Trans-splicing and operons in *C. elegans**

Thomas Blumenthal[§], Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309 USA

Table of Contents

1. Trans-splicing 2
2. Trans-splicing precursors 2
3. Mechanism of trans-splicing 3
4. The role of SL snRNP proteins 3
5. Evolution and role of trans-splicing 4
6. Summary of operons 4
7. Signals on the polycistronic pre-mRNA for SL2 trans-splicing 5
8. SL1-type and alternative operons 6
9. Function of operons 6
10. Evolution of operons in the nematodes 7
11. Acknowledgements 8
12. References 8

*Edited by Jane E. Mendel. Last revised June 15, 2012. Published November 20, 2012. This chapter should be cited as: Blumenthal T., Trans-splicing and operons in *C. elegans* (November 20, 2012), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.5.2, http://www.wormbook.org.

§To whom correspondence should be addressed. E-mail: Tom.Blumenthal@Colorado.EDU

Abstract

About 70% of *C. elegans* mRNAs are trans-spliced to one of two 22 nucleotide spliced leaders. SL1 is used to trim off the 5' ends of pre-mRNAs and replace them with the SL1 sequence. This processing event is very closely related to cis-splicing, or intron removal. The SL1 sequence is donated by a 100 nt small nuclear ribonucleoprotein particle (snRNP), the SL1 snRNP. This snRNP is structurally and functionally similar to the U snRNAs (U1, U2, U4, U5 and U6) that play key roles in intron removal and trans-splicing, except that the SL1 snRNP is consumed in the process. More than half of *C. elegans* pre-mRNAs are subject to SL1 trans-splicing, whereas ~30% are not trans-spliced. The remaining genes are trans-spliced by SL2, which is donated by a similar snRNP, the SL2 snRNP. SL2 recipients are all downstream genes in closely spaced gene clusters similar to bacterial operons. They are transcribed from a promoter at the 5' end of the cluster of between 2 and 8 genes. This transcription makes a polycistronic pre-mRNA that is co-transcriptionally processed by cleavage and polyadenylation at the 3' end of each gene, and this event is closely coupled to the SL2 trans-splicing event that occurs only ~100 nt further downstream. SL2 trans-splicing requires a sequence between the genes, the Ur element, that likely base pairs with the 5' splice site on the SL2 snRNP, in a manner analogous to the interaction between the 5' splice site in cis-splicing with the U1 snRNP. The key difference is that in trans-splicing, the snRNP contains the 5' splice site, whereas in cis-splicing the pre-mRNA does. Some operons, termed "hybrid operons", contain an additional promoter between two genes that can express the downstream gene or genes with a developmental profile that is different from that of the entire operon. The operons contain primarily genes required for rapid growth, including genes whose products are needed for mitochondrial function and the basic machinery of gene expression. Recent evidence suggests that RNA polymerase is poised at the promoters of growth genes, and operons allow more efficient recovery from growth-arrested states, resulting in reduction in the need for this cache of inactive RNA polymerase.

1. Trans-splicing

mRNAs of ~70% of *Caenorhabditis elegans* genes begin with a 22 nucleotide sequence, the spliced leader or SL, which is not associated with the gene (reviewed in Lasda and Blumenthal, 2011; Blumenthal and Steward; 1997 and Hastings, 2005). The SL is donated by a ~100 nucleotide RNA, SL RNA, by trans-splicing. This process is closely related to cis-splicing (intron removal): the 5' splice site is on the SL RNA, and the site of SL addition, the trans-splice site is the 3' splice site on the pre-mRNA (Figure 1). The reaction proceeds by way of a branched intermediate similar to the lariat of cis-splicing. Trans-splicing is catalyzed by spliceosomes, including U2, U4, U5, and U6 snRNPs but not U1 (Hannon et al., 1991). A similar reaction occurs throughout the nematode phylum as well as in some protists (e.g., trypanosomes) and many other animals including flatworms, hydra and primitive chordates (Douris et al., 2010; Agabian, 1990; Davis, 1997; Ganot et al., 2004; Stover and Steele, 2001; Vandenberghe et al., 2001). SL1 is the major spliced leader in nematodes, and it is used primarily for trans-splicing at the 3' splice sites following outrons, sequences resembling introns, but at the very 5' ends of pre-mRNAs (Conrad et al., 1991). About half the genes are estimated to have outrons and are consequently trans-spliced to SL1 (Allen et al., 2011; Zorio et al., 1994). Furthermore, a second SL, SL2, is trans-spliced to some *C. elegans* genes (Huang and Hirsh, 1989). SL2 trans-splices at trans-splice sites between genes in polycistronic pre-mRNAs from operons (Blumenthal et al., 2002; Spieth et al., 1993; Allen et al., 2011). About 15% of all *C. elegans* genes are found in operons (Allen et al., 2011). Finally, about 30% of genes specify mRNAs that are not subject to trans-splicing. In these cases, the promoter is at the 5' end of the first exon like in genes of organisms that do not trans-splice.

2. Trans-splicing precursors

The SL RNAs exist as snRNPs (Blumenthal and Steward, 1997; Hastings, 2005). They have a discrete secondary structure as do other snRNAs, they are bound to the Sm proteins, and they have a trimethylguanosine (TMG) cap like the U snRNAs. In the SL snRNPs the 5' splice sites are base paired to the upstream part of the SL sequence, resembling the U1-5' splice site base pairing. The trans-splice site consensus on the pre-mRNAs is the same as the intron 3' splice site consensus. The signal for trans-splicing is the presence of intron-like sequence, the outron, at the 5' end of the mRNA, with no functional 5' splice site upstream (Conrad et al., 1995; Conrad et al., 1993; Conrad et al., 1991). The outron contains a sequence called the Ou element that is theoretically capable of base pairing with the 5' splice site on the SL snRNP (Graber et al., 2007). Genes whose pre-mRNAs are subject to trans-splicing are distinguished from those that are not only by the presence of an outron. Thus, the choice between cis- and trans- splicing at any 3' splice site is based largely on the presence or absence of an upstream 5' splice site. Because trans-splicing is relatively efficient (like cis-splicing), it is difficult to isolate outron-containing

pre-mRNAs, so very few natural outrons have been defined. Nevertheless, the promoter locations of some trans-spliced genes have been identified, and these outrons range from 60-500 bp, with many around 300 bp (Morton and Blumenthal, 2011). Deep sequencing of RNA that has not been polyA selected should allow determination of promoter locations of many more trans-spliced genes.

