RNA (1997), 3:269–278. Cambridge University Press. Printed in the USA.

Relationship between 3′ end formation and SL2-specific *trans*-splicing in polycistronic *Caenorhabditis elegans* pre-mRNA processing

SCOTT KUERSTEN, KRISTI LEA, MARGARET MacMORRIS, JOHN SPIETH,[1] and THOMAS BLUMENTHAL

Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

ABSTRACT

About 25% of the genes in the nematode *Caenorhabditis elegans* are in operons, polycistronic transcription units in which the genes are only 100–400 bp apart. The operon pre-mRNAs are processed into monocistronic mRNAs by a combination of cleavage and polyadenylation at the 3′ end of the upstream mRNA and SL2 *trans*-splicing at the 5′ end of the downstream mRNA. To determine whether 3′ end formation and SL2 *trans*-splicing are coupled mechanistically, we tested a *gpd-2/gpd-3* operon construct driven by a *C. elegans* heat shock promoter, and measured the effects of inhibition of 3′ end formation and/or *trans*-splicing on the processing of the polycistronic RNA in vivo. The results indicate that proper 3′ end formation of the upstream mRNA in an operon is required for SL2-specificity of downstream mRNA *trans*-splicing. In contrast, *trans*-splicing of the downstream mRNA is not necessary for correct 3′ end formation of the upstream mRNA. In addition, shortening the distance between the 5′ cap and the AAUAAA of *gpd-2* (the upstream gene) decreases the efficiency of 3′ end formation and is accompanied by a replacement of SL2 with SL1 at the *trans*-splice site of *gpd-3*, the downstream gene. These results indicate that SL2 *trans*-splicing, in *C. elegans*, is coupled mechanistically to 3′ end formation in the processing of polycistronic pre-mRNAs.

Keywords: nematode; operons; polyadenylation; spliced leader; splicing

INTRODUCTION

Pre-mRNAs of most eukaryotes are processed by splicing to remove introns. This reaction is catalyzed by small nuclear ribonucleoprotein particles (snRNPs) and several associated proteins that form a large complex called the spliceosome (Moore et al., 1993). Interestingly, pre-mRNAs of the nematode *Caenorhabditis elegans* are processed by at least two additional spliceosome-catalyzed reactions that do not occur in most other eukaryotes. First, many pre-mRNAs are *trans*-spliced (Krause & Hirsh, 1987) in a reaction closely related to conventional intron removal (*cis*-splicing). In fact, the donor in *trans*-splicing, the SL RNA, is itself packaged as an snRNP. *Trans*-splicing involves the transfer of a 22-nt leader sequence from the SL snRNP to the 5′ ends of about 70% of *C. elegans* mRNAs (reviewed in Nilsen, 1993; Blumenthal, 1995). The majority of *trans*-splicing utilizes SL1 RNA and most SL1 *trans*-splicing occurs near the 5′ ends of pre-mRNAs that begin with an outron, an AU-rich intron-like sequence containing a functional 3′ splice site, but lacking a 5′ splice site (Conrad et al., 1991, 1993, 1995).

The second, and even more surprising and unusual, spliceosome-catalyzed reaction involves processing *C. elegans* operons. About 25% of *C. elegans* genes are found in closely linked clusters that are co-transcribed as polycistronic pre-mRNAs (Spieth et al., 1993; Zorio et al., 1994). Processing of these polycistronic precursors into mature transcripts involves a combination of cleavage and polyadenylation of the upstream mRNA and *trans*-splicing of the downstream mRNA, usually separated by about 100 bp. A second spliced leader RNA, SL2 (Huang & Hirsh, 1989), apparently is used exclusively for *trans*-splicing to downstream mRNAs in these polycistronic transcription units, although such *trans*-splice sites sometimes accept both SL2 and SL1 (Spieth et al., 1993; Zorio et al., 1994).

The process of 3′ end formation is best understood in vertebrates, where it involves two heterotrimeric protein complexes that bind near the site of cleavage, as well as several additional proteins required for the cleavage and the polyadenylation reactions (reviewed

Reprint requests to Thomas Blumenthal at his present address: Department of Biochemistry, Biophysics and Genetics, University of Colorado School of Medicine, 4200 E. 9th Ave., Denver, Colorado 80262, USA; e-mail: tom.blumenthal@uchsc.edu.

[1] Present address: Washington University School of Medicine, Department of Genetics, St. Louis, Missouri 63108, USA.

in Keller, 1995; Manley, 1995; O'Hare, 1995). The cleavage and polyadenylation specificity factor (CPSF) binds to the hexanucleotide sequence, AAUAAA, typically located about 15 nt upstream of the poly(A) site, whereas the cleavage stimulatory factor (CstF) binds to a U-rich or G/U-rich sequence about 30 nt downstream of the poly(A) site. The interaction of these complexes with the pre-mRNA determines the cleavage and polyadenylation site. Although 3′ end formation has not been investigated fully in *C. elegans*, it seems to use many of the same components. A survey of cDNA 3′ ends indicated that more than 50% contain an AAUAAA appropriately spaced upstream of the site of poly(A) addition, but *C. elegans* allows more sequence variation in the binding site because most of the remaining 3′ ends contain close matches to the consensus (Blumenthal & Steward, 1996). Although it has not been demonstrated that CstF plays a role in cleavage and polyadenylation in *C. elegans*, the genomic sequencing project has uncovered all three subunits of the complex (Wilson et al., 1994; C. Williams, pers. comm.), so it is likely that 3′ end formation requires CstF.

The close proximity of the SL2 *trans*-splice site of the downstream mRNA and the cleavage and polyadenylation site of the upstream mRNA suggests that the molecular "machinery" for 3′ end formation and splicing may interact in operon pre-mRNA processing. Consistent with this idea, a link between *cis*-splicing and 3′ end formation has been observed in mammalian pre-mRNA processing. Niwa et al. (1990) demonstrated that in vitro polyadenylation is stimulated by the presence of an intron upstream. Cooke and Alwine (1996) further demonstrated that this stimulation could be achieved when only a functional 3′ splice site was present upstream. In addition, it was shown in vivo and in vitro that mutating the AAUAAA reduced the efficiency of splicing of the proximal intron, but not distal introns (Chiou et al., 1991; Niwa & Berget, 1991; Nesic et al., 1993; Nesic & Maquat 1994).

A connection between 3′ end formation and *trans*-splicing has been demonstrated for the processing of polycistronic pre-mRNAs in the Trypanosomatidae, although there are distinct differences in the processing signals compared to *C. elegans*. First, trypanosomes have a single SL RNA that is different in size and sequence from either of *C. elegans*' two SL RNAs (Bruzik et al., 1988). Second, trypanosome mRNAs do not contain an AAUAAA consensus sequence and apparently no sequences upstream of the 3′ end play a role in cleavage and polyadenylation, so it is unlikely that CPSF recognizes a sequence on the pre-mRNA. Instead, it has been suggested that the site of 3′ end formation is dictated either by the site of downstream *trans*-splicing (LeBowitz et al., 1993), or by a pyrimidine-rich sequence located within the intergenic region (Matthews et al., 1994). Because *C. elegans* polycistronic pre-mRNAs do have AAUAAA just upstream of their internal poly(A) sites, 3′ end formation is predicted to occur by conventional mechanisms.

