Characterization of the self-splicing products of a mobile intron from the nuclear rDNA of *Physarum polycephalum*

Berthie Ruoff, Steinar Johansen and Volker M.Vogt*

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received September 23, 1992; Accepted October 20, 1992

GenBank accession no. L031813

ABSTRACT

We have characterized the splicing products formed *in vitro* from RNA derived from the mobile group I intron in the nuclear rDNA of *Physarum polycephalum*, Pp LSU 3. This intron is a close relative of the well known *Tetrahymena* intron Tt LSU 1, being inserted at exactly the same position in the rDNA and sharing about 90% sequence identity with Tt LSU 1 in the conserved elements characteristic of the catalytic core of all group I introns. However, Pp LSU 3 differs from Tt LSU 1 in that it encodes a site-specific endonuclease, which mediates the homing of the intron to unoccupied target sites. The endonuclease, *I-Ppo*, would appear to be a unique example of a protein encoded by an RNA polymerase I transcript. To gain clues to the splicing products formed *in vivo*, and to the nature of the messenger RNA for *I-Ppo*, we subjected Pp LSU 3 RNA to standard self-splicing conditions *in vitro*, and then analyzed the products by size, by northern blotting, and by primer extension. The results show two novel features. First, in addition to the expected 5' splice site, there is an alternative 5' splice site in the upstream exon, just preceding the first codon of the *I-Ppo* open reading frame. Second, at the position corresponding to the major circularization site in Tt LSU 1 there is an internal processing site, leading to the efficient separation of two halves of the excised intron, the 5' half encoding *I-Ppo* and 3' half containing the ribozyme. Surprisingly, this cleavage appears not to be due to circularization followed by hydrolytic opening of the circle, but rather to G addition. The formation of these products *in vitro* suggests how the messenger RNA for the *I-Ppo* endonuclease may be generated *in vivo*.

INTRODUCTION

Group I introns are defined by certain conserved sequence elements, by a common secondary and tertiary structure whose formation relies on interactions among these sequences, and by a common mechanism of splicing. Eleven classes of group I introns have been proposed on the basis of detailed structure (1). In the presence of a G-nucleotide, many group I introns are able to undergo autocatalysis *in vitro* ('self-splicing') as naked RNA.

However, *in vivo* splicing frequently depends on protein cofactors (2). The prototype for all group I introns is the well-known *Tetrahymena thermophila* rDNA intron Tt LSU 1 (see reference 3 for nomenclature conventions), the autocatalytic properties of which have been studied in detail (reviewed in reference 4). Mechanistically, splicing is accomplished by a series of transesterifications reactions, which begin with the attack at the 5' splice site by the 3' OH group of a G residue that is specifically bound to a site in the folded RNA, also called the 'ribozyme'. The linear intron is released by a second transesterification reaction involving the G residue that is found as the last residue of all group I introns. The released intron may undergo further reactions, including cyclization and hydrolytic opening of the circular RNA. All of the reactions carried out by the ribozyme are reversible under appropriate conditions. The distribution of group I introns is uneven: they are common in mitochondrial and chloroplast DNAs of lower eucaryotes, but rare in nuclear DNA. In the few known examples of nuclear group I introns, insertion of the intron is always in the genes for ribosomal RNA (rDNA), and in the best studied cases, ie. *Tetrahymena sp.* and *Physarum polycephalum*, these rDNAs exist as extrachromosomal elements instead of the more usual tandemly repeated copies on one or more chromosomes.

Some group I introns are mobile: they have the ability to transpose (or 'home') in a highly specific manner to an unoccupied homologous target site (reviewed in reference 5). Mobility is observed after mating of two strains, one carrying the intron and the other lacking it. In all known examples, mobility is mediated by a specific endonuclease encoded in the intron itself. The endonuclease cuts the target DNA near the site of intron insertion, and then a gene conversion event initiated at this break leads to replication of the intron into the unoccupied site. To date about a dozen intron-encoded endonucleases have been characterized to some extent. Most are associated with group I introns in mitochondria of yeast (6, 7, 8), chloroplasts of *Chlamydomonas ssp.* (9, 10, 11), or T-even bacteriophages (12).

The only described example of a mobile nuclear intron is Pp LSU 3, which is found in the extrachromosomal rDNA of the Carolina strain of *P. polycephalum*. When this strain is mated with other *P. polycephalum* strains, Pp LSU 3 rapidly colonizes all ca. 150 rDNA molecules lacking it (13). About half of the Pp LSU 3 sequence is closely related to that of Tt LSU 1, while

* To whom correspondence should be addressed

the other half is novel, encoding the *I-Ppo* endonuclease (14). It is not known how the gene for this endonuclease is expressed. Since the gene is embedded in rDNA, presumably the messenger RNA for *I-Ppo* is derived from the pre-ribosomal RNA, making this a unique example of a protein-coding gene naturally transcribed by RNA polymerase I. In order to gain clues to the expression of this gene, we have analyzed the self-splicing of Pp LSU 1 in vitro. The results show that this catalytic RNA can process itself to liberate an RNA fragment that we hypothesize is the messenger RNA for *I-Ppo* in vivo.