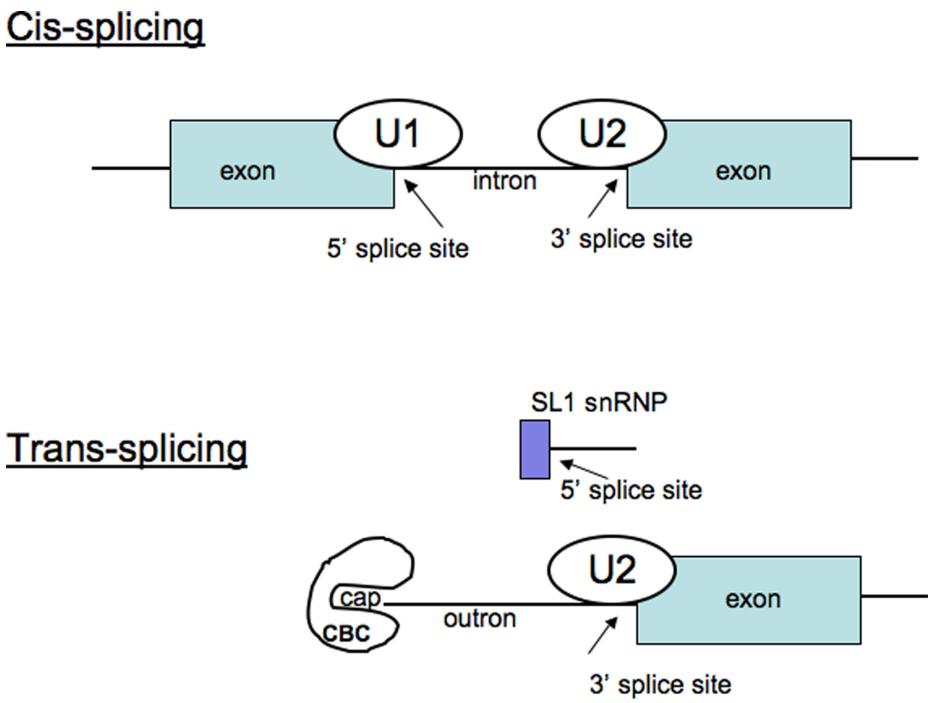

Figure 1. Comparison of cis- and trans-splicing. In cis-splicing, the U1 snRNP base pairs with the 5' splice site, and U2 snRNP base pairs with the branchpoint near the 3' splice site. The intron is excised and the two exons are spliced together. In trans-splicing there is no 5' splice site on the pre-mRNA for U1 snRNP binding. Instead, the 5' splice site is provided by the donor SL snRNP, which may base pair with a sequence in the outron, the region of the pre-mRNA between the 5' cap and the trans-splice site. The SL exon is then spliced to the first exon on the pre-mRNA. CBC: nuclear cap binding complex.

3. Mechanism of trans-splicing

The SL snRNP may be attracted to the outron by base pairing to the Ou element in a manner analogous to the base pairing between the SL2 snRNP 5' splice site/Ur element base pairing of SL2 trans-splicing (see below). Trans-splicing then follows a course similar to that of cis-splicing: in the first step, the phosphodiester bond at the 3' end of the spliced leader is transferred to an adenosine residue in the outron, resulting in the formation of a Y-branched intermediate containing the 3' splice site. This results in a free SL exon. In the second step, the SL exon is spliced to the first exon of the pre-mRNA, and the Y-branched intermediate, composed of the intron portion of the SL snRNP branched to the outron, is released and presumably recycled. It is worth noting that this intermediate presumably still contains the Sm proteins, which perform several important roles in addition to their roles in splicing (e.g. Barbee et al., 2002). It seems likely therefore that there is a mechanism for recycling these Sm proteins, since trans-splicing is a very frequent event. It has been proposed (MacMorris et al., 2007) that the proteins found associated with the SL1 snRNP may perform this role (see below).

4. The role of SL snRNP proteins

The SL snRNP was purified from *Ascaris*, a parasitic nematode, and found to contain a heterodimeric protein, neither subunit of which has known homologs outside of the nematodes (Denker et al., 2002). The larger subunit, SL-175p, is actually 95 kD, but runs anomalously large on SDS gels. In *C. elegans*, this polypeptide has a single homolog of 75 kD, SNA-2, that is required for viability and is associated with the SL1, but interestingly not the SL2, snRNP (MacMorris et al., 2007). The small *Ascaris* polypeptide, SL-30p, interacts with the branchpoint binding protein, an interaction that was shown to be important for trans-splicing (Denker et al., 2002). SL-30p has two orthologs in *C. elegans*: SNA-1, which is associated with the SL1 snRNP, and SUT-1, which is associated with

another class of snRNP first identified in Ascaris called the SmY snRNP. Even though SNA-1 and SUT-1 are associated with different classes of snRNA, they are functionally redundant: the single knockouts are viable and fertile, but the double knockout is lethal. This surprising result can be rationalized by the following hypothesis: all three of these proteins are required for recycling of the Sm proteins from the spent snRNPs, and successful recycling of Sm proteins from either SL1 snRNP or SL2 snRNP is required for viability, but failure to recycle Sm proteins from only one of the spent snRNP classes can be tolerated. Since SNA-2 is associated with both SL1 and SmY snRNPs, it makes sense that it could be required for recycling from both SL1 and SL2 spent snRNP. In contrast, SNA-1 is associated only with the SL1 snRNP, so it might be expected to be required for recycling from only the spent SL1 snRNP. Perhaps SUT-1 is required for recycling from the spent SL2 snRNP. In this idea, the SL2 recycling involves the SmY snRNP, which is theoretically capable of base pairing with the SL2 snRNP (MacMorris et al., 2007).

