

The role of the Ou element in SL1
trans-splicing in *C. elegans* nematodes

Jenna Gallegos

University of Colorado at Boulder

Molecular, Cellular, Developmental Biology

Senior Honors Thesis October 25, 2011

Thesis Advisor: Dr. Thomas Blumenthal, MCDB Chair

Direct supervisors: Dr. Erika Lasda & Dr. Jason Morton

Committee Members: Dr. Christy Fillman (MCDB),

Dr. Nancy Guild (MCDB), & Dr. William Adams (EBIO)

Abstract

The nematode *C. elegans* exhibits SL1 trans-splicing, a unique form of pre-mRNA processing seen in a variety of lower metazoans. In *C. elegans*, there are two types of trans-spliced leader that are added onto the 5' end of trans-spliced pre-mRNA: SL2 and its evolutionary predecessor, SL1. A bioinformatic analysis by Graber et al. (2007) revealed two conserved sequences associated with SL2 and SL1-accepting pre-mRNA known as the “Ur” and “Ou elements” respectively. Lasda et al. (2010) demonstrated that the Ur element is required for SL2 trans-splicing and predicted that the Ou element functions similarly for SL1 trans-splicing. The present study investigated the role of the Ou element in SL1 trans-splicing by gathering primary data using mutational analysis followed by SL1 *in vitro* trans-splicing assays. Two different Ou-mutations considerably decreased SL1 trans-splicing efficiency when compared to wild-type constructs of the same gene. These data suggest that the Ou element may play a key role in SL1 trans-splicing.

Introduction

The nematode *Caenorhabditis elegans* is commonly used in studies of cellular and molecular biology (Blumenthal & Spieth, 1996). Although it is a eukaryote, the *C. elegans* genome is small, “enabling [scientists] to view the entire array of genes required to build a complex, multicellular animal” (Blumenthal & Spieth, 1996). Like those of higher eukaryotes, *C. elegans* genes are monocistronic (each mRNA encodes only one protein); however, they express an unusual feature of pre-mRNA processing also observed in trypanosomes and other lower metazoans (Spieth et al., 1993). Many *C. elegans* genes undergo unique intermolecular ligations involving the 5' splice site of one RNA and the 3' splice site of another, termed trans-splicing (Hirsh & Krause, 1987). SL trans-splicing in *C. elegans* involves the transfer of a short, conserved leader sequence from the 5' end of a spliced leader (SL RNA) to the 5'-most exon of a pre-mRNA (Nielsen, 1993; Blumenthal, 2005; Hastings 2005). The leader sequence has been proposed to serve a number of possible functions in the mRNA preceding and including translation, but the specific purpose of SL1 trans-splicing is not known (Lall et al., 2004). SL1 trans-splicing is highly conserved in the nematode phylum. Within *C. elegans*, 70% of mRNAs are trans-spliced (Spieth et al., 1993).

Krause and Hirsch (1987) described the discovery of a unique 22 nucleotide (nt) leader sequence at the 5' end of three actin mRNA in *C. elegans*. The sequence was reminiscent of the 35 nt leader that is trans-spliced onto trypanosome mRNA. Krause and Hirsch (1987) ran several tests to eliminate alternative mechanism hypotheses, such as cis-splicing or primed transcription, and suggested that the 22 nt sequence was trans-spliced onto the ends of actin pre-mRNA. They tracked the source of the leader sequence to a novel 100 nt small nuclear RNA (snRNA). The

snRNA is part of a complex now known as the SL1 small nuclear ribonucleoprotein (snRNP) (Blumenthal, 2005). The snRNA was trans-spliced onto the 5' end of the actin mRNA. The process of replacing upstream pre-mRNA sequence with this conserved 22 nt sequence is termed SL1 trans-splicing (Blumenthal and Spieth, 1996).

Pre-mRNA transcripts that will be trans-spliced differ from those that will not by the presence of an outtron (Blumenthal, 2005). The outtron resembles an intron (non-coding pre-mRNA that is removed by cis-splicing in eukaryotes), but consists of the RNA from the transcription start site to the 5' end of the first exon and is removed by SL1 trans-splicing (Blumenthal & Spieth, 1996). The outtron is replaced by a common 22 nt sequence known as the spliced leader that is not present in the genomic DNA for that gene (Hirsh & Krause, 1987). The spliced leader is donated by a 100 nt RNA contained within the SL1 snRNP (Blumenthal, 2005). The 5' splice site on the SL RNA and the 3' splice site on the pre-mRNA closely resemble the 5' and 3' consensus intronic splice sites that function in cis-splicing (Kent and Zahler, 2000). In cis-splicing, the U1 snRNP recognizes the 5' splice site on the pre-mRNA via a base-pairing interaction (Spieth et al., 1993). Trans-splicing is mechanistically similar to cis-splicing; however, the spliced leader and, therefore, the 5' splice site is attached to the SL snRNP. In this case, it is the SL-accepting pre-mRNA that is predicted to recognize the 5' splice site by base-pairing (Lasda et al. 2010). This proposed interaction is modeled in Figure 0.

Spieth et al. (1993) discussed the gene arrangement that causes some pre-mRNAs to trans-splice to a second class of spliced leader, SL2, instead of its evolutionary predecessor SL1. Like bacteria, some *C. elegans* genes exist in clusters controlled by the same promoter known as operons; however, *C. elegans* RNA is processed prior to translation into single-gene units separated by a combination of 3' end formation and trans-splicing. Genes downstream in an

operon are trans-spliced to SL2. To prove this, Spieth et al. (1993) moved an SL1 accepting gene to a location downstream in an operon and observed that the gene was then spliced to SL2. These results suggest that “this chromosomal arrangement is necessary and sufficient for SL2 trans-splicing.” (Speith et al. 1993)

Another study suggests that while a location downstream in an operon is a prerequisite for SL2 trans-splicing, it may not be sufficient. As described above, the U1 snRNP acts in cis-splicing (removal of introns) by base pairing with a complementary sequence at the 5' splice site. Lasda et al. (2010) proposed that a U-rich region of the pre-mRNA acts by a similar mechanism in recognizing the 5' splice site on the SL2 snRNP. In 2007, Graber et al. performed a statistical analysis of the sequences of genes that are SL1 and SL2 trans-spliced in hopes of finding trends to differentiate between pre-mRNA that is not trans-spliced, SL1 trans-spliced, and SL2 trans-spliced. They further defined this evolutionarily conserved U-rich region (first identified by Huang et al. 2001) common among SL2 trans-spliced genes termed the “Ur element”. In 2010, Lasda et al. used mutational analysis and bioinformatics to more specifically describe the Ur element as a short stem loop followed by the sequence UAYYUU (where Y represents a Cytosine or Uracil) located approximately 50 nt upstream of the SL2 trans-splice site that is necessary for SL2-specific trans-splicing. The Ur element is predicted to act by base-pairing with the SL2 trans-splice site on the SL2 snRNP (Lasda et al., 2010).

Lasda et al. (2010) and Graber et al. (2007) also discussed a common outtron-specific sequence of SL1 trans-spliced pre-mRNA known as the “Ou element”. The Ou element is characterized by a UC-rich region approximately 50 nt upstream of the SL1 trans-splice site. Specifically the consensus sequence URYYYUY (where R represents an Adenine or a Guanine), or one closely related, appears to be associated with SL1 trans-spliced pre-mRNA. Lasda et al.

(2010) suggested that the Ur element exemplifies a common mechanism for SL2 trans-splicing, and hypothesized that the Ou element may be the SL1 equivalent.

The SL1 and SL2 RNAs in complex with the snRNP's are of similar length and share much of the same composition. Both contain a 22 nt SL exon and an Sm protein binding site (Blumenthal, 2005). The secondary structure of both SL RNAs is predicted to consist of three stem-loops with a 5' trimethylguanine cap (Blumenthal, 2005). SL1 RNA differs from SL2 RNA in its association with the trans-splicing specific proteins *sna-1* and *sna-2*. Further, the SL2 snRNP is unique in its interaction with 3' end formation factor CstF-64 (MacMoris et al. 2007). Nonetheless, despite their specificity to different types of gene arrangements, SL1 and SL2 trans-splicing machinery is very similar. It is, therefore, reasonable to predict that a consensus sequence exists for SL1 trans-splicing that is analogous to the Ur element in SL2 trans-splicing.

It is predicted that the Ou element base pairs with the SL1 trans-splice site on the SL1 snRNA in a necessary role similar to that of the Ur element for SL2 trans-splicing. Unlike the Ur element, however, little work has been done to experimentally define or verify the role of the Ou element in trans-splicing. Bioinformatics and the existence of a complementary sequence on the SL1 snRNP leads to the prediction that the Ou element is defined by the consensus sequence URYUY. This hypothesis is further supported by the frequency at which this combination appears in SL1 trans-spliced genes according to bioinformatics (Graber et al., 2007). Figure 0 demonstrates a model for the proposed interaction between the Ou element and the 5' trans-splice site on the SL1 RNA.