In order to investigate the possibility that 3′ end formation and SL2-specific *trans*-splicing are linked mechanistically in the processing of *C. elegans* polycistronic precursors, we have developed a system for studying operon RNA processing in transgenic worms. A construct derived from the *mai-1/gpd-2/gpd-3* operon was placed under the control of a heat shock promoter (Spieth et al., 1993) and mutations in 3′ end formation and/or *trans*-splicing signals were introduced. The results demonstrate that *trans*-splicing specificity of the downstream mRNA is dependent upon a functional AAUAAA upstream. In contrast, correct 3′ end formation of the upstream mRNA is independent of downstream *trans*-splicing. In addition, we show that, when 3′ end formation becomes less efficient because the distance between the 5′ cap and the AAUAAA is decreased, SL2 *trans*-splicing is replaced by SL1.

RESULTS

Expression of *gpd-3* mRNA from an operon under heat shock control

In order to test whether cleavage and polyadenylation of the upstream mRNA in a polycistronic pre-mRNA is connected to SL2-specific *trans*-splicing that occurs about 100 nt downstream, we have tested a synthetic construct in transgenic worms in which specific mutations in signals for 3′ end formation or *trans*-splicing were introduced. The operon construct is derived from the three-gene *mai-1/gpd-2/gpd-3* operon (Spieth et al., 1993) (Fig. 1A). The *C. elegans hsp-16* promoter (Stringham et al., 1992) was fused upstream of *gpd-2* and used to drive the expression of both *gpd-2* and a modified *gpd-3* tagged with a *vit-6* sequence at its 3′ end. This construct is referred to as HS1496 because it contains 1,496 bp of DNA upstream of the *gpd-3 trans*-splice site. We have shown previously, using a RT/PCR assay, that upon heat shock the operon is expressed and *gpd-3* mRNA is *trans*-spliced to SL2 as it normally is under the control of its own promoter (Spieth et al., 1993). In order to obtain more quantitative data, and to analyze the products of the upstream and downstream genes concurrently, we used an RNase protection assay (Fig. 1A). The probe will hybridize to the 3′ end of *gpd-2* (Fig. 1A, product *b*) formed by cleavage and polyadenylation, and the 5′ end of *gpd-3* (product *c*) formed by *trans*-splicing. In addition, this probe will protect polycistronic RNA (product *a*). Because the sequences of *gpd-2* and *gpd-3* are so similar, the probe also protects the 5′ end of *gpd-2* (product *d*), the 3′ end of *gpd-3* (product *e*), as well as part of the second exon of both genes (product *f*).

The plasmid construct shown in Figure 1A was injected into the syncytial gonad of adult nematodes as

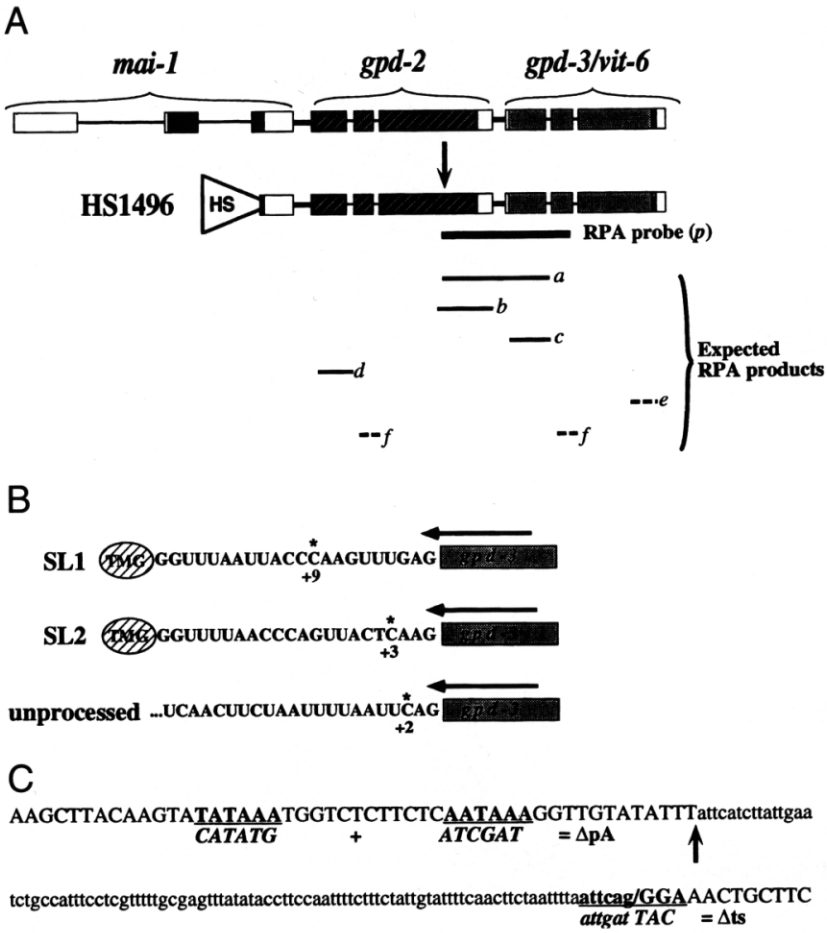

FIGURE 1. Analysis of the HS1496 operon construct. **A:** Expected RNase protection analysis (RPA) products. The HS1496 construct is as described in Spieth et al. (1993). Top diagram shows the exons and introns of the *mai-1* operon in which a small fragment of *vit-6* was fused to *gpd-3* near the 3′ end of the operon. (This feature of the construct was created to facilitate analysis of *gpd-3* by RT-PCR and northern blots. It was not used in the experiments described in this paper.) Filled bars represent coding regions; open bars, noncoding regions of exons; narrow lines, introns; wider lines, intercistronic sequences. The *C. elegans hsp-16* promoter was placed upstream of the *gpd-2* gene within the last exon of *mai-1* as shown by the triangle in the HS1496 diagram. The dark bar below the HS1496 diagram indicates the region covered by the RNase protection probe. The predicted products are shown as follows: *a*, polycistronic RNA (679 nt); *b*, *gpd-2* 3′ end (322 nt); *c*, *gpd-3* 5′ end (259 nt); *d*, *gpd-2* and polycistronic RNA 5′ ends (241 nt); *e*, *gpd-3* 3′ end (158 nt); *f*, *gpd-2* and *gpd-3* exon 2 (113 nt). Products *e* and *f* (dashed lines) are not discussed further in this report. **B:** Primer extension analysis. SL specificity is assayed by primer extension in the presence of ddGTP. The mature 5′ end of *gpd-3* is *trans*-spliced to either SL1 or SL2. ddGTP is used to terminate primer extension at the position where the reverse transcriptase encounters the first C residue. This is at +9 for SL1, +3 for SL2, and +2 for unprocessed *gpd-3*. The arrow indicates the position of the 18-nt primer that extends 1 nt beyond the *trans*-splice site. TMG represents the trimethyl guanosine cap found at the 5′ ends of both spliced leaders. **C:** Poly(A) and *trans*-splice site mutations. The sequence including the end of the last exon of *gpd-2*, the *gpd-2/gpd-3* intercistronic space, and the beginning of the first exon of *gpd-3* is shown. Exon sequences are upper case; intercistronic sequences, lower case. The signals for 3′ end formation (UAUAAA, AAUAAA) and the *trans*-splice site (attcag/GGA; / represents the *trans*-splice site) are shown in bold letters. Mutations that were made to these signals are shown in italics below the mutated bases, and their names are indicated. The arrow shows the site of cleavage and polyadenylation.