5. Evolution and role of trans-splicing

Trans-splicing occurs throughout the nematodes, and there is striking conservation of the SL sequence, whereas the portions of the SL RNAs downstream of the splice site have diverged (Blumenthal and Steward, 1997). However, recently very distantly related nematodes, *Trichinella* and *Prionchulus*, have been shown to use a variety of spliced leaders apparently unrelated to that of *C. elegans* (Pettit et al., 2008; Harrison et al, 2010). The role that SL plays in the cell, however, is not known. In *C. elegans* the SL tends to be spliced very close to the initiating methionine codon (often immediately adjacent), so it seems likely to play a role in translation initiation (Blumenthal and Steward, 1997; Lall et al., 2004). The TMG cap present at the 5’ end of the SL becomes the 5’ end of trans-spliced mRNAs. A TMG cap stimulates translation in nematodes, at least when it is present at the 5’ end of the SL sequence (Lall et al., 2004; Maroney et al., 1995). It has recently been shown that the very 5’ end of the SL forms a stem by intramolecular base pairing, and efficient translation of mRNA containing this stem requires the TMG cap (Wallace et al., 2010). In *C. elegans*, variants of the cap binding translation initiation factor, eIF4E can recognize the TMG cap (Keiper et al., 2000).

In *Ascaris*, the SL sequence in the DNA is needed for transcription of the SL RNA gene, which may be one reason why it has been so highly conserved (Hannon et al., 1990). Although the roles the SL sequence itself may perform are unknown, trans-splicing is in fact required for viability (Ferguson et al., 1996). Its required role could be a positive effect such as providing a sequence that can facilitate translation initiation, mRNA stability or localization, or it could be required for suppression of a negative effect such as inhibition of translation initiation by AUG codons in the outron.

The *C. elegans* genome contains 110 SL1 RNA genes on the 1 kb tandem repeat that also contains the genes for 5S rRNA (Krause and Hirsh, 1987). In contrast, the genome contains only18 dispersed SL2 RNA genes, which specify a variety of variant SL2 RNAs (Stein et al., 2003). Some of these have different SL2 sequences, and these have been given different names, such as SL3, SL4 etc. (Ross et al., 1995). Nonetheless, they are all variants of SL2 and they are used randomly at SL2-accepting trans-splice sites (Allen et al., 2011). The SL2 RNA genes are preceded by a conventional upstream promoter element, so the sequence of the SL2 leader is not constrained by its serving a promoter function as the SL1 sequence is (Thomas et al., 1990). The *Caenorhabditis briggsae* genome also contains 18 SL2 RNA genes, and all 36 genes from the two species descended from four primordial SL2 RNA genes present in their last common ancestor (Stein et al., 2003).

6. Summary of operons

The *gpd-3* gene and several other genes whose mRNAs receive SL2 were found to occur at downstream positions in closely-spaced clusters of genes in the same orientation (Blumenthal, 2004; Spieth et al., 1993; Zorio et al., 1994). Subsequent experimentation demonstrated that these clusters represented true operons: genes transcribed and regulated by a promoter upstream of the entire cluster (Spieth et al., 1993). Although first genes in some clusters are trans-spliced to SL1, others are not trans-spliced. It was shown that there is an overwhelming correlation between SL2 trans-splicing and presence of the trans-splice site at a downstream position in such a cluster (Blumenthal et al., 2002; Allen et al., 2011) (Figure 2). A microarray analysis of the entire genome and then deep sequencing of the transcriptome demonstrated how truly robust this correlation is. Operationally, the “operome”, the set of all operons, is now defined solely on the basis of SL2 trans-splicing. There are a few SL2-trans-spliced genes that do not appear to be downstream in a cluster. However, when some of these were investigated experimentally, it was found that they were indeed co-transcribed with the gene sometimes more than two kb upstream (Morton and Blumenthal, 2011). These instances are quite rare; the vast majority of operon genes are separated by only ~100 bp.

Furthermore, overall there is a strong negative correlation between intergenic distance and level of SL2 usage (Allen et al. 2011). With this operational definitition, the *C. elegans* operome is composed of 1255 operons containing 3193 genes, representing ~15 % of all coding genes. Half of these operons contain two genes, whereas the other half contain from 3 – 8 genes. They are found on all chromosomes, but they are rare on the X chromosome. The polycistronic pre-mRNA transcribed from operons is converted into monocistronic mRNAs by cleavage and polyadenylation at the 3' ends of the upstream genes, accompanied by SL2-specific trans-splicing at the 5' ends of the downstream genes.

Some operons, termed "hybrid operons" contain internal promoters capable of expressing the genes downstream, but without expressing the genes upstream of the internal promoter (Huang et al., 2007). This allows differential transcription of genes within the same operon, allowing the downstream genes to be expressed at a higher level or with a different developmental profile than the upstream genes. In general, these operons have longer intercistronic regions, presumably to accommodate the requirements of an independent promoter (Allen et al., 2011).