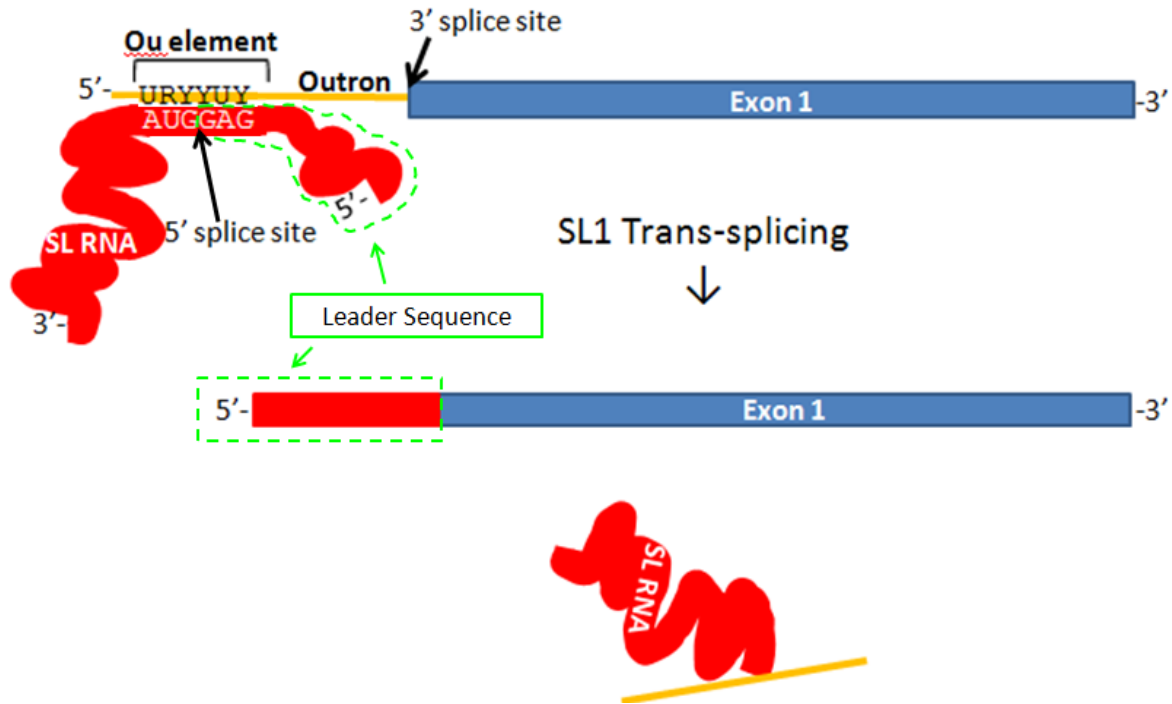


Figure 0: Model for the interaction between the Ou element in the outtron of SL1 trans-spliced mRNA and the SL1 snRNP complex.

Despite the prevalence *C. elegans* research, much still needs to be elucidated about the *C. elegans*-specific process of SL1 trans-splicing. Studies have discovered novel proteins involved in trans-splicing, and bioinformatics analysis of trans-spliced genes has identified the Ur and Ou elements, but questions remain about the specific mechanisms of trans-splicing. This project examines the role of the Ou element in SL1 trans-splicing. *C. elegans* is important as a model organism and is frequently studied with reference to human health. Thus, *C. elegans* research investigating trans-splicing will contribute to the understanding of this process and may have practical applications as well.

In this study, the role of the Ou element in SL1 trans-splicing was investigated *in vitro*. First, the outtron of an gene SL1 trans-spliced gene was defined and a potential Ou element was identified. This gene, *nduf2.2* (*T26A5.3*) is predicted to encode a mitochondrial protein. Then,

two different mutations were created in the outtron of the gene *nduf2.2* disrupting the proposed Ou consensus. An *in vitro* trans-splicing assay was used to compare the trans-splicing competency of the wild-type construct versus the two mutant constructs. A decrease in SL1 trans-spliced product in mutant constructs would indicate that the Ou element plays a role in trans-splicing, and when the consensus is interrupted, it does so less efficiently. If the Ou element is necessary for trans-splicing, a decrease in trans-splicing efficiency was predicted to occur when the proposed Ou consensus was disrupted.

It was observed from the *in vitro* reactions comparing wild-type and mutant constructs that mutating the Ou element considerably decreased the yield of SL1 trans-spliced product. These results were consistent among multiple trans-splicing reactions and two separate mutants. These data indicate that the Ou element was correctly identified for the gene *nduf2.2* and that it is in fact necessary for SL1 trans-splicing *in vitro*.

Materials and Methods

I. Defining outtron length

To analyze the importance of the Ou element, it was necessary to identify a good candidate gene for study. An ideal gene splices exclusively to SL1, has a short outtron (to limit the number of possible matches to the Ou consensus sequence), and has a close or perfect match to the sequence URYUY. The efficiency of trans-splicing and the lack of a TATA box in many *C. elegans* genes has made it difficult to identify promoters; thus, the length of most outtrons have not been determined (Blumenthal and Spieth, 1996).

To select candidate genes for this study, I aligned RNA Polymerase II ChIP-sequence data from Baugh et al. (2009) with the nematode genome in the UCSC Genome Browser. The ChIP sequence data maps the association of polymerase II with the *C. elegans* genome. Based on the data of Morton and Blumenthal (2010), it was predicted that a large polymerase peak upstream of a gene is indicative of the location where polymerase paused at the promoter and, therefore, the approximate 5' end of the outtron. A good candidate would be SL1-accepting and have a large well-defined polymerase peak very close to, but just upstream of, the trans-splice site. After selecting five candidate genes with predicted outtrons of less than 400 base pairs (bp) by this mechanism, I researched the genes on WormBase to confirm that they were trans-spliced to SL1 and not to SL2.

To verify the actual length of the outtrons of the five candidate genes (*alh-9*, *rpn-2*, *F52C9.7*, *C16A3.6*, and *nduf-2.2*), I used extracted RNA from *C. elegans* nematodes. I performed reverse transcription (RT) reactions on the mRNA from each of the five SL1 accepting genes to procure cDNA from the lysed nematode RNA. Reverse primers in the second exon and Superscript II (Invitrogen) were used. For each gene, I designed multiple forward primers to amplify the cDNA by polymerase chain reaction (PCR) at increasing distances upstream of the 5' end of the first exon (well beyond the expected length of the outtron) and reverse primers in a downstream exon. Taq polymerase (Invitrogen) was used to amplify the cDNA. Presumably, only primers within the original outtron of the gene would be extended to produce a product. Furthermore, since the primers spanned an intron, I would be able to see to what extent splicing had already occurred in the RNA samples. A separate PCR reaction was performed for each forward primer and the product was run on an agarose gel stained with ethidium bromide. For each gel, there were two negative controls, a blank containing water in

place of cDNA and one containing an RT product in which the enzyme Super Script II had been replaced with water to ensure that genomic DNA was not being amplified in the sample. A positive control containing genomic DNA instead of cDNA was also used to ensure all primers were binding specifically and efficiently.

The following Figures 1.1-1.4 contain the results for all genes not selected for study (all except *nduf2.2* which is examined in more detail in the results section). The primers are named for the distance from their 3' end to the 5' end of the first exon. Primer lengths are not to scale and only primers within the expected outtron length based on the Baugh et al. data (2009), ergo primers downstream of the polymerase peak, are correctly aligned with the figures. These genes were not chosen because their outtron length, according to the RT-PCR data, exceeded that of *nduf2.2*.

