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

MARKUS LANDTHALER AND DAVID A. SHUB*

Department of Biological Sciences and Center for Molecular Genetics, University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222

Communicated by Thomas R. Cech, University of Colorado, Boulder, CO, March 29, 1999 (received for review January 25, 1999)

ABSTRACT Analysis of RNA that can be labeled with GTP indicates the existence of group I introns in genes of at least three transcriptional classes in the genome of Staphylococcus aureus bacteriophage Twort. A single ORF of 142 amino acids (Orf142) is interrupted by three self-splicing group I introns, providing the first example of a phage gene with multiple intron insertions. Twort Orf142 is encoded in a message that is abundant 15–20 min after infection and is highly similar to a late gene product (Orf8) of the morphologically related Listeria phage A511. The introns in orf142 are spliced in vivo and contain all the conserved features of primary sequence and secondary structure of group I introns in subgroup I2 and I3 in orf142 are highly similar, and their intron insertion sites are closely spaced. The presence of transcripts with a skipped exon between these introns indicates that they may fold into a single active ribozyme resulting in alternative splicing. Alternatively, the cleaved 5′ exon preceding I2 may undergo trans splicing to the 3′ exon that follows I3. Regardless of the detailed mechanism, these results demonstrate a new means whereby a single gene can give rise to multiple messenger RNAs.

Group I introns interrupt tRNA, rRNA, and protein-coding genes. Most group I introns have been found in mitochondrial DNA, but they are also present in chloroplasts and eukaryotic viruses, in nuclear rRNA genes of various protists and other lower eukaryotes, and in eubacteria and their phages (1, 2). Most of the previously described group I introns in bacteriophages are inserted in genes coding for proteins involved in DNA synthesis. Introns in E. coli phage T4 are present in nrdB (3, 4) and nrdD/sunY (5), which code for ribonucleotide reductases and td, encoding thymidylate synthase (6). In the Bacillus phage β22, the thymidylate synthase (thy) gene is also interrupted by a group I intron that is highly similar to the T4 td intron (7), whereas the introns of the closely related Bacillus phages SP01, SP82, and φ6 are inserted in the coding regions of the DNA polymerase gene (8). Recently group I introns were also found in genes that are expressed late during infection, the terminase (terL) gene of Lactobacillus phage LL-H (9) and orf40 of Lactococcus phage r1t (10).

Group I introns share a common secondary structure and conserved sequence elements that form the catalytic core (11). The splicing pathway involves two consecutive transesterifications. The first reaction is initiated by a nucleophilic attack by the 3′ OH of exogenous guanosine at the 5′ splice site, resulting in a covalent linkage of the noncoded guanosine to the 5′ splice site, resulting in the formation of a 5′-phosphodiester bond (12). The reaction pathway can be used to detect self-splicing group I introns by incubating deproteinized RNA with radio-labeled GTP in vitro (13). GTP labeling led to the discovery of a number of group I introns in phages (3, 14) and in the tRNA genes of various eubacteria (15, 16).

Although only a core structure is required for splicing, group I introns can vary extensively in length because of the insertion of nonconserved nucleotides. In the phage T4 td and nrdD/sunY introns and in the DNA polymerase introns of phages SP01 and SP82 these extra nucleotides code for DNA endonucleases (17, 18). For the T4 td and nrdD/sunY intron, it has been demonstrated that the intron-encoded endonucleases I-Tev1 and I-TevII, respectively, are involved in the mobility (homing) of the intron into unoccupied sites (ref. 17; for a review on intron mobility, see ref. 19). On the other hand, endonuclease I-HmaII, of the SP82 intron, is required for preferential exclusion of genetic markers of the closely related phage SP01 in mixed infections (18).

Bacillus phage β22, whose intron in its thy (thymidylate synthase) gene was discovered as a result of sequencing the gene, is a member of a large family of morphologically distinct phages that infect a wide range of low G-C Gram-positive bacteria (20, 21). The initial aim of this work was to use the GTP-labeling reaction to obtain information on the distribution of group I introns in this family of bacteriophages. During this screen, we encountered a surprisingly large number of GTP-labeled products derived from staphylococcal phage Twort RNA, indicative of a large number of self-splicing introns in this phage genome. In this work, we describe cloning and characterization of three self-splicing introns that are inserted in a single late-transcribed ORF.

