbers of colonies were counted after UV irradiation and plating the samples onto LB agar plates. The plates were incubated for 20 hrs at 37℃.

Discussion

During the cloning and sequencing of the *recA* gene in *B. anthracis*, we found an intervening sequence (IVS), which turned out to be a group I self-splicing intron. The *in vivo* splicings of *B. anthracis* and *E. coli* were demonstrated by an RT-PCR experiment. It was also shown that the excised intron during splicing is labeled with [α-^{32}P]GTP (Fig. 2B). These experimental observations are consistent with the prediction of a group I-like secondary structure for the IVS. The high similarity of predicted secondary structure and sequence of the *B. anthracis recA* intron to that of the *bnrdE*-I1 intron from *B. subtilis* BSG40 may suggest a lateral transfer of these introns, most likely by transposition (7). The first identified microbial group I intron was the one in the thymidylate synthetase gene (*td*) of bacteriophage T4 (3). About seventeen introns in protein-coding genes of bacteriophages have been reported (7), besides seventy-four introns in tRNAs and rRNA of bacteria. However, the group I introns have so far never been found in a protein-coding gene of a bacterial genome. Most group I introns found on the bacterial genomes are contained in a prophage, not originated from genome. Although the group I intron of the thymidylate synthase gene was reported in *Bacillus mojavensis* s87-18 (8), it was not clearly stated whether it was from prophage or genomic origin. One could imagine that the *recA* gene presented here might be from a prophage and another genomic copy of *recA* gene exists. Through a search for *recA* sequence from *B. anthracis* (Ames strain) genome sequences in TIGR databases (www.tigr.org), we found a contig (#6264) containing the *recA* gene with its intron. Analysis of the surrounding nucleotide sequences revealed the gene order of that region as *pgsA-cinA-recA-ymdA*, which is also preserved in the *Bacillus subtilis* genome except that *B. subtilis* contains an additional *pbpX* gene between the *recA* and *ymdA* genes. The orientations of the genes are identical, suggesting that the *recA* gene with intron is a genomic copy from *B. anthracis*. Therefore, the *recA* intron characterized here would be the first case of the group I intron in a genomic protein-coding gene of bacteria.

As previously indicated, there are certain characteristics shared by the group I introns from bactriophages. Most of them are found in protein-coding genes associated with DNA metabolism and contain an ORF for an endonuclease involved in intron mobility. The group I introns found in some of tRNAs or rRNAs do not usually encode a homing endonuclease, although there are a few exceptions. The *recA* intron from *B. anthracis* appears to be associated in some way with DNA metabolism similar to the phage introns, but does not contain a homing endonuclease as found in the group I introns of tRNAs or rRNAs. Since intron mobility depends on the presence of a homing endonuclease (14), the *recA* intron of *B. anthracis* would not be mobile by this mechanism.

The presence of an intron in the *B. anthracis recA* gene seems particularly interesting because some *recA* genes from Mycobacteria are also interrupted by an intein, a protein intron that is removed after translation. There are two types of mycobacterial inteins in terms of their insertion sites. More than eight species of mycobacteria including *M. leprae* have insertions at an identical site (RecA-b, 22), which is located immediately downstream of glycine 205 in *M. leprae*, while the intein from *M. tuberculosis* is found in other location (RecA-a) further downstream. It was proposed that the inteins were acquired independently (6, 22). It is striking that the insertion site of the *B. anthracis recA* intron is nearly identical to the RecA-b site of

Mycobacterium with only one amino acid shift. Their origins might be evolutionarily related or not. The inserting an intron and intein at almost the same location with one codon difference has previously been reported from the *nrdE* gene of SPβ-related prophage of *Bacillus* (16). Currently, no role for these inteins has been assigned, and the presence of an intein does not affect RecA activity (10), pathogenecity, or the survival of the host organism. The presence of an intron in the *B. anthracis recA* gene does not itself affect the expression or activity of RecA. It is possible that the *recA* intron might play some role during evolution, providing an advantage in adaptation or survival, although its function in current species seems dispensable.