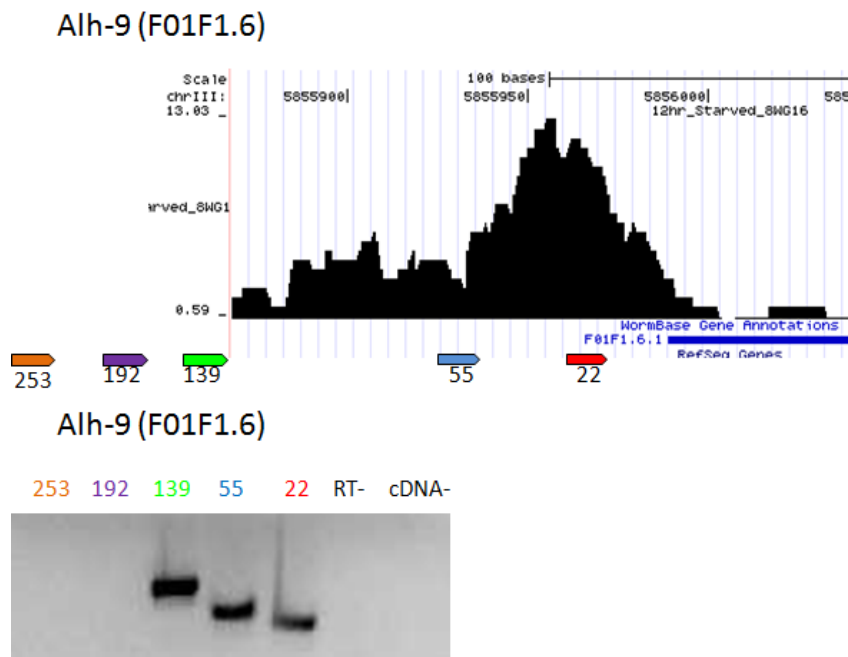


Figure 1.1. Alh-9 has an outtron greater than 139 bp in length

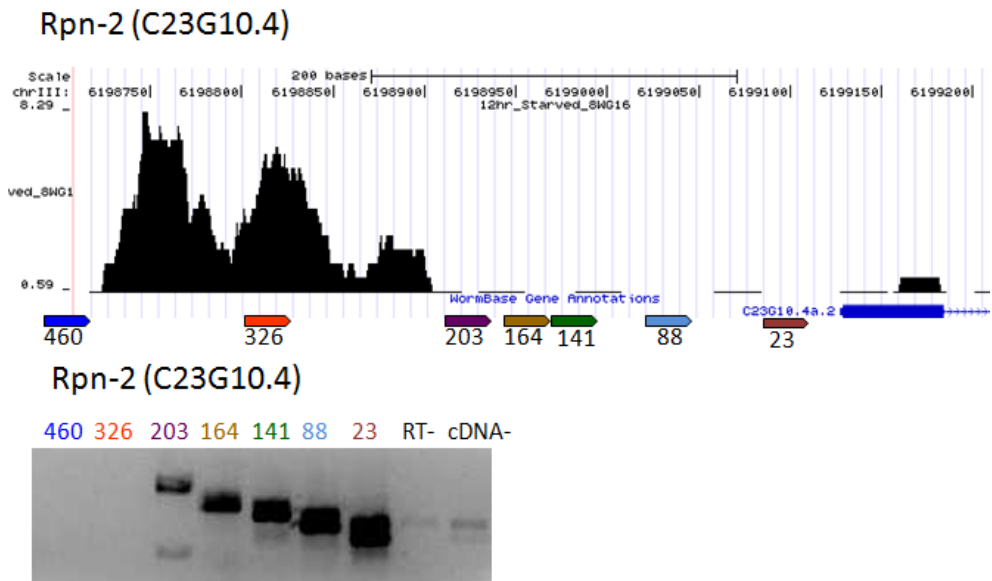


Figure 1.2. Rpn-2 has an outron greater than 203 bp in length

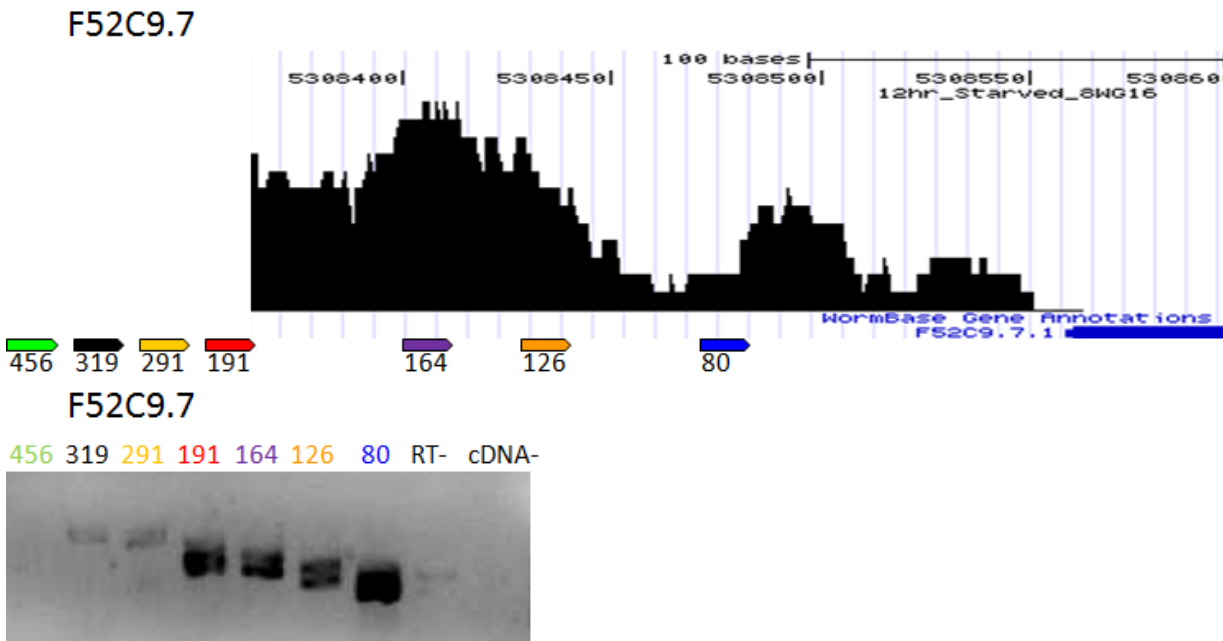


Figure 1.3. F52C9.7 has an outron greater than 191 bp in length

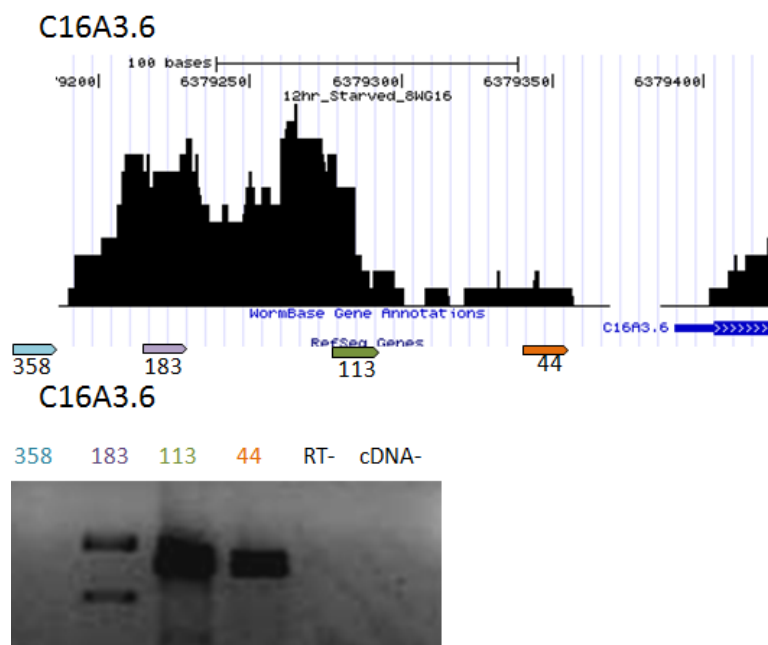


Figure 1.4. C16A3.6 has an outtron greater than 183 bp in length

II. Cloning and mutating *nduf2.2*

To create wild-type and mutant constructs, I first amplified genomic DNA corresponding to *nduf2.2* from *C. elegans* wild-type populations using pfx polymerase (Invitrogen). I then cloned a fragment from 111 nt upstream of the first exon to 31 nt downstream of the third exon into the EcoRV and KpnI restriction sites of pBluescript SK- plasmid, making a wild-type *nduf2.2* construct. I then used PCR mutagenesis to introduce substitution mutations to the prospective Ou elements using pfx polymerase (Invitrogen) and standard PCR cloning techniques. The sequences of the wild-type and mutant constructs were confirmed by DNA sequencing (wild-type and mutant constructs are summarized in Figure 5).

III. *In vitro* transcription and trans-splicing

Mutant and wild-type plasmids were linearized with XbaI and purified using a Qiagen MinElute PCR Purification Kit. Mutant and wild-type DNA constructs were transcribed *in vitro* using a T7 RNA Polymerase mMessage mMachine Kit (Ambion). Samples were then polyadenylated by an Ambion/Applied biosystems Poly(A) Tailing kit. RNA was isolated by phenol/chloroform extraction and isopropanol precipitation, resuspended, and purified by Ambion/Applied Biosystems NucAway Spin columns. The *in vitro* trans-splicing system was set up as described in Lasda et al. 2010. Briefly: 25 ng (for most samples though a gradient of 5, 10, 15, and 25 ng was also used for one trial) mutant or wild-type RNA substrate was added to a 15 μ l reaction of 50% *C. elegans* embryonic extract capable of splicing. The extract contained endogenous SL1 and SL2 complexes as well as trans-splicing machinery. ATP- samples lacking ATP, creatine phosphate, and creatine phosphokinase were set up in parallel to ensure that products obtained are ATP-dependent and, therefore, a product of the splicing reaction. Samples were incubated at 15°C for 2 hours before adding STOP buffer. Reactions were then digested by Proteinase K for 15 minutes at 37°C and the RNA was isolated by phenol/chloroform extraction and ethanol precipitation.

IV. Analysis of spliced product

20% of the RNA from the splicing reactions was used in a 10 μ l RT Reaction with Invitrogen Superscript II and a reverse primer downstream of the gene in a pBluescript SK- region. 2 μ l of the RT reaction (0.5 μ l for IC samples) was amplified in 50 μ l PCR reactions. Reverse primers corresponded to a pBluescript SK- region downstream of the gene. Forward primers specific for SL1 (SL1 leader sequence + A) were used to amplify products that have

been SL1 trans-spliced. Samples were also primed for the SL2 sequence (SL2 leader sequence + A) to determine if SL2 trans-splicing takes over when SL1 trans-splicing decreases. Products in which the SL1 leader sequence was trans-spliced to the 5' end of the second exon were isolated by the 2E2 primer which contains the SL1 leader sequence plus the first several bases of exon 2. An internal portion of the second exon (IC) was used to amplify all *nduf2.2* cDNA regardless of spliced status. Products isolated by these primers sets are examined in detail in the results section. Finally, cDNA minus controls contain no DNA in order to identify if there are any contaminants present in the PCR reaction.

12.5 ul of each PCR reaction was analyzed on a 2% agarose gel stained with ethidium bromide. Quantitative comparisons were achieved using the volume select tool on the Quantity One software used to analyze the gels. These comparisons are expressed as ratios as they are determined from band intensity and the actual volume amounts are arbitrary.

V. PCR primers

Oligo Name	Gene	Sequence (5'→ 3')	purpose
alh_RT-Rev	<i>alh9</i>	ACCTCGCACCAATCGTTATAAGCC	candidate gene selection
alh9-Rev	<i>alh9</i>	GAGATGGCGATTTCATAGTCTTGTACG	candidate gene selection
alh9_up22-For	<i>alh9</i>	CTAATACCTTCTGTTTCAAAGCGTCTAAC	candidate gene selection
alh9_up55-For	<i>alh9</i>	CTTGTTTTTTGACCACTAACCTCTTCG	candidate gene selection
alh9_up139-For	<i>alh9</i>	CGTCGTCATAAACAATCTGCCCG	candidate gene selection
alh9_up192-For	<i>alh9</i>	GATCTGCAATCTGTCCTTGAATTCC	candidate gene selection
alh9_up253-For	<i>alh9</i>	TAGCCTTGTATCATGTTCCACGTC	candidate gene selection
rpn-2_RT-Rev	<i>rpn2</i>	AAACTTTGGAGACGAGCAGTGC	candidate gene selection
rpn2-Rev	<i>rpn2</i>	GCGACCTCAAACCATGTGTTTAC	candidate gene selection
rpn2_up23-For	<i>rpn2</i>	CCTAGCTGCTCTCATTTTCGGATTG	candidate gene selection
rpn2_up88-For	<i>rpn2</i>	GACGAGAAAGCTGTTTTTCTGATAACCTAG	candidate gene selection
rpn2_up141-For	<i>rpn2</i>	GCCTGATTTATTTTCTAGACGTTTTCTAATTTGTC	candidate gene selection
rpn2_up164-For	<i>rpn2</i>	CCATTGTTGGTGTGTTGAGTAATTTTATGATTTTC	candidate gene selection
rpn2_up203-For	<i>rpn2</i>	TTGCAATCATTCTTTTCGTTCCGG	candidate gene selection
rpn2_up290-For	<i>rpn2</i>	GGATGACAATTCTGACCACAGC	candidate gene selection
rpn2_up326-For	<i>rpn2</i>	GCTGGCGCAATTATCGATTAGTTG	candidate gene selection
rpn2_up460-For	<i>rpn2</i>	GCGGTCAAACCTAGACACCAATACAC	candidate gene selection

