



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

A *Trans*-Spliced Leader Sequence on Actin mRNA in *C. elegans*

Michael Krause and David Hirsh*

Fred Hutchinson Cancer Research Center

Seattle, Washington 98104

* Synergen, Inc.

Boulder, Colorado 80301

Summary

While determining the 5' ends of *C. elegans* actin mRNAs, we have discovered a 22 nucleotide spliced leader sequence. The leader sequence is found on mRNA from three of the four nematode actin genes. The leader also appears to be present on some, but not all, nonactin mRNAs. The actin mRNA leader sequence is identical to the first 22 nucleotides of a novel 100 nucleotide RNA transcribed adjacent, and in the opposite orientation, to the 5S ribosomal gene. The evidence suggests that the actin mRNA leader sequence is acquired from this novel nucleotide transcript by an intermolecular *trans*-splicing mechanism.

Introduction

The joining of two distinct RNAs to form a complete mRNA has been demonstrated both in vivo and in vitro. It has been known for some time that influenza virus recruits 5'-capped oligonucleotides from host mRNAs as primers for viral transcription (Plotch et al., 1981; Lamb and Chopin, 1983). More recently, the mRNAs of coronaviruses have been shown to have a common 72 nucleotide leader sequence acquired apparently by leader-primed transcription (Makino et al., 1986).

Distinct from these examples of primed transcription are intermolecular ligations in which the 5' intron donor site in one RNA interacts with the 3' intron acceptor site in a separate RNA. The term *trans* splicing has been used to describe those intermolecular events. *Trans*-splicing reactions can readily occur in vitro. The efficiency of in vitro *trans* splicing increases if the two interacting RNAs share complementary regions within their respective intervening sequences (Solnick, 1985; Konarska et al., 1985). There is now substantial evidence that *trans* splicing also occurs in vivo. Trypanosome mRNAs have a 35 nucleotide 5' leader sequence encoded by a tandemly repeated mini-exon (Van der Ploeg et al., 1982; De Lange et al., 1983, 1984; Nelson et al., 1983; Guyaux et al., 1985). A 137 nucleotide mini-exon RNA is transcribed that contains, at its 5' end, the 35 nucleotide common leader sequence flanked by a 5' intron donor site sequence (Campbell et al., 1984; Kooter et al., 1984; Milhausen et al., 1984). Trypanosome message maturation involves the acquisition of the leader by a *trans*-splicing reaction analogous to typical eukaryotic intron excision in which donor and acceptor consensus sequences are utilized (Murphy et al., 1986; Sutton and Boothroyd, 1986). In addition, it has recently been observed that the chloroplast ribosomal

protein S12 is synthesized from message that apparently forms by *trans* splicing (Koller et al., 1987).

There was some reason to believe that *trans* splicing occurs in the formation of some actin mRNAs in the nematode *C. elegans*. The nematode has four actin genes; genes 1, 2, and 3 are tightly clustered on chromosome V, and gene 4 is on the X chromosome (Files et al., 1983; Landel et al., 1984; Albertson, 1985; for review, see Krause and Hirsh, 1986) (Figure 1). The DNA sequences of the four nematode actin genes are highly conserved within the coding regions. Actin genes 1 and 3 are identical, and the worst homology by pairwise nucleotide sequence comparisons of genes is 92%. The transcriptional start sites for these genes have been studied; the start site for the X-linked actin gene transcript was clear (at approximately position -43 relative to the first AUG), but the data for the clustered genes led to ambiguous start sites (Wild, Krause, Rosenzweig, and Hirsh, submitted). Part of the reason for the ambiguity was a 19 nucleotide discrepancy in the length of the 5' untranslated regions of genes 1 and 3, depending on the method of determination. The start site for gene 1 and/or gene 3 (identical genes, referred to henceforth as gene 1/3) by S1 mapping was at nucleotide -24 relative to the initiation of translation, whereas the start site by primer extension was at -43 (Wild, Krause, Rosenzweig, and Hirsh, submitted). Furthermore, the S1 start site was coincident with a consensus 3' intron splice acceptor sequence, yet no 5' donor sequence, adjacent to typical promoter elements (CAAT and TATA box sequences), was found upstream of these two genes.

To resolve this ambiguity, we used primer extension sequencing to define the 5' ends of the four actin mRNAs of *C. elegans*. Sequence from the X-linked actin gene 4 message is consistent with the earlier results of a transcriptional start site at -42 to -45. Results using gene 1/3- and gene 2-specific oligonucleotides yield a unique 22 nucleotide sequence at the 5' terminus of all three of these messages. This 22 nucleotide sequence is not present within the 15 kb region of the genome containing the actin gene cluster. An oligonucleotide homologous to the 22 nucleotide leader sequence was used as a hybridization probe and has uncovered a novel 100 nucleotide, abundant "leader" RNA. The source of this leader RNA is a repeated genomic fragment containing the 5S ribosomal RNA gene. The 5S rRNA and leader RNA genes are 176 bp apart and are transcribed in opposite directions. We propose that the actin mRNA leader sequence is acquired from the novel leader transcript by a *trans*-splicing reaction.