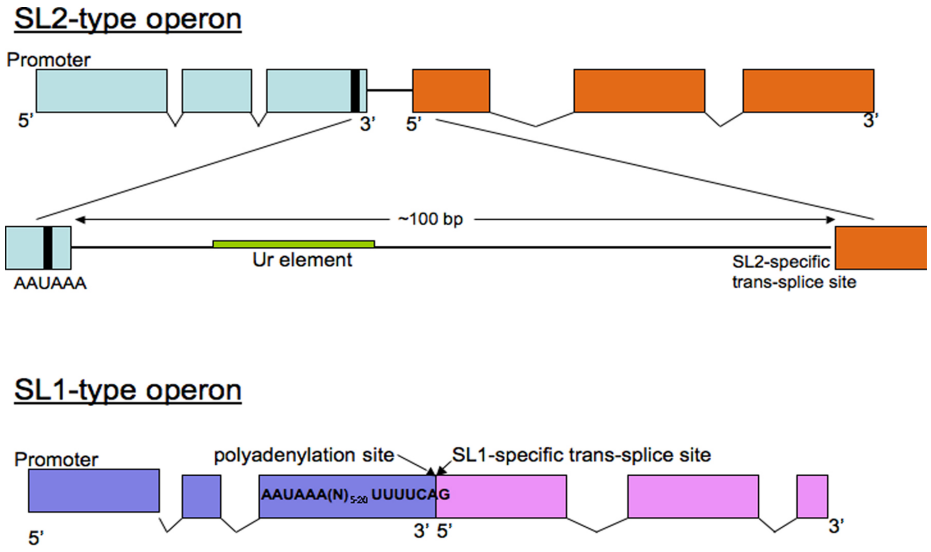

Figure 2. Two types of operons in *C. elegans*. Almost all *C. elegans* operons are of the SL2-type, shown on top. The site of 3' end cleavage and polyadenylation of the upstream gene is determined by the AAUAAA signal, indicated by the black bar. The Ur element serves to protect the downstream RNA from degradation following 3' end cleavage, as well as to attract the SL2 snRNP. Most operons of this type have about 100 bp between the cleavage and polyadenylation site of the upstream gene and the trans-splice site to which SL2 is spliced. A few *C. elegans* operons are of the SL1-type, shown on the bottom, where the site at which 3' cleavage and polyadenylation occurs is also the trans-splice site for the downstream gene, so the intercistronic distance is effectively zero. These operons are always trans-spliced by the SL1 snRNP.

7. Signals on the polycistronic pre-mRNA for SL2 trans-splicing

Besides the splice site itself, only two sequences on the pre-mRNA play an important role in SL2-specific trans-splicing: the signal for 3' end formation of the gene just upstream (Figure 2; Huang et al., 2001; Kuersten et al., 1997; Liu et al., 2001; Liu et al., 2003) and a sequence about half way between the two genes, termed the Ur element. 3' end cleavage depends on two sequences: the AAUAAA just 5' of the cleavage site that binds the cleavage and polyadenylation specificity factor (CPSF), and a U-rich sequence just 3' of the cleavage site that binds the cleavage stimulatory factor (CstF). CPSF and CstF bind cooperatively to these two sites and together position the site of cleavage. When the AAUAAA is mutated, 3' end cleavage fails to occur, and trans-splicing just downstream becomes less efficient and less specific for SL2. Nevertheless, AAUAAA is not required for SL2 trans-splicing since SL2 trans-splicing downstream still occurs in its absence. CPSF bound to AAUAAA may act by facilitating binding of CstF or by catalyzing 3' end formation, which in turn may play a role in SL2 trans-splicing. The Ur element, even further downstream, which is clearly not the CstF binding site (Graber et al., 2007), plays the major role in SL2 trans-splicing. When this sequence is mutated *in vivo*, 3' end formation still occurs, but downstream trans-spliced product fails to accumulate. Thus, either transcription terminates or the downstream RNA is degraded from the site of 3' end formation. When both 3' end formation and SL2 trans-splicing are prevented by mutating both the AAUAAA and the Ur element, downstream product is restored, but all of it is trans-spliced to SL1. The Ur element recruits the SL2 snRNP and also blocks exonucleolytic degradation of the downstream RNA

beginning at the site of 3' end cleavage (perhaps by binding the SL2 snRNP). The MS2 phage coat protein tethered to this region in place of the Ur element can also block exonucleolytic degradation of the precursor but it cannot support SL2 trans-splicing (Liu et al., 2003).

A recent report (Lasda et al., 2010) has provided significant insight into how the Ur element functions, by use of an in vitro splicing extract from *C. elegans* embryos. Several substrates were shown to trans-splice with SL1- or SL2- specificity that mimics their in vivo specificity. This allowed precise definition of the sequences required for SL2 specificity. Interestingly, in this system, only the Ur element was found to be required for SL2 trans-splicing. The Ur element was shown to be composed of a stem/loop, where its exact sequence was unimportant, followed a few bp later by the sequence element UAYYUU (where Y is any pyrimidine). Both the stem and the UAYYUU were shown to be required for SL2 specificity. Furthermore, addition of an RNA oligonuceotide identical to a Ur element from one operon was shown to both prevent SL2 trans-splicing at the normal trans-splice and to activate the SL2 snRNP for trans-splicing at inappropriate splice sites. Thus the SL2 snRNP appears to be in a state where it normally cannot splice unless activated by the Ur element just upstream of a trans-splice site. The UAYYUU sequence is capable of base pairing with the 5' splice site on the SL2 snRNP in a manner analogous to the base pairing of the 5' splice site in cis-splicing with a sequence near the 5' end of the U1 snRNP. Since the U1 snRNP plays no role in trans-splicing, it appears that interaction between the 5' splice site on the SL2 snRNP and its antisense sequence on the pre-mRNA (the UAYYUU of the Ur element) replaces the equivalent interaction of cis-splicing, but with the roles of snRNP and pre-mRNA reversed.

