A novel family of *C. elegans* snRNPs contains proteins associated with *trans*-splicing

MARGARET MACMORRIS,1,2,4 MADHUR KUMAR,1,4 ERIKA LASDA,1,2 ALISON LARSEN,1 BRIAN KRAEMER,3 and THOMAS BLUMENTHAL1,2

1Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado 80045, USA
2Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309, USA
3Geriatrics Research Education and Clinical Center, Veterans Affairs Puget Sound Health Care System, Seattle, Washington, USA

ABSTRACT

In many *Caenorhabditis elegans* pre-mRNAs, the RNA sequence between the 5′ cap and the first 3′ splice site is replaced by *trans*-splicing a short spliced leader (SL) from the Sm snRNP, SL1. *C. elegans* also utilizes a similar Sm snRNP, SL2, to *trans*-splice at sites between genes in polycistronic pre-mRNAs from operons. How do SL1 and SL2 snRNPs function in different contexts? Here we show that the SL1 snRNP contains a complex of SL75p and SL21p, which are homologs of novel proteins previously reported in the Ascaris SL snRNP. Interestingly, we show that the SL2 snRNP does not contain these proteins. However, SL75p and SL26p, a paralog of SL21p, are components of another Sm snRNP that contains a novel snRNA species, Sm Y. Knockdown of SL75p is lethal. However, knockdown of either SL21p or SL26p alone leads to cold-sensitive sterility, whereas knockdown of both SL21p and SL26p is lethal. This suggests that these two proteins have overlapping functions even though they are associated with different classes of snRNP. These phenotypic relationships, along with the association of SL26p with SL75p, imply that, like both SL21p and SL26p is lethal. This suggests that these two proteins have overlapping functions even though they are associated with different classes of snRNP. These phenotypic relationships, along with the association of SL26p with SL75p, imply that, like

Keywords: *trans*-splicing; operon; *Caenorhabditis*; SL snRNP; Sm Y

INTRODUCTION

In the nematode *Caenorhabditis elegans*, pre-mRNAs are spliced in two related but distinct ways. First, most pre-mRNAs contain introns that are spliced out, as in other organisms. Second, about half of the genes have an outron, an intron-like sequence at the 5′ end of the pre-mRNA ending with a 3′ splice site (Zorio et al. 1994). The outron is removed by *trans*-splicing. Because there is no 5′ splice site in the outron, the 3′ splice site is spliced to a mobile 5′ splice site present on a different RNA called the spliced leader RNA (SL RNA). The spliced leader is 22 nucleotides (nt) long and the RNA that donates it is ~100 nt (Blumenthal 2005).

In *C. elegans* there are two functionally distinct SL RNAs. SL1 RNA is used to replace the outron. The other, SL2 RNA, is used for splicing polycistronic pre-mRNAs (Spieth et al. 1993). Many *C. elegans* genes are contained in operons, multigene transcription units that are regulated from a single promoter at the 5′ end of the cluster. These polycistronic pre-mRNAs are processed into single-gene, mature, capped, polyadenylated mRNAs by a process involving both 3′-end formation at the poly(A) sites of upstream genes and *trans*-splicing by SL2 RNA at the 5′ ends of downstream genes, sites that are usually separated by only ~100 nt (Blumenthal and Spieth 1996).

The processes of 3′-end formation of the upstream gene transcript and *trans*-splicing of the downstream gene at these sites are closely linked (Huang et al. 2001). Sequences required for 3′-end cleavage are also required for SL2 *trans*-splicing. Following cleavage catalyzed by the cleavage and polyadenylation specificity factor (CPSF) bound to U-rich sequences in the vicinity of the cleavage site (Salisbury et al. 2006), the 3′ portion of the pre-mRNA is degraded exonucleolytically 5′ to 3′ until the nuclease reaches a block, possibly a bound protein (Liu et al. 2003). CPSF has been shown to interact with the SL2 snRNP, and mutations within the third stem–loop of SL2 RNA that eliminate this binding prevent its *trans*-splicing at operon *trans*-splice sites (Evans et al. 2001).
SL RNAs are bound to the Sm proteins that are also core proteins in spliceosomal U1, U2, U4, and U5 snRNPs (Bruzik et al. 1988; Thomas et al. 1988; Van Doren and Hirsh 1988). In addition to the Sm proteins, the SL snRNP from *Ascaris*, a nematode that is distantly related to *C. elegans*, contains two novel proteins of 95 kDa (SL95p) and 30 kDa (SL30p) (Denker et al. 1996, 2002). SL30p interacts with the splicing protein SFI/BBP that binds to the branchpoint. This interaction is thought to bring the *Ascaris* SL snRNP to trans-splice sites. Here we have examined the *C. elegans* homologs of the *Ascaris* SL-specific proteins in an effort to determine how SL1 and SL2 snRNPs function in different contexts. We demonstrate that whereas the *C. elegans* SL1 snRNP contains a complex consisting of the SL75p/SL21p proteins and SL26p, the *Ascaris* SL snRNP does not. Instead, we make the surprising discovery that SL26p, a paralog of SL21p, is complexed with SL75p, whereas the SL75p/SL21p proteins are involved in the same process. We show that SL21p and SL26p nonetheless have a synthetic phenotype, and we present a model, involving base pairing of SL RNAs to Sm Y, to account for how these proteins could be associated with different classes of snRNAs and yet be involved in the same process.