MATERIALS AND METHODS

Construction of plasmids

Plasmids pI3-941, pI3-d9, and pI3-tr742 were described previously (14). pI3-941 contains the entire 941 bp of Pp LSU 3, plus 378 bp of upstream exon and 24 bp of downstream exon, cloned between the *EcoRI* and *Clal* sites of pBluescript KS+ (Stratagene). (The GenBank accession number for the complete sequence of Pp LSU 3 is LO3183). pI3-d9 contains a deletion that eliminates the entire 5' exon and the first 8 bp of the intron. pI3-tr742 contains a truncation that deletes approximately one half of the sequences homologous to Tt LSU 1, and therefore inactivates the ribozyme. pI3-BS was constructed by cloning the *EcoRI-Clal* fragment from pI3-941 into pBluescript II KS+ (Stratagene), which contains a *BssHII* site downstream of the *Clal* site and thus can be used to extend the 3' end of the RNA used in in vitro splicing reactions. pI3-Ex153 was constructed by first introducing an *EcoRI* site 153 bp upstream of the intron by means of the polymerase chain reaction, and then cloning the resulting *EcoRI-Clal* fragment into pBluescript KS+.

In vitro transcription and splicing

Plasmids were linearized by cleavage downstream of the intron, either with *Clal* or in one case with *BssHII*. Transcription reactions contained 40 mM Tris-HCl, pH 7.5; 6 mM MgCl₂; 2 mM spermidine; 0.75 mM each of ATP, GTP, CTP, and UTP; 5 mM dithiothreitol (DTT); 1–3 µg linearized template DNA; 1 U RNasin (Promega Biotech) and 1 µl (50 units) of T7 RNA polymerase (US Biochemical Corp.) in a total volume of 50 µl. Reactions were incubated for 30 min at 37°. The transcripts were radiolabeled with the addition of 2 µl α-³²P-UTP (10 µCi/µl; Amersham). The splicing reaction was initiated by the addition of KCl to a final concentration of 1 M. The reaction was incubated at 37° for 2 h, and then the RNA was precipitated with ethanol and dissolved in 50 mM Tris-HCl, pH 7.6. A sample of 2 µl was subjected to electrophoresis in a 4% polyacrylamide, 8M urea gel and visualized by autoradiography. The G addition reaction was carried out by incubating unlabeled RNA under splicing conditions in presence of 40 µCi α-³²P-GTP.

Northern hybridization analysis

Unlabeled RNA before or after splicing was electrophoresed as described, with parallel lanes containing the same RNA labeled with ³²P-UTP as a marker. The RNA was electrophoretically transferred onto nylon membranes (Genescreen Plus, Dupont) using a Hoeffer TE 70 Semiphor transfer apparatus. The membranes were allowed to dry overnight, and then prehybridized for 3 h at 50° in hybridization buffer (5×SSC, 0.2 M NaPO₄, 7% SDS, 10×Denhardt's solution, 250 µg denatured salmon sperm DNA). An exon probe was prepared by nick translation of the *EcoRI-Clal* fragment from a strain of

Physarum polycephalum lacking Pp LSU 3. The oligonucleotide probes were prepared by polynucleotide kinase labeling of oligonucleotides complementary to position 286–267 (oligo P5'B) and position 614–595 (oligo P3') of the intron. The probes were added in fresh hybridization buffer and incubated overnight at 50°. The membranes were washed twice with 2×SSC at 40° for 45 min, and then subjected to autoradiography.

Primer extension analyses

Unlabeled RNA transcribed from pI3-941 linearized with *EcoRI* and *Clal* was submitted to splicing conditions and then electrophoresed along with ³²P-UTP labeled RNA in a parallel lane. The locations of the unlabeled bands were identified after autoradiography from the parallel radioactive samples. The individual RNAs were prepared by eluting from excised gel slices that were incubated on a rotating wheel at 37° in 0.5M ammonium acetate, 10 mM Mg acetate, 1 mM EDTA, pH 8.0, 0.1% SDS. The RNA was precipitated with ethanol and resuspended in diethylpyrocarbonate-treated water. Oligonucleotides ending at position 595 (614–595, oligo P3') and at position 42 (61–42, oligo P5'A) were used as primers in reverse transcription reactions. The primers were end labeled with γ-³²P-ATP and T4 polynucleotide kinase, and then purified on a 19% polyacrylamide gel. RNA from each gel slice was mixed with the kinased oligonucleotide P3' or P5'A, and then ethanol precipitated and resuspended in 15 µl of RT buffer (50 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 5 mM DTT, and 50 mM KCl), heated at 80° for 5 min and slowly cooled to 40°. To the samples was then added 0.4 mM of each deoxyribonucleotide triphosphate, 2 mM DTT, and 1 U RNasin, and 1–5 units of AMV reverse transcriptase. After incubation at 49° for 30 min, 3 µl of 0.5 M EDTA was added and then the products were ethanol precipitated, electrophoresed in a 7% polyacrylamide, 7 M urea gel. Dideoxy sequencing reactions of the *EcoRI-Clal* fragment were prepared in parallel, with a Sequenase kit (US Biochemicals), and electrophoresed adjacent to the primer extension products as markers.

