A Self-Splicing Group I Intron in DNA Polymerase Genes of T7-Like Bacteriophages

Richard P. Bonocora and David A. Shub*

Department of Biological Sciences, State University of New York at Albany, Albany, New York

Received 1 July 2004/Accepted 18 August 2004

Group I introns are inserted into genes of a wide variety of bacteriophages of gram-positive bacteria. However, among the phages of enteric and other gram-negative proteobacteria, introns have been encountered only in phage T4 and several of its close relatives. Here we report the insertion of a self-splicing group I intron in the coding sequence of the DNA polymerase genes of ΦI and W31, phages that are closely related to T7. The introns belong to subgroup IA2 and both contain an open reading frame, inserted into structural element P6a, encoding a protein belonging to the HNH family of homing endonucleases. The introns splice efficiently in vivo and self-splice in vitro under mild conditions of ionic strength and temperature. We conclude that there is no barrier for maintenance of group I introns in phages of proteobacteria.

The first demonstration of an intron in a bacterium was the self-splicing group I intron in the thymidylate synthase (td) gene of Escherichia coli phage T4 (1). Group I introns were subsequently discovered in tRNA genes of cyanobacteria (12, 24) and proteobacteria (20), in the tRNA genes of several Thermotoga species (16), and in a variety of phages of gram-positive bacteria and cyanobacteria (9, 15, 18). In view of the intensive investigation of E. coli phages it is surprising that, in the 20 years since their initial discovery in the td gene of T4, the only new examples in phages of E. coli or other proteobacteria are two additional introns in genes of T4 and a small number of its close relatives (the T-even phages) (8, 9, 23).

The apparent limitation of introns to phages of the T-even family might be due to a special property of these phages that would overcome a disadvantage that self-splicing group I introns might confer on other phages of proteobacteria. For example, the large genome of T-even phages (1.7 × 10⁵ bp) might accommodate introns more easily than the genomes of smaller phages, or the extensive modifications of cellular metabolism by T-even phages might overcome some unknown E. coli metabolic process that makes group I intron splicing deleterious to gene expression.

Very early in its history, a group of workers agreed to limit the study of phages of E. coli to seven isolates (phages T1 to T7) (2, 5). Although lambdoid phages and single-stranded DNA and RNA phages eventually became the subjects of intensive study, most of the literature on phages of E. coli concerns a relatively small number of phage isolates. The complete genomic sequences of phages T1, T3, T5, and T7 (GenBank accession numbers NC005833, AJ318471, NC005859, and NC001604) show no evidence of introns (7, 17, 21), supporting a preliminary study that failed to find evidence of introns in these phages (4). Even among the T-even phages, introns are relatively rare (8, 9, 23). If T2, which probably has no introns, had been selected as the model phage for genetics rather than T4, there would have been no reason to conduct a survey of the other T-even phages for introns (8), and the presence of introns in E. coli phages might still be unknown today.

Therefore, the apparent restriction of introns to T4 and its relatives might be due to a lack of comparative data. We screened several isolates of T7-like phages for variability over a 1-kb region by PCR using primers T75.0 (GACCGAAGAG ATGTCGTAGAG) and T75.7 (GTACGGAAACATGCGA GCCTC), which correspond to identically conserved sequences in genes 5 and 5.7 of phages T3 and T7. Eight phage DNAs (ΦI, ΦIIP, ΦI1W, H, and W31, which were obtained from W. F. Studier; and A1122, C21R, and W111, obtained from J. Moloney) yielded products in the PCR, with two of them (ΦI and W31) migrating slower than the others (data not shown).

A group I intron in phages ΦI and W31. In both phages the larger sizes of the PCR products are accounted for by 601-bp insertions 156 bp from the ends of their DNA polymerase genes (gene 5). These insertions can be folded into structures typical of group IA2 introns (3, 14) with a 131-codon open reading frame (ORF) inserted into stem P6a (Fig. 1). The introns are inserted between codons in a conserved region of DNA polymerases (R. P. Bonocora and D. A. Shub, submitted for publication) that specifies a metal ion binding site in the catalytic center of the T7 enzyme (6). Overall, the introns are 86% identical, while the phylogenetically conserved intron structures (without the ORFs) are 97% identical. Most of the sequence divergence occurs between the intron-encoded ORFs, which comprise more than half of the intron and are 81 and 86% identical at the nucleotide and amino acid levels, respectively.

Intron secondary structure. The introns contain helical elements P7.1 and P7.2 separated by a G-U-A sequence, features indicative of group IA2 introns (14), but lack the optional P2 helix. Although they lack a discernible P9.0 interaction at their 3’ ends, this may be compensated by long P10 pairings between the first eight nucleotides of the 3’ exon and a sequence near the 5’ end of the intron (13). All but one of the sequence

* Corresponding author. Mailing address: Department of Biological Sciences, State University of New York at Albany, 1400 Washington Ave., Albany, NY 12222. Phone: (518) 442-4324. Fax: (518) 442-4767. E-mail: shub@albany.edu.
differences in paired regions either preserve base pairing or are accompanied by a compensatory change (Fig. 1). The lone exception is in the extended P6 stem-loop region that is part of the coding sequence of the ORF.

