Nucleic Acids Research, 1993, Vol. 21, No. 4 ***913–919***

Functional analysis of a *C.elegans trans*-splice acceptor

Richard Conrad, Ruey Fen Liou+ and Thomas Blumenthal*
Department of Biology, Indiana University, Bloomington, IN 47405, USA

Received November 11, 1992; Revised and Accepted January 16, 1993

ABSTRACT

The *rol-6* gene is trans-spliced to the 22 nt leader, SL1, 173 nt downstream of the transcription start. We have analyzed splicing in transformants carrying extrachromosomal arrays of *rol-6* with mutations in the trans-splice acceptor site. This site is a close match to the consensus, UUUCAG, that is highly conserved in both trans-splice and intron acceptor sites in *C.elegans*. When the trans-splice site was inactivated by mutating the perfectly-conserved AG, trans-splicing still occurred, but at a cryptic site 20 nt upstream. We tested the frequency with which splicing switched from the normal site to the cryptic site when the pyrimidines at this site were changed to A's. Since most *C.elegans* 3′ splice sites lack an obvious polypyrimidine tract, we hypothesized that these four pyrimidines might play this role, and indeed mutation of these bases caused splicing to switch to the cryptic site. We also demonstrated that a major reason the downstream site is normally favored is because it occurs at a boundary between A + U rich and non-A + U rich RNA. When the RNA between the two splice sites was made less A + U rich, splicing occurred preferentially at the upstream site.

INTRODUCTION

In eucaryotes, initial transcripts are generally interrupted by introns that are removed by RNA splicing. This process requires the cellular machinery to recognize the intron/exon borders and splice the two exons together, excising the intron. Splicing takes place on the spliceosome, composed of five small nuclear ribonucleoprotein particles (snRNPs), U1, U2, U4, U5, and U6, and associated factors (for reviews, see 1, 2). The reaction proceeds through a branched loop or 'lariat' intermediate, in which the 5′ phosphate of the intron has been transferred to the 2′ hydroxyl of an A residue 15–40 nt upstream of the 3′ end of the intron. In many eukaryotes the region between this branch point and the 3′ end of the intron contains a tract of pyrimidines. Splice site recognition is a complex process that is not yet fully understood. U1, U2, and U5 snRNPs, as well as associated factors, all play roles in recognition of the splice-donor site, the branch point, the polypyrimidine tract, and the splice-acceptor site. The first step in splicing is apparently recognition of the 5′ splice site by the U1 snRNA, whose 5′ end base pairs with the 5′ end of the intron (3–5). The branch point is subsequently recognized by direct base pairing with U2 RNA (6–8). The intron portion of the 3′ splice site also base pairs with the U1 snRNP (9), while the exon portions of both splice sites base pair with the U5 snRNP (10). In mammalian cells, a polypyrimidine tract binding protein, U2AF, is essential for splicing and probably aids the U2-branchpoint recognition event (11).

In *C.elegans* there are both similarities to, as well as some surprising differences from, this general scheme. At the 5′ splice site the consensus sequence is G/GUAAG (12), which is nearly the same as in vertebrates and several other phyla. However at the 3′ splice site there is an unusually highly conserved consensus, UUUCAG/G, and no polypyrimidine tract other than the four pyrimidines at the splice site (12, 13). Furthermore, *C.elegans* introns have no obvious branch point consensus sequence, although the U2 RNA sequence known to base pair with the branch point in other organisms has been perfectly conserved in *C.elegans* U2 (14). *C.elegans* introns tend to be quite short: around 75% are very close to 50 nt, too short to be spliced by mammalian snRNPs (12,15). However, the *C.elegans* splicing machinery is capable of correctly recognizing long introns as well, since several have been found to exceed 1 kb (12). Finally, *C.elegans* introns are rich in A and U (about 70%), a property they share with plant introns (16, 17). In the latter case this high A + U content has been found to be a crucial element in intron recognition (17).

A related type of splicing, trans-splicing, in which a short leader is donated by a specialized snRNP (SL RNP), occurs in *C.elegans* in addition to conventional (or cis-) splicing. Trans-splicing of this type is known to occur in nematodes (18–21), trematodes (22), and in trypanosomes (23, 24). In trypanosomes, it is the sole kind of splicing. In *C.elegans* only a subset of pre-mRNAs are trans-spliced and these same pre-mRNAs are also subjected to cis-splicing (reviewed in 25). Trans-splicing and cis-splicing are closely related processes. A branched intermediate, forked instead of looped, forms (26, 27), and the spliceosomal snRNP's are required (with the possible exception of U1 (28)). Because the SL RNA is packaged in a snRNP very similar to the U-snRNPs (27–29), it is assumed that it is an integral part of the 'trans-spliceosome', and may replace U1 since it contains the splice donor site (30, 31).