**RESULTS**

**Homologs of the *Ascaris* SL proteins**

The *C. elegans* genome contains three genes homologous to the *Ascaris* SL-specific proteins identified by the Nilsen laboratory (Denker et al. 2002): SL75p is homologous to *Ascaris* SL95p, and SL21p and SL26p are homologous to *Ascaris* SL30p. The gene for SL26p is called *sna-1*. This RNA is predicted to fold into two stems separated by a single-stranded region containing a canonical Sm-binding site. Sm Y was shown to be Sm protein bound, and *Ascaris* SL75p/SL21p (underlined in Fig. 1) are candidates for interaction with a protein or RNA with which SL26p interacts, but SL21p does not.

**The *Caenorhabditis* Sm Y family**

A novel RNA, called Sm Y, associated with spliceosomal preparations from *Ascaris* extract, was identified in the Nilsen laboratory (Maroney et al. 1996). This RNA is predicted to fold into two stems separated by a single-stranded region containing a canonical Sm-binding site. The *Ascaris* Sm Y was shown to be Sm protein bound, and we demonstrate below that the *C. elegans* Sm Y RNAs are also Sm snRNPs. We identified a 12-gene family in the *C. elegans* genome that specifies Sm Y, along with 10 homologs in the *C. briggsae* genome. A dendrogram showing the relationship between these genes is shown in Figure 2.
As in the SL2 RNA gene family (see below), a few common genes in the last common ancestor of *C. elegans* and *C. briggsae* radiated to form the gene families of today. Importantly, our phylogenetic analysis indicates that there is a single evolutionarily distinct member of the gene family, Sm Y-10, in both species. The RNA specified by this gene behaves quite differently from other family members, as shown below. Previously published work designed to uncover new small RNA species in *C. elegans* has shown that many of the family members are expressed (Deng et al. 2006; He et al. 2006), and our data confirm that RNAs specified by most or all of the genes can be detected with varying levels of expression at different developmental stages (data not shown). The Sm Y RNAs range in size from 77 to 82 nt with the exception of Sm Y-10, which is distinctly larger, 88 nt.

*Caenorhabditis* Sm Y family members were identified using BLAST in both *C. elegans* (Ce) and *C. briggsae* (Cb) genomes. Gene names, smy-1–smy-12, and chromosomal locations are given in the Supplemental material available at http://mcdb.colorado.edu/faculty/blumenthal.htm. Sequences were aligned and the dendrogram made by Clustal W using MacVector (UPGMA, best tree). Note that Sm Y-10 is encoded by a distant member of the gene family. The rooting of this tree must be considered uncertain.

FIGURE 2. The *Caenorhabditis* Sm Y family. The Sm Y family members were identified using BLAST in both *C. elegans* (Ce) and *C. briggsae* (Cb) genomes. Gene names, smy-1–smy-12, and chromosomal locations are given in the Supplemental material available at http://mcdb.colorado.edu/faculty/blumenthal.htm. Sequences were aligned and the dendrogram made by Clustal W using MacVector (UPGMA, best tree). Note that Sm Y-10 is encoded by a distant member of the gene family. The rooting of this tree must be considered uncertain.

As in the SL2 RNA gene family (see below), a few common genes in the last common ancestor of *C. elegans* and *C. briggsae* radiated to form the gene families of today. Importantly, our phylogenetic analysis indicates that there is a single evolutionarily distinct member of the gene family, Sm Y-10, in both species. The RNA specified by this gene behaves quite differently from other family members, as shown below. Previously published work designed to uncover new small RNA species in *C. elegans* has shown that many of the family members are expressed (Deng et al. 2006; He et al. 2006), and our data confirm that RNAs specified by most or all of the genes can be detected with varying levels of expression at different developmental stages (data not shown). The Sm Y RNAs range in size from 77 to 82 nt with the exception of Sm Y-10, which is distinctly larger, 88 nt.

**C. elegans** SL proteins are associated with Sm snRNP

The *Ascaris* SL95p binds to SL30p (Denker et al. 2002). To determine whether one or both of the SL30p homologs, SL21p and SL26p, bind to SL75p in *C. elegans*, we made a polyclonal antibody to a peptide sequence in SL75p and, although this antibody does not recognize a specific protein on Western blots, it does effectively immunoprecipitate (IP) in vitro translated SL75p (data not shown). We also produced polyclonal antibodies to GST-fusion proteins of SL21p and of SL26p and showed that they recognize bands of the expected sizes on Western blots (Fig. 3A, lane 1), and do not significantly associate with each other or exist in a complex together (Fig. 3A, lanes 7 and 11). We immunoprecipitated SL75p and probed for SL26p and SL21p by Western blotting. Figure 3A, lane 4, shows that each of the small SL proteins is co-immunoprecipitated by anti-SL75p, whereas neither protein was precipitated by the preimmune serum (Fig. 3A, lane 3). We conclude that SL21p and SL26p each associate with SL75p. However, they do not associate with each other (Fig. 3A, lanes 7 and 11). To determine whether this interaction is dependent on RNAs with which they may interact, we degraded the RNA in the extract with RNase I (Fig. 3B) and repeated the immunoprecipitation. This treatment still allowed co-immunoprecipitation of SL21p and of SL26p with SL75p (Fig. 3A, lane 5), so we conclude that both the SL75p/SL21p and SL75p/SL26p interactions are not dependent on RNA.