RESULTS

Pp LSU 3 is a 941 nt long group I intron that is located in the large subunit RNA coding region of the extrachromosomal rDNA of *Physarum polycephalum* (Figure 1A). The intron can be divided into two parts (Fig. 1B). The RNA sequences from the 3' portion constitute the ribozyme, which is responsible for autocatalytic excision of the intron from the preribosomal RNA. Sequences from the 5' portion contain an open reading frame, which encodes the *I-Ppo* endonuclease that mediates intron homing to unoccupied target sites in DNA (13, 14). The orf begins at an AUG in the upstream exon, but there is also a second AUG just into the intron. When synthesized in a translation mixture in vitro, both the longer and the shorter forms of *I-Ppo* that were engineered to start at these AUGs are active endonucleases (14). The ribozyme part of Pp LSU 3 is highly homologous to the well known *Tetrahymena* intron Tt LSU 1, with an overall sequence identity of 71%, including 90% identity in the 120 nt of the catalytic core. This degree of homology makes Pp LSU 3 the closest known relative to Tt LSU 1 from a genus other than *Tetrahymena*. The major structural difference between Tt LSU 1 and Pp LSU 3 is that in the P1 pairing segment (see reference 3 for nomenclature) at the 5' end of the intron the *Physarum* intron contains an inserted sequence of about 550 nt

Figure 1. Schematic representation of DNA constructs and summary of RNAs formed by self-splicing. **Panel A** is a diagram of a portion of the extrachromosomal rDNA of the Carolina strain of *Physarum polycephalum*. P, promoter; SSU, coding region for the small subunit RNA, LSU, coding region for the large subunit RNA; 1, 2, and 3, introns. **Panel B** is a diagram of intron 3 and surrounding DNA. Black rectangles, exons; open rectangles, intron. The thick lines above the intron show the open reading frames that can be translated into the long and short forms of the *I-Ppo1* endonuclease. The short lines below the intron show the positions of oligonucleotides P5'A, P5'B, and P3'. RBZ, ribozyme portion of the intron; 3'UTR, untranslated region between the open reading and the ribozyme; Alt. 5'SS, alternative 5' splice site; 5'SS, 5' splice site; IPS, internal processing site; 3'SS, 3' splice site. **Panel C** shows the structure of transcripts I–V. The restriction endonuclease used to generate the 3' ends of the DNAs used for in vitro transcription are listed at the right. T7 shows the position of the phage T7 promoter. **Panel D** depicts the structures inferred for the ten major RNA species observed after self-splicing. RNA 1 is the primary transcript (Transcript I). G indicates that these species become labeled with GTP during self-splicing.

(see Figure 6). This inserted sequence includes the open reading frame.

To analyze the self-splicing products of Pp LSU 3, we synthesized ^{32}P -UTP labeled RNA transcripts in vitro with T7 RNA polymerase from several different DNA constructs (Figure 1C). Most experiments utilized RNA from p13-941 DNA linearized with *Clal*, which contains the entire intron flanked by 378 nt of 5' exon and 24 nt of 3' exon, plus 64 nt at the 5' end comprised of vector sequences (Transcript I). In some experiments an almost identical DNA template, p13-941BS, was linearized instead with *BssHII* in order to extend the length of the 3' end by 66 nt, and thus mark RNA molecules containing that end (Transcript II). Parallel reactions also were carried out with RNAs made from a *Clal*-linearized construct with a shortened 5' exon, p13-Ex153 (Transcript III). Finally, RNA was prepared from p13-d7 linearized with *Clal*, which contains a

Figure 2. Electrophoretic analysis of intron-containing RNAs. ^{32}P -labeled primary transcripts I, II, III, IV, and V were prepared and submitted to self-splicing conditions, and then electrophoresed in a polyacrylamide/8M urea gel. The autoradiograph of the gel was cut to allow insertion of the schematic representation of the fragments in the figure. Lanes a and b, transcript II; lanes c, d, f, and g, transcript I; lanes h and i, transcript IV; lanes j and k, transcript III; lanes l and m, transcript V. Lane e, schematic representation of bands seen for transcript I. Lanes a, c, f, h, j, and l, RNAs after transcription; lanes b, d, g, i, k, and m, RNAs after incubation under splicing conditions. cr, putative circular RNA.

deletion that removes the 5' splice site (Transcript IV). As a control to identify RNA species that might be generated by pausing of T7 RNA polymerase, we also synthesized RNA from *Clal*-linearized p13-tr742, a template which lacks much of the ribozyme portion of the intron and thus is unable to undergo self-splicing (Transcript V). The RNA products synthesized in vitro in the absence of added salt were adjusted to 1 M KCl, incubated under standard splicing conditions, and then analyzed by electrophoresis.