described by Mello et al. (1991) and transgenic lines were established using the *rol-6* (*su1006*) plasmid as a marker for transformation (Spieth et al., 1993). RNA from two strains containing HS1496 was isolated from untreated nematodes (−HS) or populations heat shocked at 29 °C for 2 h (+HS). The results, shown in Figure 2, indicate that, upon heat shock, both *gpd-2* and *gpd-3* are expressed and processed as expected. Cleavage and polyadenylation of *gpd-2* produces the expected 320 nt protected fragment (Fig. 2, lane 2, product *b*), whereas *trans*-spliced *gpd-3* mRNA protects a 256-nt fragment (Fig. 2, lane 2, product *c*). Because no detectable precursor RNA (product *a*) accumulates after heat shock, processing of the polycistronic precursor is apparently complete.

We expect that, if all of the polycistronic pre-mRNA produced upon heat shock is processed efficiently and correctly, *gpd-2* and *gpd-3* products will be produced in equal amounts. Therefore, the levels of products *b*, *c*, and *d* should be the same. As seen in Figure 2 (lane 2), *trans*-spliced *gpd-2* 5′ end (product *d*) and *gpd-2* 3′ end (product *b*) are present in about equal amounts, but *trans*-spliced *gpd-3* 5′ end accumulates to only about 25% of the level of *gpd-2*. In contrast, upon longer exposures of gels like that in Figure 2, endogenous *gpd-2* and *gpd-3* products are seen to be present in equal amounts (data not shown). This difference is likely due to inefficient *trans*-splicing of *gpd-3* when induced by heat shock. Cleavage and polyadenylation of *gpd-2* is predicted to create an uncapped 5′ end, thus exposing the downstream RNA to exonucleases. If *trans*-splicing were inefficient, then some *gpd-3* RNA would be degraded before being *trans*-spliced, and would result in reduced accumulation of *gpd-3* mRNA.

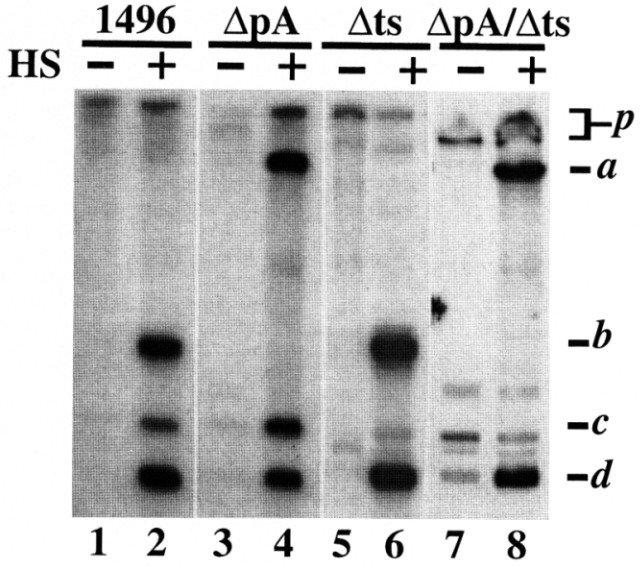

FIGURE 2. RPA of total RNA from HS1496 and various mutant transgenic strains. Transgenic worm strains were grown as described in Materials and Methods. Half of each population was subjected to a 2-h 29 °C heat shock. RNA was isolated and RPA was performed as described in Materials and Methods. RPA was performed on several strains carrying each construct, but data from only a single representative strain is shown. Each pair of lanes is labeled (−) for RNA from worms that were not heat shocked and (+) for RNA from worms that were heat shocked. Lower-case letters on the right refer to the products identified in Figure 1A. Low levels of products seen in the (−) lanes represent endogenous *gpd-2* and *gpd-3* transcripts.

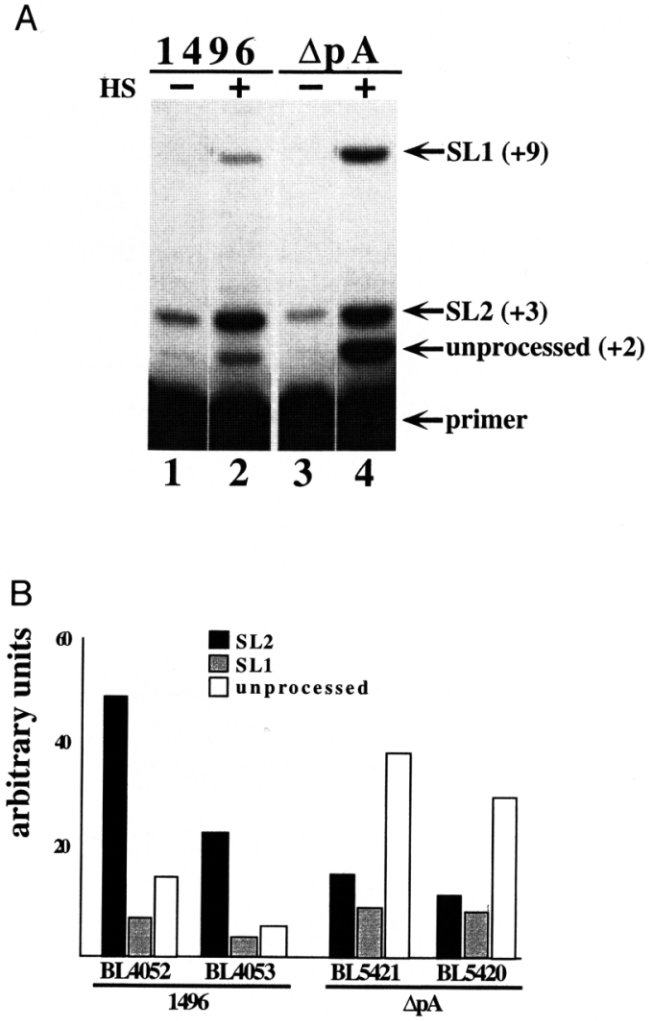

FIGURE 3. Primer extension analysis of total RNA from HS1496 and ΔpA. **A:** The same RNA preparations analyzed in Figure 2 by RPA were analyzed by primer extension in the presence of dideoxyGTP as described in Materials and Methods and Figure 1B. Expected positions of the three possible products are indicated on the right. The HS1496 strain is BL4052. The ΔpA strain is BL5421. **B:** Quantitation of the products. Gels similar to that shown in A were quantitated by phosphorimager analysis and relative levels of the three products from each strain are given in arbitrary units. Separate analysis of two strains carrying each construct is shown. Note that, although there is variability from strain to strain in the levels of expression with each construct, the relative proportions of the products are quite similar. The two HS1496 strains are BL4052 (left) and BL4053. The two ΔpA strains are BL5421 (left) and BL5420.