Trans-splice-acceptor sites in nematodes have the same consensus sequence, UUUCAG/G, as do intron acceptor sites (12). The sole difference between cis- and trans-splice acceptors

* To whom correspondence should be addressed
+Present address: PO Box 23-106, Taiwan 10098, China

Table I. Oligonucleotides used in directing mutageneses.

mutation	oligonucleotide (5′→3′)
r6AA	AAAATTTCCAAATGACCCTAA
r6Del	GAGAATATACGATTAGCCGGCCGCAATTAAAAATTTCTGGATA
r6Y/A[1]	GATCTAACTGAAAATWWWMAGATGACCCTAACTA
r6us=ds	TTCTGGATATATATTTCCAGATCTAACTGA
r6AUKO	ATATATCTTTAGGCCGAACTGCCGCTTTCCAGATGAC
r6BrKO-1	ATATCTTTAGATCTTTCTGAAAATTTCC
r6BrKO-2	ATCTTTAGTTCTAACTGTTTTTTTCCAGATGAC
r6SCS	GGTGAAGGATATTTGGTTTAATTACCCAAGTTTGAGATCTAACTGAAAATT
r6SL2A	ATATATCTTTAGCAACTTCTAATTTTATTCAGATGAC

[1]'W' refers to a 30% dATP, 70% dTTP input into the synthesis at this position; 'M' refers to a 40% dATP, 60% dCTP input.

is the presence or absence of a functional splice-donor site upstream. A conventional pre-mRNA and one destined for trans-splicing differ only by the presence in the latter of RNA recognized as 'intron' at the 5′ end (32, 33). We have named this 5′ terminal intron an 'outron'. In the present paper we report experiments in which we investigated sequences important for trans-splice acceptor site recognition in *C.elegans*. The trans-splice acceptor of *rol-6*, a cuticular collagen gene (34), was analyzed by taking advantage of a cryptic splicing event that occurs when the normal trans-splice site is disturbed. These experiments demonstrated that the four pyrimidines preceding the AG are important for activity. In addition, we present an experiment that suggests that A + U richness is also an important variable in outron recognition.

MATERIAL AND METHODS

Test genes

A 2.2 kb fragment containing the *rol-6* gene of *C.elegans* (34, 35) with ~750 bp 5′ flanking sequence was cloned into pTZ19U (36) to create pRF5. This was used as substrate for in vitro oligonucleotide-directed mutagenesis using the procedure of Nakamaye and Eckstein (37) as supplied in a kit by Amersham (Arlington Heights, IL). Splice acceptor site mutants were made in the modified plasmid, pr6AG, in which *rol-6* is marked to distinguish it from the endogenous gene in transformants (see Figure 1). Oligonucleotides are listed in Table 1. r6Y/A was made with an oligonucleotide containing four degenerate positions as shown in Figure 1. Bacterial strains transformed with each construct were grown overnight in 50 −100 ml of rich medium plus ampicillin, and DNA was purified by a large scale alkaline lysis procedure (38). The DNA was further purified by LiCl purification, digestion with RNase and proteinase K, extraction with phenol-chloroform, and precipitation with ethanol (39). Final pellets were dissolved in TE or water and the concentration determined by the OD_{260}, which was verified by comparison with standards on agarose gels.

Worm transformation

The procedure described by Mello et al. (40), with minor modifications, was used to generate transgenic strains of *C.elegans*. N2 *var* Bristol worms carrying an amber *rol-6 (n1178)* allele that is slightly dumpy in appearance (J. Kramer, unpublished) were injected with DNA at 200 mg/ml in injection buffer (41). Each strain was derived from a single F1 worm, maintained by reinoculation with 12 rolling worms on growth-plates whenever the worms exhausted their food supply. For each construct, two independent strains were maintained, and pooled immediately before RNA extraction.

Preparation of RNA

Total RNA was extracted from mixed populations of worms as described previously (32). For analysis of r6Y/A and Del-1 mutants by cDNA-PCR, 50 late L4 larvae or young adult rollers were picked from each of two strains carrying the same transgene, and their RNA extracted in a microscale preparation. This provided an amount of input RNA equivalent to 100 rollers for each transgenic type, so that the relative levels of the two PCR products for each strain, as well as the relative levels between strains could be compared. One fourth of each RNA preparation was used in PCR analysis as described below. Poly(A) RNA was prepared using the polyA$^+$tract system from Promega (Madison, WI).

RT/PCR analysis

Products of the *rol-6* transgenes were analyzed by PCR using the procedure described (32) with the following variations. 1 μg of a downstream primer (prAG, antisense) from position +37 to +18 (thus extending from the r6AG-specific region, see Figure 1a) was used in a 20 μl annealing reaction. This was then split into the reverse transcription ('+RT') and control ('−RT') reactions. The control reaction had 50 μg of DNase-free RNase A added prior to the 37° incubation. The paired 12 μl reactions were diluted to 200 μl with water, and 10 μl of each used in the subsequent PCR. 0.1 μg of each primer was used in each PCR. The downstream primer was the same as that used for reverse transcription, prAG, while the upstream primer was one with sequence equivalent to SL1 (SL1-H), and also SL2 (SL2-H) for r6SL2A. In general, PCRs were performed for 25 cycles (each cycle: 1 min at 92°, 1 min at 50°, 30 sec at 72°). However, for the r6Y/A experiment with RNA from a limited number of picked worms, 40 cycles of PCR were performed. PolyA$^+$ RNA was used as indicated in the Figure legends. The PCR products were precipitated and then dissolved in 7 μl formamide with 0.04% bromphenol blue and xylene cyanol. These were electrophoresed, after heating at 95° for 2 min, on 10% (20:1) polyacrylamide/7 M urea/TBE gels at a field density of 40 V/cm. Polyacrylamide gels were electroblotted onto Hybond-N (Amersham) in 25 mM Na_xPO_4 (pH 6.5). The blots were probed with an end-labelled oligonucleotide equivalent to sequence from +1 to +17 of *rol-6* ('r6probe' in Figure 1a).