The typical pattern of products observed after incubation of radioactive transcript I under splicing conditions consisted of ten major bands (Figure 2, lanes d, g; schematic lane e). These bands also are apparent in low amounts in the reaction before incubation (lanes c, f), indicating that some splicing occurred during transcription. We infer that the RNA molecules in these bands are genuine self-splicing products since they were absent when incubation was carried out with transcript V, derived from a template truncated in the ribozyme (lane m). Also, the same pattern of bands was observed when the primary transcript was gel-purified before incubation under splicing conditions (not shown). Incubation of transcript II containing an extended 3' exon led to a reduction in the mobility of bands 6 and 8 (lane b), consistent with an increase in size due to the extra 66 bp of

Figure 3. Northern blot analysis of RNAs. Unlabeled and ^{32}P -UTP-labeled transcripts I were incubated under splicing conditions, and then aliquots of unlabeled RNA were electrophoresed in separate wells of a polyacrylamide/8M urea gel. Parallel wells were loaded with the radioactively labeled RNA as a marker. The RNAs in the gel were transferred by blotting to Gene Screen Plus. The membrane was cut into strips, and hybridized separately with different radioactive DNA probes. Lane a, UTP-labeled RNAs; lane b, probed with oligonucleotide P5'B; lane c, probed with oligonucleotide P3'; lane d, probed with exon DNA. The oligonucleotides were labeled with γ - ^{32}P -ATP by polynucleotide kinase. The exon DNA, cloned from a strain of *Physarum* lacking intron 3, was labeled by nick translation.

sequence. Thus RNAs 6 and 8 must contain the 3' exon. Transcript II also consistently showed a high molecular weight band migrating about 230 nt larger than the intron itself (lane b). This band did not increase in intensity upon incubation under splicing conditions. We hypothesize that this RNA is an indirect result of incomplete truncation of the template with *Bss*III (evident from the label in species larger than the primary transcript), but we have not pursued the analysis of this fragment. Incubation of transcript III containing the shortened 5' exon led to an increase in mobility of RNAs 1, 6, 7, 8, and 10, implying that these species contain the 5' exon (lane k). Finally, incubation of transcript IV lacking the entire 5' exon and the 5' splice site gave a reduced level of splicing but showed what appeared to be an identical RNA 9 (lane i). All the spliced RNA mixtures showed low levels of what we presume to be circular species, as evidenced by a mobility slower than that of the primary transcript (e.g. lane d). This species was always particularly prominent after incubations of transcript IV (lane i). All of the transcripts also gave rise to other minor species that we have not tried to account for.

To gain more precise information about the molecular composition of the major RNA fragments, we used the two intron-complementary oligonucleotides P5'B and P3', as well as an exon fragment, as radioactive probes in northern blotting analysis of the spliced products. The location of the oligonucleotide probes is shown schematically in Figure 1B. The exon probe was derived by nick translation of a cloned rDNA fragment from a strain of *P. polycephalum* lacking intron 3. Parallel splicing reactions of transcript I were set up with

unlabeled transcripts, and with ^{32}P -UTP-labeled transcripts to mark the position of the ten fragments (Figure 3). Probe P5'B recognized RNAs 1, 2, 3, 4, and 5, (lane b), while probe P3' recognized RNAs 1, 2, 3, and 9 (lane c). The exon probe recognized RNAs 1, 2, 4, 6, 7, 8, and 10 (lane d). Similar blotting analyses were carried out on transcripts II and III to corroborate these results (not shown).

The proposed structure of the ten RNAs described above is summarized in Figure 1D. Further evidence for these structural assignments is presented below. The three most abundant products of self-splicing in vitro by Pp LSU 3 are the ligated exons (RNA 6), and what appear to be approximately the 5' and the 3' halves of the intron, respectively (RNA 5 and RNA 9). More minor products include two RNAs expected from in vitro splicing studies of other introns, namely the linear excised intron (RNA 3) and the 5' exon itself (RNA 7). In addition, there are two novel minor products that are related to each other, an RNA that is about 60 nt longer than the full length excised intron (RNA 2), and an RNA that is similarly longer than the 5' intron half (RNA 4). We hypothesized that these RNAs are derived from usage by the ribozyme of an alternative 5' splice site a short distance upstream in the exon.

