MOLECULAR AND CELLULAR BIOLOGY, Apr. 1991, p. 1921–1926 Vol. 11, No. 4
0270-7306/91/041921-06$02.00/0
Copyright © 1991, American Society for Microbiology

Insertion of Part of an Intron into the 5′ Untranslated Region of a *Caenorhabditis elegans* Gene Converts It into a *trans*-Spliced Gene

RICHARD CONRAD, JEFFREY THOMAS,† JOHN SPIETH, AND THOMAS BLUMENTHAL*

Program in Molecular, Cellular, and Developmental Biology and Department of Biology, Indiana University, Bloomington, Indiana 47405*

Received 14 November 1990/Accepted 10 January 1991

In nematodes, the RNA products of some genes are *trans*-spliced to a 22-nucleotide spliced leader (SL), while the RNA products of other genes are not. In *Caenorhabditis elegans*, there are two SLs, SL1 and SL2, donated by two distinct small nuclear ribonucleoprotein particles in a process functionally quite similar to nuclear intron removal. We demonstrate here that it is possible to convert a non-*trans*-spliced gene into a *trans*-spliced gene by placement of an intron missing only the 5′ splice site into the 5′ untranslated region. Stable transgenic strains were isolated expressing a gene in which 69 nucleotides of a *vit-5* intron, including the 3′ splice site, were inserted into the 5′ untranslated region of a *vit-2*/*vit-6* fusion gene. The RNA product of this gene was examined by primer extension and PCR amplification. Although the *vit-2*/*vit-6* transgene product is not normally *trans*-spliced, the majority of transcripts from this altered gene were *trans*-spliced to SL1. We termed the region of a *trans*-spliced mRNA precursor between the 5′ end and the first 3′ splice site an "outron." Our results suggest that if a transcript begins with intronlike sequence followed by a 3′ splice site, this alone may constitute an outron and be sufficient to demarcate a transcript as a *trans*-splice acceptor. These findings leave open the possibility that specific sequences are required to increase the efficiency of *trans*-splicing.

In both trypanosomes and nematodes, mRNAs are present which begin with an untranslated leader sequence acquired by *trans*-splicing between a spliced leader (SL) small nuclear ribonucleoprotein particle (snRNP) and recipient transcripts (7, 17, 20, 24, 32, 35, 36). Trypanosome pre-mRNAs do not contain introns, and all begin with the SL (21). In contrast, both *trans*-splicing and intron removal are occurring on the same nematode transcripts, and only a subset of *Caenorhabditis elegans* genes (estimated at about 15% [1]) specify transcripts that receive an SL (4). Hence, the interplay between conventional *cis*-splicing and *trans*-splicing in nematodes is an intriguing and largely unexplored area. It is clear that the two processes are quite closely related: consensus sequences for the 3′ splice site of introns and *trans*-splice sites are the same and both occur via a 2′-5′ branch (or lariat) intermediate (2, 35). Furthermore, the donor in the *trans*-splicing reaction, a 100-nucleotide RNA called SL RNA, occurs in the form of an snRNP. It has the trimethylguanosine cap typical of those snRNPs involved in the catalysis of *cis*-splicing and is bound by some of the same immunologically defined proteins (7, 35, 36). Finally, the 5′ splice site sequence contained in the SL RNA matches the consensus for intron 5′ splice sites (4). If *cis*- and *trans*-splicing are catalyzed by the same set of snRNPs and associated factors, and the same 5′ and 3′ splice sites are utilized in the two reactions, what keeps them distinct? What prevents *trans*-splicing at intron 3′ splice sites and what demarcates certain transcripts for *trans*-splicing?

The recent discovery of a second SL RNA in *C. elegans* (17) makes these questions even more interesting. Not only do transcripts have to contain the information for *trans*-splicing, they must contain information conferring specificity for one or the other SL. Where is this information? Several pairs of very similar genes have been described in which only one is *trans*-spliced (9, 19, 20, 26). In some of these cases, the coding regions of the gene pairs are nearly identical. Thus, it is likely that the information for *trans*-splicing is in the region upstream of the *trans*-splice site. However, a comparison of DNA sequences upstream from the *trans*-splice acceptor sites of 11 genes which receive SL1 (9, 13, 14, 16, 19, 20, 26) and 1 gene which receives SL2 (17) reveals no conserved sequences that are likely to serve as a signal for *trans*-splicing. Rather, the only apparent distinguishing features of these transcripts are a functional 3′ splice site without an upstream 5′ splice site and an overall A+U richness.

To test the idea that a functional 3′ splice site without an upstream 5′ splice site could supply sufficient information to specify *trans*-splicing, we constructed genes to be tested in vivo, using integrative transformation of *C. elegans* (11). Members of our laboratory have been studying regulation of the vitellogenin (*vit*) genes in stable transgenic strains. These genes are expressed exclusively in the adult hermaphrodite intestine, and they are not *trans*-spliced (29). To study the mechanism by which this developmental regulation is effected, we constructed a fusion between two *vit* genes, the 5′ end of *vit-2* joined to the 3′ end of *vit-6*. The RNA and protein products of this gene can be observed in transgenic strains. We have shown that this transgene is properly regulated in a large number of independent stable strains (31; unpublished data). In this communication, we report the construction of a modification of our *vit-2*/*vit-6* fusion to test the idea that if we put intronlike sequences into a standard gene's 5′ untranslated region (5′ UTR), it would be converted to a *trans*-spliced gene. We show that the mRNA of this construction is *trans*-spliced, suggesting that the primary signal for *trans*-splicing is as simple as a transcript that begins with an intronlike sequence ending in a 3′ splice site. This unique feature of the precursor mRNA is what we term an "outron."