An interesting historical note is provided by the fact that this phage was apparently the first to be isolated and described in the literature (22).

MATERIALS AND METHODS

Bacterial and Bacteriophage Strains and Growth Conditions. Phage Twort (HER 48) and its host Staphylococcus aureus Twort (HER1048) were obtained from the Felix d’Herelle Reference Center maintained by H.-W. Ackermann. S. aureus was grown in Nutrient Broth with supplements (10 g of Nutrient Broth/5 g of peptone/2.5 g of yeast extract/9 g of NaCl/0.3 g of KCl per liter) with shaking at 37°C. E. coli XL-1 blue (Stratagene) was used as a recipient strain for high-frequency plasmid electroporations.

Plasmids and Plasmid Constructions. After digestion of Twort DNA with HindIII and KpnI, fragments were separated on a low melting-point agarose gel and a 4.6 kilobase pair (kb) and a 5.4-kb fragment, respectively, were extracted by using

Abbreviations: RT-PCR, reverse-transcription PCR.

Data deposition: The sequence reported in paper has been deposited in the GenBank database (accession no. AF132670).

To whom reprint requests should be addressed. e-mail: shub@cnsunix.albany.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.
Oligonucleotides. 1. 5’-GGTACACGAACTATCAATA- 

RESULTS

Detection of Multiple Group I Introns by GTP Labeling of Twort RNA. RNA, isolated from S. aureus at various times after infection with phage Twort, was incubated with [α-32P]GTP under self-splicing conditions and yielded multiple GTP-labeled RNA products (Fig. 1). Interestingly, the various GTP-labeled products show maximal concentrations at different times after infection. Products of approximately 1,100 nt and 430 nt are labeled maximally in RNA isolated 3 min after infection, and a product of about 1,400 nt, appearing by 7 min, is most strongly labeled in 10-min RNA. Labeled RNAs of about 320 nt, 290 nt, and 250 nt and minor products of larger size appear by 10 min and are most intense in 15- to 20-min RNA (Fig. 1). In contrast, no labeled product was detectable in RNA derived from uninfected S. aureus. These results indicate the presence of multiple self-splicing introns in genes of at least three transcriptional classes.

Cloning and Sequencing of Late Introns. A blot of digested Twort DNA was probed with GTP-labeled 15-min RNA, resulting in hybridizing bands that differ in signal intensity (Fig. 2). For most restriction enzyme digests, a single strong and several weak signals could be detected. Some of the fragments that hybridized weakly to GTP-labeled 15-min RNA
were also detectable when GTP-labeled 7-min RNA was used for probing (data not shown), suggesting that these fragments contain introns present in RNA that is transcribed early during infection. GTP-labeled 15-min RNA hybridized intensely to a 4.6-kb \( \text{HindIII} \) and to a 5.4-kb \( \text{KpnI} \) restriction fragment (Fig. 2). Further digestion with \( \text{SacI} \) reduced the sizes of the hybridizing species to 1.2-kb and 2.0-kb, respectively. The 1.2-kb \( \text{HindIII} \)-\( \text{SacI} \) fragment was cloned (pTHS17), and sequence analysis revealed three putative group I introns of 325 nt, 242 nt, and 286 nt. The sequence surrounding the introns was obtained by cloning a combination of restriction fragments and PCR products (Fig. 3). Alignment of the sequences yielded a contiguous stretch of 2,397 bp. An annotated presentation of the first 1,800 bp is shown in Fig. 4.

**Splicing in Vivo.** We used reverse transcription and PCR to determine whether the three introns are spliced in vivo and to confirm the positions of the splice sites. RNA isolated from \( \text{S. aureus} \) 15 min after infection was reverse transcribed by using oligonucleotide 5, and the cDNA was amplified by using oligonucleotides 4 and 5 (Fig. 3). The reaction resulted in a PCR product of 130 bp, the size expected for complete excision of the introns (Fig. 5). The nucleotide sequence of the reverse–transcription PCR (RT-PCR) product confirms that the three introns were removed in vivo, precisely as predicted from the secondary structures (Figs. 6 and 7).