Results

Sequencing the 5' Ends of Actin mRNA

Two oligonucleotide primers were synthesized for actin genes 1/3 and 2, and three primers were synthesized for actin gene 4 (see Figure 3). One primer was specific for each gene, with complementarity to regions entirely within

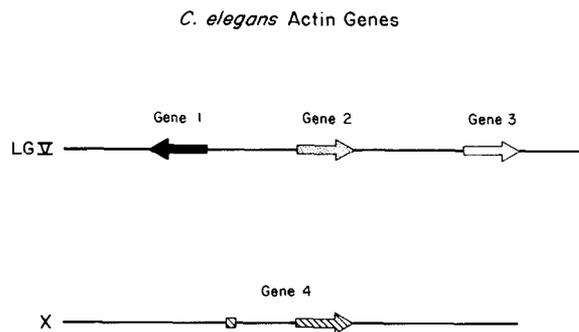


Figure 1. Genomic Organization of the Actin Genes in *C. elegans*. Genes 1, 2, and 3 are on chromosome V and gene 4 is on the X chromosome. The arrows indicate the directions of transcription for each gene.

the 5' untranslated regions of the genes. The other synthetic primer was partially specific for each gene; 11–14 nucleotides of the primer were complementary to the actin coding region (identical sequences for all four genes with the exception of 1 nucleotide in gene 2), and the remaining 6–12 nucleotides of the primer were complementary to the 5' untranslated region of each of the actin genes. In addition to the homologies among the coding regions of this second set of primers, there are short stretches of homology among the 5' untranslated regions of the genes. These subregions of homology among the second set of primers result in a lack of specificity due to their ability to cross-hybridize to other actin transcripts by looping out short regions in the mRNA.

The sequencing reactions for all six primers are shown in Figure 2, and the sequence (read as the antisense strand) is summarized in Figure 3. One primer, Act 2.1, consistently yielded regions of unreadable sequence, and two of the primers, Act 2.2 and Act 4.2, generated double sequences due to cross-annealing to the gene 1/3 mRNA. We can read the superimposed sequences by subtracting the gene 1/3 sequence, which is shown in Figures 2 and 3.

The sequences resulting from gene 1/3 and gene 2 primers were surprising; the last 22 nucleotides, representing the 5' terminus of each message, were identical (Figures 2 and 3). This common 22 nucleotide sequence is not present in the genomic DNA sequences contiguous with the gene-specific primers. However, each of these genes does have a 3' intron acceptor sequence (T_nCAG) within its 5' untranslated regions. These results, taken together, suggested the possibility that part of the 5' untranslated region was actually an intron, a notion previ-

ously pointed out by Wild, Krause, Rosenzweig, and Hirsh (submitted). Actin genes 1 and 2 are divergently transcribed with their 5' ends proximal (Figure 1). The entire region (approximately 3.5 kb) between genes 1 and 2 has been sequenced (Wild, Krause, Rosenzweig, and Hirsh, submitted), but the 22 nucleotide sequence we find on the message from genes 1/3 and 2 is not within this intergenic region.

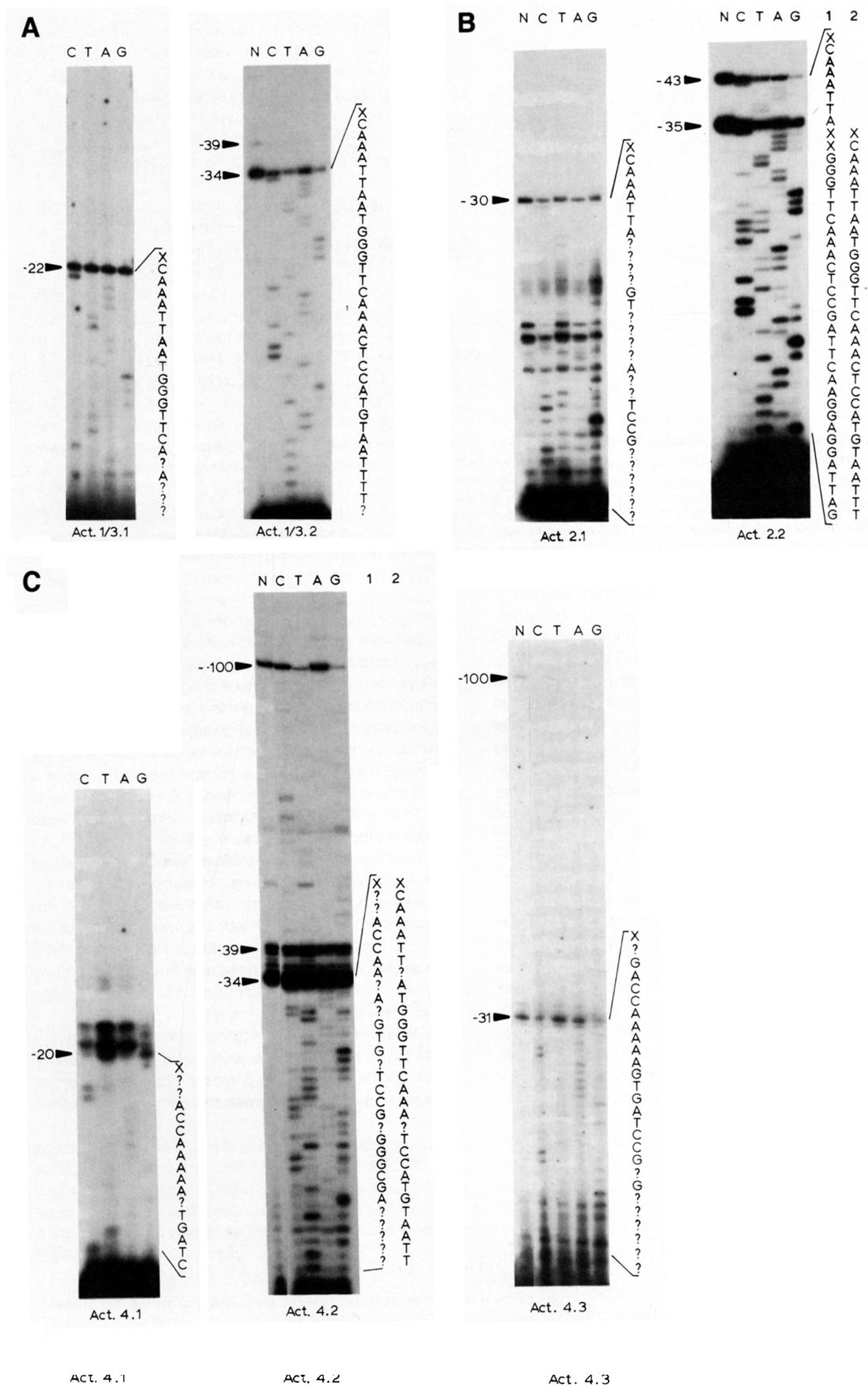
The gene-specific primer, Act 4.1, yields contiguous genomic sequence back to position –42 to –45. These positions are consistent with the transcriptional start sites for gene 4 as determined by S1 mapping (Wild, Krause, Rosenzweig, and Hirsh, submitted). However, within the 5' untranslated region of gene 4, in a region covered by the Act 4.1 primer, is a potential 3' intron acceptor sequence (TTTCAG; Figure 3). If this site is functional, the Act 4.1 primer could conceivably generate an intron sequence terminating at position –43 by using a gene 4 mRNA-processing intermediate as template. The other two primers for actin gene 4, Act 4.2 and 4.3, are downstream of the potential intron acceptor site and can be used to address the question of the possible presence of an intron. However, Act 4.2 cross-anneals to actin gene 1/3 mRNA, resulting in superimposed sequences. By subtracting the gene 1/3 mRNA sequence, we are left with a vague sequence apparently complementary to the gene 4 genomic sequence back to positions –42 and –45. In addition, there is a novel sequence from position –46 to about –100 that does not correlate with genomic sequences upstream of either gene 4 or gene 1/3. An actin mRNA transcriptional start site of –100 is not seen when an actin coding region primer is extended (Wild, Krause, Rosenzweig, and Hirsh, submitted). Our interpretation of the Act 4.2 reactions is that the novel sequence results from cross-annealing of the primer to another abundant RNA, resulting in a triple sequence from positions –1 to –45 and a single sequence from –46 to –80. The triple sequence therefore results from actin gene 1/3 RNA, actin gene 4 RNA, and an unidentified RNA. The result using the Act 4.3 primer is, again, contiguous genomic sequence to positions –42 to –45. These results demonstrate that the potential 3' intron acceptor site does not mark an intron and that the majority of gene 4 mRNA does not acquire the 22 nucleotide leader sequence.