* Corresponding author.
† Present address: National Institutes of Health, Bethesda, MD 20892.

MATERIALS AND METHODS

Worm transformation. The procedure of Fire (11) was used to generate stable transgenic worms. Worms injected were homozygous *tra-3* segregants of CB3988 {*tra-3(e1107)*/*nT1*[*unc*(*n754dom*)*let*]}, and the selection of transformants by *sup-7* supression of *tra-3* described in that report was used. The integration event in this procedure occurs at a single, random chromosomal location, and the transgene integrates in tandem repeats of three or fewer intact copies.

Construction of the reporter gene. Two vitellogenin genes, *vit-2* and *vit-6*, were used to generate a fusion gene. *vit-2* specifies a 5.1-kb mRNA which encodes a 188-kDa protein (yp170 [3]). *vit-6* specifies a 5.3-kb mRNA which encodes a 193-kDA polypeptide which is cleaved later to form yp88 and yp115 (28). Plasmids carrying the two vitellogenin genes were digested with the indicated restriction endonucleases (Fig. 1A), and the desired fragments were isolated by gel purification, joined by a polylinker, and ligated into a pUC-based vector, pAST18A (12), carrying a selectable marker, *sup-7*. This plasmid was named pJ247 (Fig. 1A). The fusion gene contains a functional promoter region (247 upstream bp) from the *vit-2* gene, the coding region from *vit-2* and *vit-6*, and the termination site from *vit-6*. It specifies a 4.0-kb mRNA which encodes a vitellogeninlike fusion polypeptide of 155 kDa (fp155), expressed exclusively in the intestine of the adult hermaphrodite (31). The 11-bp 5′ UTR of pJ247 was altered by oligonucleotide-directed mutagenesis to give a modified 16-bp 5′ UTR containing an *Eco*RV site (GAT/ATC). This construction was named pI-0. To make the experimental construct (pI-1), a 79-bp *Xmn*I-*Ava*II fragment of a third vitellogenin gene, *vit-5* (nucleotides 875 to 953 of *vit-5* [30]), was inserted into the *Eco*RV site. The inserted fragment, from the coding region of *vit-5*, contains 69 bp of a 70-bp intron, all but the 5′-most G, followed by 10 bp of the next exon (Fig. 1B).

Preparation of total RNA. Total RNA was extracted from a mostly adult population of worms grown on plates (5). Worms were extracted by vortex agitation with glass beads in a 2:1:1 mixture of lysis solution (4 M guanidine isothiocyanate, 0.13% Sarkosyl, 33 mM Tris HCl [pH 8.0], 0.5% β-mercaptoethanol, 6.7 mM EDTA)-phenol-chloroform for 4 min, after which the aqueous phase was removed and precipitated by the addition of 0.025 volume of 1 M acetic acid and 0.75 volume of ethanol (8). The RNA precipitate was further purified by the procedure of MacLeod et al. (22).

Primer extension of RNA from transformants. Synthetic oligonucleotide primers were labeled by a kinase reaction as described by Maniatis et al. (23). These were used in a modification of the primer extension procedure of Hamlyn et al. (15). Each labeled primer (50 ng) was mixed with 10 μg of total RNA from each strain in 5 μl of annealing buffer (250 mM KCl, 10 mM Tris HCl [pH 8.3]), and these mixtures were heated to 95°C and then slowly cooled to 48°C for 1.5 h to anneal the primers. To 4 μl of each mixture was added 2 μl of water and 6.6 μl of reverse transcription mix (24 mM Tris HCl [pH 8.3], 16 mM $MgCl_2$, 8 mM dithiothreitol, 100 μg of actinomycin D per ml, 0.6 mM dATP, 0.6 mM dCTP, 1.0 mM dGTP, 0.6 mM dTTP, 1 U of avian myeloblastosis virus reverse transcriptase [Boehringer Mannheim] per μl). For sequencing reactions, an equal volume of each dideoxynucleoside triphosphate (20 mM) was added in place of the water for four separate reactions. These were then incubated at 48°C for 50 min, after which 25.4 μl of TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.4) and 20 μl of 7.5 M $NH_4(CH_3COO)$ were added and mixed, followed by 120 μl of ethanol. Precipitates were removed by centrifugation after at least 4 h at −20°C, rinsed with 70% ethanol, and dried. These pellets were redissolved in formamide-dye (5 μl), and 2 μl of each was electrophoresed on a 10% polyacrylamide gel (20:1, total acrylamide-bisacrylamide; 7 M urea; TBE [89 mM Tris, 89 mM boric acid, 2 mM Na_2 EDTA]). Gels were soaked in 5% methanol–10% acetic acid and dried to filter paper before autoradiography.