References

1. Bechhofer, D. H., K. K. Hue, and D. A. Shub. 1994. An intron in the thymidylate synthase gene of Bacillus bacteriophage B22: evidence for independent evolution of a gene, its group I intron, and the intron open reading frame. Proc. Natl. Acad. Sci. USA 91:11669-11673.
2. Cech, T. R. 1990. Self-splicing of group I introns. Annu. Rev. biochem. 59:543-568.
3. Chu, F. K., G. F. Maley, F. Maley, and M. Belfort. 1984. Intervening sequence in the thymidylate synthase gene of bacteriophage T4. Proc. Natl. Acad. Sci. USA 81:3049-3053.
4. Davies, R.W., R. B. Waring, J. A. Ray, T. A. Brown, and C. Scazzocchio. 1982. Making ends meet: a model for RNA splicing in fungal mitochondria. Nature 300:719-724.
5. Davis, E. O., P. J. Jenner, P. C. Brooks, M. J. Colston, and S. G. Sedgwick. 1992. Protein splicing in the maturation of M. tuberculosis RecA protein: a mechanism for tolerating a novel class of intervening sequence. Cell 71:201-210.
6. Davis, E. O., H. S. Thangaraj, P. C. Brooks, and M. J. Colston. 1994. Evidence of selection for protein introns in the recAs of pathogenic mycobacteria. EMBO J. 13:699-703.
7. Edgell, D. R., M. Belfort, and D. A. Shub. 2000. Barriers to intron promiscuity in bacteria. J. Bacteriol. 182:5281-5289.
8. Edgell, D.R. and Shub, D.A. 2001. Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. Proc. Natl. Acad. Sci. USA 98: 7898-7903.
9. Ferat, J. L., and F. Michel. 1993. Group II self-splicing introns in bacteria. Nature 364:358-361.
10. Frischkorn, K., P. Sander, M. Scholz, K. Teschner, T. Prammananan, and E. C. Bottger. 1998. Investigation of mycobacterial recA function: protein introns in the RecA of pathogenic mycobacteria do not affect competency for homologous recombination. Mol. Microbiol. 29:1203-1214.
11. Goodrich-Blair, H., V. Scarlato, J. M. Gott, M. Q. Xu, and D. A. Shub. 1990. A self-splicing group I intron in the DNA polymerase gene of Bacillus subtilis bacteriophage SPO1. Cell 63:417-424.
12. Karlin, S., G. M. Weinstock, and V. Brendel. 1995. Bacterial classifications derived from RecA protein sequence comparisons. J. Bacteriol. 177:6881-6893.
13. Kuhsel, M. G., R. Strickland, and J. D. Palmer. 1990. An ancient group I intron shared by eubacteria and chloroplasts. Science 250:1570-1573.
14. Lambowitz, A. M., and M. Belfort. 1993. Introns as mobile genetic elements. Annu. Rev. Biochem. 62:587-622.

15. Landthaler, M., and D. A. Shub. 1999. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. Proc. Natl. Acad. Sci. USA 96:7005-7010.
16. Lazarevic, V. 2001. Ribonucleotide reductase genes of Bacillus prophages: a refuge to intones and intein coding sequences. Nucleic Acids Research 29:3212-3218.
17. Little, J. W. 1981. Purified lexA protein is a repressor of the recA and lexA genes. Proc. Natl. Acad. Sci. USA 78:4199-4203.
18. Miller, R. V., and T. A. Kokjohn. 1990. General microbiology of recA: environmental and evolutionary significance. Annu. Rev. Microbiol. 44:365-394.
19. Mullany, P., M. Pallen, M. Wilks, J. R. Stephen, and S. Tabaqchali. 1996. A group II intron in a conjugative transposon from the gram-positive bacterium, Clostridium difficile. Gene 174:145-150.
20. Reinhold-Hurek, B., and D. A. Shub. 1992. Self-splicing introns in tRNA genes of widely divergent bacteria. Nature 357:173-176.
21. Roca, A. I., and M. M. Cox. 1990. The RecA protein: structure and function. Crit. Rev. Biochem. Mol. Biol 25:415-456.
22. Saves, I., M. A. Laneelle, M. Daffe, and J. M. Masson. 2000. Inteins invading mycobacterial RecA proteins FEBS Lett. 480:221-225.
23. Shearman, C., J. J. Godon, and M. Gasson. 1996. Splicing of a group II intron in a functional transfer gene of Lactococcus lactis. Mol. Microbiol. 21:45-53.
24. van Sinderen, D., H. Karsens, J. Kok, P. Terpstra, M. H. Ruiters, G. Venema, and A. Nauta. 1996. Sequence analysis and molecular characterization of the temperate Lactococcal bacteriophage r1t. Mol. Microbiol. 19:1343-1355.
25. Xu, M. Q., S. D. Kathe, H. Goodrich-Blair, S. A. Nierzwicki-Bauer, and D. A. Shub. 1990. Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. Science **250**:1566-1570.