

Spliced Leader RNA–Mediated *trans*-Splicing in Phylum Rotifera

Natalia N. Pouchkina-Stantcheva and Alan Tunnacliffe

Institute of Biotechnology, University of Cambridge, Cambridge, United Kingdom

In kinetoplastids, *Euglena*, and four metazoan phyla, *trans*-splicing has been described as a mechanism for the generation of mature messenger RNAs (mRNAs): 5′-ends of precursor mRNAs are replaced by a short spliced leader (SL) exon from a small SL RNA. Although the full phylogenetic range is unknown, *trans*-splicing has not been found in vertebrates, insects, plants, or yeast. In animal groups where it does occur, i.e., nematodes, cnidarians, platyhelminths, and primitive chordates, SL RNAs do not show sequence relatedness across phyla. The apparently sporadic phylogenetic distribution and the lack of SL RNA homology have led to opposing hypotheses on its evolution, involving either an ancient origin followed by loss in multiple lineages or independent acquisition in several taxa. Here we present evidence for the occurrence of *trans*-splicing in bdelloid rotifers (Bdelloidea, Rotifera). A common 23-nt sequence, representing the SL exon—diagnostic of SL RNA–mediated *trans*-splicing—was found at the 5′-end of at least 50%–65% of mRNAs from *Adineta ricciae* and *Philodina* sp. The *trans*-splicing pattern in bdelloid rotifers can be unusually complex, as observed in transcripts from a heat shock protein gene, *hsp82-1*, where the SL exon was spliced to three alternative positions. Bdelloid rotifer SL RNAs were found to be 105 or 106 nt long and comprised the SL sequence, a conserved splice donor site and an intron containing a putative spliceosome–binding motif. Intriguingly, some similarity of rotifer SL RNA sequence and predicted secondary structure was seen to that of the predominant SL1 RNA of nematodes, although it is unlikely that this demonstrates homology. In addition, sequence corresponding to the rotifer SL exon was found at the 5′-end of a number of full-length complementary DNA (cDNA) clones in a rice (*Oryza sativa*) database. None of these cDNAs gave a close match with homologous plant genes, suggesting that a small but significant portion of the rice expressed sequence tag database represents sequences derived from rotifers. In summary, the description of SL-mediated *trans*-splicing in Rotifera extends its representation to at least five metazoan phyla, making it increasingly probable that this is a phylogenetically widespread and therefore ancient phenomenon.

Introduction

Trans-splicing is an mRNA processing event which accurately joins sequences derived from separately transcribed RNAs. In one form of *trans*-splicing, a leader sequence is spliced from the 5′-end of a small RNA (spliced leader RNA [SL RNA]) to pre-mRNA and constitutes the 5′ terminal exon of the mature mRNA. It was first discovered in trypanosomes (Murphy, Watkins, and Agabian 1986; Sutton and Boothroyd 1986) and later reported in flagellated protozoa (*Euglena*) (Tessier et al. 1991), nematodes (Krause and Hirsh 1987), flatworms (Rajkovic et al. 1990; Davis et al. 1994), hydra (Cnidaria) (Stover and Steele 2001), and primitive chordates, *Ciona intestinalis* (Vandenbergh et al. 2001) and *Oikopleura dioica* (Ganot et al. 2004). However, this mRNA-processing pathway is apparently absent in vertebrates, insects, yeast, and plants. In the nematode *Caenorhabditis elegans*, for example, over 80% of pre-mRNAs are *trans*-spliced to the SL, SL1 (Krause and Hirsh 1987); the same, almost invariant sequence is also used by many other nematode species (Blaxter and Liu 1996). Approximately 25% of genes in *C. elegans* are arranged in clusters and are transcribed as polycistronic pre-mRNAs (Zorio et al. 1994): a second major *trans*-SL, SL2, is used during processing of these precursors into mature mRNAs. Several other minor SLs have also been described in *C. elegans* (Blaxter and Liu 1996). The addition of the SL sequence to mRNA contributes a m⁷G cap in kinetoplastids (Campbell, Sturm, and Yu 2000) or a trimethylguanosine cap in cnidarians and nematodes (Blumenthal 1995; Stover and Steele 2001) and

influences mRNA stability and translational efficiency of *trans*-spliced mRNA (Maroney et al. 1995; Zeiner et al. 2003).