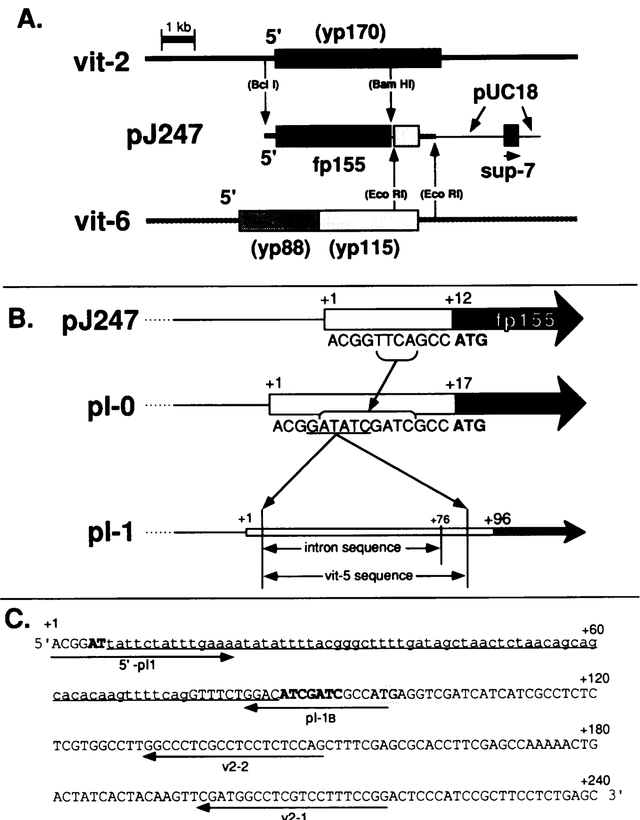

FIG. 1. Generation of test plasmids. (A) Construction of pJ247. A *Bcl*I-*Bam*HI fragment of *vit-2* was joined to an *Eco*RI fragment of *vit-6* by a polylinker and ligated into a pUC-based vector carrying a selectable marker, *sup-7* (12). The *vit-2* upstream flanking region (247 bp) efficiently drives expression of the 4.0-kb fusion mRNA, exclusively in the intestine of the adult hermaphrodite (31; unpublished observations). This mRNA encodes a vitellogeninlike fusion polypeptide (fp155) of a unique size of 155 kDa, recognizable by antibodies to both the *vit-2* and *vit-6* gene products. (B) Construction of pI-1. The pI-0 construct was made by a nine- for four-nucleotide switch in the 5′ UTR of the pJ247 fusion gene, generating an *Eco*RV site (underlined). The pI-1 construct was made by an insertion into this site of a 79-bp fragment of a third gene, *vit-5* (hatched). Protein-coding regions are indicated by filled arrows. (C) Sequence of 5′ end of pI-1. The sequence is of bases +1 to +240 of the 5′ region of the pI-1 transcript. Lightface uppercase letters indicate *vit-2* sequence (29); boldface uppercase letters indicate synthetic sequence; and underlined letters indicate *vit-5* sequence (intron, lower case [30]). The translation initiation site is shaded. Antisense primers for reverse transcription (pI-1B and v2-2) and PCR (v2-1) are indicated by arrows.

PCR amplification of pI-1 cDNA. A 1.65-μg sample of a 21-mer complementary to positions +197 to +217 from the transcriptional start site of pI-1 (v2-1 in Fig. 1) was used to prime 10 μg of total RNA from either the pI-1 strain or a control strain (N2, nontransformed wild-type *C. elegans*). Annealing was performed by heating to 95°C and incubating

at 42°C for 45 min in 10 μl of 50 mM Tris HCl (pH 8.3)–60 mM NaCl–10 mM dithiothreitol–1.33 mM each deoxynucleoside triphosphate. To this mixture was added 2 μl of a 5-U/μl avian myeloblastosis virus reverse transcriptase solution in 50 mM Tris HCl (pH 8.3)–60 mM NaCl–30 mM $Mg(CH_3COO)_2$–10 mM dithiothreitol. This was incubated at 42°C for 45 min and then heated to 95°C to inactivate the enzyme. A standard polymerase chain reaction (PCR) mix (Cetus Perkins-Elmer) (27) was added, with an additional 0.066 μg of the downstream oligonucleotide and 0.165 μg of one of the two upstream oligonucleotides. These were identical to either (i) the 22-nucleotide SL1 sequence (SL1 homolog) or (ii) the 22-nucleotide SL2 sequence (SL2 homolog). PCRs were performed with a 90-s denaturation step (92°C), 90-s annealing step (50°C), and 90-s polymerization step (72°C). Ramps between each step took approximately 40 s. Products were chloroform extracted to remove the mineral oil overlay and ethanol precipitated, using sodium acetate as a co-salt. Pellets were rinsed with 70% ethanol, dried, and redissolved in 5 μl of distilled H_2O. To identify the PCR products by Southern blotting, we added 5 μl of 50% glycerol with tracking dyes and 5 μl of this was loaded onto a 5% polyacrylamide gel in TBE. Electrophoresis was performed at 12 V/cm until the bromophenol blue reached the bottom of the gel. The gel was then electroblotted onto a nylon membrane (Hybond-N; Amersham) and probed with an end-labeled oligonucleotide complementary to positions +131 to +150 of the nascent transcript from pI-1 (v2-2 in Fig. 1).