To determine whether SL26p is associated with snRNAs bound by Sm proteins, we immunoprecipitated snRNP with anti-Sm serum, extracted the precipitate for RNA and protein, and analyzed them on Northern and Western blots. Figure 3C shows that SL26p co-immunoprecipitates with Sm proteins (lane 3) and is not detected in the precipitate when the RNA is destroyed by RNase I (lane 4), indicating that it is associated with an Sm-bound RNA. Thus, like the SL75p/SL21p complex, the SL75p/SL26p complex is bound to an Sm snRNP.

**FIGURE 3.** Interactions between SL snRNP proteins. (A) SL75p interacts with SL21p and SL26p independent of RNA. Western blot of proteins immunoprecipitated from embryonic extract. Some samples were treated with RNase (+) before IP. The blot was probed with anti-SL26p or anti-SL21p, as indicated on the left. IP antibodies are shown at the top. In each case, input and IP samples shown on the gels are approximately equal percentages of the total. (B) Destruction of snRNAs by RNase 1. Northern blot of RNAs isolated from extract used in panel A. RNAs were detected by hybridization with 32P-labeled oligonucleotides complementary to SL2 RNA, SL1 RNA, and Sm Y. The Sm Y probe recognizes a number of the Sm Y family members and appears on the Northern blot as a doublet. (Lane 1) control; (lane 2) RNase-treated. (C) SL26p is part of an Sm snRNP. Western blot probed with anti-SL26p antibody. RNase treated (+). IP: lanes 3 and 4, Kung patient serum (anti-Sm); lanes 5 and 6, nonimmune rabbit serum.
The small SL proteins are associated with different snRNPs

Anti-Sm serum immunoprecipitates SL1 RNA, SL2 RNA, Sm Y family members (two bands), Sm Y-10 (Fig. 4A, lane 2), and the expected U snRNAs (data not shown). In order to determine which snRNA each protein complex is associated with, we immunoprecipitated each protein and probed Northern blots for the SL and other snRNAs. The SL21p IP contains SL1 RNA, but it does not contain SL2, Sm Y, or Sm Y-10 (Fig. 4A, lane 3), indicating that SL21p is a component of the SL1 snRNP. This observation is not surprising given the close relationship between C. elegans SL21p and Ascaris SL30p, which is associated with the only SL snRNP in Ascaris, SL1 snRNP.

To analyze the role of SL26p and its associations we used a mutant C. elegans strain containing a premature stop codon near the 5' end of the gene encoding SL26p, sut-1. Anti-SL26p blots of worms carrying this mutation demonstrate the lack of SL26p protein (Fig. 4B, lanes 4–6). In spite of the fact that SL26p is clearly snRNP-associated based on its precipitability with anti-Sm (Fig. 3C, lane 3), SL26p antibody failed to IP either SL1 or SL2 RNA (Fig. 4C, lane 3) or any of the U snRNAs (data not shown). As Sm Y had previously been identified as a novel spliceosomal Sm snRNP in Ascaris, we asked whether this RNA was precipitated with anti-SL26p antibody. We probed Northern blots of RNAs extracted from the SL26p IPs with a probe that recognizes most of the Sm Y family members (but not Sm Y-10). The association was clear (Fig. 4C, lane 3): SL26p is complexed with Sm Y RNAs, but a probe to the more distantly related Sm Y-10 shows that this RNA is not bound by SL26p. As shown in Figure 4C, lane 4, antibody to SL26p no longer IPs Sm Y in the sut-1 mutant strain lacking SL26p, which confirms the association of SL26p with Sm Y.

We analyzed the RNA extracted from anti-SL75p IP and found both SL1 and Sm Y, but not SL2 RNA or Sm Y-10 (Fig. 4D, lane 4). Furthermore, we were able to use the sut-1 strain to ask if SL75p association with Sm Y requires the presence of SL26p. The results (Fig. 4E, lane 4) demonstrate that SL75p can associate with Sm Y RNA even in the absence of SL26p.

Gradient mobility shift assays also demonstrate SL1 RNA/SL21p and Sm Y/SL26p interactions

To investigate sizes of SL and Sm Y snRNP complexes, whole-cell embryo extracts were sedimented on 10–40% glycerol gradients (Fig. 5A). The SL1 snRNP would be expected to have a molecular weight of 217 kDa, based on only its known components of SL1 RNA, the Sm proteins, and SL75p/SL21p. Northern blots demonstrated a peak of SL1 RNA in fractions 3–5, consistent with a complex of the expected size. SL21p is present in fractions 4 and 5, consistent with its presence in some SL1 snRNPs. The SL2 snRNP is expected to have a molecular weight of only 124 kDa based on its known components, SL2 RNA and the Sm proteins. However, it co-sediments with SL1 snRNP, suggesting it contains at least one unknown component. Sm Y has an expected molecular weight of 215 kDa, based on its known components: Sm Y, the Sm proteins, and SL75p/SL26p. However, the Sm Y is found in somewhat heavier fractions than the SL snRNPs, suggesting it may also associate with unknown components. The highest levels of SL26p are found in the same fractions as Sm Y, consistent with its presence in the Sm Y snRNP. Sm Y-10 is found in the same fractions as the SL snRNPs (at ~200 kDa), indicating Sm Y-10 is present in smaller complexes than the other Sm Y family members. Sm Y-10 snRNP has an expected molecular weight of only 117 kDa based on its known components, Sm Y-10 and the Sm proteins, suggesting that it, too, is likely to associate with at least one unknown component.