A second, but less likely, possibility is that the transgenic *trans*-spliced *gpd-3/vit-6* fusion mRNA is less stable than the *gpd-2* mRNA following heat shock.

Trans-splicing specificity of *gpd-3* following heat shock

If the operon pre-mRNA is processed properly from the heat shock promoter, we would expect the majority of *gpd-3* product to be *trans*-spliced to SL2. Previously, it was determined that the *gpd-3* mRNA produced following induction was *trans*-spliced properly to SL2 using an RT/PCR assay (Spieth et al., 1993). In order to confirm that result with a more quantitative assay, we performed primer extensions using an oligonucleotide that spans the *trans*-splice site from +17 to −1 in the presence of dideoxyGTP (Fig. 1B). This assay will distinguish by size the three predicted products, *gpd-3* precursor, SL1 *trans*-spliced mRNA, and SL2 *trans*-spliced mRNA, because the reverse transcriptase encounters the first C residue at different positions with each RNA template (as shown in Fig. 1B). The results demonstrate that most HS1496 product is *trans*-spliced to SL2 (Fig. 3A, lanes 1 and 2). As determined previously, endogenous *gpd-3* is *trans*-spliced only to SL2 in worms that are not heat shocked (Fig. 3A, lane 1) (Huang & Hirsh, 1989; Spieth et al., 1993). The small amount of unprocessed transgenic *gpd-3* RNA that accumulates upon heat shock indicates that processing is nearly complete. Quantitative analysis of two independent HS1496 transgenic strains (Fig. 3B) indicates that about 90% of the *trans*-spliced *gpd-3* product is spliced to SL2. This result demonstrates that an operon expressed from a transgenic array following a 2-h heat shock is processed correctly, confirming our previous result (Spieth et al., 1993). This allows introduction of mutations in the sequences required for 3′ end formation and/or *trans*-splicing in order to assay their effects on polycistronic pre-mRNA processing.

Trans-splicing specificity is dependent upon 3′ end formation

In eukaryotes, a protein complex called the cleavage and polyadenylation specificity factor (CPSF) recognizes and binds to the hexanucleotide sequence AAUAAA. This step constitutes the initial recognition of the site of 3′ end formation (reviewed in Keller, 1995; Manley, 1995; O'Hare, 1995). In order to test whether 3′ end formation of *gpd-2* is required for *trans*-splicing specificity of *gpd-3*, we mutated the AAUAAA (and a cryptic UAUAAA sequence located just upstream) in *gpd-2* to the sequences shown in Figure 1C and produced transgenic lines that carry this plasmid (called ΔpA). RNA was isolated and analyzed as described above. RNase protection assays indicate that proper *gpd-2* 3′ end formation does not occur in the absence of an AAUAAA; no *gpd-2* 3′ end product (*b*) accumulates upon heat shock (Fig. 2, lane 4). However, a large amount of polycistronic RNA (product *a*) accumulates, which presumably represents the *gpd-3* product that would have been degraded because of failure of *trans*-splicing following heat shock.

Trans-spliced *gpd-3* mRNA (product *c*) also accumulates, indicating that *trans*-splicing can occur when upstream 3′ end formation is prevented. Because no 3′ end formation occurs, this *trans*-spliced product must result from *trans*-splicing of the polycistronic pre-mRNA. Which spliced leader, SL1 or SL2, is used in this unexpected intercistronic processing event? Surprisingly, both SL1 and SL2 are *trans*-spliced to *gpd-3* in ΔpA RNA (Fig. 3A, lane 4). The quantitation of this data, shown in Figure 3B, indicates that, in ΔpA, ~60% of *trans*-splicing is to SL2, compared to ~90% in the HS1496 control. Thus, preventing 3′ end formation leads to a large accumulation of polycistronic pre-mRNA, which is *trans*-spliced, but with loss of specificity. These results suggest that SL2-specific *trans*-splicing is indeed connected to cleavage and polyadenylation. When cleavage and polyadenylation are prevented, *trans*-splicing is aberrant even though the *trans*-splice site is intact.

Mutation of the *trans*-splice site prevents accumulation of *gpd-3* mRNA

To test the possibility that *trans*-splicing of the downstream mRNA might have an influence on 3′ end formation of the upstream mRNA, we eliminated the *trans*-splice site of *gpd-3* (Δts) (Fig. 1C). Analysis of RNA isolated from transgenic strains carrying this construct demonstrates that this mutation eliminates all properly processed *gpd-3* product, presumably because *trans*-splicing does not occur. However, 3′ end formation apparently is unaffected. *Gpd-2* product (*b*), with the correct 3′ end, forms at the expected level, and no polycistronic RNA (*a*) accumulates (Fig. 2, lane 6). This result demonstrates that *trans*-splicing does not influence upstream 3′ end formation. The downstream RNA resulting from cleavage at the 3′ end of *gpd-2* is *gpd-3* plus the intercistronic space. This product would migrate just above product *b*, but was never observed. We hypothesize that, when *gpd-2* 3′ end formation occurs, the downstream product remains uncapped and, because it cannot be *trans*-spliced, is degraded rapidly.

When both the CPSF binding sites and the *trans*-splice site are eliminated (ΔpA/Δts), both 3′ end formation and *trans*-splicing are prevented (Fig. 2, lanes 7 and 8). As expected, the only RNA that accumulates upon heat shock is the polycistronic RNA that gives both products *a* and *d*.

These results indicate that, whereas preventing 3′ end formation of the upstream mRNA does influence *trans*-splicing of the downstream mRNA, the converse is not true. Inhibition of *trans*-splicing has no effect on 3′ end formation upstream.