RESULTS

Requirements for trans-splicing: inactivation of the normal acceptor site in *rol-6* leads to trans-splicing at a cryptic site

All of the experiments reported in this paper were performed with a dominant allele of the *rol-6* gene that is in general use in *C.elegans* labs as a selectable marker for transformation studies (40). We utilized a plasmid containing *rol-6* , altered as shown in Figure 1 to allow specific detection of products of the transgene. The *rol-6* pre-mRNA is normally trans-spliced to SL1 at a site 173 nt downstream of the transcription initiation site (35). In an effort to determine whether trans-splicing is required for function, we attempted to inactivate the trans-splice site by mutating the final base in the outron from G to A. This is an invariant position, and in other systems has been shown to be absolutely required for splicing (42, 43). The worms transformed with this construct alone expressed the roller phenotype, indicating either that trans-splicing is dispensible or takes place in spite of the mutant trans-splice site. Analysis of the RNA from this transformed strain indicated that trans-splicing had occurred

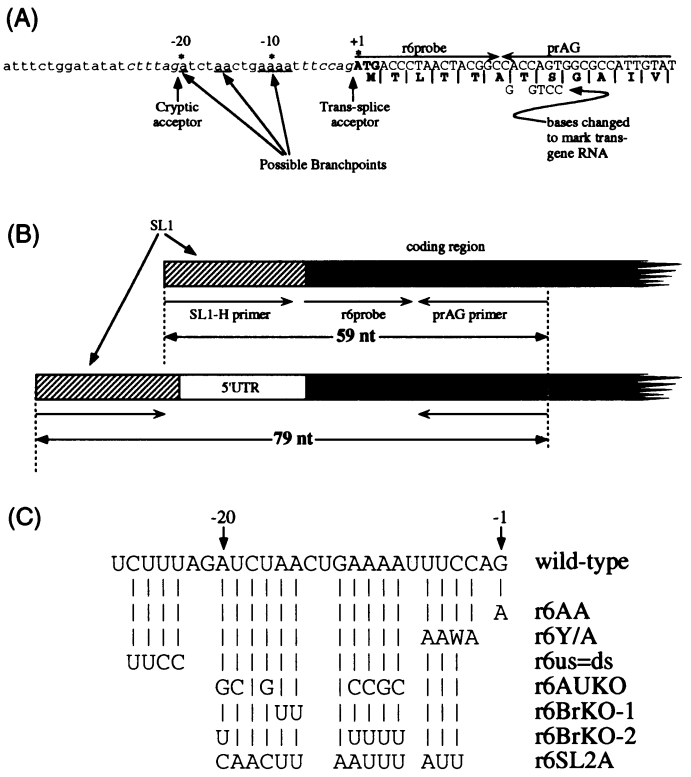

Figure 1. Summary of *rol-6* constructs. (A) Sequence of the region from −40 to +37. The wild-type *rol-6* sequence is shown, with the translation start site at +1, the normal trans-splice site at −1/+1, the cryptic trans-splice site at −21/−20, the encoded amino acid sequence, and the bases changed in pr6AG to distinguish it from the endogenous gene. Note that the amino acid identities are preserved. Overhead arrows indicate sense and sequence of oligonucleotides. Possible branchpoints are indicated (underlined). (B) Diagramatic representations of two different 5′ ends produced by splicing at the normal (+1) and cryptic (−20) sites. Black areas indicate the coding region, white areas indicate the 5′ untranslated region, and cross-hatched areas indicate the spliced leader. The two PCR primers (SL1-H and prAG) and the probe oligonucleotide are represented by arrows. (C) RNA sequence at the 3′ end of the *rol-6* outron and base changes in each mutant. The 'r6Y/A' row indicates potential changes for each of four positions; 'W' indicates 'A or U'.

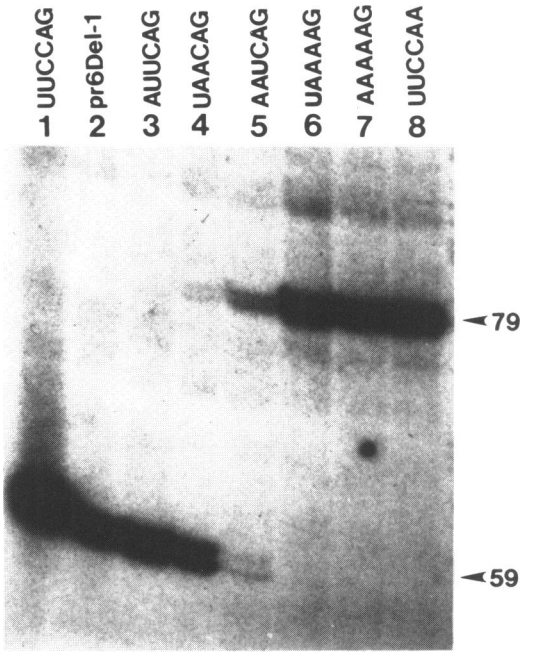

Figure 2. Splicing at a cryptic site in trans-splice site mutants. RNA was isolated from 25 roller transformants containing wild-type and mutant *rol-6* genes and subjected to RT-PCR. SL1-H was used as upstream primer. Positions for products from trans-splicing at the normal (59 nt) and cryptic (79 nt) sites are indicated. All bands were reverse transcriptase-dependent; reverse-transcriptase-minus lanes are not shown. pr6Del-1 represents a 100 bp deletion within the outron from position −50 to −150. The sequences above lanes 1 and 3−8 represent the last 6 bases of the outron.