F52C9.7_RT-Rev	F52C9.7	GAGGATTATCCATGATTTCCCGACG	candidate gene selection
F52C9.7-Rev	F52C9.7	TTTCTTCTCGAACCTTGACGGC	candidate gene selection
F52C9.7_up80-For	F52C9.7	GCGCAACGCGTTTTGAAGTGTA	candidate gene selection
F52C9.7_up126-For	F52C9.7	CCATGGCCGAGTTTTGACCTATATTG	candidate gene selection
F52C9.7_up164-For	F52C9.7	CACCAACAATGGCCTAGAACCAC	candidate gene selection
F52C9.7_up191-For	F52C9.7	CCTTCTCCATTTGAAATACCGCG	candidate gene selection
F52C9.7_up291-For	F52C9.7	GCTGCTCCTTAACACCAAATTCG	candidate gene selection
F52C9.7_up319-For	F52C9.7	GTITTTTGGTTTATATATTAGTATGAGAGCGATGG	candidate gene selection
F52C9.7_up456-For	F52C9.7	CCCAGCAACACATTTTCAGCC	candidate gene selection
C16A3.6_RT-Rev	C16A3.6	GGCTAAGAGCTCTTTCTCAATAGCG	candidate gene selection
C16A3.6_RT	C16A3.6	GCAACAAGTGCCTTCTCTTCAC	candidate gene selection
C16A3.6_UP5-For	C16A3.6	CTCAAAAAATTCGAATTTTTCAGAAACTGGG	candidate gene selection
C16A3.6_UP44-For	C16A3.6	CTTTTTCCTTTTATCGAAGTAATTCTCGTGTG	candidate gene selection
C16A3.6_UP87-For	C16A3.6	CAATTTTTTTGCCCCGTTCTTCTCAATC	candidate gene selection
C16A3.6_UP119-For	C16A3.6	GAAGACGCACACTTTCGGTTC	candidate gene selection
C16A3.6_UP177-For	C16A3.6	GAAAATAGAGAGAAGGCGCCTTAAGG	candidate gene selection
C16A3.6_UP242-For	C16A3.6	GTGGCCCAGTTATTTCTGTG	candidate gene selection
nduf2.2_RT-Rev	nduf2.2	CCCTCATAATTGTCGCTGAGCAC	candidate gene selection
nduf2.2_Rev	nduf2.2	GGATACGCGTTCTGACATCCAAAG	candidate gene selection
nduf2.2_UP0-For	nduf2.2	CAAGTTTAACTGTAAATTATTACTGTAAATTTGTTTCAGAG	candidate gene selection
nduf2.2_UP15-For	nduf2.2	CTCCTTGAATTTTCAAGTTTAACTGTAAATTATTACTG	candidate gene selection
nduf2.2_UP28-For	nduf2.2	CGTTTTCTCCTTGAATTTTCAAGTTTAACTG	candidate gene selection
nduf2.2_UP36-For	nduf2.2	GTTTTACACGTTTTCTCCTTGAATTTTCAAG	candidate gene selection
nduf2.2_UP49-For	nduf2.2	GAAATTAGCAATATTTGTTTACACGTTTTCTCC	candidate gene selection
nduf2.2_UP66-For	nduf2.2	CGACTACGGTAGTATTTGAAATTAGCAATATTTG	candidate gene selection
nduf2.2_UP91-For	nduf2.2	CAACGATTGAAAGCGACTACGG	candidate gene selection
nduf2.2_UP120-For	nduf2.2	GCAAGGAGCAAGTCTCGAGTAG	candidate gene selection
nduf2.2_UP149-For	nduf2.2	GCGCTACCTTAATCTTTGAATAGCC	candidate gene selection
nduf2.2_UP256-For	nduf2.2	GGCGTTTGAGAAAGATTA AAAAGCCC	candidate gene selection
nduf2.2_KpnI-For	nduf2.2	CGATTGAAAGCGACTACGGTAGTATTTG	cloning nduf2.2
nduf2.2_EcoRV-Rev	nduf2.2	CAAGACCTACACACAGGTTGAGTCGTAA	cloning nduf2.2
T3-Rev	pBSK	TCCCTTTAGTGAGGGTTAAT	mutating nduf2.2
T7-For	pBSK	GTAATACGACTCACTATAGGG	mutating nduf2.2
n2uf2.2_mut1-For	nduf2.2	GCAATAAAACAATTACACGTTTTCTTCTTGAATTTTCAAG TTTAACTG	mutating nduf2.2 (mut 1)
n2uf2.2_mut1-Rev	nduf2.2	GAAAACGTGTAATTGTTTTATTGCTAATTTCAAATACTAC CGTAGTCG	mutating nduf2.2 (mut 1)
n2uf2.2_mut2-For	nduf2.2	GCAATATTTACCTCACACGTTTTCTCCTTGAATTTTCAAG TTTAACTG	mutating nduf2.2 (mut 2)
n2uf2.2_mut2-Rev	nduf2.2	GAAAACCTGTGAGGTAAATATTGCTAATTTCAAATACTAC CGTAGTCG	mutating nduf2.2 (mut 2)
pBSK_RT-Rev	nduf2.2	ACTAGTGGATCCCCCGGGCTG	RT of spliced RNA
pBSK-Rev	pBSK	GAATTCCTGCAGCCCCGGGGGATC	PCR of spliced cDNA
SL1 + A	n/a	GGTTTAATTACCCAAGTTTGAGA	PCR of spliced cDNA
SL2 + A	n/a	GGTTTAAACCCAGTTACTCAAGA	PCR of spliced cDNA
nduf2.2_2E2-For	nduf2.2	GGTTTAATTACCCAAGTTTGAGACTG	PCR of spliced cDNA
nduf2.2_IC-For	nduf2.2	GCGATCCGGTGCTCAGC	PCR of spliced cDNA

Results

I. Defining the outtron length of *nduf2.2*

I examined the outtron length of five genes and determined *nduf2.2* to be the best candidate. Figure 2 shows Chromatin Immunoprecipitation sequence data of RNA Polymerase from Baugh et al.(2009) aligned with the UCSC *C. elegans* genome browser for the gene *nduf2.2*. I predicted from the Baugh et al. (2009) data that *nduf2.2* has an outtron of approximately 80 bp.

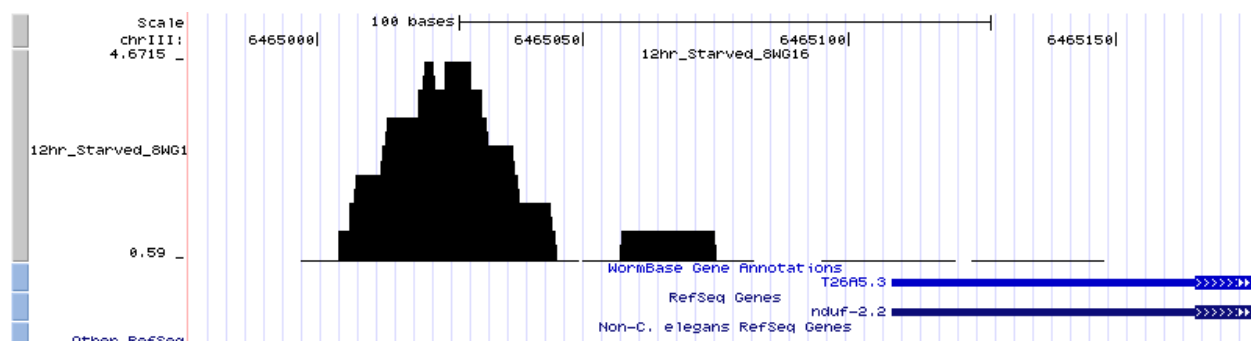


Figure 2. ChIP sequence data from Baugh et al. (2009) aligned with the UCSU genome browser (<http://genome.ucsc.edu/>). The center of the polymerase peak occurs approximately 80 bp upstream of the first exon.

In order to confirm the length of the outtron, I reverse transcribed *nduf2.2* RNA and amplified it using seven primers each at a different location upstream of the trans-splice site (Figure 4). RT-PCR data presented in Figure 4 confirm that the outtron is approximately 80 bp in length. The lanes are labeled according to the distance of the 3' end of the primer used to amplify the *nduf2.2* cDNA from the trans-splice site. As explained in materials and methods, there are three samples for each primer: cDNA minus (negative controls containing no template DNA), the *nduf2.2* cDNA sample, and genomic DNA (positive control to ensure primer precision). The

last primer to successfully amplify the *nduf2.2* cDNA is 49 bp upstream of the first exon at its 3' end and 84 bp upstream of the first exon at its 5' end. This supports my prediction that the distance between the transcription start site and the SL1 trans-splice site of *nduf2.2* is approximately 80 nt.

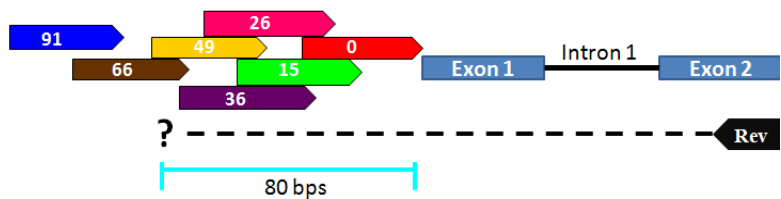


Figure 3: Map of primers used to determine outtron length. Oligos are named for the distance between their 3' end and the trans-splice site. Oligo lengths are not to scale.

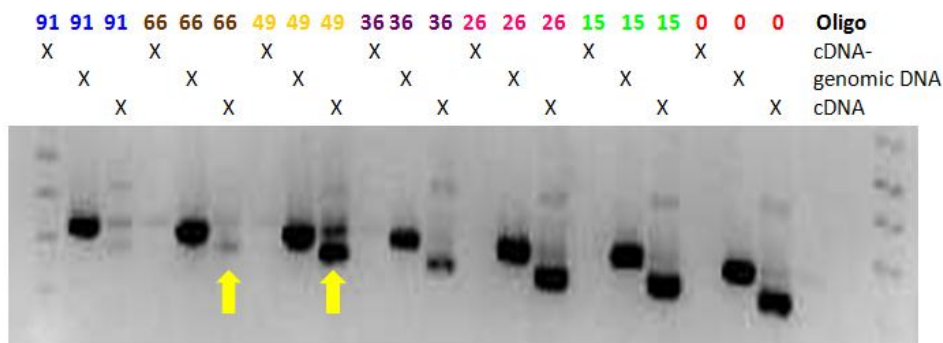


Figure 4. Results from RT-PCR analysis of outtron length of *nduf2.2*. The genomic DNA bands run slightly higher as they contain introns. Double bands in the cDNA lanes (as for oligo 49) indicate cis-spliced and non cis-spliced cDNA. cDNA samples are detectable by oligo 49 but not oligo 66 (as shown by arrows) indicating the 5' end of the outtron lies between these primers.