Detection of a Novel Transcript

An oligonucleotide complementary to 20 of 22 nucleotides within the leader sequence (see Figure 5) was used as a probe on Southern blots of *C. elegans* DNA, digested with a variety of restriction enzymes, to determine the genomic

Figure 2. Actin 5' mRNA Sequencing Gels

Oligonucleotide primers (also see Figure 3) were used to sequence the complementary mRNAs. The primers used in the reactions are indicated at the bottom of each set of sequencing lanes and the sequence is shown at the right. Lanes labeled N were extended with deoxyribonucleotides only to determine termination sites. The lengths of extended fragments are given relative to the 5' end of the primer. (A) Actin gene 1/3 sequencing reactions. (B) Actin gene 2 sequencing reactions. Question marks indicate uncertainties in the sequence. Act 2.2 results in a double sequence due to annealing to both gene 2 RNA and gene 1/3 RNA (presumably by looping out 3 nucleotides in the gene 1/3 mRNA). Lane 1 is the gene 2-derived sequence and lane 2 is the gene 1/3-derived sequence. (C) Actin gene 4 sequencing reactions. Question marks indicate uncertainties in the sequence. Act 4.2 results in multiple sequences. Lane 1 is the gene 4-derived sequence and lane 2 is the gene 1/3-derived sequence. The unlabeled sequence present above position –45 is from an unidentified source.