Figure 3. Splice site choice when the two sites have the same sequence. RT-PCR was performed as decribed in the legend to Figure 2 on RNA isolated from strains carrying: Lane 1, r6us=ds; Lane 2, r6AG. All bands were reverse transcriptase-dependent; reverse-transcriptase-minus lanes are not shown.

at a cryptic site 20 nt upstream (Figure 1). This was determined by primer extension (data not shown) and cDNA-PCR (Figure 2, lane 8). Although this cryptic site (CUUUAG) is a good match to the consensus (12), no trans-splicing occurred at the cryptic site in transformants carrying the wild-type site (Figure 2, lane 1).

How important is the trans-splice site consensus?

When the splice site was completely inactivated, trans-splicing nevertheless occurred, but at a novel site, and there was no obvious effect on expression of the gene. This finding enabled us to test constructs in which the splice site might be only partially inactivated. It seemed likely that the degree of inactivation of the splice site would be reflected in the ratio of the two spliced products. Those variations of the splice site that partially or fully inactivate it should partially or completely shift splicing to the cryptic site. This is easily measurable by cDNA-PCR, which uses the same primers for both products. We used a degenerate oligonucleotide to direct replacements of the pyrimidines in the UUCCAG at the normal site to A's. All constructs generated rolling transformants and analysis of each is shown in Figure 2. The single distal base substitution, AUUCAG, had no observable effect on site usage. Thus, even though the −6 position is very highly conserved in *C.elegans* splice acceptor sites, mutation of the conseved U to an A did not appreciably inactivate the site. In a separate experiment (not shown) another single-site change, UUUAAG, also showed no change in site preference, indicating that the pyrimidine at −4 is also not required for this splice site. However, the double-substitutions, UAACAG and AAUCAG, did shift usage to the upstream site to a considerable extent. It is particularly interesting that the double mutation at −5 and −6 had a more severe effect than did the double mutation at −4 and −5, since the U at −6 is more highly conserved in *C.elegans*

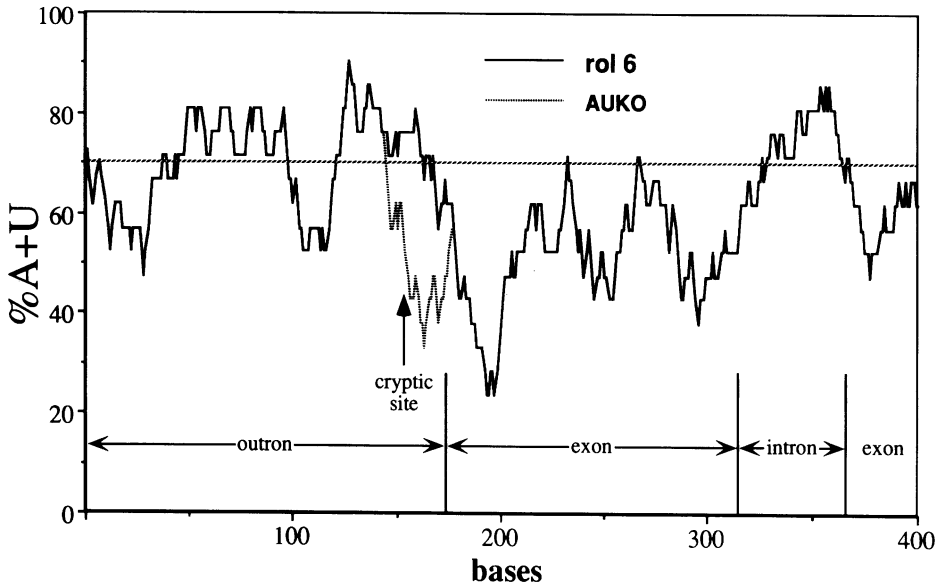

Figure 4. A+U content within the 5′ portion of the *rol-6* pre-mRNA. The percentage of A+U was calculated in a 20 nt window from position 1 to 400 of the initial transcript (translational start and trans-splice site are at nt173). Horizontal line is at 70% A+U. Dotted plot line indicates change in profile in AUKO mutation.

introns than the U at −4 (12). The triple and quadruple substitutions effectively inactivated the trans-splice site, at least in the context of the alternative site, since splicing was shifted completely to the upstream site. We conclude that, outside of the invariant AG, the single-base changes we tested weren't sufficient to inactivate the site, but overall match to the consensus is clearly a crucial variable in splice site choice. Importantly, the presence of pyrimidines at positions −3 to −6 is required for effective utilization of this splice site.

Why is the upstream site normally cryptic?

The cryptic splice site differs from the consensus only at positions −6 and −3, and at both of these positions pyrimidines are still present. Could this small deviation from the consensus result in its inactivity? To answer this question, we changed the sequence at the cryptic site to match that of the downstream site. RNA from transformants carrying this gene was analyzed by PCR as above (Figure 3). The results clearly show that splicing took place exclusively at the downstream site. The failure of an identical extended splice site to function in the cryptic site position indicates that the sequence of the splice acceptor is not responsible for the bias against this site.