The identity of the ligated exon fragments (RNA 6 and RNA 8) was verified by specific endonuclease cleavage of PCR products. RNA from a preparative in vitro splicing reaction was reverse transcribed using an oligonucleotide primer complementary to a sequence 7 to 24 nt downstream of the intron insertion site. The resulting cDNA was amplified with PCR and a second oligonucleotide, extending from 134 to 153 nt upstream of the intron insertion site. The sizes of the DNA fragments generated in these reactions were found to be those expected of RNA 6 and RNA 8. These DNA bands were submitted to cleavage with two different endonucleases, *I-Ppo* and *Afl*III. The recognition sequence for *I-Ppo* comprises the ca. 15 base pairs spanning the insertion site for Pp LSU 3 (E. Ellison and V.M. Vogt, unpublished results). The recognition site for *Afl*III forms the central 6 base pairs of this sequence. Thus both *I-Ppo* and *Afl*III sites should be accurately represented in the PCR product of RNA 6, the major expected species. By contrast, the *Afl*III but not the *I-Ppo* sites should be recreated in the PCR product of RNA 8, the minor unexpected species. When the PCR products were incubated with these enzymes, the cleavage pattern found was exactly as expected (data not shown).

To precisely define the 5' ends of RNAs 4, 5, and 9, we used oligo P5'A and P3' in primer extension analyses of these RNA fragments eluted from 8M urea polyacrylamide gels (Figure 4). The unlabeled fragments were located by means of a parallel gel track containing ^{32}P -UTP labeled, spliced RNA. Primer extensions also were carried out in parallel with total RNA after self-splicing. In order to identify the sequence at the 5' end, dideoxy sequencing reactions were performed with the same oligonucleotides. RNA 5 yielded a primer extension product that corresponds to the normal 5' splice site at the junction of the upstream exon and the intron (panel A, lanes a and g). As expected, the extension product migrated slightly slower than the dideoxy chain terminated product ending at the first base of the intron, consistent with the added non-coded G that initiates splicing. The minor RNA 4 yielded an extension product about 63 nt longer than that generated from RNA 5 (lanes a and b). In the reaction with the gel-purified fragment, a smaller extension product identical to that described above also is visible, presumably due to contamination of the excised RNA 4 by some

Figure 4. Primer extension analysis of RNAs 4, 5, and 9. Unlabeled transcript I was incubated under splicing conditions and electrophoresed on a polyacrylamide/8M urea gel. A parallel well was loaded with UTP-labeled RNA to mark the positions of the RNA fragments. Gel slices corresponding to RNAs 4, 5, and 9 were excised and the RNAs were eluted. RNAs 4 and 5 were incubated in separate tubes with oligonucleotide P5'A as a primer, plus reverse transcriptase and four unlabeled deoxyribonucleoside triphosphates. RNA 9 was similarly incubated, but with oligo P3' as a primer. Both primers were previously labeled at their 5' ends with γ - ^{32}P -ATP and polynucleotide kinase. The resulting complementary DNAs were electrophoresed on a polyacrylamide/8M ureas gel, together with the products of standard dideoxy sequencing reactions generated with the same primers in the presence of the cloned intron 3 DNA. Panel A, RNAs 4 and 5; panel B, RNA 9. Lane a, primer extension products from total, unfractionated RNA after self-splicing; lane b, primer extension products from RNA 4; lane g, primer extension products from RNA 5; lane l, primer extension products from RNA 9; lanes c, d, e, and f, and lanes h, i, j, and k, DNA sequencing reactions terminated by A, C, G, and T, respectively. Note that the sequence read from the primer extension reactions is complementary to that of the RNA. The horizontal markings at the sides indicate the major extension products.

of the more abundant RNA 5 from lower in the preparative gel. From the ratio of the primer extension products from unfractionated RNA (lane a), and from the ratio of ^{32}P -UTP label in RNA 4 to that in RNA 5 in several experiments, we estimate that RNA 4 is about 10% as abundant as RNA 5. The end of RNA 9 was located by a similar experiment, by means of primer P3' (panel B, lane l). From comparison with the parallel sequencing track, we infer that this RNA species has a 5' end corresponding in position to nucleotide 556 of the intron. The relative position of the 5' end of RNA 9 is nearly the same as that of the major circularization site of Tt LSU 1 (15).

A distinguishing feature of self-splicing by group I introns is that an exogenous G moiety (usually as a nucleotide) becomes covalently attached to the 5' end of the excised intron. To further characterize the self-splicing of Pp LSU 3, we submitted unlabeled RNA to self-splicing conditions in the presence of α - ^{32}P -GTP. All molecules containing the 5' end of the intron thus should become labeled. The results of this experiment confirmed that RNAs 4 and 5 indeed carry a radioactive G (Figure 5, lanes b, d, and f). RNA 9 also became labeled with

Figure 5. G-labeling of RNAs during autocatalysis. ^{32}P -UTP-labeled or unlabeled transcripts I, II, III, and IV were prepared by in vitro transcription, and then the RNAs were separated from free nucleotides by two ethanol precipitations. The RNAs then were incubated under splicing conditions, with the unlabeled RNA being in the presence of α - ^{32}P -GTP. The products were analyzed by autoradiography as in Figure 2. Lanes a, c, e, and g, UTP-labeled RNAs; lanes b, d, f, and h, GTP-labeled RNAs. Lanes a and b, transcript I; lanes c and d, transcript II; lanes e and f, transcript III; lanes g and h, transcript IV.