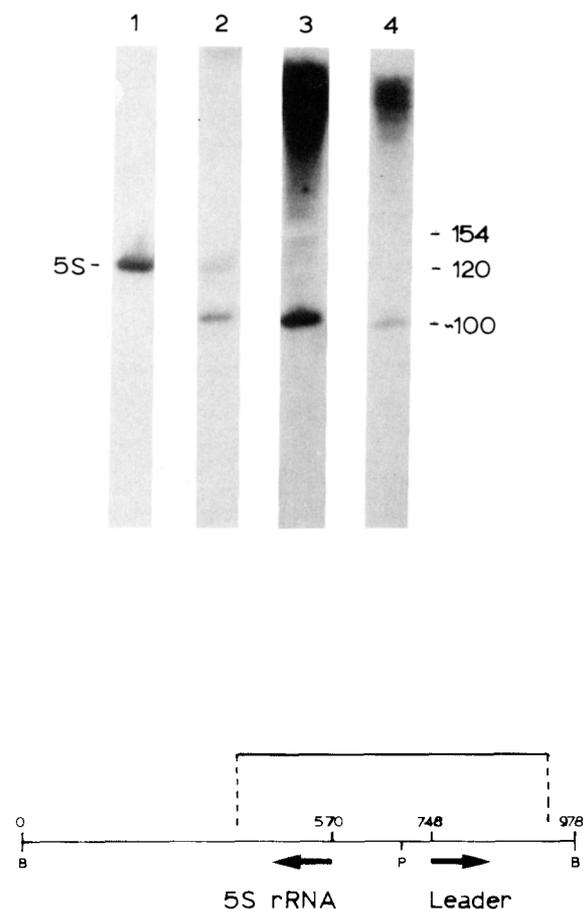


Figure 4. Detection of a Novel Transcript

Top: Ten microgram fractions of total RNA were separated on 5% polyacrylamide-urea gels, electroblotted to nitrocellulose membrane, and hybridized to the following labeled probes: lane 1, a 695 bp, gel-purified, BamHI-PvuII restriction fragment from the 1 kb ribosomal repeat containing the 5S rRNA gene; lane 2, a 295 bp, gel-purified, BamHI-PvuII restriction fragment representing the remainder of the 5S rRNA repeat and not containing the 5S rRNA gene; or lane 3, leader oligonucleotide, a 20-mer complementary to the actin leader sequence (see Figure 5). Lane 4 is presented as a shorter exposure of lane 3 to show multiple large RNAs hybridizing with the leader oligonucleotide probe. Sizes are in nucleotides. Bottom: Schematic of the leader RNA and 5S rRNA within the 1 kb repeat. BamHI (B) and PvuII (P) restriction sites are indicated. The sequence of the bracketed region is depicted in Figure 5.

rRNA repeat on Southern blots using the leader sequence oligonucleotide as a probe. Only the subcloned insert of the 5S repeat contained sequences complementary to the leader oligonucleotide (data not shown).

Is This Leader Present on Other mRNAs?

The actin mRNA sequencing results suggest that actin gene 4 message does not have the leader sequence, showing that this leader is not on all *C. elegans* mRNAs. To determine whether any other messages had the 22 nucleotide leader sequence, the leader oligonucleotide was used to probe Northern blots of poly(A)⁺, poly(A)⁻, and total RNA (Figure 6). The leader probe hybridizes to the leader RNA and abundant messages 1400 to 1600

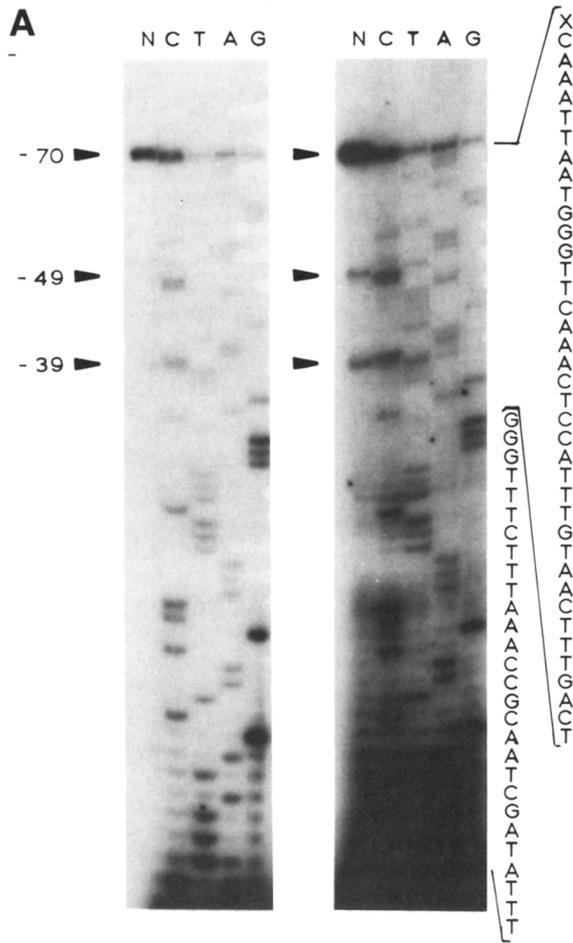
nucleotides long, which is the size of actin mRNAs observed on Northern blots (Krause and Hirsh, 1986). The leader RNA is present in the poly(A)⁻ fraction of RNA and is undetected in the poly(A)⁺ fraction, indicating that the leader RNA is not polyadenylated. There are an additional five to seven bands that potentially represent nonactin mRNAs containing the leader sequence. Actin mRNAs larger than 1600 nucleotides have never been observed on Northern blots, making it more likely that those bands are other mRNAs with spliced leader. The background smear of hybridization within the lane could reflect lower abundance mRNAs with the leader sequence, degradation of the abundant messages, or nonspecific hybridization.

The actin mRNAs on which the leader sequence is present encode proteins that are involved in the thin filaments of the nematode body-wall muscle (Waterston et al., 1984; Landel et al., 1984; Krause and Hirsh, 1986). Perhaps the leader represents a muscle-specific signal sequence. The gene *unc-54* encodes a body wall muscle-specific myosin heavy chain (Epstein et al., 1974; Karn et al., 1983, 1985). We have sequenced the 5' terminus of the *unc-54* mRNA by extension of a 5' untranslated region oligonucleotide, and no leader sequence is detected (data not shown). The *unc-54* mRNA sequence is homologous to the genomic sequence and indicates a predominant 5' termination site for primer extension at approximately position -80 relative to the initiator AUG codon.

Discussion

We have found a 22 nucleotide 5' splice leader sequence on three of the four actin mRNAs in *C. elegans*. The source of this leader appears to be an abundant 100 nucleotide, poly(A)⁻ RNA. The leader RNA is transcribed from a tandemly repeated genomic fragment that also encodes 5S rRNA. It is not known whether each element of the repeat is actively transcribed. The 5S rRNA gene begins 176 bp downstream of, and is transcribed in the opposite orientation to, the 100 nucleotide leader RNA. The leader RNA is not flanked by any obvious homologies to known promoter or terminator sequences, and in vitro transcription assays suggest that this RNA is not a polymerase III transcript (Nelson and Honda, personal communication). It is interesting to note that the sequence corresponding to what we believe is the 3' end of the leader RNA contains dyad symmetry (Figure 5). Dyad symmetry within the 3' ends of eukaryotic histone mRNAs is thought to be involved in processing and/or stability of these non-polyadenylated polymerase II transcripts (Birnstiel et al., 1985).