This result implies that, in addtion to the closeness of the match to the consensus, there exists a second criterion on the basis of which splice sites are chosen in *C.elegans*. One of the characteristics of introns in *C.elegans* is a high percentage of A+U bases (~70%) (13, 25). To test whether this is a functionally important characteristic we replaced 7 of the 14 nt upstream of the consensus site and downstream of the cryptic site so that G's and C's replaced A's and U's (Figure 1). This effectively moved the transition from high to moderate A+U content closer to the cryptic site and further upstream of the normal site (Figure 4). Analysis of this mutation, called AUKO (for 'AU KnockOut'), is shown in Figure 5. While the wild-type control (AG) was spliced entirely at the normal site and the strain carrying the AA mutation is spliced entirely at the cryptic site,

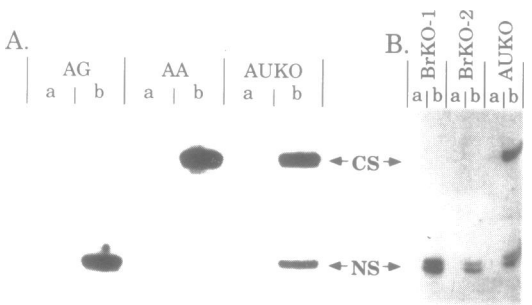

Figure 5. Effect of alteration of A+U content and potential branch point mutations. Analyses were performed on total RNA isolated from transformants. 'CS' indicates position for product of trans-splicing at cryptic site, 'NS' splicing at normal site. 'a' lanes are controls with reverse transcriptase omitted and RNase A added in the first step, 'b' lanes are with reverse transcriptase. (A) RT-PCR of AUKO-containing strain with AG and AA strains as controls. (B) RT-PCR of r6BrKO-1, r6BrKO-2 strains, and r6AUKO strain as control.

AUKO was spliced at both sites. Thus the cryptic site had been activated, but the normal site was still used at a reduced level. This result suggests that the A+U content in the immediate vicinity of the splice site is an important variable in trans-splice site selection in *C.elegans*.

The AUKO result is not due to interference with branch site formation

Previous work (26, 27) demonstrated the presence of a branched intermediate in trans-splicing analogous to the lariat intermediate in cis-splicing. The one documented branchpoint for trans-splicing in nematodes was determined using an in vitro splicing system from the parasitic nematode, *Ascaris lumbricoides* (44). For the particular gene examined, a branch is formed at the two underlined A's in the sequence AGUAAAA, where the A's are at positions −18 and 119 from the splice site. Note that there

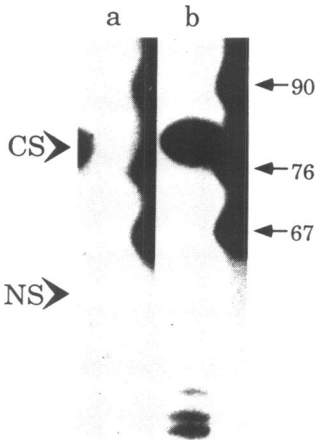

Figure 6. Failure to trans-splice pre-mRNA with the SL1 sequence at the cryptic site. Analysis was performed on poly(A) RNA from two pooled SCS-transformed worm strains. 'a' lane is control with reverse transcriptase omitted and RNase A added in the first step, 'b' lane is with reverse transcriptase. 'CS' indicates position of unprocessed RNA. 'NS' indicates position at which a trans-spliced product would have been expected.

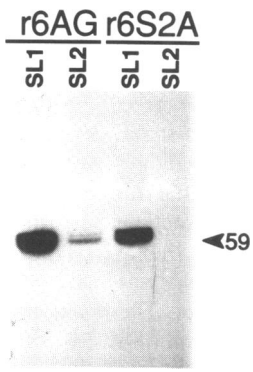

Figure 7. Trans-splicing of SL1 to r6SL2A mRNA. RNA from transformants carrying the r6SL2A construction was tested for SL1- and SL2-specific trans-splicing by RT-PCR as described in the legend to Figure 2. 'SL1' and 'SL2' lanes refer to the identity of the upstream primers used in PCR (SL1-H or SL2-H). All bands were reverse transcriptase-dependent; reverse-transcriptase-minus lanes are not shown.

is no match to the consensus mammalian branchpoint sequence, PyNPuAPy, consistent with the the observation that no consensus has been found an appropriate distance upstream from *C.elegans* 3′ splice sites (13). In fact, the only indication of where branchpoints might occur in *C.elegans* is an increased frequency of A's 15 to 17 nt upstream of the splice acceptor site. This and the nearly universal usage of A's as branchpoints in other organisms suggest that the A's at postions −15 and −16 in *rol-6* are the best candidates to act as a branch site. For this reason, these A's were not mutated in creation of the AUKO mutation. However, the lone A at −19 falls within a mammalian consensus, and the 4 A's at positions −8 to −11 provide a tetra-A sequence similar to that used in the *ascaris* in vitro system, which could be involved in a unique nematode branchpoint definition. To test whether any of these sites are needed for branchpoint formation, two separate mutations were created. In the first, BrKO-1, the two A's at −15 and −16 were changed to U's, while in the second, BrKO-2, all A's originally changed to G's or C's in AUKO were changed to U's. Changing these bases to U's maintained the A+U-richness, but presumably reduced the chance that branchpoints could form at these sites since U's do not generally function as branchpoint sites. The results in Figure 5B indicate that neither one of these mutations affects the decision to splice at the downstream site. No products from splicing at the cryptic site were observed, even on overexposure of the autoradiogram (not shown). We believe that either the branchpoint is further upstream or, more likely, the selection of the branchpoint is fluid. Such fluidity in branchpoint selection is known to occur in higher eucaryotes where eliminating branchpoints leads to utilization of cryptic branch sites (45). Whichever explanation is correct, these results demonstrate that interference with the trans-splice branchpoint is not the cause for the shift in splice site selection in the AUKO mutant strain. They also provide evidence that no sequence between the two potential splice sites defines the downstream site as the preferred site, since the same bases are changed as in AUKO, but when they are changed to U's, splicing at the normal site is not interfered with.