SL RNAs are less than 150 nt in length and, in addition to the SL exon, consist of a conserved 5′ splice donor site and an intron, which includes a putative spliceosome (Sm)–binding motif essential for SL RNA association with the spliceosomal complex and for the *trans*-splicing process itself (Denker et al. 1996; Ferguson, Heid, and Rothman 1996; Sturm, Yu, and Campbell 1999). To date, no significant nucleotide sequence conservation among SL RNAs from the various phyla has been described. However, in all known SL RNAs, the exon-intron boundary region is predicted to form part of a stem-loop structure (Greenbaum et al. 1996).

The lack of sequence conservation and the apparently sporadic occurrence of *trans*-splicing among different phyla have prompted debate about its evolutionary origin. The two opposing hypotheses are that either *trans*-splicing arose in an ancient eukaryote and that this facility has subsequently been lost in multiple lineages or that *trans*-splicing has evolved independently in several taxa. Current information on the phylogenetic distribution of *trans*-splicing does not allow one hypothesis to be conclusively favored over the other, but it has been argued that as the phenomenon is identified in more phyla, an ancient origin becomes increasingly more likely (Nilsen 2001; Vandenbergh et al. 2001). Similarity between primary sequences of SL RNAs from different phyla would also favor an ancestral origin, but such similarity has not been observed to date. Here we present evidence for *trans*-splicing in the phylum Rotifera, obtained during studies on the molecular basis of anhydrobiosis (extreme desiccation tolerance) in two species of bdelloid rotifers, *A. ricciae* and *Philodina* sp.

Key words: bdelloid rotifer, spliced leader, *trans*-splicing, spliced leader RNA, *hsp82* (heat shock protein), anhydrobiosis.

E-mail: at10004@biotech.cam.ac.uk.

Mol. Biol. Evol. 22(6):1482–1489. 2005

doi:10.1093/molbev/msi139

Advance Access publication March 23, 2005

Methods

Adineta ricciae (a gift of Claudia Ricci) was maintained largely as described for *Adineta vaga* (Lapinski and Tunnacliffe 2003). *Philodina* sp. was obtained from Carolina Biological Supply Company (Burlington, N. C., www.carolina.com). Upon receipt, cultures were transferred to tissue culture flasks and collected during the week following receipt. Cultures were washed in water extensively prior to harvest. Culture medium was carefully removed and replaced with either carbonated spring water (pH 7.8) or medium with elevated NaCl concentration (50–100 mM) to cause rotifers to detach from the flask. Animals were collected by centrifugation (3 min at 4,000 rpm; Heraeus Multifuge 3 S-R) in 50-ml Falcon tubes, washed, and resuspended by gentle pipetting in 500 μ l distilled water (dH₂O). Rotifers were left for a few hours to settle to minimize stress to the animals and then collected on a Nitex nylon filter (20 μ m pore size), covered with a second Nitex filter and placed between two wet Whatman 3MM filter papers. Samples were air-dried in a petri dish at 26°C for 6 or 8 h for *A. ricciae* and *Philodina* sp., respectively. After drying, the Nitex membrane was transferred to a test tube, frozen in liquid nitrogen and stored at –80°C until required.

Total rotifer RNA was isolated using TRI Reagent (Sigma-Aldrich, Poole, UK) according to manufacturer's instructions. Between 20 and 40 ng of total RNA was used per first-strand complementary DNA (cDNA) synthesis reaction performed in the presence of SMART oligonucleotide (Super SMART PCR cDNA Synthesis Kit; BD Bioscience, Cowley, UK). As a result, cDNA populations were enriched in full-length cDNAs. Resulting libraries were subcloned into pCRII-TOPO vector (Invitrogen, Paisley, UK); cDNAs were sequenced by MWG Biotech (Edersberg, Germany). SL-primed cDNA libraries were constructed by conventional first-strand cDNA synthesis followed by PCR amplification with SL-based (GGCT-TATTACAACCTTACCAAG) and oligo(dT) primers. Total *A. ricciae* RNA, isolated from animals dehydrated for 6 h, served as a template in the first-strand synthesis reaction. A number of heterogeneous products in the range ~0.5 to 2.5 kb were cloned and sequenced.