Effect of heat shock promoter location on 3′ end formation and *trans*-splicing specificity

Complete elimination of 3′ end formation results in expression of a polycistronic pre-mRNA that is *trans*-spliced without apparent SL specificity. It has been demonstrated previously that efficiency of 3′ end formation can be reduced by decreasing the distance between the 5′ cap and the AAUAAA (Iwasaki & Temin, 1990; Sanfaçon & Hohn, 1990; S. Flaherty & G. Gilmartin, pers. comm.). In order to determine whether SL specificity correlates with efficiency of 3′ end formation, we placed the heat shock promoter at several locations within *gpd-2* (thus reducing the distance from the cap to the AAUAAA) and measured 3′ end formation by RNase protection and SL specificity by primer extension as above. We predicted that shortening the distance between the cap and the AAUAAA of *gpd-2* would result in less efficient 3′ end formation and might thus result in reduction of SL2-specific *trans*-splicing to *gpd-3*.

The deletions tested are shown in Figure 4A. In HS872, the heat shock promoter is placed at the 5′ end of the third exon of *gpd-2*; in HS229, it is just upstream of the *gpd-2* translation termination codon; and in HS78, it is in the intercistronic space, 78 bp upstream of the *gpd-3 trans*-splice site. The results of the *gpd-2* deletions on 3′ end formation are shown in Figure 4B. First, it can be seen that the pre-mRNAs are processed efficiently in all cases, because no polycistronic precursors accumulate. Second, as predicted, 3′ end formation is less efficient when the cap is moved closer to the AAUAAA of *gpd-2*. A comparison of the ratio of product *b* (after subtracting the value for the no heat shock control, which represents endogenous *gpd-2* mRNA) to product *c* in the HS872 mutant shows clearly that product *b*, resulting from *gpd-2* 3′ end formation, is reduced (Fig. 4B). Presumably, when *trans*-splicing

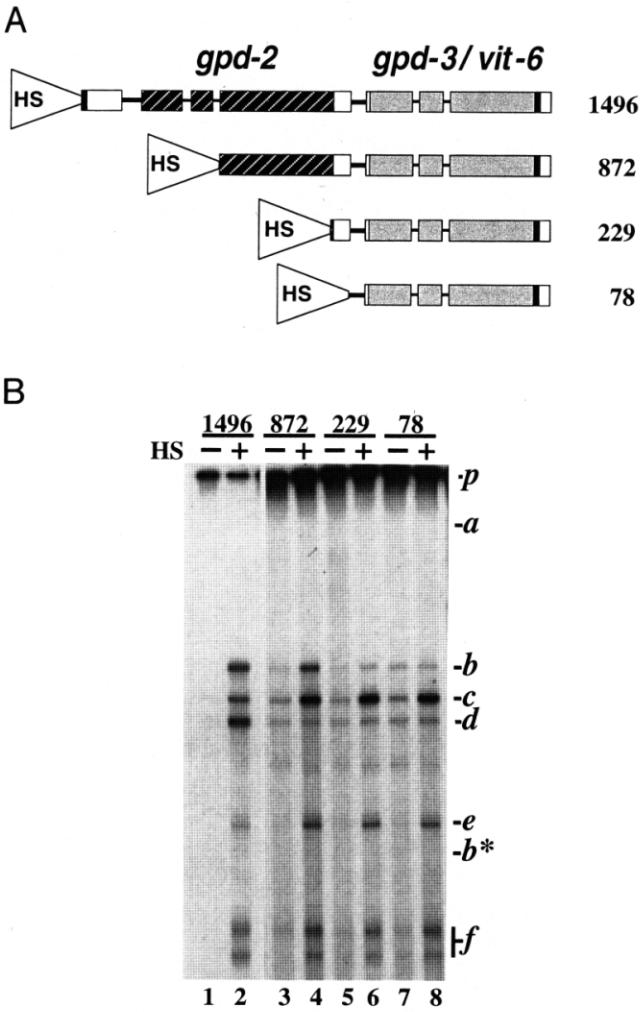

FIGURE 4. RPA of RNA from strains carrying deletion mutant constructs. **A:** Deletion constructs are diagrammed and were made as described in Spieth et al. (1993). Triangle at the left of each bar represents the location of the heat shock promoter. Numbers to the right indicate the distance (in base pairs) from the heat shock promoter to the *gpd-3 trans*-splice site. **B:** RPA of total RNA from transgenic strains carrying the constructs in A. Lower-case letters on the right refer to the products identified in Figure 1A, except *b**, which corresponds to the location expected from a product resulting from protection of the probe by the 3′ end of the remaining *gpd-2* sequence in the 229 mutant construct.

of *gpd-3* is not accompanied by 3′ end formation of *gpd-2*, the upstream product, left unprotected by lack of a poly(A) tail, is degraded. In the HS229 mutant, the 3′ end of *gpd-2* would produce a protected fragment of a smaller size (product *b** would be 135 nt in length). Because we detected no product at this position, we conclude that 3′ end formation in the HS229 mutant is either very inefficient or does not occur at all. These results suggest that, as the distance between the cap and the AAUAAA is shortened, the efficiency of cleavage and polyadenylation decreases, although the possibility that 3′ end formation did occur, but that the products were unstable, cannot be eliminated.

Was there a concomitant loss in SL specificity? The effects on *trans*-splicing specificity were assayed by dideoxyguanosine primer extension (Fig. 5A) and indicate that, when the cap is closer to the AAUAAA, SL2 specificity is lost. In HS1496, virtually all of the *trans*-splicing is to SL2 (Fig. 5A, lane 2), whereas, with HS872, a significant level of SL1 *trans*-splicing occurs. When even more of the *gpd-2* gene is deleted, the amount of SL2 *trans*-spliced product decreases further, until it reaches background levels and is totally replaced by SL1 *trans*-spliced product in HS78 (where no 3′ end formation is possible). Quantitation of similar data with two strains carrying each construct is shown in Figure 5B. In HS1496 strains, 92% of the *gpd-3* product is *trans*-spliced to SL2. This figure decreases to