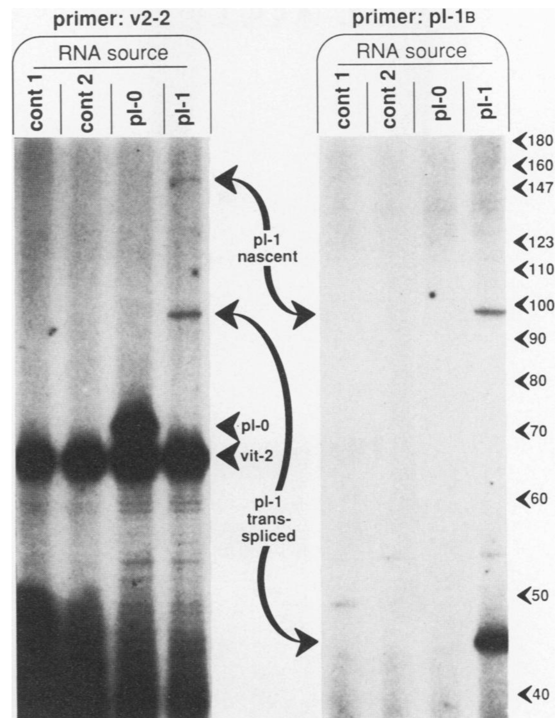

FIG. 2. Primer extension of pI-0 and pI-1 RNA with *vit-2*- and pI-1-specific primers. Left panel: Primer is complementary to *vit-2* mRNA from position +36 to +55 relative to the start of translation (v2-2, Fig. 1B). Right panel: Primer is specific for the pI-1 mRNA, complementary to the region from position −14 to +2 relative to the start of translation (pI-1B, Fig. 1B). Cont 1 is control RNA from a wild-type *C. elegans* strain, N2. Cont 2 is RNA prepared from a pJ247-transformed strain. Band sizes (in nucleotides) were determined by standards on the same gel, indicated at right of figure.

RESULTS

Construction of a gene with an artificial outron. The reporter gene is a fusion between two yolk protein genes, *vit-2* and *vit-6* (31) (Fig. 1A). Numerous strains containing this reporter gene demonstrated correct developmental regulation at levels of expression nearly as high as that of the endogenous *vit-2* gene. This construct was modified by replacement of four nucleotides of the 5′ UTR with nine nucleotides containing an *Eco*RV site (pI-0). Into this site we inserted an *Xmn*I-*Ava*II fragment from a third vitellogenin gene, *vit-5* (30), containing 69 nucleotides of a 70-nucleotide intron with flanking 3′ exon (Fig. 1B). This final construct, called pI-1, possessed all of a normally used intron, except the 5′ G nucleotide, contained within the transcript's 5′ UTR (Fig. 1C). This created a 5′ splice site-deficient intron in the 5′ UTR of the reporter gene's putative initial transcript. Our hypothesis predicts such a sequence will be equivalent to a naturally occurring outron, so the pI-1 transcript should undergo *trans*-splicing at the 3′ *cis*-splice site provided by the *vit-5* intron sequence.

Generation of transformed worm lines. Transformations were performed as described in Materials and Methods, with the initial selection provided by the *sup-7* suppression of *tra-3*. If the reporter gene is being expressed properly, transformants with both pI-0 and pI-1 should yield a protein of a unique molecular weight detectable by Western immunoblots (31). Transformants generated with pI-0 and pI-1 were screened for expression of this fusion protein (data not shown), and two transformants containing each transgene were verified in this way. Three of the four strains were inviable in the homozygous state but could be maintained as heterozygous lines. Genomic Southern and genetic analyses (data not shown) indicated that each of the four strains had the transgene integrated at a single chromosomal location at low copy number (less than or equal to three copies per haploid genome). Both of the strains carrying the pI-1 plasmid grew very poorly even as heterozygotes and were very difficult to maintain. Although the precise reason why these strains were so sick is unknown, we hypothesize that this may be due to the high level of expression from the *vit-2* promoter. Perhaps the high level of unnatural substrate for the *trans*-splicing machinery (see below) overwhelms it, radically reducing the essential *trans*-splicing of other intestinal transcripts.

Primer extension of transgene transcripts. Total RNA was prepared from about 0.2 g of each transformant strain for examination by reverse transcription primer extension. Two primers were used, one specific for *vit-2* (v2-2), and one specific for the pI-1 insert (pI-1B; Fig. 1B). Figure 2 shows the products obtained by primer extension on the pI-0 heterozygous strain and one of the pI-1 heterozygous strains. The other pI-0 strain, as well as the other pI-1 strain, gave the same results and are not shown. Priming with the *vit-2*-specific primer created a strong *vit-2* signal with all RNA samples. Furthermore, the RNA samples from both the pI-0 and pI-1 strains showed additional bands. Primer extension of RNA from the pI-0 strain yielded a product of the length expected for the extended 5′ UTR of the pI-0 transgene product (five bases longer than the *vit-2* product). The RNA sample from the pI-1 strain showed products not only from the predicted initial transcript of the transgene, 84 nucleotides longer than the *vit-2* message (labeled nascent RNA in Fig. 2), but also a more prominent band of the