Total RNA isolated from nondehydrated *A. ricciae* served as template for first-strand synthesis which was driven by a gene-specific primer (TCAGTAGAGATATC-TTCGGGA), followed by second-strand synthesis using an SL primer (see above) plus a nested 3' primer (GACAA-GATCGGCTTTTGTACATACC). PCR fragments were subcloned into pCRII-TOPO, and clones with insert size corresponding to bands visualized on agarose gels were sequenced. For comparison with the *hsp82* gene sequence, *A. ricciae* genomic DNA was isolated using Qiagen DNAeasy Tissue Kit following manufacturer's protocol and ~10 ng DNA was used per PCR reaction. The 5' primer design was based on 5' untranslated region (UTR) sequence of the long *hsp82-1* transcript (III-L) identified in this study. Primers used were 5' primer, CCGAACG-TATTCTTCCATAAG and 3' primer, CATCTTCTGCA-CCGTGATCCA.

Rotifer SL RNA was cloned using a modified protocol for RNA 3'-end ligation-mediated PCR (LM PCR) (Elbashir, Lendeckel, and Tuschl 2001; Hitchcock et al. 2004): 5 μ M of chimeric RNA-DNA adapter 5'-(5'-P)rCr-ArGdCdTdCdCdAdGdTdAdAdCdCdTdAdCdC dideoxycytidine (ddC; Dharmacon, Dallas, Tex.) was ligated to 50–100 ng of total rotifer RNA using 20 U of T4 RNA ligase (New England BioLabs, Hitchin, UK) in a final volume of 20 μ l. The 5'-end of the hybrid oligonucleotide was phosphorylated, and 3' modification (ddC) was used to block undesirable 3'-end ligations. The ligation reaction was incubated at 15°C for 1 h. The ligated RNA was phenol-chloroform purified, ethanol precipitated, and resuspended in 8 μ l of diethylpyrocarbonate H₂O. The whole ligation was used in a first-strand synthesis reaction driven by Invitrogen SuperScript III reverse transcriptase with 2.5 μ g reverse adapter primer (GGC TAG GTT ACA TGG AG). The reaction was performed at 50°C for 30 min; 2 μ l of RT mix served as template during second-strand synthesis, where 10 μ M each of reverse adapter primer and SL-specific 21-bp nested forward primer (GGC TTA TTA CAA CTT ACC AAG) were added. PCR products were resolved on 2% agarose gels. Products in the region of 100 bp were isolated with Qiaquick Gel Extraction kit (Qiagen, Crawley, UK) and cloned into pCRII-TOPO. Inserts were sequenced at the Department of Genetics, University of Cambridge. Lasergene sequence analysis software (DNASTAR, www.dnastar.com) and the Mfold server (www.bioinfo.rpi.edu/applications/mfold/old/rna; Zuker 2003) were used for sequence analysis. Pairwise alignments were performed using ClustalW (Thompson, Higgins and Gibson 1994) on the San Diego Supercomputer Center Workbench Web site (<http://workbench.sdsc.edu>) with gap-opening penalty set to 15 and gap-extension penalty set to 6.66. The nucleotide sequence data reported in this paper have been submitted to the EMBL/GenBank Data Libraries with accession numbers AY823992, AY823993, AY823994, and AY942832 through AY942842.