We have examined the prevalence of this leader on *C. elegans* mRNA by Northern blot analysis of poly(A)⁺, poly(A)⁻, and total RNA using an oligonucleotide complementary to the leader sequence. In addition to a band comigrating with the actin mRNAs, the poly(A)⁺ RNA lane has five to seven prominent, nonactin bands of hybridization along with a background smear. The smear of hybridization may result from additional, less abundant RNAs containing the leader, degradation of the abundant messages, or nonspecific hybridization.



The actin leader could be acquired by one of three mechanisms: primed transcription, *cis* splicing, or *trans* splicing. The 5' end of the 100 nucleotide leader RNA comprises the 22 nucleotide actin leader sequence adjacent to a consensus 5' intron donor site (Figure 5). Although no 3' intron acceptor site is present within the leader transcript, an acceptor site is present upstream of the three actin genes encoding messages that acquire the 5' leader. The presence of unpaired intron splice sites within the leader RNA and actin genes suggests leader acquisition by a splicing mechanism rather than priming by the leader sequence alone.

Cis splicing represents typical intron excision and would require contiguous transcription between the leader RNA gene and the actin genes. One possibility is that the leader RNA gene and actin genes are physically linked, providing a continuous template for transcription. In fact, the actin genes and the 5S ribosomal gene cluster have been genetically and cytologically mapped to chromosome V (Files et al., 1983; Albertson, 1985; Nelson and Honda, 1986). However, the two gene clusters map to different positions on the chromosome. It is known that a small number of 5S ribosomal repeats (and presumably leader RNA repeats) are not associated with the mapped cluster (Nelson and Honda, 1985). By Southern blot analysis we have found no copies of the leader RNA in the 15 kb region of the genome containing the actin gene cluster. Thus, given the enormous distances that would be involved, we do not think that conventional transcription followed by *cis* splicing could account for our results.

Alternatively, the leader RNA gene and actin genes could serve as discontinuous templates for transcription. The entire leader RNA (or a major portion of it) could serve as a primer for transcription of actin genes 1, 2, and 3. Extension of the leader RNA primer would generate a continuous transcript containing intron splice sites that could

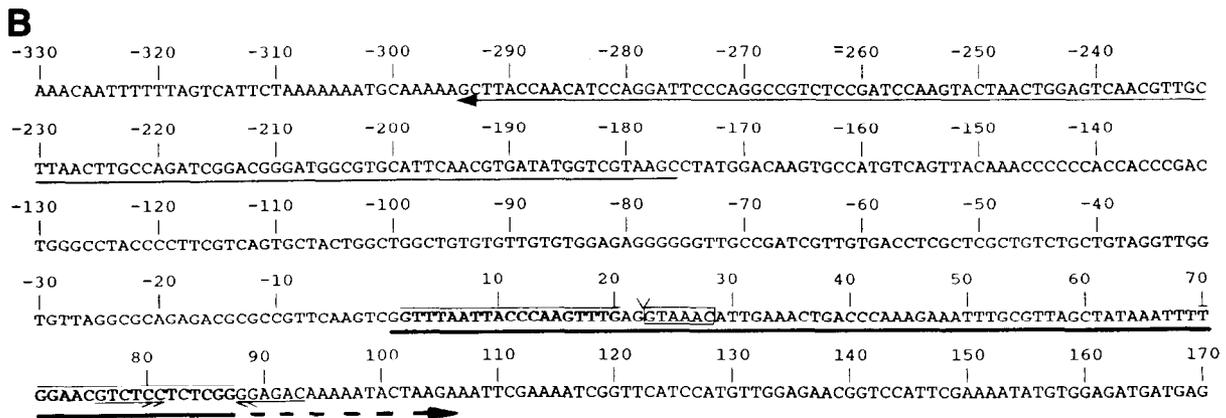


Figure 5. Sequence of the Leader RNA

(A) Sequencing reactions using an oligonucleotide primer complementary to nucleotides 70 to 88 of the leader RNA. The lane labeled N was extended with deoxynucleotides to determine termination sites. Extension fragment sizes are given relative to nucleotide position 70. (B) Genomic sequence of the region overlined at the bottom of Figure 4 containing the leader RNA (bold underline) and 5S rRNA (narrow underline) (from Nelson and Honda, 1985). Oligonucleotides used for sequencing (nucleotide positions 70 to 88) and hybridizations (nucleotide positions 2 to 21) are overlined. Dyad symmetry in the putative 3' end of the leader RNA is underlined with double-barbed arrows. A splice donor consensus sequence is boxed, with a carat designating the putative splice donor site.

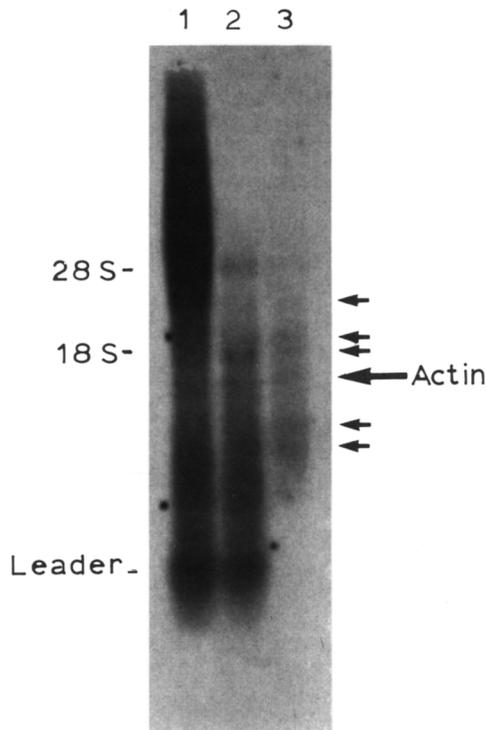


Figure 6. Northern Analysis of Leader-Containing RNAs

Total RNA (10 μ g), poly(A)⁻ (10 μ g), and poly(A)⁺ (0.5 μ g) RNA fractions were separated on 1.5% agarose-formaldehyde gels and blotted to nitrocellulose membrane. Filters were hybridized with labeled leader oligonucleotide, a 20-mer as described in Figure 5. The large arrow designates the size of the actin mRNAs.