The high similarity of the second and the third introns (Fig. 7) and the close proximity (23 bp) of their insertion sites (Fig. 4) further suggested that these two introns might fold to form a single functional ribozyme, resulting in skipping of an exon.

**Intron Structures and Sequences.** Fig. 8 shows the proposed secondary structure of \( \text{orf142} \) intron 2 (I2) as a representative structure. The three \( \text{orf142} \) introns (designated \( \text{orf142}-\text{I1}, \text{I2}, \text{and I3} \)) follow the consensus for group I introns closely.

---

**FIG. 2.** Southern blot hybridization of GTP-labeled Twort RNA to genomic DNA. Twort RNA was digested with \( \text{PstI} \) (P), \( \text{XbaI} \) (X), \( \text{KpnI} \) (K), \( \text{NsiI} \) (N), \( \text{EcoRI} \) (E), \( \text{SalI} \) (S), \( \text{SacI} \) (Sc), and \( \text{DraI} \) (D) and transferred. The blot was probed with 50 \( \mu \)g 15-min Twort RNA that was end-labeled with \( [\alpha-^{32}\text{P}] \) GTP under self-splicing conditions. Sizes of \( \text{BstEII} \)-digested DNA fragments are indicated to the right.

**FIG. 3.** Genomic organization and restriction map. Relevant restriction sites are given. Predicted ORFs are shown as boxes. \( \text{orfA} \) and \( \text{orfC} \) are incomplete. Introns in gene \( \text{orf142} \) are shown as half-sized boxes. Positions of the oligonucleotides used for PCR and RT-PCR are represented with arrows, with the arrowhead indicating the 3' end of the primer. Location of the cloned DNA fragments is shown and plasmid names are given.

**FIG. 4.** Nucleotide sequence of the cloned region of the Twort genome. Predicted amino acid sequences are given below. Stop codons are indicated with asterisks. Presumed ribosomal binding sites (RBS) and restriction sites are underlined. Inverted repeat downstream of \( \text{orf142} \) is indicated by arrows. Intron splice sites are shown as downward-pointing arrows.

The possibility of alternative splicing was examined by amplifying the cDNA with oligos 4 and 5, primer 5 having been end-labeled with T4 polynucleotide kinase. The PCR products were separated on a native 8% polyacrylamide gel and visualized by autoradiography. An extremely weak band at about 110 bp was detectable, indicating a PCR product that migrated faster than the major product of 130 bp (data not shown). The smaller PCR product was eluted from the gel, reamplified, cloned, and sequenced. The nucleotide sequence of this product revealed an exact deletion of exon 3, as expected for an amplified cDNA generated from an alternatively spliced transcript (Fig. 6).
containing all the conserved sequence elements and base-paired regions (Fig. 7) that are typical for introns of subgroup IA2, comprising the bacteriophage introns (11, 23). Unlike most other bacteriophage introns, the Twort orf142 introns show no large insertions of nonconserved nucleotides in terminal loop regions indicative of intron-encoded ORFs.

The orf142 introns I2 and I3 are more closely related to each other in sequence and structure than each of them is to intron I1. The overall identity of introns I2 and I3 is 85% over 212 aligned nucleotides, whereas I2 and I3 show 69% and 68% identity to intron I1 (Table 1). The 71 nucleotides at the 3' end, including structural element P9.1 (whose sequence is not phylogenetically conserved), are identical between introns I2 and I3 (Fig. 7).