Results

To further investigate the phylogenetic distribution of trans-splicing, we examined cDNAs from two species of bdelloid rotifer (Bdelloidea, Rotifera). Interest in these organisms has been stimulated by their ability to reproduce asexually and to undergo anhydrobiosis. Analysis of full-length cDNA libraries from bdelloid rotifers *A. ricciae* and *Philodina* sp. revealed a number of cDNAs with a common 23-nt sequence (fig. 1). This finding suggested that rotifer mRNAs undergo SL addition at their 5'-ends. The putative SL sequences were found on cDNAs encoding a variety of proteins, including structural proteins, metabolic enzymes, and regulatory proteins. However, not all cDNAs seemed to contain the SL sequence; of 14 nonredundant, nuclear-encoded 5' cDNA ends from *A. ricciae*, 7 lacked the leader sequence. Because some of the latter group of cDNAs might represent incomplete mRNA copies, we performed reverse transcriptase-polymerase chain reaction (RT-PCR) using a primer based on SL sequence in combination with a gene-specific primer for each case. Two of the

```

GGCTTATTCAACTTACCAAGAGaaccattgtctacttctgttttctcttcaacgtagttatcaataATG... Ar1
GGCTTATTACAACCTACCAAGAG-----aaataatcaataagaagtacaaaatATG... Ar2
GGCTTATTACAACCTACCAAGAG-----attcatatataatcccccaatcttcatcgataaatATG... Ar3
GGCTTATTACAACCTACCAAGAG-----ataaatcttctatcgatctcggttgaaaccagcATG... Ar4
GGCTTATTACAACCTACCAAGAG-----atcatagaatttcttctgctgcaacaATG... Ar5
GGCTTATTACAACCTACCAAGAG-----attgtttgactaacctatttctgtcaacaATG... Ar6
GGCTTATTACAACCTACCAAGAGataaattatataaataatcaaatgtttataa...->+109...tcaATG... Ar7
GGCTTATTACAACCTACCAAGAG-----atcgaaatccggttgaaagacggataaagaaaaATG... P1
GGCTTATTACAACCTACCAAGAG-aagtttacaagtttagacgacaaatcttcttctattcaatcaattATG... P2
GGCTTATTACAACCTACCAAGAG-----aaagataaaacaagacataaaagcaattaaaATG... P3
GGCTTATTACAACCTACCAAGAG-----ATG... P4

```

FIG. 1.—A common sequence at the 5'-ends of multiple bdelloid rotifer mRNAs. The common 23-nt sequence at the 5' termini of bdelloid rotifer (*Adineta ricciae*: Ar₍₁₋₇₎) and *Philodina* sp.: P₍₁₋₄₎) mRNAs is shown in upper case; divergent downstream sequences are in lower case. The translation start codon is indicated in bold upper case. Single nucleotide transitions within the SL sequence are highlighted. Closest matches of SL-containing transcripts were determined by BlastX; the nucleotide sequence data presented in the figure have been submitted to the EMBL/GenBank Data Libraries with following accession numbers: Ar1 (AY942832), similar to RAD23 homolog B (*Danio rerio*); Ar2 (AY942833), similar to unnamed protein product CAF98644 (*Tetraodon nigroviridis*); Ar3 (AY942834), similar to AAS91553/AmphiHMG1/2 (*Branchiostoma belcheri tsingtaunense*); Ar4 (AY942835), similar to ribosomal protein L18a (*Canis familiaris*); Ar5 (AY942836), similar to 60S ribosomal protein L37 (*Ixodes pacificus*); Ar6 (AY942837), similar to ribosomal protein S7 (*Petromyzon marinus*); Ar7 (AY942838), similar to a DnaJ-like protein (AAH75905) from *D. rerio*; P1 (AY942839), similar to S14e ribosomal protein (*Dascillus cervinus*); P2 (AY942840), similar to fba1 (*Schizosaccharomyces pombe*); P3 (AY942841), similar to guanine nucleotide-binding protein beta polypeptide 2-like 1 (*D. rerio*); and P4 (AY942842), similar to putative 42-9-9 protein (*Mus musculus*).