G (lanes b, d, f, and h), at an efficiency similar to that of RNA 5. Thus it appears that G addition at the internal processing site near the middle of the intron leads to the generation of this new 5' end. This is an unexpected finding. In the case of *Tetrahymena*, the excised linear intron RNA circularizes at the equivalent site, leading to the release of a short oligonucleotide fragment of 15 nt (including the non-coded 5' G residue), and then the resulting circle opens again by hydrolysis of the strained bond, not by G addition (15,16).

In group I introns, structural features determine at which location splicing will occur. In Pp LSU 3, the P1 pairing segment can be drawn in a pairing scheme similar to that of Tt LSU 1, to explain the major observed 5' splice site that leads to generation of the 5' end of the intron (Figure 6B). An alternative pairing can be drawn to explain the observed minor splice site 63 nt upstream in the exon (Figure 6C). In this case the sequence CUCU replaces the sequence CUCUCU, potentially forming a helix of slightly lesser stability but still containing the essential G:U pair. In *Tetrahymena* rDNA the same upstream sequence CUCU also occurs, and yet to our knowledge a significant level of an alternative spliced product generated at this site is not observed. This difference suggests to us that the tertiary structure of the RNA in this region, which may be different in Tt LSU 1 and Pp LSU 3 due to the large P1 insertion, plays an important role in splice site selection. The chance placement of the same sequence CUCU in transcript IV could explain why G addition also occurs near the 5' end of this RNA, despite the deletion of the normal 5' splice site and the entire upstream exon. This sequence is found in the polylinker into which the partially deleted intron was cloned, at a position that is within a few nt from where the normal splice site would be.

Figure 6. Structural models of intron RNA. **Panel A**, secondary structure of the group I intron ribozyme (RBZ) region. Arrows indicate the 5' and 3' splice sites as well as the internal processing site (IPS). The P1-extension containing the *I-Ppo* coding region and the 3' untranslated region (3' UTR) is shown. **Panel B**, proposed secondary structure of the 3' UTR as well as the major splicing site. **Panel C**, structure of the alternative P1 segment (alt-P1) that includes the alternative 5' splice site (alt-5'SS) in the upstream exon.

While modeling the folding of the Pp LSU 3 RNA, we noticed a prominent potential secondary structure in the 125 nt between the end of the *I-Ppo* open reading frame and the internal processing site (Figure 6B). The free energy of this structure is calculated to be about -40 kcal/mol (17). The sequencing reactions with primer P3' showed a strong stop or pause site for the Sequenase enzyme in the first helix of this structure (Fig. 4, lanes h–k), at about nucleotide 535. Taken together, these observations suggest that this or a similar structure also forms in vivo. It is well known that in histone mRNAs, which are not polyadenylated, the 3' untranslated regions of the RNAs (3' UTR) fold into stable secondary structures which bind specific proteins as well as the U7 ribonucleoprotein (18, 19), and which are essential for synthesis of histones. This analogy leads us to hypothesize that the predicted 3' UTR just downstream of the *I-Ppo* open reading frame acts in a similar manner, to stabilize the putative *I-Ppo* messenger or to direct its transport from the nucleus.

DISCUSSION

We have mapped the major products generated from self-splicing of RNA transcribed in vitro from Pp LSU 3, the mobile group I intron in the extrachromosomal rDNA of *Physarum polycephalum*. In addition to the species expected from similar studies with other self-splicing introns, in particular the closely related Tt LSU 1, RNA species are generated via two novel

reactions of the ribozyme. One is the utilization of an alternative 5' splice site 63 nt upstream in the exon. The other is an internal processing event that leads to the efficient separation of the 3' half of the intron RNA, which contains the ribozyme, from the 5' half of the intron RNA, which encodes the *I-Ppo* endonuclease. This internal processing reaction, which appears to be dependent on prior excision of the intron, is unexpectedly based on a G addition reaction.