How does the the Ur element attract specifcally the SL2 snRNP? Antibodies to CstF have been found to immunoprecipitate the SL2 snRNP from *C. elegans* embryo extracts, and the region of SL2 RNA required for SL2 identity is also required for the CstF interaction, although it is not required for snRNP function *per se* (Evans and Blumenthal, 2000; Evans et al., 2001). The current model for polycistronic pre-mRNA processing involves first 3' end cleavage at the upstream gene by conventional mechanisms. This sets in motion the chain of events that leads to SL2 trans-splicing: the free 5' phosphate at the cleavage site is attacked by an exonuclease, presumably XRN-2, which attacks all RNA that follows 3' end cleavage (Luo et al., 2006). XRN-2 is stopped at the Ur element, which attracts by base pairing the SL2 snRNP in collaboration with the CstF needed for 3' end formation just upstream.

8. SL1-type and alternative operons

The *C. elegans* genome also contains a second type of operon, different in three significant ways from those described above (Figure 2; Williams et al., 1999). First, the mRNA of the downstream gene is trans-spliced to SL1, rather than SL2. Second, there is no intercistronic sequence; the site of polyadenylation of the upstream gene and the trans-splice site are at adjacent nucleotides. Third, the 3' UTRs of the upstream genes are characterized by very long pyrimidine-rich sequences. In the one case where the requirement for this pyrimidine-rich sequence has been investigated experimentally, it was found to be required for the RNA processing events. The possibility that the the free 3' end created by the trans-splicing is used for polyadenylation of the upstream mRNA is supported by the fact that polyA addition often or always occurs immediately following the AG of the trans-splice site of the downstream gene (T. Blumenthal, unpublished observations). On the other hand, when the trans-splice site was mutated, 3' end formation was nonetheless able to occur in the immediate vicinity, demonstrating that at least in the one operon investigated experimentally, 3' end formation does not absolutely depend on trans-splicing for cleavage. Only 23 operons of this type appear to be present in the genome (T. Blumenthal, unpublished).

There are also several cases of what has been termed "alternative operons" (Morton and Blumenthal, 2011; Jan et al., 2011). These are single genes that can be processed in either of two ways: as a single gene or as an operon. In these cases, 3' end formation accompanied by SL2 trans-splicing at a cis-splice site ~100 bp downstream can occur within what is otherwise an intron. This can result in either a full-length mRNA or two mRNAs representing the 5' and the 3' parts of the gene, where the 3' part is trans-spliced to SL2. The functional significance of these alternative operons is not yet clear, although some interesting genes including *smg-6*, *sma-9* and *pcf-11* (which encodes a key 3' end formation factor) are all alternative operons.

9. Function of operons

Do the *C. elegans* operons exist to assure coordinate regulation of genes whose products function together? There is no question that the genes in *C. elegans* operons are often co-expressed (Land et al., 1994), but in many cases they appear not to be (K. Seggerson-Gleason and T. Blumenthal, unpublished observations; Baugh et al., 2003). This could be because the mRNAs have very different stabilities or because processing is modulated. One reasonable idea is that the operons have accumulated genes that are regulated at some level after transcription

initiation. In this model the operons are transcribed from promiscuously expressed promoters because the genes in them are regulated at the level of RNA stability or translation. Consistent with this idea, genes that encode proteins that act in mRNA degradation are the most frequent class of genes to be contained in operons (80% are in operons; Blumenthal and Gleason, 2003). These may be autogenously regulated at the level at which they act. This idea is quite similar to gene regulation in trypanosomes, where all transcription is polycistronic and genes are regulated primarily or entirely post-transcriptionally (Clayton, 2002).

However, there are unquestionably examples among the operons where genes of related function are co-expressed due to their presence in an operon under the control of a single promoter. For instance, the two *lin-15* genes are contained in an operon; these two unrelated proteins collaborate in an aspect of signal transduction required for formation of the vulva (Clark et al., 1994; Huang et al., 1994). A second example is the *des-2/deg-3* operon, which encodes both subunits of the acetylcholine receptor channel (Treinin et al., 1998). Clearly, the co-expression of these two genes in the same transcription unit allows their products to be co-expressed. In another instance a protein function was hypothesized based on the two genes being present in an operon together: one gene was found to encode a modifier of the gating of an ion channel encoded by another gene in the same operon (Furst et al., 2002). There are also several examples of operons that encode both a basic transcription factor for RNA polymerase I, II or III, along with a subunit of that polymerase. So co-expression of genes of related function is clearly a feature of at least some of the operons (Blumenthal and Gleason, 2003).

Certain classes of genes are dramatically overrepresented in operons while other classes are missing or nearly so from the operon list (Blumenthal and Gleason, 2003). In general, tissue-specific genes are not transcribed in operons. The most frequently operon-included genes are those that encode mitochondrial proteins and the basic machinery of gene expression: transcription, splicing and translation. It has been reported that genes with high expression in the female germ line are dramatically over-represented in operons and that this pattern of expression may constitute the driving force for operon inclusion (Reinke and Cutter, 2009). A competing, but not nessarily exclusive, idea was recently suggested by Zaslaver and co-workers (Zaslaver et al., 2011). They suggest that the operons may exist to allow rapid co-induction of genes needed for recovery from growth-arrested states without requiring as much RNA polymerase as would be required if each of these genes had a poised RNA polymerase II molecule at its own promoter. *C. elegans* is in fact very sensitive to restriction of RNA polymerase levels under rapid growth conditions (Zaslaver et al., 2011). A possible rationalization of the two hypotheses for explaining selection for operon expression might be that there is a strong correlation between genes needed for rapid growth and for those expressed in the female germ line.

10. Evolution of operons in the nematodes

The *C. elegans* operons are probably not ancient but are instead an innovation, perhaps having evolved to handle situations where rapid growth is advantageous (Zaslaver et al., 2011). The only molecule known to be specific for operon pre-mRNA processing, SL2 RNA, appears to be evolving rapidly, consistent with the idea that it is a relatively recent innovation (MacMorris et al., 2007). Once formed, the operons appear to be relatively stable: most *C. elegans* operons are present in *C. briggsae*, a nematode that is ~50-100 million years diverged from *C. elegans* (Stein et al., 2003). Nonetheless, analysis of the operon content of several *C. elegans* relatives has demonstrated that both formation and loss of operons can be documented over this time scale (Qian and Zhang, 2008; Cutter et al., 2009; Cutter and Agrawal, 2010).