Splicing at the cryptic site is not followed by subsequent splicing at the downstream site

It is possible that splicing normally does occur at the cryptic site, but that this spliced product is then spliced at the downstream site, so that the original splicing product is not detected. Splicing at the upstream site would be unmasked only when splicing at the downstream site was made inefficient enough to allow export of upstream-spliced RNA's to the cytoplasm. In order to test for this possibility, the sequence from the start of transcription to the cryptic site was replaced with the SL1 sequence. This gene construction, called SCS, did not generate rolling transformants, implying a signal essential for proper expression of the gene might be missing. Because of this, the SCS mutant gene was introduced along with unmodified *rol-6* (allele *su1006*). PCR analysis of trace amounts of the SCS RNA was still feasible since the co-injected *rol-6* was not marked with the r6AG tag and would not prime with the prAG oligonucleotide. The results of this experiment are shown in Figure 6. PCR from poly(A) RNA without reverse transcription, which would represent product from any contaminating DNA, shows no product (lane a), while a reverse transcriptase-dependent band is clearly seen at the position of the cryptic site (Figure 6, CS), the position expected for unprocessed product. The fact that no product is observed at the position expected for splicing at the normal site (Figure 6, NS) indicates that RNA with the SL1 sequence at the upstream site is not processed further.

The region immediately upstream of the *rol-6* trans-splice site does not contain specificity determinants

In order to determine whether the outron must remain intact for correct trans-splicing to occur, we deleted positions −50 to −150 (measured from the splice site) by oligonucleotide-directed mutagenesis to create r6Del-1. Figure 2, lane 2, shows that splicing is either unaffected or only slightly decreased by this large deletion. Thus nothing in this region of the *rol-6* outron is required for specification of trans-splicing at the correct site or for function of the *rol-6* gene.

In *C.elegans*, a second spliced leader, SL2, is found on the mRNAs of certain genes (46). In these studies and in previous work with other genes (32), SL1 is the primary spliced leader

attached by trans-splicing; PCR indicates little or no SL2 is spliced onto any of the sites we have created. It appears as though an extra signal is required for SL2 specificity, either to signal the SL2 snRNP directly or indirectly, or to block the SL1 snRNP. To determine whether this signal is present in the region from −20 to −1 from the splice site, we eliminated this region of *rol-6* and replaced it with that of *gpd-3*, an SL2-accepting gene, creating SL2A. RT/PCR reactions performed on this RNA using the upstream primer for SL1 (SL1-H) and SL2 (SL2-H) are shown in Figure 7. The RNA was analyzed in tandem with r6AG to provide a positive control. The reaction with SL2-H indicates that although a small amount of SL2 was found on mRNA from the control strain, little or no SL2 was trans-spliced on the r6SL2A transcript. This indicates that the information needed to specify SL2 splicing is not contained entirely within this part of the outron.

DISCUSSION

Properties of *C.elegans* splice acceptor sites

In *C.elegans* trans-splice acceptor sites have the same consensus sequence as cis-splice acceptors, UUUCAG/G. Furthermore, neither one has an obvious polypyrimidine tract nor a branch point consensus (12, 13). The only other known distinguishing feature to demarcate introns and outrons relative to exons is their relatively high A+U content. The results of previous experiments in our laboratory (32, 33) have led us to the conclusion that no special sequence is required to signal trans-splicing to SL1, outside of those sequence elements normally found in introns. Rather, any gene that has an intron, now called an outron, at the 5′ end of its pre-mRNA will be trans-spliced to SL1. The fact that deletion of 100 nt from the 173 nt outron of *rol-6* did not affect trans-splicing provides additional support for this idea.

In this paper we investigated whether the two distinguishing features of *C.elegans* splice acceptor sites, the UUUC just upstream of the perfectly-conserved AG and the high A+U content of the outron, were necessary elements for trans-splice site recognition. The observation that elimination of the natural trans-splice site by conversion of the AG to AA did not eliminate trans-splicing or gene function made this study possible. Because trans-splicing now occurred at a formerly silent site 20 nt upstream, we were able to exploit the novel competition between these two sites to investigate what parameters influence the choice between the two sites. Any changes that result in changing splice site choice from the natural to the silent site must affect a component of the recognition signal.