function in conventional *cis*-splicing reactions. Such a mechanism would predict primary actin transcripts with 5' untranslated regions longer than 55 nucleotides (the sum of the lengths of the 5' leader sequence plus intron donor and acceptor sites and the shortest actin 5' untranslated region). We do not detect any extension products in this size range using primers specific to actin genes 1/3 or 2, which suggests that either the abundance of these putative primary transcripts in total RNA preparations is very low or they do not exist. Moreover, we have been unable to detect complementary regions between the leader RNA and genomic sequences upstream of actin genes 1, 2, or 3. One might expect such complementarity if the leader RNA serves as a primer for actin gene transcription. These considerations, coupled with the lack of precedence for such splicing mechanisms, make us view this possibility as unlikely; however, this point clearly needs further study.

Trans splicing, therefore, seems to be the most plausible explanation for the generation of the mature mRNA from actin genes 1, 2, and 3. *Trans* splicing implies that the primary transcripts from these three actin genes are initiated by independent promoters. Each of the three genes has a reasonable upstream promoter as defined by consensus CAAT and TATA box sequences (Wild, Krause, Rosenzweig, and Hirsh, submitted). The spacing of these promoter sequences, relative to the initiator methionine

codon, is similar in all four actin genes and is typical of eukaryotes. If these putative upstream promoters for actin genes 1, 2, and 3 are functional, then we would predict a transcriptional start site for the primary transcripts at about position -45 relative to the AUG. The primary transcripts might therefore have the same size 5' untranslated regions as do the actin mRNAs containing the leader sequence, making it difficult for us to detect precursors of the spliced leader mRNAs.

If the actin leader sequence is acquired by a *trans*-splicing reaction with the 100 nucleotide leader RNA, one might expect to find a population of truncated leader RNAs due to either the loss of the 5' terminal 22 nucleotides or the formation of a branched RNA by-product resulting from *trans* splicing. (Branched RNAs predicted by *trans* splicing have been detected in trypanosomes; Murphy et al., 1986; Sutton and Boothroyd, 1986.) An RNA of the appropriate length is detected in the primer extension reactions for sequencing the leader RNA from its 3' end (Figure 5). The lane extended with deoxynucleotides shows three major stop sites; one corresponds to the full-length RNA, terminating 70 nucleotides upstream of the primer. There are also two other stops at 39 and 49 nucleotides upstream of the primer. The stop at position 49 corresponds to the 5' donor site sequence within the leader RNA, suggesting the presence of the RNA species predicted by a *trans*-splicing model. We have no explanation for the stop at position 39.

All four actin genes have a putative 3' intron acceptor site within the 5' untranslated region, yet apparently only RNA from the three clustered actin genes acquires the leader sequence. A comparison of the 5' untranslated region sequences shows that genes 1, 2, and 3 each have a conserved element not found in actin gene 4. This element consists of the sequence CTAATC located 12 bp downstream from the 3' intron acceptor site (Figure 3); we have no evidence that this element has any role in splicing.

The splicing efficiency of the leader might be related to the relative abundance of actin gene primary transcripts and spliced products. We are unable to detect any sequence corresponding to the putative primary transcripts from actin genes 1 or 3 on long exposures of the sequencing gels, which suggests that for these two mRNAs the acquisition of the leader sequence is highly efficient and/or that the 5' ends of the primary transcripts are very short lived. It is possible that the minor stop at -39 observed with the Act 1/3.2 primer represents primary transcripts from genes 1 and 3. The inefficient sequencing reactions resulting from actin gene 2 primers make it difficult to assess leader splicing efficiency for gene 2 transcripts.

Trans splicing has been shown to be the mechanism of mRNA formation in trypanosomes (Murphy et al., 1986; Sutton and Boothroyd, 1986). Nearly all trypanosome mRNAs have a 35 nucleotide leader sequence. The 35 nucleotide leader sequence of trypanosomes is also present in an abundant 135 nucleotide transcript derived from a tandemly repeated genomic fragment (Nelson et al., 1983; De Lange et al., 1984; Kooter et al., 1984; Milhausen et al., 1984). The 5' terminal end of this 135 nucleotide

transcript consists of the 35 nucleotide leader adjacent to an unpaired 5' intron donor site. One difference between the trypanosome and nematode mRNAs is that the leader in *C. elegans* is not on all mRNAs. Another difference is that *C. elegans* also carries out *cis* splicing.