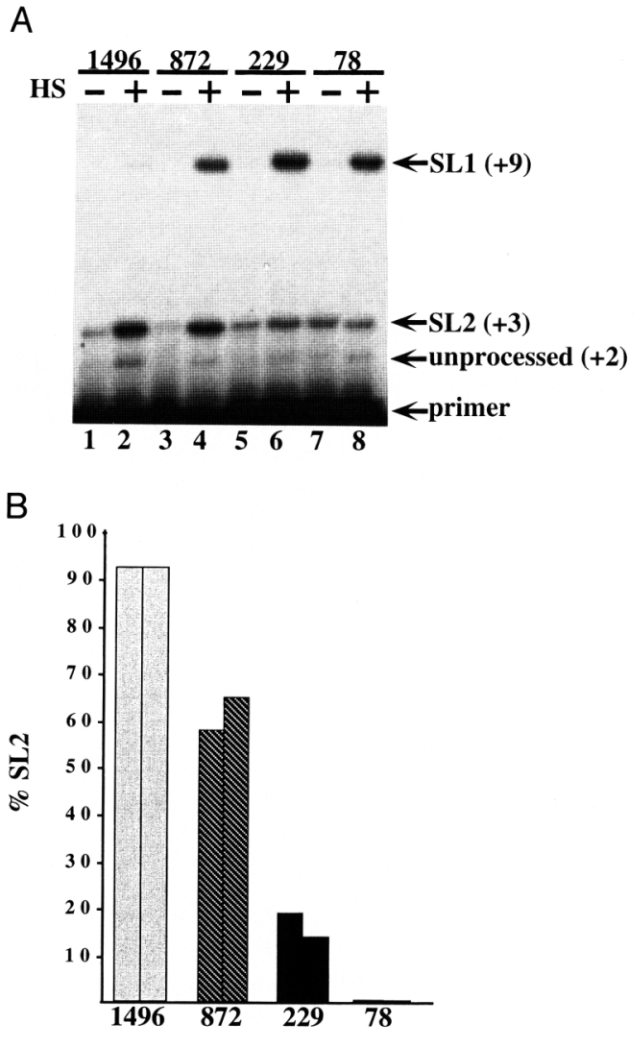

FIGURE 5. Primer extension analysis of total RNA from HS1496 and deletion mutant constructs. **A:** ddG primer extension of RNA from the constructs in Figure 4A. HS− and HS+ lanes are indicated. Identities of the products are shown to the right. The strains are: HS1496, BL4052; HS872, BL5200; HS229, BL4021; and HS78, BL5031. **B:** Quantitation of the products. Gels similar to that shown in A were quantitated by phosphorimager analysis and relative levels of the SL1 and SL2 products calculated. Separate analysis of two strains carrying each construct is shown. Data are plotted as percent of total *trans*-spliced product. Strains tested were: HS1496-BL4052 and BL4053; HS872-BL5200 and BL5201; HS229-BL4021 and BL4022; HS78-BL5031 and BL5032.

57–64% in HS872, to 14–20% in HS229 strains, and to zero in HS78 strains.

These results are consistent with the hypothesis that, in the processing of a polycistronic pre-mRNA, SL2-specific *trans*-splicing of the mRNA for the downstream gene is mechanistically linked to cleavage and polyadenylation of the mRNA for the upstream gene. As the distance between the cap and the *gpd-2* 3′ end is shortened, cleavage and polyadenylation efficiency decreases, accompanied by a loss in SL2 specificity.

DISCUSSION

AAUAAA is needed for 3′ end formation and SL specificity

In order to test what role, if any, 3′ end formation has in specifying SL2 *trans*-splicing to downstream products of polycistronic pre-mRNAs in *C. elegans*, we performed a series of experiments in which 3′ end formation was either eliminated completely by mutating the CPSF binding sites of the *gpd-2/gpd-3* operon construct, or made less efficient by decreasing the distance between the cap and the AAUAAA. In both of these experiments, loss or reduction of 3′ end formation was accompanied by a decrease in SL2-specific *trans*-splicing to the downstream mRNA, *gpd-3*. When the *gpd-2* AAUAAAs were mutated (ΔpA), 3′ end formation failed to occur (Fig. 2, lane 4). In addition, we observed a reduction in SL2 specificity; both SL1 and SL2 were *trans*-spliced to *gpd-3* in the ΔpA strains (Fig. 3B), whereas only SL2 was utilized when the CPSF binding sites were present. This data indicates that the AAUAAA is not required for *trans*-splicing, but does influence its specificity.

Another way to perturb cleavage and polyadenylation efficiency is to decrease the distance between the 5′ cap and the AAUAAA (Iwasaki & Temin, 1990; Sanfaçon & Hohn, 1990; S. Flaherty & G. Gilmartin, pers. comm.). We constructed a series of *gpd-2* deletions by placing the heat shock promoter at various positions within *gpd-2* (Fig. 4A). Upon heat shock, these constructs produced pre-mRNAs with 5′ caps at various distances from the AAUAAA, which should result in a length-dependent reduction in 3′ end formation efficiency. The results in Figure 4B confirm this prediction. The amount of *gpd-2* 3′ end product relative to *trans*-spliced *gpd-3* decreased as the cap was moved closer to the AAUAAA (Fig. 4B). In addition, as cleavage and polyadenylation decreased, *trans*-splicing efficiency was not reduced, but SL2 specificity was lost; SL2 *trans*-splicing to *gpd-3* was replaced by SL1 *trans*-splicing. Thus, SL1 *trans*-splicing appeared to become a more effective competitor when the distance between the cap and the AAUAAA was shortened.

Decreasing the distance between the 5′ and 3′ end of the upstream mRNA gave results very similar to those observed with the ΔpA mutant. In both cases, 3′ end formation was interfered with and SL2 *trans*-splicing was replaced with SL1 *trans*-splicing. However, with ΔpA, only 40% of *trans*-splicing was to SL1, whereas, when the cap was moved closer to the AAUAAA, SL2 *trans*-splicing was replaced completely by SL1. Perhaps this difference can be explained by hypothesizing that the distance between the cap and the *trans*-splice site may be an important determinant of SL1 *trans*-splicing efficiency. We suggest that polycistronic RNA may be a relatively poor substrate for SL1 *trans*-splicing due to the fact that the cap is nearly 1.5 kb upstream from the *trans*-splice site. In contrast, in HS229, where the cap is 229 nt from the *trans*-splice site, SL1 *trans*-splicing can compete more effectively for use of the normally SL2-specific *trans*-splice site. An intriguing possibility is that the SL1 snRNP may interact with the cap binding complex (CBC) (Lewis et al., 1996), thus explaining why close proximity of the cap and the *trans*-splice site increases the competitiveness of SL1 *trans*-splicing.

These experiments clearly show a relationship between SL2-specific *trans*-splicing and 3′ end formation. When cleavage and polyadenylation were either eliminated or reduced, there was a concomitant loss of SL2 *trans*-splicing to a downstream mRNA. However, in the inverse experiment, we did not find an effect of loss of *trans*-splicing on 3′ end formation. When the *trans*-splice site was mutated (Δts), cleavage and polyadenylation were not inhibited and polycistronic RNA did not accumulate (Fig. 2, lane 6). Therefore, in *C. elegans* polycistronic pre-mRNA processing, although 3′ end formation is needed for SL2-specific *trans*-splicing downstream, the converse is not true; *trans*-splicing appears to play no role in 3′ end formation at the site 100 bp upstream.