Importance of the four pyrimidines

The consensus for the 3′ splice site presumably reflects the preferences for each position in the sequence. A survey of *C.elegans* introns (12) found the percent identity to UUUCAG to be 87, 99, 76, 81, 100, 100, respectively. Out of the 139 introns analyzed, only 2 did not have a U at position −5, for instance. Although we did not isolate a single-site mutation at −5, all the mutations that shifted the splicing in part or in whole to the upstream site do have the −5 position changed to an A, consistent with the idea that this is an important component of the recognition signal. When only one other base was changed to an A, as in UAACAG and AAUCAG, there was only a partial shift to the upstream site. Interestingly, the synergistic effect of the additional change reflects the level of conservation at the sites: alteration of the −6 position, which is more highly conserved than position −4, showed a more profound effect than did alteration of the −4 position. Alteration of any additional bases in the consensus shifted splicing completely to the upstream site. These observations lead us to hypothesize the existence of some component of the splicing machinery in *C.elegans* that recognizes this sequence, with exact matches not required for recognition, but with an overall 'fit' necessary. When this fit is poor, the presence of an AG is ignored or activated very slowly, so that nearby sites are recognized. This could be evolutionarily equivalent to the component responsible for recognition of the polypyrimidine tract adjacent to the 3′ splice site in higher eucaryotes (47−49). In any case, our experiments clearly demonstrate that the high conservation of the UUUC just upstream of the AG at which splicing occurs can be attributed to their role in splice acceptor site recognition.

Why isn't the upstream site used?

The upstream site is cryptic unless the normal site is inactivated. Its sequence from −6 to −1, CUUUAG, might be expected to contribute substantially to its lower activity, since C is rare at −6, occurring in only 3% of introns (vs 87% U), and U is rare at position −3, occurring in 18% of introns (vs 81% C) (12). However our experiment in which the cryptic site was replaced with a sequence identical to the downstream site (in r6us=ds) indicates that it is not the sequence of the cryptic site *per se* that results in its not being chosen under normal circumstances. Clearly, the upstream site functions perfectly well when the downstream site is inactivated. As the stepwise inactivation documented in Figure 2 indicates, the upstream site is a perfectly good splice site when the normal one is inactivated. These observations imply that something else favors the downstream site. An analysis of *C.elegans* genes indicates that introns and outrons are easily identifiable because of their high A+U content relative to the surrounding regions (~70% vs ~50%). A plot of the A+U-richness of *rol-6* shows that the normal trans-splice acceptor occurs at the boundary between A+U-rich and A+U-average RNA, while the cryptic site precedes this boundary by 20 nt (Figure 4). By changing 7 of the 14 bases between −6 and the cryptic site, we altered this 'A+U topology' as shown by the dotted line in Figure 4. In AUKO the cryptic site is located at the site of transition from A+U-richness to A+U-moderation. The fact that this alteration was sufficient to activate the cryptic site suggests strongly that the location of the A+U-richness transition is a second important parameter in splice site choice in *C.elegans*, as it has previouly been shown to be in plants (17). We do not know why the downstream site is not completely deactivated by the AUKO mutation. Perhaps, although the A+U-transitional aspect would favor the upstream site, a 5′ to 3′ exon scanning mechanism as described by Niwa et al. (50) would recognize the downstream site in the process of defining the smallest exon.

Fluidity of Branchpoint selection

From the results in the *Ascaris* in vitro system (44) and the high frequency of A's at −15 to −17, it seems likely that the *rol-6* branchpoint is within 20 nt of the splice acceptor site. BrKO-1 was created to determine if the A's retained in AUKO served as branchpoint for the normal splice site by changing them to less favorable U's. By a similar method, BrKO-2 was created to determine if any of the 5 A's altered in AUKO were involved in branchpoint recognition. We were able to show that none of these nucleotides are required for utilization of the normal splice

site. This suggests that the site of branchpoint formation in *C.elegans* may be quite fluid, since no A between the two splice acceptor sites is required for utilization of the downstream site.

SL1 vs. SL2 trans-splicing

In previous work we determined that changing an intron into an outron or creating an outron *de novo* resulted in SL1 trans-splicing onto the recipient transcript with, perhaps, only a trace amount of SL2 trans-splicing (32, unpublished results). This led us to conclude that SL1 splicing is a default mode, and that there is some specific signal required for SL2 splicing. Replacment of the region between the cryptic site and normal site in *rol-6* with the same region of the SL2-trans-spliced *gpd-3* gene (46, 51) allowed us to conclude that the signal for SL2 specificity does not lie solely in this region.

ACKNOWLEDGEMENTS

We thank John Spieth for his critical reading of this manuscript, and Jim Kramer for his kind gift of the cloned *rol-6 (su1006)* gene. This work was supported by grants GM30870 and GM42432 from the National Institute of General Medical Sciences.