When the nucleotide database was searched by using BLASTN (24) with the sequence of intron I2, the highest scoring match (identity of 26/27 bases, E value = 0.002) was to the phage T4 nrdB intron in a sequence that includes the structural element P7.2 (whose sequence is not conserved among the group I introns). Further pairwise comparison of orf142 introns I2 and I3 to the T4 nrdB intron indicates 78% and 80% identity, respectively, from the beginning of the intron to the 5' portion of P4 (Fig. 7), and 70% and 72% identity overall (Table 1). Coding Regions. An analysis of the nucleotide sequence after intron removal revealed two partial ORFs, designated orfa (position 1–281) and orfc (position 1741–2397), and a complete ORF of 142 amino acids, designated orf142, which is interrupted by three self-splicing group I introns. Translation of the putative alternatively spliced transcript results in an ORF of 130 amino acids, in which the last 13 amino acids of Orf142 are replaced by an arginine (Fig. 4). We used the deduced amino acid sequences for database searches with gapped BLASTP (using standard parameters) provided by the National Center for Biotechnology Information network service (24). Statistically significant alignments were obtained for Orfa and Orf142 to late cotranscribed genes of Listeria phage AS11 (25) belonging to the same taxonomic group (A1) as Twort (26). Orfa shows 86% similarity (70% identity) to the 93 carboxyl-terminal amino acids of the major tail sheath protein (Tsh) of phage AS11, and Orf142 shows 82% similarity (71% identity) to Orf8, located immediately downstream of tsh, over 138 amino acids that were aligned. Although database searches provide no indication for a function of Orf142, it is likely to be a late gene product that is not involved in DNA synthesis.

**DISCUSSION**

End labeling of RNA *in vitro* with GTP provides an unbiased assay for the presence of group I introns. Although there are many reasons why an intron might not be detected by this method (it may be present in a poorly transcribed gene; protein or other factors may be required for splicing, etc.) GTP-labeled RNA can be used to characterize self-catalytic introns without preconceived ideas of their likely genomic locations. This assay has revealed an unexpected richness of group I introns in phage Twort. Introns are present in RNA molecules representing at least three temporal periods of infection (Fig. 1).

**Are Phage Introns Useful?** The four introns initially described in *E. coli* phage T4 (3–5, 27) and *B. subtilis* phage SPO1 (14) are all in genes concerned with DNA synthesis, stimulating the suggestion that splicing might be involved in a global regulatory mechanism involving DNA replication (3, 14). The recent finding of a group I intron and an intein in the nrdE gene of the *B. subtilis* prophage SPβ (28) provoked a reexamination of this idea (29). However, our description of three introns in orf142 and recent reports of introns in late genes of other phages (9, 10) weaken the hypothesis that introns are retained in the genome to act in the regulation of DNA synthesis.

The question that remains is, why are introns maintained in bacteriophage genomes, in particular in Twort, with multiple introns in different transcriptional classes? Introns could be beneficial to the phage in several ways. For example, they could provide noncoding targets for recombination leading to shuffling of exons (30). Also, the presence of multiple introns in a single gene may allow synthesis of several products of the same gene by incomplete or alternative splicing of the mRNA. We have demonstrated here that some transcripts of Twort orf142 mRNA precisely lack the exon between the highly similar introns I2 and I3. Although we cannot be certain of the biological significance of these shorter transcripts [they could have been generated by aberrant splicing events *in vitro*, by intramolecular trans splicing *in vivo* (31), or by folding of the highly similar introns I2 and I3 into a single hybrid ribozyme], their existence demonstrates the feasibility of programmed exon skipping by using self-splicing introns. Previous studies have suggested a different kind of alternative splicing of some
group I introns, whereby fungal mitochondrial introns bring their upstream exons into phase with their internal ORFs by occasional use of an alternative 3′ splice site located between the ORF and the conserved intron core (11, 32–34).

On the other hand, the phage introns may be solely selfish DNA elements with invasive potential because of the intron-encoded DNA endonucleases that promote the mobility of the intron (35). The GTP-labeled products of about 1,100 nt and 1,400 nt in early Twort RNAs (Fig. 1) suggest the presence of group I introns with sizes that are sufficient to contain intron ORFs with potential to encode endonucleases. However, unlike most other bacteriophage introns, the three orf142 introns do not contain ORFs, which argues against a “parasitic” nature of these introns.

Goodrich-Blair and Shub (18) have described an intermediate situation, where an intron is responsible for the selective propagation of neighboring genes during mixed infection with a related phage carrying its own version of the intron. Although providing an interesting case of competition among introns, this example does not resolve the issue of whether introns can have a selective value at the organism level (36).