3′ End formation and *trans*-splicing in *C. elegans* and trypanosomes

Although both trypanosomes (protists), and *C. elegans* (an animal) process polycistronic pre-mRNAs by *trans*-splicing and 3′ end formation, our results demonstrate that there are distinct differences in the mechanism by which these processing events occur in these two distantly related organisms. In trypanosomes, apparently all mRNAs are processed from polycistronic precursors, whereas in *C. elegans*, only about 25% of genes are co-transcribed with other genes. Trypanosomes contain only a single SL snRNP, which is used for all splicing events, including processing at the 5′ end of the transcript and at *trans*-splice sites between genes (trypanosome pre-mRNAs do not contain introns). In contrast, in *C. elegans*, there are two classes of SL snRNPs: one that donates SL1 at outrons (Conrad et al., 1991, 1993, 1995), and one that donates SL2 at positions between genes in polycistronic transcripts

(Spieth et al., 1993; Zorio et al., 1994). Most importantly, however, the mechanism of 3′ end formation in *C. elegans* appears to be like that of other metazoans, involving CPSF binding to the consensus AAUAAA and CstF binding to a U-rich sequence downstream of the cleavage site. Our evidence suggests that the role of the AAUAAA in 3′ end formation between genes in operons is not different from those at the ends of monocistronic mRNAs. In sharp contrast, neither AAUAAA nor any other consensus is found near the 3′ ends of trypanosome mRNAs. Thus, poly(A) site selection may occur in a novel fashion without using CPSF and CstF. In support of this idea, LeBowitz et al. (1993) reported that selection of the site of cleavage and polyadenylation in *Leishmania* was controlled by the location of the downstream *trans*-splice site. When the position of the splice site was altered, the poly(A) site followed it, remaining at a position 200–500 nt upstream of the new site. In *Trypanosoma brucei*, Matthews et al. (1994) demonstrated that the site of both 3′ end formation and *trans*-splicing was controlled by a pyrimidine-rich sequence located near the *trans*-splice site. Elimination of this polypyrimidine tract resulted in the use of cryptic sites for both poly(A) formation and *trans*-splicing. In both of these studies, the site at which 3′ end formation occurred was controlled by sequences located downstream, either by the *trans*-splice site or by the polypyrimidine tract that functions in *trans*-splicing, rather than by an AAUAAA, as in animals. In contrast, our data indicate that, in *C. elegans*, the AAUAAA is required for 3′ end formation and is also an important determinant for *trans*-splicing specificity in the processing of polycistronic RNAs.

This distinct difference in how 3′ ends are formed in trypanosomes and *C. elegans* suggests that operons may have arisen independently in the two phyla. It is possible that, in trypanosomes, operons are a primitive characteristic and conventional signals for 3′ end formation were not needed because the *trans*-splicing machinery was able to supply the information for poly(A) site selection. In *C. elegans*, operons may have evolved by combining previously independent genes that already used CPSF and CstF for 3′ end formation. Once operons formed in *C. elegans*, *trans*-splicing became adapted to function, in combination with the already existing 3′ end formation machinery, to allow efficient processing of downstream mRNAs. The result is that, in trypanosomes, processing of polycistronic pre-mRNAs depends on signals that are located downstream of the site of 3′ end formation because no other signals existed for poly(A) site selection, whereas in *C. elegans*, those signals were already in place and only needed to acquire the ability for the conventional 3′ end formation machinery to interact with the *trans*-splicing machinery. In this model, SL2 evolved to facilitate this interaction.

Significance of the link between 3′ end formation and *trans*-splicing

The fact that SL2 *trans*-splicing still occurs in the ΔpA strains indicates that the AAUAAA is probably not involved *directly* in specifying SL2 *trans*-splicing to downstream mRNAs. Other factors or signals may be present that promote the use of SL2 over SL1. It does, however, indicate that the AAUAAA is needed to insure that all *trans*-splicing to the downstream site is to SL2, and this may have important functional significance. The connection between 3′ end formation and SL2 *trans*-splicing may have evolved to permit efficient expression of the downstream mRNA in polycistronic transcription units. It has been demonstrated in vertebrates that, following cleavage and polyadenylation, transcription terminates at a relatively undefined location downstream of the AAUAAA. After cleavage, the downstream RNA is uncapped; it has a free 5′ phosphate and is hence a substrate for 5′-3′ exonucleases. Termination is thought to occur when the exonuclease "catches up" to the transcribing polymerase and somehow causes it to fall off the DNA template (Connelly & Manley, 1988; Proudfoot, 1989). In the context of a *C. elegans* polycistronic pre-mRNA, this could potentially cause significant problems for expressing downstream mRNAs, unless *trans*-splicing provides a cap for the otherwise unprotected downstream mRNA, thus preventing degradation and transcription termination. The Δts mutation provides support for this idea. When *trans*-splicing was prevented (Δts), 3′ end formation occurred, but downstream *gpd-3* product failed to accumulate (Fig. 2, lane 6). It seems likely that, when *trans*-splicing was prevented, the *gpd-3* pre-mRNA that formed following cleavage and polyadenylation was degraded, presumably because it was uncapped. This may also explain why the distance between genes in operons is typically only 100 bp. If the distances were greater, there would be more opportunity to terminate transcription before *trans*-splicing could occur, thereby preventing efficient expression of downstream genes. Of course, it is also possible that variations in the distance between genes in operons could serve an important role in determining the levels of mRNAs of downstream genes in operons by balancing degradation and termination with stabilization by SL2 *trans*-splicing.

MATERIALS AND METHODS

Worm strains and RNA preparation

The maintenance and growth of worms was as described by Brenner (1974) and Sulston and Hodgkin (1988). Transgenic worm strains carrying extrachromosomal arrays were generated by the method of Mello et al. (1991) and Spieth et al. (1993). Worms were heat shocked in a 29 °C water bath for 2 h on floating petri plates spread with bacteria. Total RNA

was prepared as described in Conrad et al. (1991) from heat shocked or untreated populations of mixed-stage cultures containing transgenic worms and those lacking the array.

Plasmid construction

Transgenic strains of HS1496, HS229, and HS78 were constructed as described in Spieth et al. (1993). HS872 was constructed by inserting a *Bam*H I site at the beginning of the last exon of *gpd-2* in the HS1496 plasmid by using an oligonucleotide-directed in vitro mutagenesis kit (Amersham). The plasmid was then digested with *Bam*H I immediately downstream of the *hsp-16* promoter and at the newly created *Bam*H I site in *gpd-2* and then religated. The ΔpA construct was made by recombinant PCR as described by Higuchi (1990) using the oligonucleotide *Cla* I/*Nde* I-polyA-ko: 5′-GATGAATAAATATACAACCATCGATGAGAAGAG ACCACATATGTACTTGTAAGCTTCTAGG-3′ and anti-*Cla* I/*Nde* I-polyA-ko:5′-CCTAGAAGCTTACAAGTACATATGT GGTCTCTTCTCATCGATGGTTGTATATTTATTC-3′, which insert *Cla* I and *Nde* I sites in place of the TATAAA and AATAAA of *gpd-2*, respectively. The Δts construct was generated using an oligonucleotide-directed in vitro mutagenesis kit (Amersham).