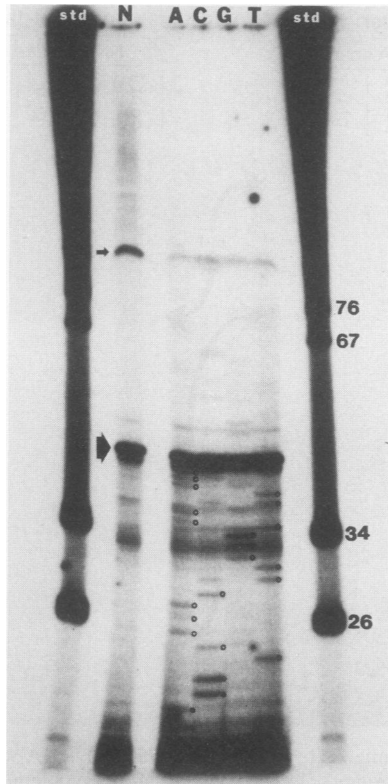

FIG. 3. Dideoxynucleotide sequencing of 5′ end of pI-1 transcript. Primer extension was performed with a pI-1-containing strain as described in the legend to Fig. 2 but in the presence of dideoxynucleotides (lane N is reaction without addition of dideoxynucleotide). The primer was pI-1B. The wide arrow indicates the size predicted for the 5′ end of the *trans*-spliced product (44 nucleotides). The thin arrow indicates the size predicted for the initial transcript (97 nucleotides). The bands for SL1 are indicated by open circles. Bands visible but not indicated are consistent with the sequence of the initial transcript. The sequence indicated for SL1 reverse complement (CTCAAACTTGGGTAATTAAACC) should be contrasted with that for the reverse complement of SL2 (CTTGAGTA ACTGGGTTAAAACC). SL2-specific bands are not present. std, sizes-standards (numbers indicate nucleotides).

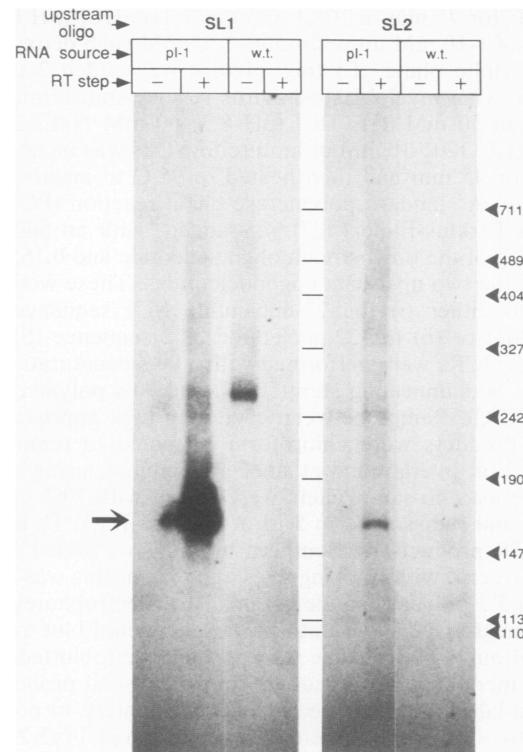

FIG. 4. PCR amplification of RNA from strain carrying pI-1. Amplifications were performed with v2-2 and either SL1 or SL2 homologous oligonucleotides (upstream oligo). Southern blots of nondenaturing gels were probed with pI-1B labeled with polynucleotide kinase. Band sizes (in base pairs) were determined by standards run on the same gel (visualized by ethidium bromide staining), indicated at right. RNA sources were either wild-type (w.t.) or pI-1-transformed (pI-1) strains. RT step refers to the presence or absence of the initial reverse transcription step, to confirm that the signal seen is RNA dependent. Arrow at left indicates the predicted position of the amplification product produced from the *trans*-spliced transcript.

precise size expected for a product produced by *trans*-splicing the 22-nucleotide SL onto the inserted 3′ splice site. These products were much less abundant than the pI-0 or the *vit-2* products because most of the animals in the pI-1 population were pseudomales which lacked the transgene. Pseudomales were much more frequent in the pI-1 strain because most of the hermaphrodites, which carried one copy of the gene, either died or produced few progeny. Primer extension with the second oligonucleotide, pI-1B, showed no major products with either control or pI-0 RNA preparations, as expected. However, with the RNA isolated from the pI-1-containing strain, both a minor product of the size expected for the initial pI-1 transcript and a major product of the size expected for the *trans*-spliced message were again observed. Dideoxynucleotide sequencing of RNA from the pI-1 strain with this primer gave bands indicating splicing of SL1 to the inserted 3′ splice site (Fig. 3, open circles). Although bands consistent with the sequence of the unspliced precursor could also be seen, no bands consistent with the SL2 sequence were observed.