In the expression of Pp LSU 3 in vivo, what could be the biological function of an intron-derived RNA that includes 63 nt of exon sequence at its 5' end? The placement of the alternative 5' splice site is provocative. It occurs just 13 nucleotides upstream from the AUG codon that begins the open reading frame for the *I-Ppo* endonuclease, suggesting that the RNA species resulting from this splicing event might function as a messenger for the longer form of *I-Ppo*. Such splicing presumably would lead to loss of function of the corresponding preribosomal RNA molecule, because of the resulting 63 nt deletion in the exon, but a minor amount of this aberrant product might be tolerated in vivo. The mobile group I intron Sc LSU 1 in yeast mitochondrial rDNA provides a direct precedent for such alternative processing. In this case the putative messenger RNA for the encoded endonuclease, *I-Sce 1*, is derived by a minor processing event about 1500 nucleotides upstream in the exon (20, 21). Although both the long form of *I-Ppo* initiated from the AUG in the exon, and the shorter form initiated just downstream of the exon-intron border, are active endonucleases

when translated in vitro (14), it is possible that only one of these two proteins is translated in vivo, or that they serve different roles. For example, the maturases in the *coxI* and *cob* genes of yeast mitochondria, which are required in vivo for proper splicing of group I introns in these genes, are encoded in part in the intron and in part in the upstream exon (22). An alternative model to explain the significance of the alternative 5' splice site in Pp LSU 3 postulates that the sequence of 63 nt of exon-derived RNA serves as a required leader for the *I-Ppo* messenger. In *E. coli*, the corresponding LSU sequence contains one of the binding sites for ribosomal protein L1 (23). Perhaps by binding to this sequence on RNA molecules derived by the alternative splicing event, the equivalent *Physarum* protein protects the RNA from degradation or plays a role in its transport to the cytoplasm. At present we favor the latter model because of results from recent experiments with Pp LSU 3 expressed in yeast (D. Muscarella and V. Vogt, submitted for publication). We have derived yeast strains in which all the ca. 140 rDNA repeats carry Pp LSU 3. Extracts of these strains show evidence of *I-Ppo* activity, despite the fact that yeast rDNA does not have an AUG initiation codon at the equivalent upstream position to yield a long form of *I-Ppo*. Thus at least in these slow-growing heterologous cells, the longer form of *I-Ppo* is not essential either for enzymatic activity or for correct splicing.

The existence of an efficient internal processing site near the middle of the Pp LSU 3 RNA, between residues G₅₅₅ and U₅₅₆, raises mechanistic questions. This site is very close to the position of the major circularization site in the closely related Tt LSU 1. In this circularization reaction in *Tetrahymena*, the 3' terminal G residue of the excised linear intron attacks the bond between nucleotide 15 and 16, leading to circle formation and release of an oligonucleotide of 15 nt corresponding to the 5' end of the intron (15). Because the newly formed bond is strained, hydrolysis occurs at the same sequence, reopening the circle (16). We originally expected that a similar reaction would take place in Pp LSU 3, leading to the release of a 556 nt linear RNA (including the non-coded G residue) similar to the *Tetrahymena* 15mer. Indeed, we observed low levels of circular RNA species. However, when gel purified and submitted to splicing conditions, the major circular species gave rise to a linear RNAT closely similar in size to that of RNA 3, and not to that of RNA 9 (not shown), implying that there is a circularization site close to or at the major 5' splice site. While it is not excluded that a second circularization site also exists in the middle of Pp LSU 3 RNA, and that hydrolysis does occur there, the intensity of the G labeling of RNA 9 implies that the primary reaction that is responsible for the majority of the 5' ends of RNA 9 is G addition and not simple hydrolysis. An internal G-addition site also has been reported in the self-splicing $\alpha 3$ intron of yeast mitochondria (24). In that case the internal processing site resembles the 5' splice site, and G-addition results in the separation of the upstream open reading frame from the downstream group I intron ribozyme. However, the sequence at internal processing site in Pp LSU 3 has a 3' terminal G; it does not match either known circularization sites or G addition sites, both of which are located 3' of a U. G-addition after a G residue has been reported at the 5' splice site of a Tt LSU 1 mutant, but at a rate more than 1000-fold lower than that of the wild type intron (25). To account for this apparent anomaly, we hypothesize that internal processing in Pp LSU 3 is mechanistically related to the G-exchange reaction observed in vitro for several group I introns, including Tt LSU 1 (26), and for another mobile nuclear group I intron, Di SSU

1 (S. Johansen, and V. Vogt, unpublished). Here the 3' terminal G of the intron RNA attacks the bond 3' of the 5' terminal G of the intron RNA, resulting in a release of the 5' terminal G and circularization of the intron RNA (see discussion in reference 27). If an exogenous G nucleotide were to similarly attack the bond 3' of G₅₅₅ in Pp LSU 3, the reaction products would be those we have observed. A more detailed analysis of Pp LSU 3 self-splicing will be required to address this hypothesis.