II. Comparing splicing for mutant and wild-type constructs

After identifying the outtron for *nduf2.2*, Dr. Blumenthal, Dr. Morton, and I then searched the sequence for potential Ou elements. There is a perfect match for the predicted Ou element

(UGUUUU) 60 bp upstream of the trans-splice site as well as several nearby and/or over-lapping sequences that are one to two base mismatches from the Ou consensus. One or all of these sequences may interact independently or redundantly with the SL1 trans-splicing machinery and be required for SL1 trans-splicing. We then designed two different mutations to disrupt the Ou consensus. We also mutated two base-pairs just upstream of the proposed Ou element in order to also disrupt a one base mismatch to the consensus overlapping and just upstream with the sequence UAUUUG. Figure 5 shows the proposed Ou element and the sequence of both mutants.

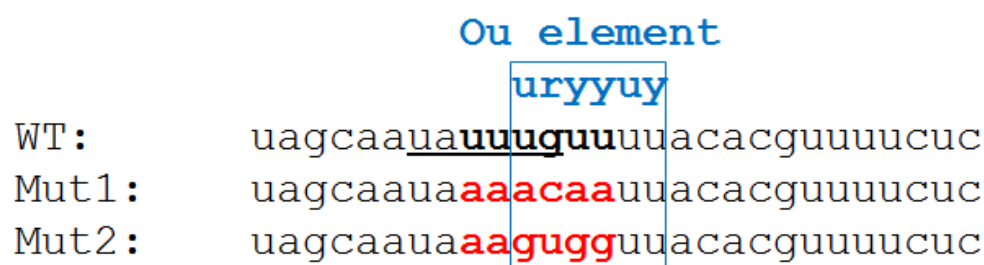


Figure 5: Sequences of wild-type (WT) and mutant (Mut) constructs across the potential Ou element from 49 to 76 bp upstream of the SL1 trans-splice site. Mutated regions are in bold. The mutated bases are in red, the region of the potential Ou element is surrounded by a blue box, and the overlapping one base mismatch is underlined.

In order to define the role of the Ou element in *nduf2.2*, as Lasda et al. (2010) have for the Ur element, I have employed many of the same methods but focused on SL1 trans-splicing. To test the significance of mutating the Ou element of *nduf2.2* on trans-splicing efficiency, I assayed trans-splicing using mutant and wild-type RNA constructs treated with the same *in vitro* splicing reaction. The resulting RNA was reverse transcribed. The cDNA was PCR amplified with the SL1 primer, thus isolating all SL1 trans-spliced products.

Figure 6 shows a schematic of the SL1 primer aligned with its complementary region of SL1 trans-spliced mRNA. The red box symbolizes the SL1 leader sequence and the dotted line represents the expected product for the primer system. *In vivo* cis-splicing out-competes trans-

splicing at intronic splice sites. *In vitro*, however, splicing occurs post-transcriptionally and trans-splicing machinery is in higher abundance relative to cis-splicing machinery than *in vivo*. As a result, a number of trans-spliced products are observed that are not seen *in vivo*. In addition to trans-splicing to the first exon, trans-splicing to the second and third exons is observed *in vitro*. Cis-splicing is also not 100% efficient. As a result products with varying degrees of cis-splicing (cDNA in which no introns are present, both introns are present, or only one intron is present) are observed. Figure 6 outlines all of the potential SL1 trans-spliced products that can occur *in vitro*. Products in which only intron 1 has been removed and products in which only intron 2 has been removed are equivalent in size (533 and 532 nt respectively) so in agarose gels (Figures 7-9 and 11-13) both products appear as the same band. A schematic of the mRNA in which only intron 2 has been removed is used to represent both products.

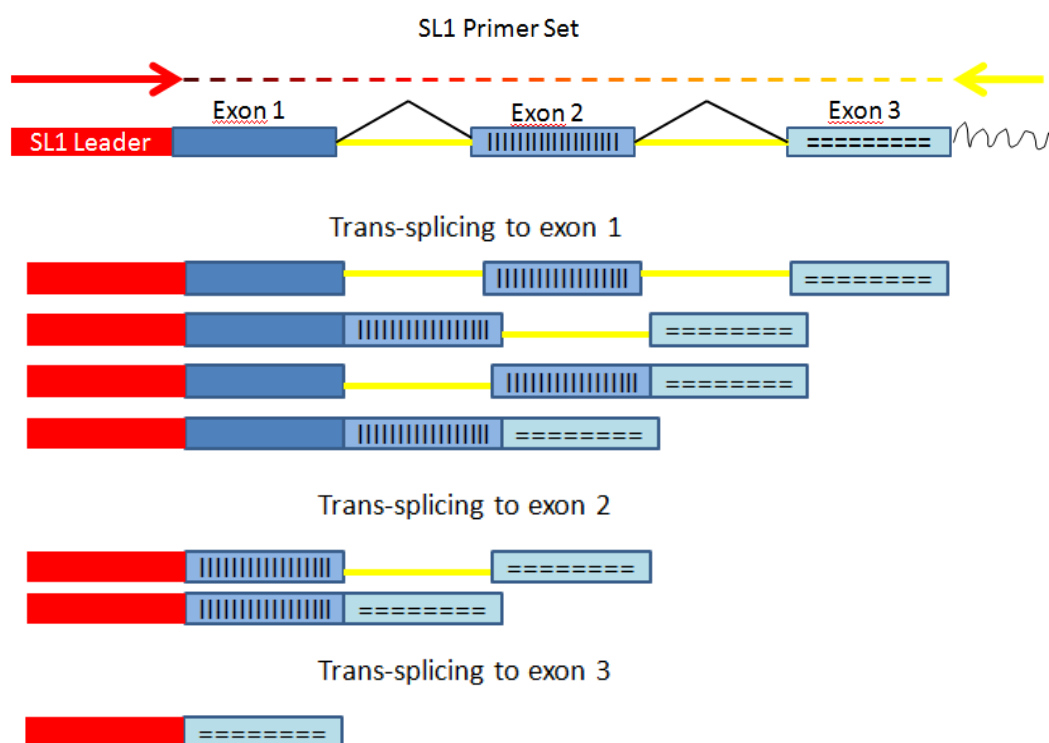


Figure 6: *In vitro* trans-spliced products of *nduf2.2*

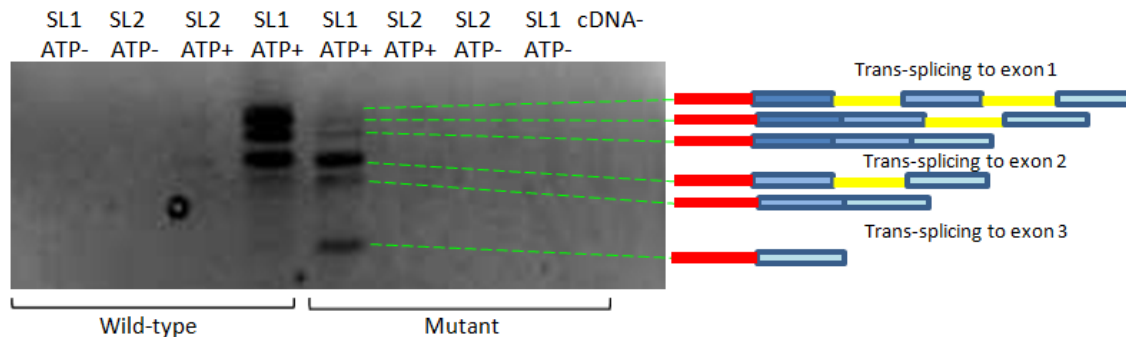


Figure 7. Mutating the Ou element may reduce trans-splicing efficiency *in vitro*. First trans-splicing reaction comparing SL1 trans-spliced RNA in wild-type and mutant.

Figure 7 demonstrates the results of this first trans-splicing reaction comparing wild-type and mutant 1 substrates. As expected, all ATP minus lanes are blank indicating that ATP is required for splicing and the products observed in the SL1 ATP+ lanes have been spliced. The cDNA minus lane is also blank confirming there were no contaminants present in the PCR reactions. All SL2 lanes are also blank indicating that no splicing of SL2 to *nduf2.2* occurs *in vitro* regardless of the presence or absence of the Ou element.

In Figure 7 the only observable products are SL1 trans-spliced in the ATP positive reactions for both wild-type and mutant 1 constructs. Focusing first on the products in which SL1 trans-splicing to exon 1 occurred (the top three bands); these results clearly demonstrate that mutating the Ou element decreases trans-splicing efficiency *in vitro*. The intensity of the bands representing trans-splicing to exon 1 is greatly decreased in the mutant 1 reaction. As expected, trans-splicing to exon 2 (the middle two bands) does not seem to be affected by the mutation. Trans-splicing to exon 3 (the lower band in the second lane), however, is evident for mutant 1 but not for wild-type samples. These results suggest that mutating the Ou element considerably decreased trans-splicing efficiency at the expected SL1 trans-splice site and that exons may

compete for SL1 trans-splicing *in vitro*. The effect of the mutation on trans-splicing to the first exon supports my hypothesis that the Ou element plays an important role in SL1 trans-splicing.

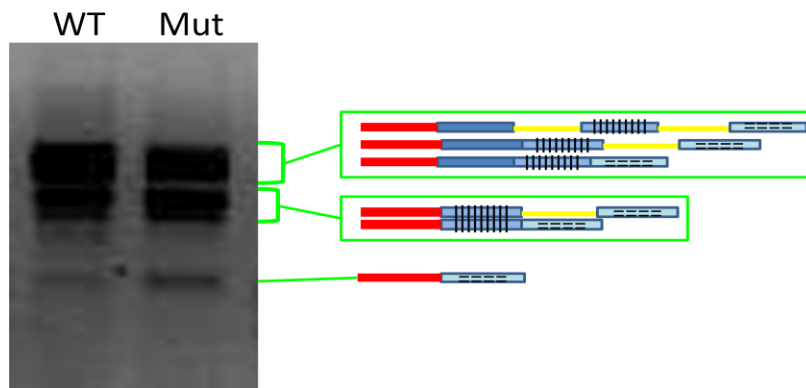


Figure 8. Second trans-splicing reaction comparing SL1 trans-spliced RNA of wild-type and mutant samples.