**The Origin of Phage Introns.** All of the phage introns are very closely related, belonging to a single subgroup (IA2) of the group I introns (11), suggesting that they are all derived from a relatively recent common ancestor. The high similarity between the thy intron of phage b22 and the td intron of phage T4 (7) and between orf142 introns I2 and I3 of phage Twort and the nrdB intron of T4 is particularly noteworthy. It is intriguing to note that (with the exception of integrated prophage genes) this subgroup has not been found in eubacterial chromosomal genes. Group I introns in bacteria have been found only in tRNA genes, and all of them belong to subgroup IC3 (15, 16, 37, 38).

Although the origin of the phage introns remains in doubt, their distribution is decidedly skewed. Although introns were found in three genes of E. coli phage T4, a survey of a large number of close relatives (T-even phages), either obtained from various culture collections or newly isolated from different natural sources, yielded only a few that had even one of the introns found in T4 (39). Moreover, introns have never been observed in other phages of E. coli or Gram-negative bacteria.

On the other hand, group I introns appear to have been very successful among the phages of low G-C Gram-positive bacteria. Introns have been found in a wide selection of genes, in both temperate and virulent phages of a number of different genera (Bacillus, Lactobacillus, Lactococcus, Staphylococcus). Moreover, when independent isolates of relatives of intron-containing phages SPO1 (8) or b22 (this work and our unpublished results) were examined, a majority were found to contain introns. To explain this disparity, either introns were relatively recently introduced into T4 from a Gram-positive phage, or conditions in Gram-positive bacteria are particularly favorable for the propagation and maintenance of introns in their viruses.

The presence of highly similar group I introns in the thymidylate synthase genes of Bacillus phage b22 and E. coli phage T4 suggested a lateral transfer of introns between phages of Gram-positive and Gram-negative bacteria (7). However, different locations of the introns in these homologous genes, and the different locations of the ORFs within the introns, this example does not resolve the issue of whether introns can have a selective value at the organism level (36).

**Table 1.** Sequence identity between Twort orf142 introns and the T4 nrdB intron

<table>
<thead>
<tr>
<th>orf142-I2</th>
<th>orf142-I3</th>
<th>T4 nrdB</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf142-I2</td>
<td>147/212 (67%)</td>
<td>131/212 (62%)</td>
</tr>
<tr>
<td>orf142-I3</td>
<td>144/212 (68%)</td>
<td>148/212 (70%)</td>
</tr>
<tr>
<td>orf142-I3</td>
<td>181/212 (85%)</td>
<td>152/212 (72%)</td>
</tr>
</tbody>
</table>

Number of identical nucleotides of 212 compared positions are given.
introns, made it difficult to explain the mobility of the intron by a simple homing event. Similarly, insertion of the T4 orfB intron and the closely related Twort orfI42 introns I2 and I3 in nonallelic sites is also consistent with intron transposition. Transposition of an intron to a nonallelic site probably occurs by a mechanism related to homing of the intron into unoccupied sites of homologous genes. This could be accomplished by means of homing endonucleases (40), whose genes reside in introns, or by reverse splicing followed by reverse transcription (41).

Although we cannot be certain which mechanism is responsible for intron spread among the bacteriophages, the properties of orfI42 introns I2 and I3 are consistent with an RNA-mediated transposition event. These introns, which lack ORFs for homing endonucleases, are 85% identical in nucleotide sequence and have identical PI pairings with the last four nucleotides of their 5’ exons. Roman and Woodson (42) have shown that the Tetrahymena LSU rRNA intron reverse splices successfully in vivo into the 23S rRNA of E. coli, requiring only a tetranucleotide recognition sequence that interacts with the internal guide sequence of the intron to form PI. Bacteriophage Twort may be a useful biologically relevant system to study intron transposition by reverse splicing.

Concluding Remarks. Low G-C Gram-positive bacteria (Bacilli) were the subjects of pioneering studies of bacteria by Koch and Tyndall, as is known by every student of microbiology. Twort, the first bacteriophage to be recognized and described in the literature (22), is easily propagated to high titer by using standard methods. It is amusing to contemplate that if, instead of E. coli, the low G-C Gram-positive bacteria and their phages had become the paradigms for bacterial genetic research, split genes and self-splicing RNA might have been discovered initially as properties of bacteria.

We thank David Edgell and Bruno Paquin for discussions and critical reading of the manuscript and Thomas Cech for helpful comments. This work was supported by grant GM 37746 from the National Institutes of Health.