PCR amplification of pI-1 RNA. To demonstrate that mRNA containing spliced leader is present, we used PCR to amplify single-stranded cDNA made from total RNA isolated from the same pI-1 strain. Two different upstream primers were used in two separate reactions. These reactions utilized upstream primers equivalent to the two known SL sequences, SL1 (20) and SL2 (17). The results, shown in Fig. 4, indicated that the majority of the pI-1 transcript is *trans*-spliced to SL1, with a small amount *trans*-spliced to SL2. When SL1 was used as the upstream oligonucleotide, an intense band was seen at the position expected for *trans*-splicing to the inserted 3′ splice site (arrow in Fig. 4). This band was not seen if reverse transcriptase was omitted during step 1 of the reaction or if RNA from a control strain was used as the template. (The band seen with SL1 as an upstream primer in the absence of reverse transcriptase is of unknown origin and was not seen in repeats of this experiment.) When the SL2 oligonucleotide was used under identical reaction conditions, a faint band was observed with the pI-1 RNA as the template. While these PCR results are not rigorously quantitative, both this experiment and the primer extension sequencing of the RNA product of pI-1 are consistent with the conclusion that virtually all *trans*-splicing of the pI-1 transcript is to SL1. We conclude that the intron 3′ splice site placed into the 5′ UTR of the reporter gene is

being *trans*-spliced and that SL1 snRNP is the major donor. In our hands the SL2 oligonucleotide gave a PCR product only when a transcript was *trans*-spliced to SL2, although the amount of SL2-spliced transcript may be overrepresented due to the plateau effect for common products (from SL1-spliced transcripts) relative to rare products that frequently occurs in PCR. We also determined that the SL2 oligonucleotide does efficiently prime to give PCR products with known SL2-containing mRNAs (data not shown). For these reasons, we believe it is also fair to conclude that a small fraction of the transgene product is *trans*-spliced to SL2.

DISCUSSION

Relationship between *cis*- and *trans*-splicing. *trans*-splicing exists side-by-side with *cis*-splicing in *C. elegans*. Our results suggest that the basic signal differentiating between these two events could simply be the existence near the 5′ end of the transcript of an incomplete intron lacking a 5′ splice site, which we are terming an outron. Processing of precursor mRNAs in *C. elegans* differs from splicing in other metazoans in several ways in addition to *trans*-splicing. The sequence of 5′ splice sites in *C. elegans* is similar to that in other metazoans, but the average intron length in *C. elegans* is much shorter (about 50 nucleotides) and all *C. elegans* introns are extremely A+U-rich. In addition, a polypyrimidine tract is not present, and there is a strong preference for UUUCAG/Pu as the 3′ splice site (4, 10). *trans*-splicing itself appears to be quite similar to *cis*-splicing. First, the consensus sequence for the intron 5′ splice site (G/GUAAG [10]) is matched quite closely by the 5′ splice sites found on SL1 RNA (20) and SL2 RNA (17). Second, the 3′ splice site consensus found for introns and *trans*-spliced outrons is the same (9, 13, 14, 16, 19, 20, 26). Third, outrons are also very rich in A and U (about 70%). Fourth, both *cis*- and *trans*-splicing occur through branched intermediates (2, 4, 35). Finally, *C. elegans* has the usual complement of spliceosomal snRNAs (U1, U2, U4, U5, and U6 [34, 35]), and presumably these participate in both kinds of splicing, although U1 may be replaced in *trans*-splicing by the SL snRNPs (6, 7).

Possible signals for *trans*-splicing? What features does a transcript contain that indicate to the splicing machinery that it is a substrate for *trans*-splicing? The most obvious possibility is a sequence at or near the *trans*-splice site. However, comparison of the sequences of outrons receiving SL1 or SL2 reveals no candidate sequences which might specify *trans*-splicing. So we hypothesized that an intron with no 5′ splice site, at the 5′ end of a transcript, alone could specify *trans*-splicing. To test this, we created an artificial outron from DNA which normally specifies an intron, which we altered in two ways: it was inserted in a new context and its 5′ splice site was eliminated. The changed context, from a position within the coding region to one in the 5′ UTR, cannot be solely responsible for the *trans*-splicing we observe, since the presence of an intron in the 5′ UTR of a naturally occurring mRNA, a similar distance from the 5′ end, has been documented (26). This suggests that the key difference between *cis*- and *trans*-splicing in *C. elegans* is the presence or absence of a functional 5′ splice site on the recipient mRNA. Simply stated, conventional RNAs begin with an exon; transcripts that get *trans*-spliced begin with an intron.

Is the story this simple? *trans*-splicing is normally very efficient; steady-state mRNA contains little if any detectable precursor with natural *trans*-spliced mRNAs. However, although most of the transgene product is *trans*-spliced, we also observed significant amounts of non-*trans*-spliced precursor in RNA from the pI-1 strain. We suggest two explanations for the presence of precursor. First, expression of a *trans*-spliced substrate under control of the powerful *vit-2* promoter could overwhelm the intestinal *trans*-splicing machinery (i.e., the supply of SL1 and SL2 snRNPs). In that case, export of the non-*trans*-spliced precursor from the nucleus might compete effectively with the *trans*-splicing reaction. The second possibility is that *trans*-splicing of the transgene product is not as efficient as are natural *trans*-spliced mRNAs, perhaps because a signal required for efficient *trans*-splicing is not present in the intron used to create the artificial outron. While our experiment suggests that there is no obligatory *trans*-splicing-specific signal in the outron, it does not eliminate this possibility; such a signal could be present fortuitously in the *vit-5* intron used to create the pI-1 outron. However, lack of any obvious sequence similarity between known outrons makes the existence of such a sequence less likely. Furthermore, the extreme similarity between the exon regions of non-*trans*-spliced and *trans*-spliced gene pairs mentioned previously indicates that sequence information for *trans*-splicing is unlikely to be contained in the exons. For all these reasons, we believe that a key signal for *trans*-splicing could be as simple as the presence of a functional 3′ splice site with no upstream functional 5′ splice site.