To examine these relationships further, we added either SL21p or SL26p antibodies to the extracts before sedimenting the mixtures on identical glycerol gradients. This experiment is functionally equivalent to a gel supershift experiment. Antibody to SL21p shifted some of the SL1 RNA and all of the SL21p from fractions 4 and 5 to fractions 6–9 (Fig. 5B). The rest of the SL1 snRNA is not shifted by the antibody; it remains in fractions that contain no SL21p, suggesting the possibility that these embryo extracts contain some SL1 snRNP that is not bound to SL21p or the complex dissociates during sedimentation. Neither the SL2 RNA, the Sm Y-10, nor the Sm Y peaks were affected by anti-SL21p antibody. We conclude that
further, we began by characterizing the SL2 RNA gene that might involve base pairing. To examine this question, families might suggest an interrelationship between them. Thought the coordinate expansion of these two RNA RNAs each form a small family of related RNAs, we SL12) form a gene family. Since the SL2 RNAs and Sm Y mRNAs. SL2 RNA and its sequence variants (called SL3–2006) and only a single Sm Y (T. Blumenthal, unpubl.). In each case, interspecies base pairings were not found. Sm Y-10 is the only Sm Y that is not associated with SL26p, so it would be interesting if it interacts with SL1 RNA, the only snRNA associated with SL21p.

SL21p and SL26p show a synthetic phenotype

In order to examine the functional significance of the small snRNP-associated proteins, SL21p and SL26p, we attempted knockdown by RNAi. RNAi of SL75p was 100% lethal, indicating that this protein is required for viability. RNAi of SL75p, however, had no obvious phenotype. SL26p, on the other hand, produces a strong developmental phenotype. Similar to the SL26p RNAi knockdown, embryonic lethality and slightly reduced numbers of F1 progeny compared to the wild-type (WT) strain N2 (Table 1). However, the sut-1 mutant strain cannot be maintained at 15°C; embryos hatch into larvae that develop into 100% sterile adults. Sut-1 mutant adults shifted to 15°C produce

SL21p is a component of the SL1 snRNP, as in Ascaris, and not of SL2 or Sm Y snRNPs.

In contrast, the data in Figure 5C demonstrate that antibody to SL26p shifts all of SL26p and most of the Sm Y to much higher molecular weight fractions, whereas this antibody has little effect on the SL RNAs or Sm Y-10. This observation is consistent with the conclusion that SL26p associates exclusively with Sm Y in these embryo extracts.

SL RNAs and Sm Y RNAs

Caenorhabditis species possess SL1 RNA, a family of SL2 RNAs, and a family of Sm Y RNAs (Deng et al. 2006; He et al. 2006). In contrast, Brugia malayi and Ascaris lumbricoides, two nematode species that are very distantly related to C. elegans, have only SL1 RNA (Nilsen et al. 1989; Blaxter and Liu 1996; Lall et al. 2004; Giuliano and Blaxter 2006) and only a single Sm Y (T. Blumenthal, unpubl.). The multiplicity of SL RNAs in the rhabditid nematodes has apparently evolved for trans-splicing operon pre-mRNAs. SL2 RNA and its sequence variants (called SL3–SL12) form a gene family. Since the SL2 RNAs and Sm Y RNAs each form a small family of related RNAs, we thought the coordinate expansion of these two RNA families might suggest an interrelationship between them that might involve base pairing. To examine this question further, we began by characterizing the SL2 RNA gene families of C. elegans and C. briggsae (Fig. 6). We identified all of the SL2 RNA genes in both species by repeatedly subjecting the two genomes to BLAST analysis with all SL2 RNA genes as they were uncovered in previous rounds (Stein et al. 2003). We found a total of 18 SL2 RNA genes in each species. The 18 C. elegans genes specify 11 different SL2 variants, and the 18 C. briggae genes specify six variants, four of which are also found in C. elegans. A dendrogram of all 36 genes shows that they descended from a few ancestral genes found in the last common ancestor of the two species.