In a second trans-splicing reaction repeated exactly as before, SL1 trans-splicing levels appear to be equivalent at all exons for wild-type and mutant 1 samples (Figure 8). One possible explanation for this is that PCR amplification levels had reached a maximum value for both samples, thus rendering differences between the reactions indiscernible. It is possible that the second trans-splicing reaction was slightly more efficient than the first due to such factors as slightly different incubation or setup times/temperatures such that 35 cycles of amplification may have been enough to cause the concentration of products to reach the same level and then stop being amplified appearing equivalent regardless of the relative amounts of SL1 trans-spliced product in each sample.

To test this hypothesis, results from the same trans-splicing reaction were compared at varying cycles of amplification. Figures 9 and 10 confirm this hypothesis by demonstrating that mutant 1 and wild-type samples differ more drastically in observable trans-spliced product when the PCR cycle number is reduced. cDNA samples from the second trans-splicing reaction were re-amplified at 30, 31, and 32 cycles. For these samples, differences in trans-splicing efficiency

at the trans-splice site between wild-type and mutant 1 samples are evident for all, but most obvious at 30 cycles of amplification. This difference is more easily observed in Figure 10. This graph represents trans-splicing to exon 1 in mutant 1 samples as a percentage of trans-splicing to exon 1 in wild-type samples. Figures 9 and 10 show data indicating that the mutant 1 and wild-type constructs in the second splicing reaction are not equal in SL1 trans-splicing efficiency. In fact, results from the second trans-splicing reaction agree with those of the first trans-splicing reaction (Figure 7). SL1 trans-splicing levels are considerably decreased when the Ou element is mutated. This once again suggests that the Ou element is important for SL1 trans-splicing efficiency.

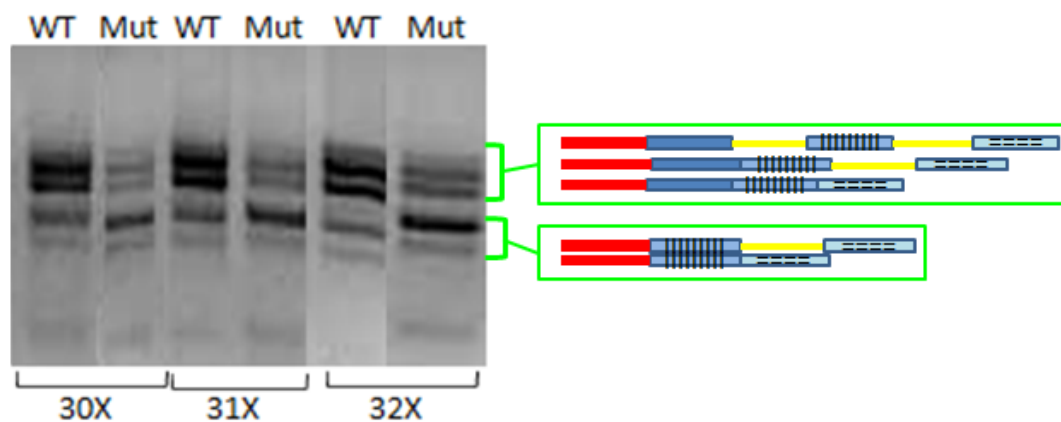


Figure 9. SL1 trans-spliced RNA from the second trans-splicing reaction amplified at a varying number of cycles.

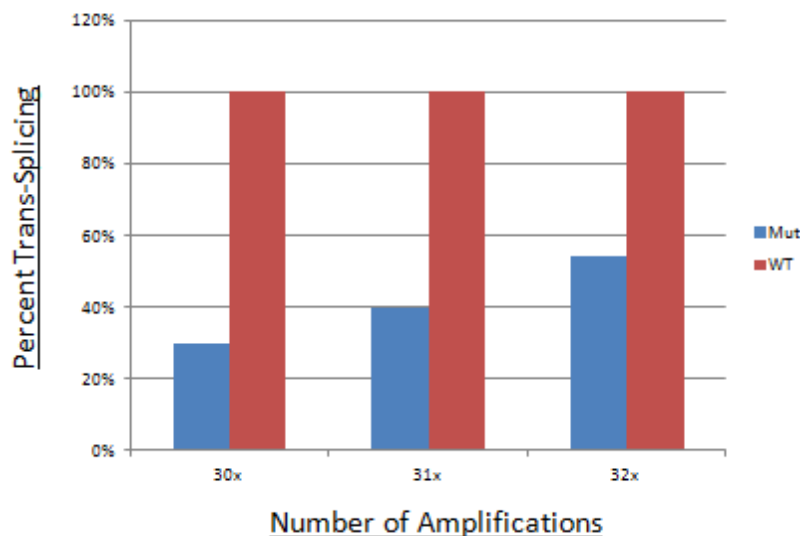


Figure 10. Graphical representation comparing levels of SL1 trans-splicing to exon 1 using data from Figure 9. The intensity of the bands representing trans-spliced RNA is expressed as a relative percentage with wild-type trans-splicing levels set at 100%.

The results from the first two splicing reactions clearly demonstrate that when mutant 1 and wild-type constructs are compared, mutating the Ou element greatly decreases SL1 trans-splicing at the trans-splice site. Unlike the first reaction, however, the data shown in Figure 9 suggests that levels of trans-splicing to exon 2 (demonstrated by the lower two bands) may actually be higher in mutant 1 than wild-type samples. It is possible that trans-splice sites compete *in vitro* and, when trans-splicing to the exon 1 is decreased in mutant 1 samples, trans-splicing to exon 2 (and/or 3) increases. An alternative explanation, however, is that exon 2 trans-spliced product is being preferentially amplified in the absence of exon 1 trans-spliced product in mutant 1 reactions.

To differentiate between these hypotheses, I designed a PCR primer that selects for and only amplifies product in which trans-splicing occurred to the second exon. This primer, called “2E2” (for “to exon 2”), contains the SL1 leader sequence plus the first several bases of the second exon. A second PCR primer was created to examine base levels of *nduf2.2* construct

cDNA present in the samples. This forward primer lies within the second exon and is designated “IC” (for internal control). The IC primer amplifies all *nduf2.2* cDNA present in the sample, thus affirming that the same amount of total sample is present in both mutant and wild-type reactions. As with the SL1 primer set, both 2E2 and IC primers were used in conjunction with a reverse primer downstream of the gene *nduf2.2* in the pBluescript plasmid. Schematic diagrams of the products amplified by the 2E2 and IC primers are included in Figure 11.

Figure 11 shows results from the third trans-splicing reaction comparing wild-type and mutant 1 samples amplified by the SL1, 2E2, and IC primer sets. According to these results, not only is trans-splicing to exon 1 decreased when the Ou element is mutated (first four lanes labeled SL1), but trans-splicing to exon 2 may also be affected (lanes labeled 2E2). The last two lanes (labeled IC) in Figure 11 confirm that base levels of *nduf2.2* cDNA present in mutant 1 versus wild-type samples are equivalent by demonstrating that an internal control independent of splicing appears at the same intensity in mutant and wild-type samples. This control is significant as it demonstrates that decreased levels of trans-splicing in mutant 1 samples are not attributable to a decrease in total RNA present in the sample. In conclusion, differences in trans-splicing to exon 1 between wild-type and mutant 1 constructs are attributable to the mutation in the Ou element. Further, mutating the Ou element may also affect trans-splicing to exon 2 *in vitro*. Results regarding splicing to exon 2, however, are inconsistent among the three splicing reactions, making it difficult to draw conclusions.

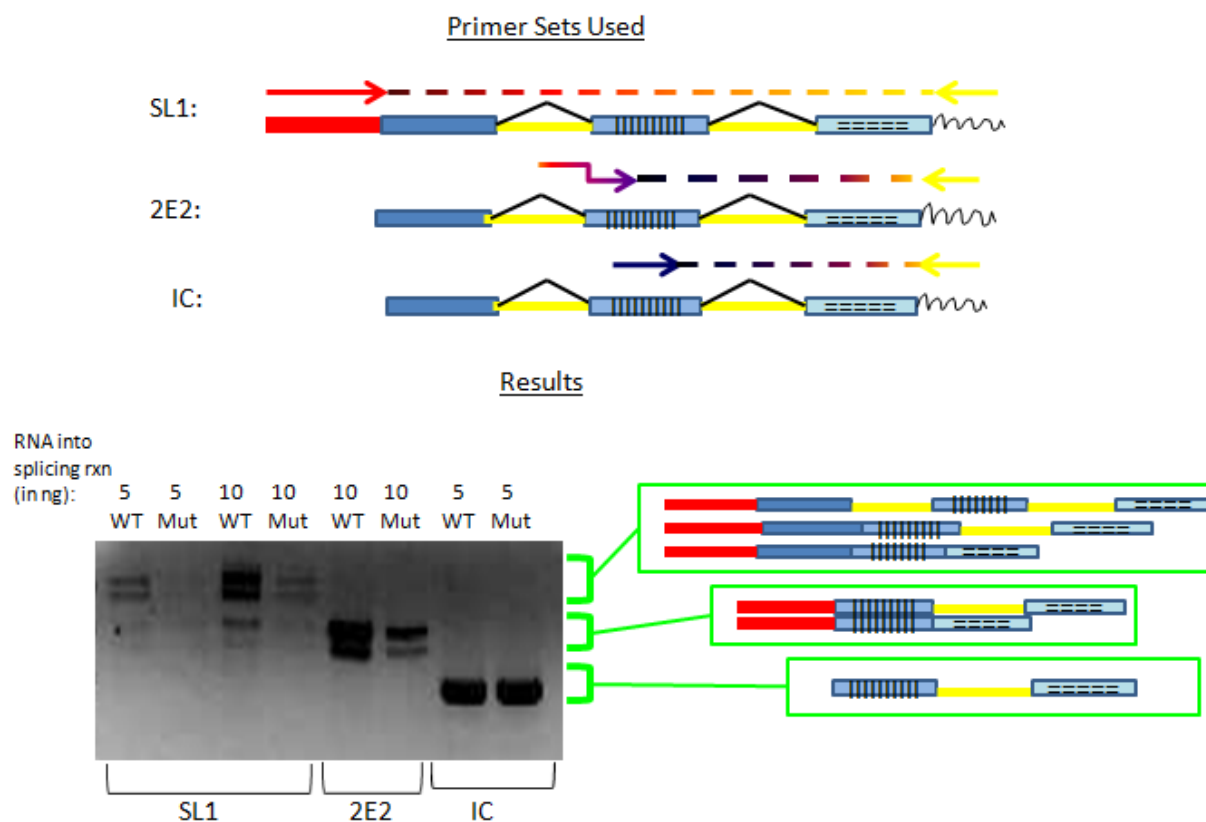


Figure 11. Third trans-splicing reaction comparing SL1 trans-spliced RNA in wild-type and mutants confirms that mutating the Ou element decreases trans-splicing to exon 1 and indicates it may also affect trans-splicing to exon 2.