The SL2 snRNP has been found in several members of the rhabditid nematode group including *Caenorhabditis*, *Oscheius*, *Haemonchus* and *Pristionchus*, many of which have been shown to use SL2 to trans-splice downstream genes in operons (Evans et al., 1997; Lee and Sommer, 2003; Redmond and Knox, 2001; Laing et al., 2011). Distantly related nematodes may not contain a specialized snRNP for processing operons, but they appear to contain operons nonetheless (Whitton et al., 2004). It seems likely that operons developed early in nematode evolution because they had trans-splicing to process downstream mRNAs. The same snRNP could serve for either outron or operon processing. However, during nematode evolution, a new snRNP, SL2, evolved from SL1, specialized for processing operon pre-mRNAs more efficiently. It appears this was accomplished at least partly by evolving a site for CstF interaction (Evans et al., 2001). Once this snRNP evolved, it was able to process operons efficiently, so it provided the opportunity for further evolution of operons. These additional operons then provided the selective pressure to improve SL2 function in operon processing.

Evolution away from dependence on trans-splicing and operons should theoretically be difficult. Trans-splicing removes selective pressure contraints on regions upstream from the splice acceptor site, presumably allowing upstream AUG codons to accumulate in the region that is removed by trans-splicing in most mRNAs. If

trans-splicing did not occur, out of frame AUG codons would presumably interfere with translation of the correct open reading frame. Similarly, operons may be stable because downstream genes become dependent on the upstream gene's promoter. Many other distantly related animal phyla (e.g., flatworms (Davis and Hodgson, 1997) and primitive chordates (Ganot et al., 2004), also have both trans-splicing and operons, so the evolutionary scenario hypothesized above may have occurred several times in different animal lineages (Lasda and Blumenthal, 2011).

11. Acknowledgements

This work was supported by research grant R01 GM42432 from the National Institute of General Medical Sciences.

12. References

Agabian, N. (1990). Trans splicing of nuclear pre-mRNAs. Cell *61*, 1157–1160. Abstract Article

Allen, M.A., Hillier, L., Waterston, R.H. and Blumenthal, T. (2011). A global analysis of trans-splicing in *C. elegans*. Genome Res. *21*, 255-264. Abstract Article

Barbee, S.A., Lublin, A.L. and Evans, T.C. (2002). A novel function for the Sm proteins in germ granule localizaion during *C. elegans* embryogenesis. Curr. Biol. *12*, 1502-1506. Abstract Article

Baugh, L.R., Hill, A.A., Slonim, D.K., Brown, E.L., and Hunter, C.P. (2003). Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. Development *130*, 889–900. Abstract Article

Blumenthal, T. (2004). Operons in eukaryotes. Briefings in Funct. Genomics Proteomics *3*, 199–211. Abstract Article

Blumenthal, T., Evans, D., Link, C.D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W.L., Duke, K., Kiraly, M., and Kim, S.K. (2002). A global analysis of *Caenorhabditis elegans* operons. Nature *417*, 851–854. Abstract Article

Blumenthal, T., and Gleason, K.S. (2003). *Caenorhabditis elegans* operons: form and function. Nat. Rev. Genet. *4*, 112–120. Abstract Article

Blumenthal, T., and Steward, K. (1997). RNA processing and gene structure. In *C. elegans* II, D. L. Riddle, ed. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 117–145.

Clark, S.G., Lu, X., and Horvitz, H.R. (1994). The *Caenorhabditis elegans* locus *lin-15*, a negative regulator of a tyrosine kinase signaling pathway, encodes two different proteins. Genetics *137*, 987–997. Abstract

Clayton, C.E. (2002). Life without transcriptional control? From fly to man and back again. EMBO J. *21*, 1881–1888. Abstract Article

Conrad, R., Lea, K., and Blumenthal, T. (1995). SL1 trans-splicing specified by AU-rich synthetic RNA inserted at the 5’ end of *Caenorhabditis elegans* pre-mRNA. RNA *1*, 164–170. Abstract

Conrad, R., Liou, R.F., and Blumenthal, T. (1993). Conversion of a trans-spliced *C. elegans* gene into a conventional gene by introduction of a splice donor site. EMBO J. *12*, 1249–1255. Abstract

Conrad, R., Thomas, J., Spieth, J., and Blumenthal, T. (1991). Insertion of part of an intron into the 5’ untranslated region of a *Caenorhabditis elegans* gene converts it into a trans-spliced gene. Mol. Cell. Biol. *11*, 1921–1926. Abstract

Cutter, A.D., Dey, A. and Murray, R.L. (2009). Evolution of the *Caenorhabditis elegans* genome. Mol. Biol. Evol. *26*, 1199-1234. Abstract Article

Cutter, A.D. and Agrawal, A.F. (2010) The evolutionary dynamics of operon distributions in eukaryotic genomes. Genetics *185*, 685-693. Abstract Article

Davis, R.E. (1997). Surprising diversity and distribution of spliced leader RNAs in flatworms. Mol. Biochem. Parasitol. *87*, 29–48. Abstract Article

Davis, R.E., and Hodgson, S. (1997). Gene linkage and steady state RNAs suggest trans-splicing may be associated with a polycistronic transcript in *Schistosoma mansoni*. Mol. Biochem. Parasitol. *89*, 25–39. Abstract Article

Denker, J.A., Zuckerman, D.M., Maroney, P.A., and Nilsen, T.W. (2002). New components of the spliced leader RNP required for nematode trans-splicing. Nature *417*, 667–670. Abstract Article