We next searched for possible base pairing between members of the SL and Sm Y gene families. Many such possible base-pairing interactions were predicted, and all involve pairing between the second stem–loop of a SL RNA and the second stem–loop of a Sm Y (circled in the diagram of Fig. 7). The proposed base-pairing interactions for the C. elegans RNAs are shown in Figure 7. Similar interactions can be drawn for C. briggae (T. Blumenthal, unpubl.). In each case, docking would occur by interaction of the two loop sequences, and then the base pairing could extend additional base pairs formerly part of the stems. As many as 14 contiguous base pairs in the case of Sm Y-10 may be involved in this base pairing. It is intriguing that such base pairing can be drawn for Sm Y-10 of C. elegans with SL1 RNA of C. elegans, with corresponding structures for Ascaris and Brugia (Fig. 7, box), but interspecies base pairings were not found. Sm Y-10 is the only Sm Y that is not associated with SL26p, so it would be interesting if it interacts with SL1 RNA, the only snRNA associated with SL21p.
50% fewer F1 progeny (per P₀) and no F2 progeny (Table 1). The same 15°C-shifted sut-1 mutant adults show an increase in the number of unfertilized oocytes laid compared to similar-staged WT controls, before the stage when adults would have normally exhausted their limited supply of sperm.

Interestingly, RNAi of the other member of this gene family, SL21p, shows a cold-sensitive phenotype very similar to that shown by the sut-1 strain (Table 1). Whereas SL21p RNAi-treated adults are normal at 20°C, those kept at 15°C have reduced numbers of F1 progeny (60% of untreated controls) and reduced fertility of the F1 progeny, i.e., fewer F2 progeny. Table 1 shows that 63% of the RNAi-treated adults have low F1 fertility (presented as F2 progeny/P₀). It therefore appears that reduction or loss of either of the two small proteins, SL21p or SL26p, results in a cold-sensitive defect in fertility.

Furthermore, reduction in the levels of both of these proteins as a result of SL21p RNAi in the sut-1 mutant strain produced dramatic reductions in viability and fertility at both 15°C and 20°C (Table 1). At 20°C embryonic lethality increased to nearly 20%, but none of the progeny matured to adulthood, leading to a complete absence of F2 progeny. At 15°C the effect was enhanced. In this case the offspring of the RNAi-treated worms died as embryos or young larvae. Embryonic lethality increased to 80% (Table 1). In many cases individual RNAi-treated animals produced broods with 100% embryonic lethality, demonstrating that the fully penetrant phenotype for knockdown of both proteins is embryonic lethality.

While the number of unfertilized oocytes laid by sut-1 adults is abnormally elevated at both 15°C and 20°C, SL21p RNAi treatment of sut-1 exacerbates this effect as well.

These observations support the idea that SL21p and SL26p are at least partially functionally interchangeable.

DISCUSSION

SL1 and SL2 snRNPs have different roles, and we show here that they have distinct protein components in addition to different RNAs. SL1 snRNP trans-splices at outrons, near the pre-mRNA 5’ end. It contains the SL75p/SL21p FIGURE 6. The Caenorhabditis SL2 RNA genes. The SL2 RNA genes from both Caenorhabditis species were identified by BLAST and the alignment and dendrogram produced as in Figure 2. C. elegans (ce); C. briggsae (cb). We have adopted the terminology of Thierry-Mieg (pers. comm.) in which the C. elegans genes are designated SL2–SL12 depending on their frequency in the C. elegans EST database, with SL2 being the most abundant and SL12 the least. Note that these names are dependent on the donated leader sequence, rather than the entire SL RNA sequence, so a single spliced leader is specified by several different genes. For example, SL2 is specified by four genes in C. elegans and eight genes in C. briggsae that contain sequence differences in the SL RNA, but no differences in the spliced leader. In the case of SL3 RNA, identical SLs are specified by distantly related genes (CeSL3-1 closely related to CeSL12; and CeSL3-2 and CeSL3-3 that tree together). The C. briggsae genes are named based on the sequences of their SL, but with two sequences not found in C. elegans called SL13 and SL14. Thus C. briggsae has the following spliced leaders: SL1, SL2, SL3, SL4, SL10, SL13, and SL14.

FIGURE 7. Possible base pairing between Sm Y and SL RNA family members. The diagram shows a representation of Sm Y or Sm Y-10 of each species.
complex that in *Ascaris* interacts with SF1 to bring it to *trans*-splice sites (Denker et al. 2002). SL2 RNA evolved from SL1 RNA, but it has lost the ability to bind to the SL75p/SL21p complex. The only proteins we have identified bound to SL2 RNA are the Sm proteins. However, the sedimentation mobility of the SL2 snRNP suggests it is bound to proteins not yet identified. SL2 snRNP does interact with the 3′-end formation protein, CstF, and may be brought to *trans*-splice sites between genes in polycistronic pre-mRNAs by virtue of this interaction (Evans et al. 2001). However, we do not know whether there is a direct interaction between CstF and the SL2 snRNP or whether there is a more extensive network of proteins and/or RNAs. The SL75p/SL26p complex is bound to a different class of Sm snRNP, the Sm Y snRNPs. This snRNP has the potential to interact transiently with the SL2 snRNP, as shown by the predicted base pairing in Figure 7 and thus could bring the SL75p/SL26p complex into a splicing reaction. If this scenario is correct, SL1 and SL2 *trans*-splicing utilizes a related set of proteins, which are brought to the reaction by different snRNPs.