The determinant of a functional 3′ splice site. Our results indicate a basic equivalence between 3′ *cis*-splice sites and *trans*-splice sites, but what parameters define functional 3′ splice sites? More than the UUUCAG/Pu must be required to signal an outron/exon or intron/exon boundary, since some functional 3′ splice sites are only weak matches to this consensus, while some perfect matches are nonfunctional (unpublished observations). What additional information could define a region as a functional 3′ splice site? The most obvious characteristic is that a 3′ splice site must have a region immediately upstream from it that is recognized as intronlike by the cell. The extension of this premise is that if the intronlike region begins in a 5′ end (presumably capped), it is an outron. There are two additional characteristics shared by *C. elegans* introns and outrons: a branch point and a very asymmetric base composition. Although there is no obvious conserved branch point sequence, a requirement presumably exists for a particular nucleotide or nucleotides some distance upstream from the 3′ splice site. Second, all introns and outrons have an approximately 70% A+U composition in the sequence preceding the 3′ splice site (4). In addition, there is a minimum intron length found in *C. elegans* introns (38 nucleotides in *unc-54* [18]), and it is possible that there is a similar requirement for outrons. This is suggested by the existence of a putative *trans*-splice acceptor site which is not utilized in the 5′ UTR of the *act-4* gene, perhaps due to an inadequate length of RNA upstream from the splice site (30 nucleotides). Thus, any or all of these parameters—A+U richness, length, proximity to the cap site, or presence of a branch site—may play a role in recognition of a *trans*-splice acceptor.

SL1- versus SL2-specific splicing. If the presence of an outron is the primary signal for *trans*-splicing, where is the information for specifying SL1 versus SL2? In our artificial construct, the vast majority of the transcript was spliced to SL1. This could be explained by the coincidental presence of an SL1-specific signal in the *vit-5* intron, added to the 5′ UTR. Alternatively, it is conceivable that the ratio of SL1 to

SL2 spliced onto the transgene transcript simply reflects the ratio of SL1 to SL2 snRNPs in the intestinal cells (currently unknown). However, it seems most likely to us that SL1 splicing represents a default mode. While SL1 is found in a wide variety of free-living and parasitic nematodes (1, 25, 33, 37), SL2 has been found only in *Caenorhabditis* species (17). Furthermore, more than 20 genes have been found whose products are *trans*-spliced to SL1 (9, 13, 14, 16, 19, 20, 26; unpublished observations), but only 3 are known to be spliced to SL2 (11a, 17, 18a). Although there is no obvious sequence homology between the outrons of these SL2-accepting genes, it seems likely that this region contains information that results in SL2 splicing in place of SL1.

ACKNOWLEDGMENTS

We are grateful to Kristi Lea for expert technical assistance and to Susan Strome for critically reviewing an early version of the manuscript. We also thank Craig Hunter for suggesting the word outron.

This work was supported by Public Health Service grant GM30870 from the National Institute of General Medical Sciences. J.T. was supported by a National Research Service award from N.I.G.M.S.

REFERENCES

1. **Bektesh, S., K. Van Doren, and D. Hirsh.** 1988. Presence of the *Caenorhabditis elegans* spliced leader on different mRNAs and in different genera of nematodes. Genes Dev. **2**:527–535.
2. **Bektesh, S. L., and D. I. Hirsh.** 1988. *C. elegans* mRNAs acquire a spliced leader through a *trans* splicing mechanism. Nucleic Acids Res. **16**:5692.
3. **Blumenthal, T., M. Squire, S. Kirtland, J. Cane, M. Donegan, J. Spieth, and W. J. Sharrock.** 1984. Cloning of a yolk protein gene family from *Caenorhabditis elegans*. J. Mol. Biol. **174**:1–18.
4. **Blumenthal, T., and J. Thomas.** 1988. Cis and trans mRNA splicing in *C. elegans*. Trends Genet. **4**:305–308.
5. **Brenner, S.** 1974. The genetics of *Caenorhabditis elegans*. Genetics **77**:71–94.
6. **Bruzik, J. P., and J. A. Steitz.** 1990. Spliced leader RNA sequences can substitute for the essential 5′ end of U1 RNA during splicing in a mammalian in vitro system. Cell **62**:889–899.
7. **Bruzik, J. P., K. Van Doren, D. Hirsh, and J. A. Stietz.** 1988. Trans splicing involves a novel form of small nuclear ribonucleoprotein particles. Nature (London) **335**:559–562.
8. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
9. **Cummins, C., and P. Anderson.** 1988. Regulatory myosin light-chain genes of *Caenorhabditis elegans*. Mol. Cell. Biol. **8**:5339–5349.
10. **Emmons, S. W.** 1988. The genome, p. 47–79. *In* W. B. Wood (ed.), The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
11. **Fire, A.** 1986. Integrative transformation of *Caenorhabditis elegans*. EMBO J. **5**:2673–2680.

11a. **Fire, A.** Personal communication.