What might be the function of the actin leader sequence? Actin is an abundant cellular protein. One might speculate that the leader enhances message stability and/or affinity for polysomes and would therefore be present on other mRNAs encoding similarly abundant proteins. Several genes encoding abundant mRNAs and proteins have been characterized in *C. elegans*, including the *unc-54* body-wall myosin (Schachat et al., 1977; Karn et al., 1985), collagens (Kramer et al., 1985; Cox, Fields, Kramer, Rosenzweig, and Hirsh, submitted), vitellogenins (Spieth et al., 1985), and the major sperm proteins (Klass et al., 1984). None of the mRNAs transcribed from these genes appears to involve unusual 5' untranslated region intron splicing. In most cases, the sequences of the 5' ends of the mRNAs have not been compared to the corresponding genomic sequences. However, the 5' end sequences of the vitellogenin mRNAs (Spieth et al., 1985; T. Blumenthal, personal communication) and of the major transcript from the *unc-54* myosin gene (Krause, unpublished data) are colinear with the genomic DNA sequences.

Stabilization of polysome interactions with the actin mRNAs could be mediated by protein-leader interactions and/or secondary structure between the leader sequence and rRNA. We have no evidence for or against a protein-leader interaction. A search of the *C. elegans* rRNA sequences (Ellis et al., 1986) did not reveal any obvious sequence complementarity between the leader and the ribosomal RNAs. The absence of the leader sequence on the *unc-54* myosin mRNA also indicates that the leader does not mark all muscle-specific messages.

Regardless of the function or origin of the actin mRNA leader sequence, the apparent utilization of *trans* splicing in the nematode raises the possibility that the machinery for *trans* splicing is widespread in eukaryotes. *Trans*-spliced leader sequences could provide a means to affect the stability, location, transport, and translational efficiency of a heterogeneous population of mRNAs and might therefore be a regulatory mechanism in gene expression.

Experimental Procedures

RNA Isolation and Analysis

Total RNA and poly(A)⁺ RNA were isolated as previously described (Kramer et al., 1985; Cox et al., 1985). Large RNAs were separated on 1.5% agarose-formaldehyde (2.2 M) gels and blotted to GeneScreen (DuPont) or Biotodyne (Pall Laboratories) nitrocellulose membrane by capillary action with 20× SSC (3 M NaCl, 0.3 M NaCitrate). Small RNAs were separated on 5% polyacrylamide-urea (7.8 M) gels and transferred to GeneScreen by electroblotting for 2.5 hr at 30 volts in 0.25× TBE (1× TBE: 100 mM Tris, pH 6.5; 89 mM Boric acid; 89 mM EDTA). Oligonucleotide probes were kinased with T4 kinase (New England BioLabs) by standard procedures and hybridized to filters in 50 mM Tris, pH 6.5; 1 M NaCl; 1 mM EDTA; 10 mM PO₄, 6.5; 0.1% SDS; 10× Denhardt's; and 200 µg/ml sheared salmon sperm DNA. Hybridizations were done at 22°C or 37°C overnight. Hybridized filters were

washed in 6× SSC, 10 mM PO₄, 0.1% SDS at 37°C–55°C. Double-stranded DNA probes were hybridized as previously described (Landel et al., 1984).

Primer Extension Sequencing

Labeled oligonucleotides (18–20 nucleotides in length) were annealed to 5–50 µg of total RNA and were extended essentially as described by Zaug et al. (1984). Annealing reactions were carried out at 60°C–65°C for 5 min and extension reactions with AMV Reverse Transcriptase (P-L Chemicals) performed at 37°C–45°C for 20–40 min. Extension products were separated on 8% polyacrylamide-urea (7.8 M) gels and detected by autoradiography.

Acknowledgments

We are extremely grateful to Harold Weintraub who generously provided space and support for some of this work to be finished in his laboratory. We are also appreciative of his helpful criticism of the manuscript. We thank Don Nelson and Barry Honda for 5S rRNA information and clones, and Susan Bektesh for technical suggestions and criticism of the manuscript. This work was supported by National Institutes of Health grant GM37823 to D. H. and, in part, by NIH grant CA42506 to Harold Weintraub. M. K. was supported by NIH Postdoctoral Fellowship GM11303.

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

Received March 6, 1987; revised April 10, 1987.

References

- Albertson, D. G. (1984). Localization of ribosomal genes in *C. elegans* chromosomes by *in situ* hybridization using biotin-labeled probes. *EMBO J.* 3, 1227.
- Albertson, D. G. (1985). Mapping muscle protein genes by *in situ* hybridization using biotin-labeled probes. *EMBO J.* 4, 2493–2498.
- Birnstiel, M. L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3' end processing: the end is in site! *Cell* 4, 349–359.
- Campbell, D. A., Thornton, D. A., and Boothroyd, J. C. (1984). Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. *Nature* 311, 350–355.
- Cox, G. N., and Hirsh, D. (1985). Stage-specific patterns of collagen gene expression during development of *Caenorhabditis elegans*. *Mol. Cell. Biol.* 5, 363–372.
- De Lange, T., Liu, A. Y. C., Van der Ploeg, L. H. T., Borst, P., Tromp, M. C., and Van Boom, J. H. (1983). Tandem repetition of the 5' min-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription. *Cell* 34, 891–900.
- De Lange, T., Berkuens, T. M., Veerman, H. J. G., Frasc, A. C., Barry, J. D., and Borst, P. (1984). Comparison of the genes coding for the common 5' terminal sequence of messenger RNA in three trypanosome species. *Nucl. Acids Res.* 12, 4431–4443.
- Ellis, R. E., Sulston, J. E., and Coulson, A. R. (1986). The rDNA of *C. elegans*: sequence and structure. *Nucl. Acids Res.* 14, 2345–2364.
- Epstein, H. F., Waterston, R. H., and Brenner, S. (1974). A mutant affecting the heavy chain of myosin in *Caenorhabditis elegans*. *J. Mol. Biol.* 90, 291–300.
- Files, J. G., Carr, S., and Hirsh, D. (1983). Actin gene family of *Caenorhabditis elegans*. *J. Mol. Biol.* 164, 355–375.
- Guyaux, M., Cornelissen, A. W. C. A., Pays, E., Steinert, M., and Borst, P. (1985). *Trypanosoma brucei*: a surface antigen mRNA is discontinuously transcribed from two distinct chromosomes. *EMBO J.* 4, 995–998.
- Honda, B. M., Devlin, R. H., Nelson, D. W., and Khosla, M. (1986). Transcription of class III genes in cell-free extracts from the nematode *Caenorhabditis elegans*. *Nucl. Acids Res.* 14, 869–881.
- Karn, J., Brenner, S., and Barnett, L. (1983). Protein structural domains in the *Caenorhabditis elegans* myosin heavy chain gene are not separated by introns. *Proc. Natl. Acad. Sci. USA* 80, 4253–4257.