**Evolution of SL2 RNA and SL26p**

SL1 RNA appears to be much more ancient than SL2 RNA. SL1 RNA is found throughout the nematode phylum (Blaxter and Liu 1996), whereas SL2 RNA has been found only in several rhabditid nematodes (Evans et al. 1997; Redmond and Knox 2001; Lee and Sommer 2003). Although operons are found in other nematode species, it appears there is no specialized snRNP for processing downstream genes in these operons (Guiliano and Blaxter 2006). Instead, they are *trans*-spliced to SL1. It seems most likely that SL2 RNA evolved from SL1 RNA as a means to *trans*-splice at splice sites between genes in operons more efficiently than SL1 RNA could (Blumenthal 2005). At the same time, evolutionary change in the snRNP-associated proteins may have involved duplication of the gene for the small SL protein, followed by specialization of each protein. The newer protein, SL26p, has maintained its ability to interact with SL75p, but is changed in some way such that the SL75p/SL26p interacts with Sm Y instead of SL1 RNA. Unlike SL1, SL2 RNA does not appear to interact with either of the small proteins. We note that the C-terminal third of SL26p is quite distinct from that of SL21p or *Ascaris* SL30p, but is nonetheless highly conserved between *C. elegans* and *C. briggsae*, suggesting it plays an important role. We hypothesize that this role could be interaction with Sm Y.

**What role does Sm Y play?**

Sm Y was discovered in *Ascaris*, where it was found associated with spliceosomal preparations (Maroney et al. 1996), implicating it in some aspect of the splicing process. Like other spliceosomal snRNPs, Sm Y has a classical Sm binding site and is bound to the Sm proteins. However, this RNA has not been found outside the nematode phylum. Since *cis*-splicing occurs in all animals, but *trans*-splicing is found only in a more limited group, it seems unlikely that Sm Y plays a critical role in *cis*-splicing, even in nematodes. Rather, we hypothesize that Sm Y is a component of the *trans*-splicing machinery. In support of this idea, we find it bound to SL75p and SL26p, a paralog of SL21p. Both SL75p and SL21p are components of the SL1 snRNP. Thus, even though the Nilsen laboratory found association of Sm Y with both *cis*- and *trans*-spliceosomes, we believe it is more likely that the role of Sm Y may be exclusive to *trans*-splicing. It could be a direct component of the *trans*-spliceosome, required for function, or it could be a chaperone for the SL snRNP, keeping it from splicing inappropriately.

**How do the Sm proteins get recycled following splicing?**

*Trans*-splicing presents the cell with unique problems, not experienced by cells that engage only in *cis*-splicing. In *cis*-splicing the Sm proteins are packaged into stable U-snRNPs that catalyze the splicing of other RNAs. However, in *trans*-splicing the Sm proteins are also packaged with the SL RNAs that are consumed in the process of...
splicing. Furthermore, trans-splicing occurs on most pre-mRNAs in *C. elegans*, so this process presumably utilizes a lot of Sm proteins, compared to the amount required for cis-splicing. Thus, it seems probable that the Sm proteins, left on the downstream part of the SL snRNA molecule following trans-splicing (the Y-branched intermediate), are utilized, rather than being discarded along with that portion of the SL RNA. We hypothesize that this could be a role of the Sm Y snRNP. Perhaps the base pairing does not occur between the two mature snRNPs, but instead occurs between Sm Y and the downstream portion of the SL snRNP following trans-splicing. This might destabilize the Sm binding site on the spent SL snRNP portion of the molecule, thus allowing transfer of the Sm proteins to Sm Y and subsequent transport of this snRNP to a new assembly site. In this model, we do not propose how the Sm Y snRNP would disassemble the Sm proteins for reuse, but the role of SL75p/SL21p and SL75p/SL26p could be to facilitate that portion of the recycling events envisaged here. That fact that loss of these proteins results in a cold-sensitive phenotype implies they may be required for destabilizing RNA secondary structure, consistent with the above proposition. The fact that SL21p and SL26p show a synthetic phenotype suggests they may be performing related roles. We argue that, based on the proposed base-pairing interactions, most members of the Sm Y family interact with members of the SL2 RNA family, whereas Sm Y-10 interacts with SL1. If this recycling idea is correct, loss of SL21p might result in failure to recycle the Sm proteins from SL1 trans-splicing, but the Sm proteins from SL2 trans-splicing would still be recycled, allowing survival. Similarly, loss of SL26p in the *sut-1* mutant worms would still allow the recycling of the Sm proteins from SL1 trans-splicing. However, loss of both would prevent recycling of all Sm proteins used in trans-splicing, resulting in lethality. This scenario could explain why the two paralogous proteins, SL21p and SL26p, show a synthetic phenotype even though they are bound to different classes of snRNP in embryo extract.

**Evolution of trans-splicing**

We have argued previously that it is likely that trans-splicing has arisen from cis-splicing repeatedly during the evolution of animals (Blumenthal 2005). This argument is based on the sporadic appearance of SL-type trans-splicing in various lineages, as well as the fact that trans-splicing is likely to be self-perpetuating since it can result in accumulation of upstream AUGs in the outrons of many pre-mRNAs. These AUGs could suppress translation of the authentic AUG at the beginning of the real open reading frame if they were not removed by trans-splicing. Furthermore, if trans-splicing were an ancient phenomenon, we would expect to find the proteins and RNAs that catalyze the reaction in all species that perform SL-type trans-splicing. However, SL RNAs from different phyla, while bearing an overall structural resemblance, show no obvious sequence similarities. Finally, relatives of neither Sm Y nor the SL proteins have been found outside the nematode phylum, even though the complete sequences of other trans-splicing organisms have been searched (T. Blumenthal, unpubl.).