The ribosome binding sites (RBS) of the intronic ORFs are sequestered in the secondary structure. Similarly, the ORFs in the three introns in phage T4 each have an RBS sequestered in stable base-paired helices with intron sequences. These structures are involved in temporal regulation of ORF expression, preventing translation from the early-transcribed introns. Instead, the ORFs are independently expressed from transcripts under control of T4 late promoters located immediately upstream of the RBS (11). However, no T7-like promoter can be identified upstream of the intronic ORFs of \( \Phi 9021 \) or W31.

We have shown elsewhere that the intronic ORFs of both phages encode an endonuclease (I-TslI) that cleaves the intronless allele 1 bp from the intron insertion site, activating the introns for homing (Bonocora and Shub, submitted). We suggest that, rather than being involved in temporal regulation of gene expression, sequestering the RBS serves primarily to reduce expression of I-TslI while perhaps also contributing to stability of intron RNA structure. Expression of I-TslI may need to be maintained at levels much lower than those of DNA polymerase (whose transcript it shares), since at high protein concentrations secondary cleavages may occur elsewhere in the genome. In fact, we have observed a secondary cleavage site just 216 bp upstream of the intron insertion site in gene 5 (data not shown). Since this potential cleavage site is not protected by intron insertion, high in vivo concentrations of I-TslI could be very deleterious.

**Intron splicing in vivo.** *E. coli* \( B_{E} \) was grown in Luria-Bertani medium (22) at 37°C to an optical density at 590 nm of 0.3 to 0.35 (2 \( \times \) \( 10^{9} \) CFU/ml) and infected with phage \( \Phi 1 \) at a multiplicity of 4 per cell. RNA was isolated at various times after infection, using an RNAeasy mini-prep kit and RNase-Free DNase (QIAGEN) and assayed for splicing by reverse transcription. Primer \( \Phi 5.0_{5} \) (CAGAAGACACCGGAAGTTCC) was annealed to isolated RNA and incubated at 25°C for 5 min, 37°C for 60 min, and 75°C for 15 min with IMPROMPII reverse transcriptase (Promega) in 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 10 mM dithiothreitol, and 3 mM MgCl\(_{2}\). One-tenth of the reverse transcriptase reaction mixture was used as

![FIG. 1. Secondary structure of the \( \Phi I \) group I intron. Exon sequences are in lowercase letters, and intron sequences are in uppercase letters. Filled arrows indicate splice sites (ss). Open arrows indicate sequence differences between the \( \Phi I \) and W31 introns. Conserved structural elements P1 to P9 are shown. Boxed regions designate the P10 pairing. Filled circles indicate the start and stop codons of the 1-Tsl coding sequence. The putative RBS in stem P6a is indicated by a thick vertical bar. Additionally, the 329-nucleotide sequence of the intron-encoded ORF is indicated.](http://jb.asm.org/)

![FIG. 2. Splicing of the \( \Phi I \) group I intron. A. In vivo splicing. RNA isolated from \( \Phi I \)-infected *E. coli* was analyzed by reverse transcription-PCR. The times of RNA isolation (in minutes) after infection are indicated. The sizes (in kilobases) of prominent bands of the DNA size standard (lane M) are indicated on the left. Lane U contains RNA isolated from uninfected *E. coli*. PCR products generated from \( \Phi I \) and T7 DNA are shown on the right. B. In vitro self-splicing. GTP labeling was performed on the RNA preparations shown in panel A. DNA size standards were 5’ end labeled with \( ^{32} \)P.](http://jb.asm.org/)
template for PCRs using Taq DNA polymerase with primers \( \Phi I5.0rev \) and \( T75.0b \) (GAAGATTGGNCAGATTGTGG), and products were analyzed by electrophoresis in a 0.8% agarose gel (Fig. 2A).

The size of the amplification product from unspliced transcripts is expected to be approximately 1.1 kb, whereas the size from the spliced transcript is 0.4 kb. At 5 min postinfection and continuing until 30 min the 0.4-kb product predominates, indicating that the intron is spliced from the transcript in vivo. The virtual absence of a product from the unspliced precursor is likely due to the difficulty of reverse transcription through the highly structured intron RNA.

**Self-splicing in vitro.** RNA preparations can be screened for self-splicing introns in vitro by exploiting the group I intron splicing mechanism, in which the attacking guanosine nucleophile becomes covalently attached to the 5' end of the excised intron (10). When the same RNA preparations were incubated with \([\alpha-^{32}P]GTP (19)\) followed by electrophoresis in 4% de

**addendum in proof**

We have recently learned (D. Savalia, I. Molineux, and K. Severinov, unpublished results) that FP1, a 17-like podophage, has an insertion in its DNA polymerase gene in the same position as, and highly similar to, the intron and homing endonuclease gene reported here.

**REFERENCES**

lineux. 2002. Complete nucleotide sequence and likely recombinatorial ori
ERRATUM

A Self-Splicing Group I Intron in DNA Polymerase Genes of T7-Like Bacteriophages

Richard P. Bonocora and David A. Shub

Department of Biological Sciences, State University of New York at Albany, Albany, New York

Volume 186, no. 23, p. 8153–8155, 2004. Page 8155: In the second line of the Addendum in Proof, “FPI” should be “FP1” and “17-like” should be “T7-like.”