- Karn, J., Dibb, N. J., and Miller, D. M. (1985). Cloning nematode myosin genes. In *Cellular and Molecular Biology of the Cytoskeleton*, Volume 6, J. Shay, ed. (New York: Plenum Corporation).
- Klass, M., Kinsley, S., and Lopez, L. C. (1984). Isolation and characterization of a sperm-specific gene family in the nematode *C. elegans*. *Mol. Cell. Biol.* 4, 529–537.
- Koller, B., Fromm, H., Galun, E., and Edelman, M. (1987). Evidence for in vivo *trans* splicing of pre-mRNAs in tobacco chloroplasts. *Cell* 48, 111–119.
- Konarska, M. M., Padgett, R. A., and Sharp, P. A. (1985). *Trans* splicing of mRNA precursors in vitro. *Cell* 42, 165–171.
- Kooter, J. M., De Lange, T., and Borst, P. (1984). Discontinuous synthesis of mRNA in trypanosomes. *EMBO J.* 3, 2387–2392.
- Kramer, J. M., Cox, G. N., and Hirsh, D. (1985). Expression of the *Caenorhabditis elegans* collagen genes *col-1* and *col-2* is developmentally regulated. *J. Biol. Chem.* 260, 1945–1951.
- Krause, M., and Hirsh, D. (1986). The actin genes in *Caenorhabditis elegans*. In *Cellular and Molecular Biology of the Cytoskeleton*, Volume 7, J. Shay, ed. (New York: Plenum Corporation).
- Lamb, R. A., and Choppin, P. W. (1983). The gene structure and replication of influenza virus. *Ann. Rev. Biochem.* 52, 467–506.
- Landel, C. P., Krause, M. W., Waterston, R. H., and Hirsh, D. (1984). DNA rearrangements in the actin gene cluster in *C. elegans* accompany reversion of three muscle mutants. *J. Mol. Biol.* 180, 497–513.
- Makino, S., Stohman, S. A., and Lai, M. M. C. (1986). Leader sequences of murine coronavirus mRNAs can be freely reassorted: evidence for the role of free leader RNA in transcription. *Proc. Natl. Acad. Sci. USA* 83, 4204–4208.
- Milhausen, M., Nelson, R. G., Sather, S., Selkirk, M., and Agabian, N. (1984). Identification of a small RNA containing the trypanosome spliced leader: a donor of the shared 5' sequences of trypanosomatid mRNAs? *Cell* 38, 721–729.
- Murphy, W. J., Watkins, K. P., and Agabian, N. (1986). Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. *Cell* 47, 517–525.
- Nelson, D. W., and Honda, B. M. (1985). Genes coding for 5S ribosomal RNA of the nematode *Caenorhabditis elegans*. *Gene* 38, 245–251.
- Nelson, D. W., and Honda, B. M. (1986). Genetic mapping of the 5S rRNA gene cluster of the nematode *C. elegans*. *Can. J. Gen. Cytol.* 28, 545–553.
- Nelson, R. G., Parsons, M., Barr, P. J., Stuart, K., Selkirk, M., and Agabian, N. (1983). Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphans in *Trypanosoma brucei*. *Cell* 34, 901–909.
- Plotch, S. H., Bouloy, M., Ulmanen, I., and Krug, R. M. (1981). A unique cap(m⁷GpppXm)-dependent influenza virion endonuclease cleaves capped mRNAs to generate the primers that initiate viral RNA transcription. *Cell* 23, 847–858.
- Schachat, F. H., Harris, H. E., and Epstein, H. F. (1977). Two homogeneous myosins in the body-wall muscle of *Caenorhabditis elegans*. *Cell* 10, 721–728.
- Solnick, D. (1985). *Trans* splicing of mRNA precursors. *Cell* 42, 157–164.
- Spieth, J., Denison, K., Kirtland, S., Cane, J., and Blumenthal, T. (1985). The *C. elegans* vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes. *Nucl. Acids Res.* 13, 5283–5295.
- Sutton, R. E., and Boothroyd, J. C. (1986). Evidence for *trans* splicing in trypanosomes. *Cell* 47, 527–535.
- Van der Ploeg, L. H. T., Bernardis, A., Rijsewijk, F. A. M., and Borst, P. (1982). Characterization of the DNA duplication–transposition that controls the expression of two genes for the variant glycoprotein in *Trypanosoma brucei*. *Nucl. Acids Res.* 10, 593–609.
- Waterston, R. H., Hirsh, D., and Lane, T. R. (1984). Dominant mutations affecting muscle structure in *Caenorhabditis elegans* that map near the actin gene cluster. *J. Mol. Biol.* 180, 473–496.
- Zaug, A. J., Kent, J. R., and Cech, T. R. (1984). A labile bond at the ligation junction in a circular intervening sequence RNA. *Science* 224, 574–578.