There are two possible explanations for the decrease in SL1 trans-splicing observed in mutant 1 samples. The mutation either disrupted an important sequence (presumably the Ou element), or introduced an inhibitory sequence (ACAAUU) that interferes with the process of trans-splicing by an unknown mechanism. In order to distinguish between these possibilities, I designed a second mutant that also disrupted the Ou consensus but has the sequence GUGGUU (see Figure 2 for summary of wild-type and mutant constructs). As with mutant 1, mutant 2 was transcribed spliced *in vitro*. I then performed an RT-PCR of mutant 2 and wild-type reactions, amplifying samples with SL1, 2E2, and IC primers. Two different splicing reactions were assayed comparing SL1 trans-splicing efficiency for mutant 2 and wild-type constructs. The

results from both are consistent. Mutant 2 also decreased trans-splicing efficiency supporting the hypothesis that the Ou element is necessary for trans-splicing. The results for mutant 2 reactions are shown in Figures 12 and 13.

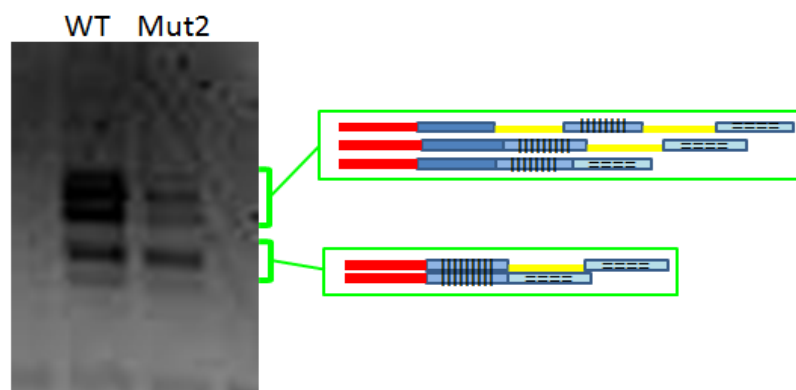


Figure 12. Mutant 2 also decreases SL1 trans-splicing. Results from first trans-splicing reaction comparing SL1 trans-spliced RNA in the wild-type and mutant 2.

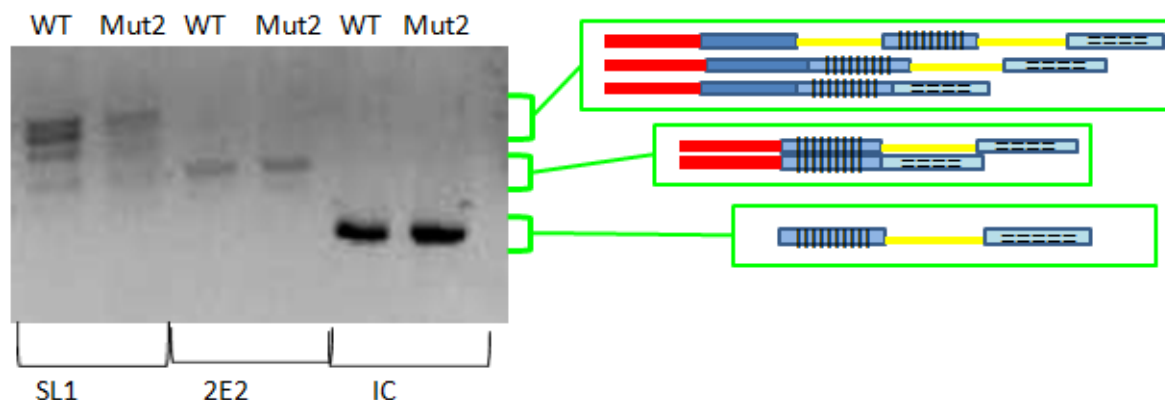


Figure 13. Results from a second trans-splicing reaction comparing SL1 trans-spliced RNA in wild-type and mutant 2 constructs including 2E2 and IC primed samples.

Figure 13 confirms that a second mutation in the Ou element of *nduf2.2* also decreased SL1 trans-splicing. According to 2E2 PCR reactions, SL1 trans-splicing to exon 2 is comparable for wild-type and mutant 2 samples. Results regarding trans-splicing to exon 2 are inconsistent and therefore inconclusive. Results comparing trans-splicing to the trans-splice site in wild-type, mutant 1, and mutant 2 samples, however, are consistent. Mutating the Ou element decreases

trans-splicing efficiency *in vitro*. These results indicate that the predicted Ou element was correctly identified and that it is important for SL1 trans-splicing.

Discussion

To test the relevance of the predicted Ou element for SL1 trans-splicing *in vitro*, I introduced mutations to the proposed Ou consensus in an *nduf2.2* construct. I then subjected the mutant and wild-type samples to an *in vitro* trans-splicing assay with embryonic extract capable of splicing. A cDNA copy was made of the spliced RNA samples and those containing the SL1 leader sequence were selected for and amplified. The amount of trans-spliced product for mutant versus wild-type samples was compared. After this initial examination I designed a second mutant with the similar goal of testing the importance of the predicted Ou element in SL1 trans-splicing. The purpose of the second mutant is to ensure that the absence of the Ou element is the variable affecting trans-splicing, and not the introduction of a new sequence that is somehow deleterious. Control PCR reactions were included to demonstrate that the same amount of DNA was present in both mutant and wild-type samples and that cis-splicing levels were consistent from wild-type to mutant, thus allowing any differences observed to be attributed to the effect of the mutations on trans-splicing efficiency.

Results comparing SL1 trans-splicing of mutant 1 and wild-type samples were consistent. In three different trans-splicing reactions, mutant samples exhibited considerably less product in which the SL1 leader was trans-spliced onto the first exon of *nduf2.2* (Figures 7, 9, and 11). These results suggest that *in vitro* trans-splicing to the *nduf2.2* trans-splice site was considerably less successful for *nduf2.2* constructs lacking an Ou consensus sequence than those with the

proposed Ou element still intact. It must also be taken into consideration, however, whether or not the mutation introduced a sequence that is inhibitory to SL1 trans-splicing. To account for this, the same procedures used to test mutant 1 were repeated for a second mutant, mutant 2. Once again in two different trans-splicing reactions, trans-splicing to exon 1 was decreased in mutant 2 samples as compared to wild-type (Figures 12 and 13). These data support the hypothesis that the decreases in trans-splicing efficiency observed *in vitro* are attributable to the disruption of the proposed Ou consensus sequence. In conclusion, various samples from five different splicing reactions comparing SL1 trans-spliced products in mutant versus wild-type samples all agree. Mutating the sequence UGUUUU in the gene *nduf2.2* significantly reduces SL1 trans-splicing efficiency *in vitro*.

It is important to note that trans-splicing was decreased in mutant reactions, but never completely knocked down. One reason for this might be the redundancy of the Ou element. As previously mentioned, in addition to the proposed Ou element (UGUUUU), there were several 1-3 base mismatches to the URYUY consensus overlapping or nearby. Lasda et al. (2010) noticed similar patterns in redundancy among Ur elements in the intercistronic region (ICR) of SL2 trans-spliced genes. When all copies of the UR consensus were mutated, however, SL2 trans-splicing was almost entirely lost. It is possible that if a similar experiment was repeated for the Ou element, results would indicate that it too acts redundantly.

In vivo experiments could also be conducted to test the necessity of the Ou element in SL1 trans-splicing. Such a project would involve creating transgenic *C. elegans* strains in which the Ou element was mutated. Huang et al. (2001) used such techniques to demonstrate the necessity of the Ur element in SL2 trans-splicing *in vivo*. They introduced linker scan (LS) mutations to the ICR between genes in an operon. They found that for LS mutants in which the

Ur element (which was previously unidentified) was disrupted, accumulation of mRNA from genes downstream in the operon was significantly reduced. They also found that when the Ur element was mutated and 3' cleavage of the upstream gene was prevented, SL1-trans-splicing increased. It is now known that the UR element includes the sequence UAYYUU as well as a stem-loop just upstream (Lasda et al., 2010). When cleavage of the 3' end of the upstream gene occurs, a 5' phosphate in the ICR is exposed. In the absence of a Ur element, this uncapped mRNA is rapidly degraded by XRN2 exonuclease until transcription is halted. When the Ur element is intact, however, the stem loop may prevent degradation, allowing SL2 trans-splicing to occur. This schematic is demonstrated in Figure 14.

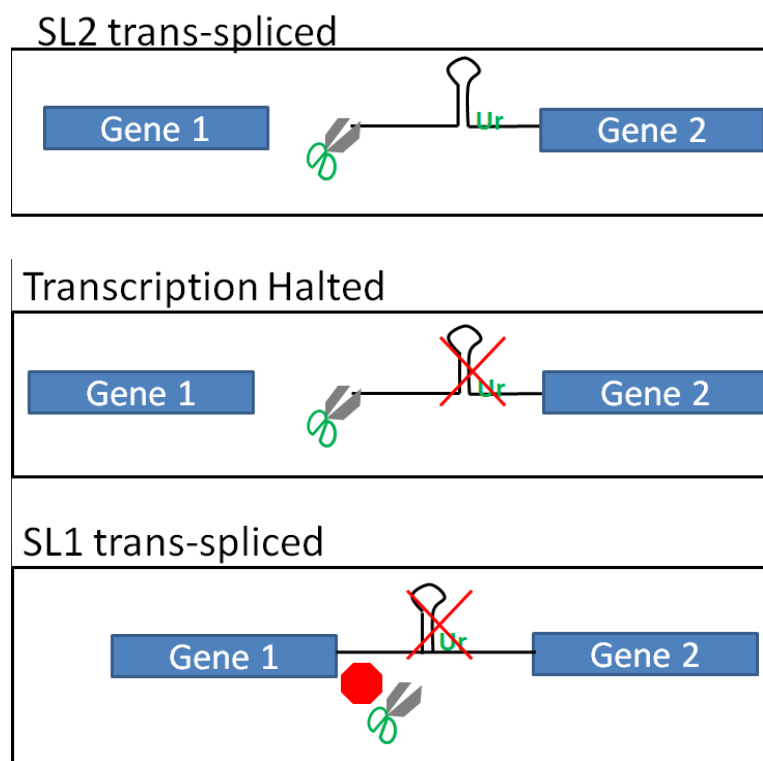


Figure 14. Model for the relationship between the Ur element, SL1, and SL2 trans-splicing

Unlike ICR's, outrons have not been shown to assume any particular secondary structure. In fact, UA rich intron-like sequence in which the 5' splice site has been removed is sufficient for SL1 trans-splicing *in vivo* (Conrad et al., 1991 & 1993). It may be that in the absence of the stem-loop structure the ICR acts like an intron. Interestingly, the Ur element (UAYYUU) also satisfies the Ou consensus (URYYUY). Thus SL1 trans-splicing may act as a default in the absence of the stem-loop structure.