**CONCLUSIONS**

This work presents an interesting conundrum: The SL1 and SL2 RNAs are functionally distinct, but structurally quite similar. Nevertheless they have different protein compositions: the SL1 snRNP contains the SL75p/SL21p complex, while the SL2 snRNP does not. Also, SL21p and SL26p show a synthetic phenotype, and loss of either causes a similar cold-sensitivity, implying that SL26p is also involved in trans-splicing. Surprisingly, it is bound to a different class of snRNA, the Sm Y snRNP, which implies these snRNPs in trans-splicing as well. However, what role these novel snRNAs play in trans-splicing is still unknown. They may base pair with the SL RNAs, and the SL snRNP proteins may be involved with the interaction between the two classes of snRNP, thereby rationalizing the fact that the proteins show a synthetic phenotype even though they are bound to different classes of snRNP.

**MATERIALS AND METHODS**

**Strains and extracts**

Maintenance and growth of worms was as described by Brenner (1974) and Sulston and Brenner (1974). Embryonic extracts were made from synchronous cultures of worms grown in liquid culture, hypochlorite treated to isolate embryos, and broken using a Dounce homogenizer as described (Hollins et al. 2005). The strain C121 contains *sut-1* mutant allele (*bk79*), which has a UAA stop codon at amino acid position 40 of the SL26p gene, T13B5.8, now named *sut-1*.

**Antibody preparation**

For SL21p and SL26p polyclonal antibodies, cDNAs were prepared by reverse-transcribing N2 RNA and amplifying with gene-specific primers for T13B5.8 (*sut-1*) and W02F12.6 (*sna-1*) (Wormbase). Full-length translational gene fusions were prepared by cloning the cDNAs into pGEX-2T. The GST-SL21p (pPM164) and GST-SL26p (pPM167) fusion proteins were expressed in *E. coli* and eluted with glutathione. Rabbits were inoculated with GST-protein eluates (Pocono Rabbit Farm & Laboratory; Josman Laboratory). The anti-SL75p antibody was made by injecting into rabbits a synthetic peptide corresponding to amino acids CTKASGVVRANDDGSKKI from the middle portion of the T13F2.7 (*sna-2*) gene, the *C. elegans* ortholog of the *Ascaris* SL95p gene. The resulting antiserum was affinity purified against the peptide (Biosource). Anti-Sm patient serum (Kung) was kindly provided by the Mattaj laboratory (EMBL).
Immunoprecipitation

Aliquots of protein A sepharose (Amersham) were suspended and washed three times in PBS, then incubated for 1 h rocking at room temperature with antisera or preimmune sera as a control (approximately equal volumes of packed beads and antisera) suspended in a total of 0.5 mL PBS. Antiserum used for IP shown in Figure 4C was affinity purified using GST-SL26p fusion protein. Beads were blocked for at least 30 min in 0.1% nonfat dry milk in 1 mL of PBS, then the unbound antibody was removed and the beads were washed three times with PBS, followed by three washes with the IP buffer A (20 mM Tris at pH 8.0, 100 mM KCl, 0.2 mM EDTA, 0.1% NP40, 1 mM DTT) or buffer B (20 mM Tris at pH 7.4, 100 mM KCl, 0.2 mM EDTA, 0.1% NP40), with added protease inhibitor tablet (1 complete per 50 mL; Roche) and RNase inhibitor (~20 U/IP reaction with either RNasin [Promega] or RNaseOUT [Invitrogen]). Twenty to 40 μL of extract were added to beads per IP reaction with an equal volume of packed beads in a total volume of 0.5 mL IP buffer. Each tube was rocked at 4°C for 2 h to overnight, then the supernatant was collected and the beads washed three or four times with IP buffer. If protein and RNA analyses were performed, the beads were subdivided after washing. Protein samples were prepared by mixing beads with SDS sample buffer, heating, and loading on 12.5% acrylamide, SDS gels. RNA samples were treated with proteinase K, phenol/chloroform extracted, ethanol precipitated, 12.5% acrylamide, SDS gels. RNA samples were treated with proteinase K, phenol/chloroform extracted, ethanol precipitated, and resuspended in formamide dye loading buffer and electro-phoresed on 12% acrylamide, 7 M urea denaturing gels. Gels were subdivided after washing. Protein samples were prepared by mixing beads with SDS sample buffer, heating, and loading on 12.5% acrylamide, SDS gels. RNA samples were treated with proteinase K, phenol/chloroform extracted, ethanol precipitated, and resuspended in formamide dye loading buffer and electro-phoresed on 12% acrylamide, 7 M urea denaturing gels. Gels were blotted to Hybond N (Amersham) for RNA or nitrocellulose (Schleicher and Schuell) for protein transfer using a semi-dryblotted to Hybond N (Amersham) for RNA or nitrocellulose phoresed on 12% acrylamide, 7 M urea denaturing gels. Gels were subdivided after washing. Protein samples were prepared by mixing beads with SDS sample buffer, heating, and loading on 12.5% acrylamide, SDS gels. RNA samples were treated with proteinase K, phenol/chloroform extracted, ethanol precipitated, and resuspended in formamide dye loading buffer and electro-phoresed on 12% acrylamide, 7 M urea denaturing gels. Gels were blotted to Hybond N (Amersham) for RNA or nitrocellulose (Schleicher and Schuell) for protein transfer using a semi-dry transfer apparatus. Transfer buffer for RNA was 0.5× TBE and Towbin buffer for protein.