RNase protection

Probes were made by PCR amplifying the region shown in Figure 1A using various mutant plasmids as templates. PCR products were then cloned using the TA cloning kit from Invitrogen. Plasmids were then linearized by digesting with *Xba* I and transcribed with SP6 RNA polymerase in the presence of 1 mM each of ATP, CTP, and GTP, 0.1 mM UTP, and 4 μL ^{32}PαUTP (3,000 Ci/mmol). The radiolabeled RNA probes were then gel purified on 4% polyacrylamide gels containing 7 M urea. After gel purification, ~50,000–100,000 cpm of probe was used for each reaction. Probe was mixed with 10–20 μg of total RNA and dried. Pellets were resuspended in 30 μL of hybridization buffer (80% deionized formamide, 40 mM pipes, pH 6.4, 400 mM NaOAc, 1 mM EDTA), heated to 85 °C for 5 min and then placed in a 45 °C bath for 4–6 h. After hybridization, 300 μL of digestion mix (10 mM Tris, pH 7.4, 5 mM EDTA, 200 mM NaOAc, and 0.1 units of Promega RNase One) was added and incubated at 37 °C for 1 h. Reactions were stopped by adding 0.15% SDS and 50 μg *Escherichia coli* tRNA and then precipitated with ethanol. Pellets were resuspended in denaturing dye and electrophoresed on 8% polyacrylamide gels containing 7 M urea.

Primer extension

Primer extensions with RNA purified from transformant strains were performed essentially as described in Conrad et al. (1991), except dGTP was omitted and 0.25 mM dideoxyguanosine triphosphate (Pharmacia) was included. The products were electrophoresed on 20% polyacrylamide gels containing 7 M urea. The primer was gpd3-5′ end: 5′-GCGTTGAAGCAGTTTCCC-3′ and was gel purified prior to phosphorylating with T4 polynucleotide kinase. Quantitation was performed on a Molecular Dynamics PhosphorImager using ImagQuant software.

ACKNOWLEDGMENTS

This work was supported by research grant GM42432 from the National Institute of General Medical Sciences. We are grateful to our colleagues, Carol Williams and Atul Despande, for helpful suggestions on the manuscript.

Received November 19, 1996; returned for revision December 16, 1996; revised manuscript received December 20, 1996

REFERENCES

Blumenthal T. 1995. Trans-splicing and polycistronic transcription in *C. elegans*. *Trends in Genetics 11*:132–136.

Blumenthal T, Steward K. 1997. RNA processing and gene structure. In: Riddle D, Blumenthal T, Meyer B, Priess J, eds. *C. elegans II*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp. 117–145.

Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics 77*:71–94.

Bruzik J, Van Doren K, Hirsh D, Steitz J. 1988. *Trans*-splicing involves a novel form of small nuclear ribonucleoprotein particles. *Nature 335*:559–562.

Chiou H, Dabrowski C, Alwine J. 1991. Simian virus 40 late mRNA leader sequences involved in augmenting mRNA accumulation via multiple mechanisms, including increased polyadenylation efficiency. *J Virol 65*:6677–6685.

Connelly S, Manley J. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & Dev 2*:440–452.

Conrad R, Lea K, Blumenthal T. 1995. SL1 *trans*-splicing specified by AU-rich synthetic RNA inserted at the 5′ end of *Caenorhabditis elegans* pre-mRNA. *RNA 1*:164–170.

Conrad R, Liou RF, Blumenthal T. 1993. Functional analysis of a *C. elegans* trans-splice acceptor. *Nucleic Acids Res 21*:913–919.

Conrad R, Thomas J, Spieth J, Blumenthal T. 1991. Insertion of part of an intron into the 5′ untranslated region of a *Caenorhabditis elegans* gene converts it into a trans-spliced gene. *Mol Cell Biol 11*:1921–1926.

Cooke C, Alwine JC. 1996. The cap and 3′ splice site similarly affect polyadenylation efficiency. *Mol Cell Biol 16*:2579–2584.

Higuchi R. 1990. Recombinant PCR. In: Innis MA, Gelfand DH, Sninsky JJ, White T, eds. *PCR protocols: A guide for methods and applications*. San Diego: Academic Press, Inc. pp 177–183.

Huang XY, Hirsh D. 1989. A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. *Proc Natl Acad Sci USA 86*:8640–8644.

Iwasaki K, Temin H. 1990. The efficiency of RNA 3′-end formation is determined by the distance between the cap site and the poly(A) site in spleen necrosis virus. *Genes & Dev 4*:2299–2307.

Keller W. 1995. No end yet to messenger RNA processing. *Cell 8*:829–832.

Krause M, Hirsh D. 1987. A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell 49*:753–761.

LeBowitz J, Smith H, Rusche L, Beverly S. 1993. Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes & Dev 7*:996–1007.

Lewis JD, Izaurralde E, Jarmolowski A, McGuigan C, Mattaj IW. 1996. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5′ splice site. *Genes & Dev 10*:1683–1698.

Manley J. 1995. A complex protein assembly catalyzes polyadenylation of mRNA precursors. *Curr Opin Genet Dev 5*:222–228.

Matthews K, Tschudi C, Ullu E. 1994. A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes & Dev 8*:491–501.

Mello C, Kramer J, Stinchcomb D, Ambros V. 1991. Efficient gene transfer in *C. elegans*: Extrachromosomal maintenance and integration of transforming sequences. *EMBO J 10*:3959–3970.

Moore M, Query CC, Sharp P. 1993. Splicing of precursors to mRNAs by the spliceosome. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 303–358.

Nesic D, Cheng J, Maquat L. 1993. Sequences within the last intron function in RNA 3′ end formation in cultured cells. *Mol Cell Biol 13*:3359–3369.

Nesic D, Maquat L. 1994. Upstream introns influence the efficiency of final intron removal and RNA 3′ end formation. *Genes & Dev 8*:363–375.

Nilsen TW. 1993. Trans-splicing of nematode premessenger RNA. *Ann Rev Microbiol 47*:413–440.

Niwa M, Berget S. 1991. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes & Dev 5*:2086–2095.

Niwa M, Rose S, Berget S. 1990. In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes & Dev 4*:1552–1559.

O'Hare K. 1995. mRNA 3′ ends in focus. *Trends Genet 11*:255–257.

Proudfoot N. 1989. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci 14*:105–110.

Sanfaçon H, Hohn T. 1990. Proximity to the promoter inhibits recognition of Cauliflower Mosaic Virus polyadenylation signal. *Nature 346*:81–84.

Spieth J, Brooke G, Kuersten S, Lea K, Blumenthal T. 1993. Operons in *C. elegans*: Polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell 73*:5121–5132.

Stringham E, Dixon D, Jones D, Candido E. 1992. Temporal and spatial expression patterns of the small heat shock (*hsp16*) genes in transgenic *Caenorhabditis elegans*. *Mol Biol Cell 3*:221–233.

Sulston J, Hodgkin J. 1988. Methods. In: Wood WB, ed. *The nematode Caenorhabditis elegans*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 587–606.

Wilson R et al. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature 368*:32–38.

Zorio DAR, Cheng NN, Blumenthal T, Spieth J. 1994. Operons as a common form of chromosomal organization in *C. elegans*. *Nature 372*:270–272.