Prior to this study, the sequence of the Ou element had been loosely defined. The hypothesized sequence URYYUY was derived for its ability to base pair with the 5' splice site on the SL1 spliced leader. A bioinformatic analysis by Graber et al. (2007) confirmed the existence of an outtron-specific UC-rich sequence. Previously, however, no work had been done to experimentally test the significance of the sequence. The results presented here on trans-splicing to exon 1 of *nduf2.2* suggest that the Ou element for this gene has been correctly identified and that it is in fact necessary for efficient SL1 trans-splicing *in vitro*.

In addition to providing evidence for the role of the Ou element in SL1 trans-splicing, this study raises interesting questions about the trans-splicing dynamics observed in the *in vitro* splicing system used. The results achieved by using this splicing system agree with those reported by Lasda et al. (2010) who first reported the successful use of *C. elegans* embryonic extract to splice pre-mRNA *in vitro*. Splicing specificity seen *in vivo* was recapitulated in my *in vitro* analysis. The wild-type construct of the gene *nduf2.2* was spliced exclusively to SL1 leader sequence. Furthermore, it was found that even when SL1 trans-spliced product decreased (in the case of the mutant constructs), no SL2 spliced product was apparent. This means that SL2 splicing is not compensating for the decrease in SL1 trans-splicing achieved by mutating the Ou

element *in vitro*. As previously described, this may be due to the lack of the stem-loop structure required for SL2 trans-splicing.

Interestingly, there was an unexpected fluctuation in trans-splicing to exons 2 and 3 when the Ou element was mutated. Lasda et. al. (2010) observed similar cases of trans-splicing to cis-splice sites in the *in vitro* system, but did not specifically isolate these products. In the first trans-splicing reaction, trans-splicing to exon 3 was apparent for mutant 1 but not wild-type constructs (Figure 7). One possible explanation suggests that exons may be competing for trans-splicing in this *in vitro* system. *In vivo*, *nduf2.2* is SL1 trans-spliced exclusively at the first exon. The *in vitro* system used, however, may have trouble distinguishing between the actual trans-splice site and the intronic cis-splice sites since splicing is occurring post-transcriptionally. Interestingly, there are sequences that satisfy the Ou consensus in both the first and second intron (sequences TACTTT and TATTTC respectively). Especially when the Ou element in the outtron is mutated, the *in vitro* trans-splicing machinery may be recognizing the downstream Ou consensus sequences by mistake; however, these sequences are very near the 3' splice site while the Ou consensus is typically found approximately 50 bp upstream of the splice site (Graber et al. 2007).

In the first trans-splicing reaction, trans-splicing to exon 2 appears to be occurring at equal frequencies for wild-type and mutant 1 samples (Figure 7). The second trans-splicing reaction, on the other hand, demonstrates an increase in trans-splicing to exon 2 in mutant 1 samples (Figures 8 and 9). Furthermore, the first trans-splicing reaction of mutant 2 suggests that trans-splicing to exon 2 is decreased for mutant samples (Figure 12). These inconsistencies regarding levels of SL1 trans-splicing to exon 2 make it difficult to draw conclusions.

I designed the 2E2 primer to test the hypothesis that trans-splicing to the second exon was not occurring at greater or equal levels in the mutant, but that these cDNA species were

being preferentially amplified in mutant samples in the absence of cDNA in which trans-splicing to exon 1 occurred. Results from the third trans-splicing reaction were analyzed to compare trans-splicing to exon 2 exclusively between wild-type and mutant 1 samples. The 2E2 PCR reactions, revealed that there was actually more product in which the SL1 leader sequence was trans-spliced to exon 2 for wild-type than for mutant samples (Figure 11). It is therefore possible, that trans-splicing to exon 2 is also reduced in mutant samples but to a lesser degree than trans-splicing to exon 1. One possible explanation for this result is that trans-splicing to exon 2 is also affected by the presence versus absence of the Ou element. This explanation is plausible as the Ou element is only 223 bases from the start of exon 2.

When mutant 2 was similarly analyzed with the 2E2 primer, however, trans-splicing to exon 2 was approximately equivalent for wild-type and mutant 2 samples (Figure 13). Thus, even when trans-splicing to exon 2 was specifically isolated results were inconsistent. These conflicting data make it impossible to determine from these results whether trans-splicing to the second exon seen *in vitro* is affected by the presence or absence of the Ou element. Nonetheless, these data raise interesting questions regarding the range of influence of the Ou element. At what position upstream of the trans-splice site is the Ou element most effective at mediating trans-splicing? In order to answer this question, *in vitro* trans-splicing reactions in which the Ou element occupied various different locations could be compared.

Further research to investigate the correlation between the Ou element and SL1 trans-splicing might also include exploration of additional mutants. The results presented here suggest that the Ou element is necessary for SL1 trans-splicing, but say nothing of its sufficiency. If the Ou element is also sufficient in directing trans-splicing, then a different sequence from the wild-type that still satisfies the Ou consensus sequence would be predicted to restore SL1 trans-

splicing efficiency to the same level as wild-type. Additionally a synthetic outtron containing and lacking an Ou element should demonstrate stark differences in levels of SL1 trans-splicing.

To identify the specific function of the Ou element, analogous experiments with the SL1 snRNP are necessary. If the Ou element acts by base pairing with the 5' splice site on the SL1 RNA as predicted, then mutating the sequence complementary to the Ou (GAGGUA) should exhibit similar decreases in SL1 trans-splicing efficiency. Likewise, introducing a compensatory mutation to the Ou element of the mRNA to be trans-spliced should restore trans-splicing levels. Although the results of this study have greatly advanced current understanding of the requirements for SL1 trans-splicing *in vitro*, various other such experiments are necessary to reveal more about the direct role of the Ou element.

References

- Baugh, L., Demodena, J., & Sternberg, P. (2009). RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science*, 324: 92-94.
- Blumenthal, T. (2005). Trans-splicing and operons. *WormBook*, The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.5.1, <http://www.wormbook.org>
- Blumenthal, T., & Spieth, J. (1996). Gene structure and organization in *Caenorhabditis elegans*. *Current Opinion in Genetics and Development*, 6:692-698.
- Conrad, R., Liou, R., & Blumenthal, T. (1993). Conversion of a trans-spliced *C. elegans* gene into a conventional gene by introduction of a splice donor site. *The European Molecular Biology Organization Journal*, 12(3), 1249-1255.
- Conrad, R., Liou, R., & Blumenthal, T. (1993). Functional analysis of a *C. elegans* trans-splice acceptor. *Nucleic Acids Research*, 21(4), 913-919.
- Conrad, R., Liou, R., 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. *Molecular and Cellular Biology*, 11(4), 1921-1926.
- Graber, J. H., Sallsbury, J., Hutchins, L. N., & Blumenthal, T. (2007). *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing. *RNA*, 13(9):1409-1426.
- Huang, T., Kuersten, S., Deshpande, A.M., Spieth, J., MacMorris, M., & Blumenthal, T. (2001). Intercistronic region required for polycistronic pre-mRNA processing in *Caenorhabditis elegans*. *Molecular and Cellular Biology*, 21: 1111-1120.
- Kent, W., & Zahler, A. (2000). Conservation, regulation, synteny, and introns in a largescale *C. briggsae*-*C. elegans* genomic alignment. *Genome Research*, 10:1115-1125.
- Krause, M., & Hirsh, D. (1987). A Trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell Press*, 49: 753-761.
- Lall, S., Friedman, C., Jankowski-Anyszka, M., Stepinski, J., Darzynkiewicz, E., & Davis, R. (2004). Contribution of trans-splicing, 5'-leader length, cap poly(A) synergism, and initiation factors to nematode translation in an *Ascaris suum* embryo cell-free system. *Journal of Biological Chemistry*, 279: 45573-45585
- Lasda, E., Allen, M., & Blumenthal, T. (2010). Polycistronic pre-mrna processing *in vitro*: snRNP and pre-mRNA. *Genes and Development*, 24(15): 1645-1658.
- Li, H., Wang, J., Mor, G., & Sklar, J. (2008). A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science*, 321:1357-1361.
- MacMorris, M., Kumar, M., Lasda, E., Larsen, A., Kraemer, B., & Blumenthal, T. (2007). A novel family of *C. elegans* snRNPs contains proteins associated with trans-splicing. *RNA*, 13(4):511-520.
- Mingxue, C., Allen, M. A., Larsen, A., MacMorris, M., Han, M., & Blumenthal, T. (2008). Genes involved in pre-mRNA 3'-end formation and transcription termination revealed by a *lin-15* operon Muv suppressor screen. *Proceedings of the National Academy of Sciences of the United States of America*, 105(43): 16665-16670.

- Morton, J., & Blumenthal, T. (2010). Identification of transcription start sites of trans-spliced genes: Uncovering unusual operon arrangements. *RNA*, doi: 10.1261/rna.2447111
- Nilsen, T. (1993). Trans-splicing of nematode premessenger RNA. *Annual Review of Microbiology* , 47: 413-440.
- Speith, 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 Press*, 72:521-532.
- Ucsc genome bioinformatics*. Retrieved from <http://genome.ucsc.edu/>
- Wormbase the biology and genome of C. elegans*. Retrieved from <http://www.wormbase.org/db/gene/gene?name=WBGene00006463;class=Gene>