12. **Fire, A., K. Kondo, and R. Waterston.** 1990. Vectors for low copy transformation of *C. elegans*. Nucleic Acids Res. **18**:4269–4270.
13. **Georgi, L. L., P. S. Albert, and D. L. Riddle.** 1990. *daf-1*, a *C. elegans* gene controlling dauer larva development, encodes a novel receptor protein kinase. Cell **61**:635–645.
14. **Graham, R. W., D. Jones, and E. P. M. Candido.** 1989. *ubiA*, the major polyubiquitin locus in *Caenorhabditis elegans*, has unusual structural features and is constitutively expressed. Mol. Cell. Biol. **9**:268–277.
15. **Hamlyn, P. H., G. G. Brownlee, C. C. Cheng, M. J. Gait, and C. Milstein.** 1978. Complete sequence of constant and noncoding regions of an immunoglobulin mRNA using the dideoxynucleotide method of RNA sequencing. Cell **15**:1067–1075.
16. **Huang, X.-Y., L. A. M. Barrios, P. Vonkhorporn, S. Honda, D. G. Albertson, and R. M. Hecht.** 1989. Genomic organization of the glyceraldehyde-3-phosphate dehydrogenase gene family of *Caenorhabditis elegans*. J. Mol. Biol. **206**:411–424.
17. **Huang, X.-Y., and D. Hirsh.** 1989. A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **86**:8640–8644.
18. **Karn, J., S. Brenner, and L. Barnett.** 1983. Protein structural domains in the *Caenorhabditis elegans unc-54* myosin heavy chain gene are not separated by introns. Proc. Natl. Acad. Sci. USA **80**:4253–4257.

18a. **Kimble, J.** Personal communication.

19. **Kramer, J. M., R. P. French, E.-C. Park, and J. J. Johnson.** 1990. The *Caenorhabditis elegans rol-6* gene, which interacts with the *sqt-1* collagen gene to encode organismal morphology, encodes a collagen. Mol. Cell. Biol. **10**:2081–2089.
20. **Krause, M., and D. Hirsh.** 1987. A trans-spliced leader sequence on actin mRNA in *C. elegans*. Cell **49**:753–761.
21. **Laird, P. W.** 1989. *Trans* splicing in trypanosomes: archaism or adaptation? Trends Genet. **5**:204–208.
22. **MacLeod, A. R., J. Karn, and S. Brenner.** 1981. Molecular analysis of the *unc-54* myosin heavy-chain gene of *Caenorhabditis elegans*. Nature (London) **291**:386–390.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 122–126. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Murphy, W. J., K. P. Watkins, and N. Agabian.** 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. Cell **47**:517–525.
25. **Nilsen, T., J. Shambaugh, J. Denker, G. Chubb, C. Faser, L. Putnam, and K. Bennett.** 1989. Characterization and expression of a spliced leader RNA in the parasitic nematode *Ascaris lumbricoides* var. *suum*. Mol. Cell. Biol. **9**:3543–3547.
26. **Park, Y.-S., and J. M. Kramer.** 1990. Tandemly duplicated *Caenorhabditis elegans* collagen genes differ in their modes of splicing. J. Mol. Biol. **211**:395–406.
27. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–489.
28. **Spieth, J., and T. Blumenthal.** 1985. The *Caenorhabditis elegans* gene family includes a gene encoding a distantly related protein. Mol. Cell. Biol. **5**:2495–2501.
29. **Spieth, J., K. Denison, S. Kirtland, J. Cane, and T. Blumenthal.** 1985. The *C. elegans* vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes. Nucleic Acids Res. **13**:5283–5295.
30. **Spieth, J., K. Denison, E. Zucker, and T. Blumenthal.** 1985. The nucleotide sequence of a nematode vitellogenin gene. Nucleic Acids Res. **13**:7129–7138.
31. **Spieth, J., M. MacMorris, S. Broverman, S. Greenspoon, and T. Blumenthal.** 1988. Regulated expression of a vitellogenin fusion gene in transgenic nematodes. Dev. Biol. **130**:285–293.
32. **Sutton, R. E., and J. C. Boothroyd.** 1986. Evidence for *trans* splicing in trypanosomes. Cell **47**:527–535.
33. **Takacs, A. M., J. A. Denker, K. G. Perrine, P. A. Maroney, and T. W. Nilsen.** 1988. The 22-nucleotide spliced-leader sequence in the human parasite nematode *Brugia malayi* is identical to the trans-spliced leader exon in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **85**:7932–7936.
34. **Thomas, J., K. Lea, E. Zucker-Aprison, and T. Blumenthal.** 1990. The spliceosomal snRNPs of *Caenorhabditis elegans*. Nucleic Acids Res. **18**:2633–2642.
35. **Thomas, J. D., R. C. Conrad, and T. Blumenthal.** 1988. The *C. elegans* trans-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. Cell **54**:533–539.
36. **Van Doren, K., and D. Hirsh.** 1988. Trans-spliced leader RNA exists as small nuclear ribonucleoprotein particles in *C. elegans*. Nature (London) **335**:556–559.
37. **Zeng, W., C. M. Alarcon, and J. E. Donelson.** 1990. Many transcribed regions of the *Onchocerca volvulus* genome contain the spliced leader sequence of *Caenorhabditis elegans*. Mol. Cell. Biol. **10**:2765–2773.