Douris, V., Telford, M.J. and Averof, M. (2010). Evidence for multiple independent origins of trans-splicing in metazoa. Mol. Biol. Evol. *27*, 684-693. Abstract Article

Evans, D., and Blumenthal, T. (2000). *trans* splicing of polycistronic *Caenorhabditis elegans* pre-mRNAs: analysis of the SL2 RNA. Mol. Cell. Biol. *20*, 6659–6667. Abstract Article

Evans, D., Perez, I., MacMorris, M., Leake, D., Wilusz, C.J., and Blumenthal, T. (2001). A complex containing CstF-64 and the SL2 snRNP connects mRNA 3' end formation and trans-splicing in *C. elegans* operons. Genes Dev. *15*, 2562–2571. Abstract Article

Evans, D., Zorio, D., MacMorris, M., Winter, C.E., Lea, K., and Blumenthal, T. (1997). Operons and SL2 trans-splicing exist in nematodes outside the genus *Caenorhabditis*. Proc. Natl. Acad. Sci. U. S. A. *94*, 9751–9756. Abstract Article

Ferguson, K.C., Heid, P.J., and Rothman, J.H. (1996). The SL1 trans-spliced leader RNA performs an essential embryonic function in *Caenorhabditis elegans* that can also be supplied by SL2 RNA. Genes Dev. *10*, 1543–1556. Abstract Article

Furst, J., Ritter, M., Rudzki, J., Danzl, J., Gschwentner, M., Scandella, E., Jakab, M., Konig, M., Oehl, B., Lang, F., et al. (2002). ICln ion channel splice variants in *Caenorhabditis elegans*: voltage dependence and interaction with an operon partner protein. J. Biol. Chem. *277*, 4435–4445. Abstract Article

Ganot, P., Kallesoe, T., Reinhardt, R., Chourrout, D., and Thompson, E.M. (2004). Spliced-leader RNA trans splicing in a chordate, *Oikopleura dioica*, with a compact genome. Mol. Cell. Biol. *24*, 7795–7805. Abstract Article

Graber, J. H., Salisbury, J., Hutchins, L.N. and Blumenthal, T. (2007). *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing. RNA *13*, 1409-1426. Abstract Article

Hannon, G.J., Maroney, P.A., Ayers, D.G., Shambaugh, J.D., and Nilsen, T.W. (1990). Transcription of a nematode trans-spliced leader RNA requires internal elements for both initiation and 3' end-formation. EMBO J. *9*, 1915–1921. Abstract

Hannon, G.J., Maroney, P.A., and Nilsen, T.W. (1991). U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing *in vitro*. J. Biol. Chem. *266*, 22792–22795. Abstract

Harrison N, Kalbfleisch A, Connolly B, Pettitt J, Müller B. (2011). SL2-like spliced leader RNAs in the basal nematode *Prionchulus punctatus*: New insight into the evolution of nematode SL2 RNAs. RNA *16*, 1500-1507. Abstract Article

Hastings, K.E. (2005). SL trans-splicing: easy come or easy go? Trends Genet. *21*, 240–247. Abstract Article

Huang, L.S., Tzou, P., and Sternberg, P.W. (1994). The *lin-15* locus encodes two negative regulators of *Caenorhabditis elegans* vulval development. Mol. Biol. Cell *5*, 395–411. Abstract

Huang, T., Kuersten, S., Deshpande, A.M., Spieth, J., MacMorris, M., and Blumenthal, T. (2001). Intercistronic region required for polycistronic pre-mRNA processing in *Caenorhabditis elegans*. Mol. Cell. Biol. *21*, 1111–1120. Abstract Article

Huang, X.Y., and Hirsh, D. (1989). A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U. S. A. *86*, 8640–8644. Abstract Article

Huang P, Pleasance E.D., Maydan J.S., Hunt-Newbury R., O'Neil N.J., Mah A., Baillie D.L., Marra M.A., Moerman D.G., and Jones S.J. (2007). Identification and analysis of internal promoters in *Caenorhabditis elegans* operons. Genome Res. *17*, 1478–1485. Abstract Article

Jan, C.H., Friedman, R.C., Ruby, J.G. and Bartel, D.P. (2011). Formation, regulation and evolution of *Caenorhabditis elegans* 3' UTRs. Nature *469*, 97-101. Abstract Article

Keiper, B.D., Lamphear, B.J., Deshpande, A.M., Jankowska-Anyszka, M., Aamodt, E.J., Blumenthal, T., and Rhoads, R.E. (2000). Functional characterization of five eIF4E isoforms in *Caenorhabditis elegans*. J. Biol. Chem. *275*, 10590–10596. Abstract Article

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

Kuersten, S., Lea, K., MacMorris, M., Spieth, J., and Blumenthal, T. (1997). Relationship between 3' end formation and SL2-specific trans-splicing in polycistronic *Caenorhabditis elegans* pre-mRNA processing. RNA *3*, 269–278. Abstract

Laing, R., Hunt, M., Protasio, A.V., Saunders, G., Mungall, K., Laing, S., Jackson, F., Quail, M., Beech, R., Berriman, M., and Gilleard, J.S. (2011). Annotation of two large contiguous regions from the *Haemonchus contortus* genome using RNA-seq and comparative analysis with *Caenorhabditis elegans*. PLoS One *6*, e23216. Abstract Article

Lall, S., Friedmann, C.C., Jankowska-Anyszka, M., Stepinksi, J., Darzynkiewicz, E., and Davis, R.E. (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. J. Biol. Chem. *279*, 45573-45585. Abstract Article

Land, M., Islas-Trejo, A., and Rubin, C.S. (1994). Origin, properties, and regulated expression of multiple mRNAs encoded by the protein kinase C1 gene of *Caenorhabditis elegans*. J. Biol. Chem. *269*, 14820–14827. Abstract