RNase treatment

Embryonic extract was treated with RNase ONE (Promega) at 1 U/2 μL extract in 0.25× RNase I buffer with 1× complete protease inhibitor (Roche) at 20°C for 18 min followed by freezing and thawing to stop the reaction. Control samples were treated identically except RNase was omitted.

Northern blot analysis

Membranes with RNA were prehybridized in 6× SSC, 0.2% SDS, 10× Denhardt’s at 65°C for 1 h, and hybridized overnight with 32P-labeled oligonucleotide probes at 65°C in 6× SSC, 0.2% SDS, 5× Denhardt’s. Oligonucleotides used as probes for hybridization included SL1-3’ (CCCCGAGAGGAGACGTTCC), SL2-3’ (TTT GCTTACCGAGATGACCC), SL2-PE (CAGAATCCTACCG TACCTTG), Sm Y-1 (GTCITTTAT/TGCAGGGCCCTGA), and Sm Y-10 (ACCAAGTCCTGATTTACGG). After hybridization, membranes were washed in 6× SSC, 0.2% SDS and exposed to PhosphorImager screens for image capture with a Molecular Dynamics Storm PhosphorImager.

Western blot analysis

Nitrocellulose membranes with transferred proteins were blocked with 3% milk in PBS, incubated with primary antibodies for 1 h to overnight, washed in PBS, reacted with protein A: HRP (Sigma) for 1 h, and reacted with Western Lightning Chemiluminescence Reagent Plus (PE), before X-ray film exposure.

RNA interference

Double-stranded RNA was prepared by transcribing with T3 or T7 polymerase from cDNA for W02F12.6 cloned into pGEM. Each transcript was electrophoresed to assay its integrity, the two strands annealed, diluted, and used for injection. Worms were injected, allowed to recover overnight, and then placed on individual plates in order to monitor their progeny. Injected P0 were moved to new plates each day to enable counting of newly laid embryos and their subsequent hatching and survival to adulthood. Unfertilized oocytes laid by P0 within the first 2 d of the experiment were counted at the same time that the embryos were counted.

Gradient analysis

Crude embryonic N2 extract (100 μL) was preincubated with or without 100 μL anti-serum against SL21p or SL26p in a total volume of 400 μL for 2 h rocking at 4°C (final concentration: 13 mM Tris–HCL at pH 8.0, 53 mM KCl, 0.6 mM MgCl2, 0.3 mM PMSF, 1 mM DTT, 0.75× Roche protease inhibitor tablet, 0.4 U/μL Invitrogen RNaseOUT). Samples were loaded onto 10%–40% glycerol gradients. Ultracentrifugation was performed at 29,000 rpm for 22 h at 4°C in a Beckman SW41 rotor. Twelve 1 mL samples were collected from each gradient from top to bottom using a Biocomp Gradient Station ip. Fractions were divided for RNA and protein analysis. RNA was isolated and analyzed by Northern blot. Protein was precipitated with nine volumes 100% ethanol and resuspended in SDS sample buffer prior to analysis on 12.5% polyacrylamide, SDS gels for Western blotting. A marker gradient with chicken ovalbumin (~44 kDa; Bio-Rad), β amylase (~200 kDa; Sigma), and apoferritin (~443 kDa; Sigma) was run in parallel and analyzed by SDS-PAGE and Coomassie staining.

Bioinformatics

Sequence alignments and distance trees were performed by the Clustal W Program using MacVector. Trees were calculated using the UPGMA/Best Tree method.

ACKNOWLEDGMENTS

We are grateful to Dick Davis, David Bentley, and Ravinder Singh for helpful discussions and for suggestions on the manuscript. We thank Donnie Evans and Scott Kuersten for their contributions to early aspects of this project. This work was supported by Research Grant R01 GM42432 from the National Institute of General Medical Sciences to T.B., a Department of Veterans Affairs Merit Review Entry Program Grant to B.K., and NIA grant PO1 AG17586.

Received December 5, 2006; accepted January 2, 2007.

REFERENCES


Thomas, J.D., Conrad, R.C., and Blumenthal, T. 1988. The *C. elegans* trans-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. *Cell* 54: 533–539.


A novel family of *C. elegans* snRNPs contains proteins associated with trans-splicing

Margaret MacMorris, Madhur Kumar, Erika Lasda, et al.

*RNA* 2007 13: 511-520 originally published online February 5, 2007
Access the most recent version at doi:10.1261/rna.426707

**References**
This article cites 26 articles, 12 of which can be accessed free at: http://rnajournal.cshlp.org/content/13/4/511.full.html#ref-list-1

**License**

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.