REFERENCES

1. Smith,C.W.J, Patton,J.G., and Nadal-Ginard,B. (1989) *Annu. Rev. Genet.*, **23**, 527–577.
2. Green,M. R. (1991) *Annu. Rev. Cell Biol.*, **7**, 559–599.
3. Ruby,S.W. and Abelson,J. (1988) *Science*, **242**, 1028–1035.
4. Seraphin,B., Kretzner,L., and Rosbash,M. (1988) *EMBO J.*, **7**, 2533–2538.
5. Rosbash,M. and Seraphin,B. (1991) *Trends in Biochem. Sci.*, **16**, 187–190.
6. Parker,R.A., Siliciano,P.G., and Guthrie,C. (1987) *Cell*, **49**, 229–239.
7. Zhuang,Y, Goldstein,A.M., and Weiner,A.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2752–2756.
8. Wu,J. and Manley,J. (1989) *Genes Dev.*, **3**, 1553–1561.
9. Reich,C.I., VanHoy,R.W., Porter,G.L., Wise,J.A. (1992) *Cell*, **69**, 1159–1169.
10. Newman,A.J. and Norman,C. (1992) *Cell*, **68**, 743–764.
11. Zamore,P.D. and Green,M.R. (1989) *Proc. Natl. Acad. Sci.USA*, **86**, 9243–9247.
12. Fields,C. (1990) *Nucleic Acids Res.*, **18**, 1509–1512.
13. Blumenthal,T. and Thomas,J. (1988) *Trends in Genet.*, **4**, 305–308.
14. Thomas,J. , Lea,K., Zucker-Aprison,E., and Blumenthal,T. (1990) *Nucleic Acids Res.*, **18**, 2633–2642.
15. Ogg,S.C., Anderson,P. and Wickens,M.P. (1990) *Nucleic Acids Res.*, **18**, 143–149.
16. Hanley,B.A. and Schuler,M.A. (1988) *Nucleic Acids Res.* **16**, 7159–7176.
17. Goodall,G.J. and Filipowicz,W. (1991) *EMBO J.*, **10**, 2635–2644.
18. Krause,M. and Hirsh,D. (1987) *Cell*, **49**, 753–761.
19. Bektesh,S., Van Doren,K., and Hirsh,D. (1989) *Genes Dev.*, **2**, 1277–1283.
20. Nilsen,T.W., Shambaugh,J., Denker,J., Chubb,G., Faser,C., Putnam,L., and Bennett,K. (1989) *Mol. Cell. Biol.*, **9**, 3543–3547.
21. Zeng,W., Alarcon,C.M., and Donelson,J.E. (1990) *Mol. Cell. Biol.*, **10**, 2765–2773.
22. Rajkovic,A., Davis,R.E., Simonsen,J.N. and Rottman,F.M. (1990) *Proc. Natl. Acad. Sci, USA*, **87**, 8879–8883.
23. Sutton,R.E. and Boothroyd,J.C. (1986) *Cell*, **47**, 527–535.
24. Murphy,W.J., Watkins,K.P. and Agabian.,N. (1986) *Cell*, **47**, 517–525.
25. Huang,X-Y. and Hirsh,D. (1992) In Setlow,J.K. (ed.) *Genetic Engineering*, Plenum Press, NY. vol. 14, pp. 211–229.
26. Bektesh,S.L. and Hirsh,D.I. (1988) *Nucleic Acids Res.*, **16**, 5692.
27. Thomas,J.D., Conrad,R.C., and Blumenthal,T. (1988) *Cell*, **54**, 533–539.
28. Hannon,G.J., Maroney,P.A. and Nilsen,T.W. (1991) *J. Biol. Chem.*, **266**, 22792–22795.
29. Van Doren,K. and Hirsh,D. (1988) *Nature*, **335**, 556–559.
30. Bruzik,J.P., Van Doren,K., Hirsh,D., and Steitz,J. (1988) *Nature*, **335**, 559–562.
31. Bruzik,J.P. and Steitz,J. (1990) *Cell*, **62**, 889–899.
32. Conrad,R., Thomas,J., Spieth,J. and Blumenthal,T. (1991) *Mol. Cell. Biol.*, **11**, 1921–1926.
33. Conrad,R.,Liou,R.F., and Blumenthal,T. (1993) *EMBO J.*, **12**, in press.
34. Cox,G.N., Laufer,J.S., Kusch,M. and Edgar,R.S. (1980) *Genetics*, **95**, 317–339.
35. Kramer,J.M., French,R.P., Park,E.-C. and Johnson,J.J. (1990) *Mol. Cell. Biol.*, **10**, 2081–2089.
36. Mead,D.A., Szczesna-Skorupa,E. and Kemper,B. (1986) *Protein Engineering*, **1**, 67–74.
37. Nakamaye,K.L. and Eckstein,F. (1986) *Nucleic Acids Res.*, **14**, 9679–9698.
38. Birnboim,H.C. (1983) *Methods Enzymol.*, **100**, 243–255.
39. Fire,A. and Waterston,R.H. (1989) *EMBO J.* **8**, 3419–3428.
40. Mello,C.C., Kramer,J.M., Stinchcomb,D. and Ambros,V. (1991) *EMBO J.*, **10**, 3959–3970.
41. Fire,A. (1986) *EMBO J.*, **5**, 2673–2680.
42. Ruskin,B. and Green,M.R. (1985) *Nature*, **317**, 732–734.
43. Reed,R. and Maniatis,T. (1985) *Cell*, **41**, 95–105.
44. Hannon,G.J., Maroney,P.A., Denker,J.A., and Nilsen,T.W. (1990) *Cell*, **61**, 1247–1255.
45. Ruskin,B., Greene,J.M., and Green,M.R. (1985) *Cell*, **41**, 833–844.
46. Huang,X-Y. and Hirsh,D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8640–8644.
47. Gil,A., Sharp,P.A, Jamison,S.F., Garcia-Blanco,M.A. (1991) *Genes Dev.*, **5**, 1224–1236.
48. Patton,J.G., Mayer,S.A., Tempst,P., and Nadal-Ginard,B. (1991) *Genes Dev.*, **5**, 1237–1251.
49. Zamore,P.D. and Green,M.R. (1991) *EMBO J.*, **10**, 207–214.
50. Niwa,M., MacDonald,C.C., and Berget,S.M. (1992) *Nature*, **360**, 277–280.
51. Huang,X-Y., Barrios,L.A.M., Vonkhorporn,P., Honda,S., Albertson,D.G., and Hecht,R.M. (1989) *J. Mol. Biol.*, **206**, 411–424.