Lasda, E., Allen, M.A. and Blumenthal, T. (2010). Polycistronic pre-mRNA processing in vitro: snRNP and pre-mRNA role reversal in trans-splicing. Genes Dev. *24*, 1645-1658. Abstract Article

Lasda, E. and Blumenthal, T. (2011). Trans-splicing. Wiley Interdiscip. Rev.: RNA *2*, 417-434. Abstract

Lee, K.Z., and Sommer, R.J. (2003). Operon structure and trans-splicing in the nematode *Pristionchus pacificus*. Mol. Biol. Evol. *20*, 2097–2103. Abstract Article

Liu, Y., Huang, T., MacMorris, M., and Blumenthal, T. (2001). Interplay between AAUAAA and the trans-splice site in processing of a *Caenorhabditis elegans* operon pre-mRNA. RNA *7*, 176–181. Abstract Article

Liu, Y., Kuersten, S., Huang, T., Larsen, A., MacMorris, M., and Blumenthal, T. (2003). An uncapped RNA suggests a model for *Caenorhabditis elegans* polycistronic pre-mRNA processing. RNA *9*, 677–687. Abstract Article

Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription: implications for a unified allosteric-torpedo model. Genes Dev. *20*, 954-965. Abstract Article

MacMorris, M., Kumar, M., Lasda, E., Larsen, A., Kraemer, B. and Blumenthal, T. (2007). A novel family of *C. elegans* snRNPs contains proteins associated with trans-splicing. RNA *13*, 511-520. Abstract Article

Maroney, P.A., Denker, J.A., Darzynkiewicz, E., Laneve, R., and Nilsen, T.W. (1995). Most mRNAs in the nematode *Ascaris lumbricoides* are trans-spliced: a role for spliced leader addition in translational efficiency. RNA *1*, 714–723. Abstract

Morton, J.J. and Blumenthal, T. (2011). Identification of transcription start sites of trans-spliced genes: uncovering unusual operon arrangements. RNA *17*, 327-337 Abstract Article

Park, Y.S., and Kramer, J.M. (1990). Tandemly duplicated *Caenorhabditis elegans* collagen genes differ in their modes of splicing. J. Mol. Biol. *211*, 395–406. Abstract Article

Pettitt J, Müller B, Stansfield I, and Connolly B. (2008). Spliced leader trans-splicing in the nematode *Trichinella spiralis* uses highly polymorphic, noncanonical spliced leaders. RNA *14*, 760-770. Abstract Article

Qian, W. and Zhang, J. (2008). Evolutionary dynamics of nematode operons: Easy come, slow go. Genome Res. *18*, 412-421. Abstract Article

Redmond, D.L., and Knox, D.P. (2001). *Haemonchus contortus* SL2 trans-spliced RNA leader sequence. Mol. Biochem. Parasitol. *117*, 107–110. Abstract Article

Reinke, V. and Cutter, A.D. (2009). Germline expression influences operon organization in the *Caenorhabditis elegans* genome. Genetics *181*, 1219-1228. Abstract Article

Ross, L.H., Freedman, J.H., and Rubin, C.S. (1995). Structure and expression of novel spliced leader RNA genes in *Caenorhabditis elegans*. J. Biol. Chem. *270*, 22066–22075. Abstract Article

Spieth, J., Brooke, G., Kuersten, S., Lea, K., and Blumenthal, T. (1993). Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. Cell *73*, 521–532. Abstract Article

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003). The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS. Biol. *1*, E45. Abstract Article

Stover, N.A., and Steele, R.E. (2001). Trans-spliced leader addition to mRNAs in a cnidarian. Proc. Natl. Acad. Sci. U. S. A. *98*, 5693–5698. Abstract Article

Thomas, J., Lea, K., Zucker-Aprison, E. and Blumenthal, T. (1990). The spliceosomal snRNAs of *C. elegans*. Nucleic Acids Res. *18*, 2633-2642. Abstract Article

Treinin, M., Gillo, B., Liebman, L., and Chalfie, M. (1998). Two functionally dependent acetylcholine subunits are encoded in a single *Caenorhabditis elegans* operon. Proc. Natl. Acad. Sci. U. S. A. *95*, 15492–15495. Abstract Article

Vandenberghe, A.E., Meedel, T.H., and Hastings, K.E. (2001). mRNA 5'-leader trans-splicing in the chordates. Genes Dev. *15*, 294–303. Abstract Article

Wallace, A., Filbin, M.E., Veo, B., McFarland, C., Stepinski, J., Jankowska-Anyszka, M., Darzynkiewicz, E., and Davis, R.E. (2010). The nematode eukaryotic translation initiation factor 4E/G complex works with a trans-spliced leader stem-loop to enable efficient translation of trimethylguanosine-capped RNAs. Mol. Cell. Biol. *30*, 1958-1970. Abstract Article

Whitton, C., Daub, J., Quail, M., Hall, N., Foster, J., Ware, J., Ganatra, M., Slatko, B., Barrell, B., and Blaxter, M. (2004). A genome sequence survey of the filarial nematode *Brugia malayi*: repeats, gene discovery, and comparative genomics. Mol. Biochem. Parasitol. *137*, 215–227. Abstract Article

Williams, C., Xu, L., and Blumenthal, T. (1999). SL1 trans splicing and 3'-end formation in a novel class of *Caenorhabditis elegans* operon. Mol. Cell. Biol. *19*, 376–383. Abstract

Zaslaver, A., Baugh, L.R. and Sternberg, P.W. (2011). Metazoan operons accelerate recovery from growth-arrested states. Cell *145*, 981-992. Abstract Article

Zorio, D.A., Cheng, N.N., Blumenthal, T., and Spieth, J. (1994). Operons as a common form of chromosomal organization in *C. elegans*. Nature *372*, 270–272. Abstract Article