If the internal processing site is utilized in vivo as well as it is in vitro, then the probable messenger RNA for the *I-Ppo* endonuclease has its 3' end at this site. As a consequence, the prominent secondary structure downstream of the open reading frame would be at the 3' end of the messenger, at the same position as that occupied by the higher order structures formed in histone mRNAs. The hypothesis that this structure plays an important biological role is strengthened by the recent characterization of another mobile group I intron that is also found in nuclear rDNA, Di SSU 1 from the acellular slime mold *Didymium iridis* (S. Johansen and V. Vogt, unpublished). Although the open reading frame for the putative endonuclease is positioned very differently in this intron, a similar secondary structure of the 3' UTR can be drawn just downstream. Expression and mutational analyses of Pp LSU 3 and Di SSU 1 in an organism amenable to molecular genetics will be required to decipher the roles that these secondary structures play in expression of these genes embedded in pre-ribosomal RNA transcripts.

ACKNOWLEDGMENTS

We acknowledge James Varnum for numerous pilot experiments on self splicing of Pp LSU 3. This work was supported in part by USPHS grant GM 31577. S.Johansen holds a fellowship from the Norwegian Research Council.

REFERENCES

1. Michel, F. and Westhof, E. (1990) *J. Mol. Biol.*, **216**, 585–610.
2. Lambowitz, A.M. and Perlman, P.S. (1990) *Trends Biochem.*, **15**, 440–444.
3. Burke, J.M., Belfort, M., Cech, T.R., Davis, R.W., Schweyen, R.J., Shub, D.A., Szostak, J.W. and Tabak, H.F. (1987) *Nucleic Acids Res.*, **15**, 7217–7221.
4. Cech, T.R. (1990) *Annu. Rev. Biochem.*, **59**, 543–568.
5. Dujon, B. (1989) *Gene*, **82**, 91–114.
6. Jacquier, A. and Dujon, B. (1985) *Cell*, **41**, 383–394.
7. Delahodde, A., Goguel, V., Becam, A.M., Creusot, F., Perea, J., Banroques, J. and Jacq, C. (1989) *Cell*, **56**, 431–441.
8. Wenzlau, J.M., Saldanha, R.J., Butow, R.A. and Perlman, P.S. (1989) *Cell*, **56**, 421–430.
9. Colleaux, L., Michel-Wolwertz, M.-R., Matagne, R.F. and Dujon, B. (1990) *Mol. Gen. Genet.*, **223**, 288–296.
10. Dürrenberger, F. and Rochaix, J.-D. (1991) *EMBO J.*, **10**, 3495–3501.
11. Gauthier, A.T., Turmel, M. and Lemieux, C. (1991) *Curr. Genet.*, **19**, 43–47.
12. Quirk, S.M., Bell-Pedersen, D. and Belfort, M. (1989) *Cell*, **56**, 455–465.
13. Muscarella, D.E. and Vogt, V.M. (1989) *Cell*, **56**, 443–454.
14. Muscarella, D.E., Ellison, E.L., Ruoff, B.E. and Vogt, V.M. (1990) *Mol. Cell. Biol.*, **10**, 3386–3396.
15. Zaug, A.J., Grabowski, P.J. and Cech, T.R. (1983) *Nature*, **301**, 578–583.
16. Zaug, A.J., Kent, J.R. and Cech, T.R. (1985) *Biochemistry*, **24**, 6211–6218.
17. Zuker, M. and Stieger, P. (1981) *Nucleic Acid Res.*, **9**, 133–148.
18. Pandey, N.B., Sun, J.-H. and Marzluff, W.F. (1991) *Nucleic Acids Res.*, **19**, 5653–5659.
19. Melin, L., Soldati, D., Mital, R., Streit, A. and Schümperli, D. (1992) *EMBO J.*, **11**, 691–697.
20. Zhu, H., Macreadie, I.G. and Butow, R.A. (1987) *Mol. Cell. Biol.*, **7**, 2530–2537.

21. Zhu, H., Conrad-Webb, H., Liao, X.S., Perlman, P.S. and Butow, R.A. (1989) *Mol. Cell. Biol.*, **9**, 1507–1512.
22. Perlman, P.S. and Butow, R.A. (1989) *Science*, **246**, 1106–1109.
23. Brimacombe, R., Greuer, B., Mitchell, P., Osswald, M., Rinke-Appel, J., Schüler, D. and Stade, K. (1990) In Hill, W.E., Dahlberg, A., Garret, R.A., Moore, P.B., Schlessinger, D. and Warner, J.R. (eds.), *The Ribosome: Structure, Function and Evolution*. ASM, Washington, D.C., pp. 217–235.
24. Tabak, H.F., Van der Horst, G., Kamps, A.M.J.E., and Arnberg, A.C. (1987) *Cell* **48**, 101–110.
25. Barford, E.T. and Cech, T.R. (1989) *Mol. Cell. Biol.*, **9**, 3657–3666.
26. Been, M.D. and Cech, T.R. (1985) *Nucleic Acids Res.*, **13**, 8389–8408.
27. Thompson, A.J. and Herrin, D.L. (1991) *Nucleic Acids Res.*, **19**, 6611–6618.