7 sequences were successfully amplified, showing that at least 9 out of 14 *A. ricciae* mRNAs contain the SL sequence. Similar estimates were obtained for *Philodina* sp. cDNAs, where four out of eight nonredundant clones apparently represented *trans*-spliced mRNAs. Therefore, at least 50%–65% of bdelloid rotifer mRNAs were found to contain a SL. In three cases, single nucleotide transitions were detected within the SL sequence (fig. 1). This might represent some variation within the SL exon in multiple copies of the SL RNA gene but will require further analysis for confirmation. To date, however, we have not observed a major SL variant, like SL2 of *C. elegans*, in rotifers.

Intriguingly, Blast searches revealed that a number of full-length cDNAs derived from rice (*Oryza sativa*) cDNA libraries (Kikuchi et al. 2003) contained a sequence identical to that found at the 5'-end of rotifer cDNAs (table 1). Ten such clones were identified; they encode a range of proteins with excellent matches to sequences from various species, the best of which is listed in the column "Organism" of table 1. However, none of these sequences matched closely with homologous plant genes. Because a *trans*-splicing pathway has not been reported in plants, it is likely that the SL-containing cDNAs represent foreign—probably rotifer—material in rice libraries. In support of this, we compared one of the sequences of table 1 (no. 4), encoding a 60S ribosomal protein L37, with *A. ricciae* cDNA for the

same protein (Ar5; fig. 1). ClustalW analysis showed a percent identity of 85.5% at the nucleotide level, indicating that the expressed sequence tag in question is likely to originate from a rotifer, although probably a different species. If correct, this would suggest the occurrence of *trans*-splicing in a third (or more) rotifer species. Interestingly, one of the "rice" cDNAs (no. 10, table 1) contains a SL variant identical to that seen in *A. ricciae* clone Ar1 (fig. 1).

As first realized for nematodes, a conserved 5' sequence can be used to construct full-length cDNA libraries (Bektesh and Hirsh 1998). To demonstrate this principle in rotifers, RT-PCR was performed on poly(A)-containing mRNA using both SL and oligo(dT) primers. A selection of complete cDNAs was generated in this way, as determined by the presence of a major open reading frame just downstream of the SL sequence and a poly(A) tract at the 3'-end, and included sequences encoding a variety of proteins (data not shown).

To provide additional evidence for *trans*-splicing in bdelloid rotifers, we analyzed transcript copies from a *hsp82* (82 kDa heat shock protein) gene in *A. ricciae*. A single partial *hsp82* genomic sequence from *A. ricciae* is recorded in sequence databases (accession number AY394701); the corresponding gene is termed here *hsp82-1*. To map the 5'-ends of its transcripts, we performed gene-specific RT-PCR using the SL sequence as 5' primer; gene-specific 3' primers were designed from

Table 1
Nearest BlastX Matches ("Hit ID") with Nonredundant Sequences from a Rice Expressed Sequence Tag Database Containing Bdelloid Rotifer SL-like 5' Additions

Clone ID	Leader Sequence	Hit ID (BlastX)	e Value	Organism	
1	AK100559	GGCTTATTACAACCTACCAAGAG	Glycogen synthase	0.0	<i>Anopheles gambiae</i>
2	AK071809	GGCTTATTACAACCTACCAAGAG	Hypothetical protein MGC75638	1×10^{-167}	<i>Xenopus tropicalis</i>
3	AK069174	GGCTTATTACAACCTACCAAGAG	Unnamed protein product	3×10^{-8}	<i>Tetraodon nigroviridis</i>
4	AK063177	GCTTATTACAACCTACCAAGAG	Probable 60S ribosomal protein L37	1×10^{-28}	<i>Drosophila melanogaster</i>
5	AK063233	-GCTTATTACAACCTACCAAGAG	Cytochrome c oxidase subunit VIc	1×10^{-5}	<i>Saimiri sciureus</i>
6	AK063098	-GCTTATTACAACCTACCAAGAG	Proliferating cell nuclear antigen	8×10^{-92}	<i>Rattus norvegicus</i>
7	AK062639	-GCTTATTACAACCTACCAAGAG	No significant similarity		
8	AK062512	-GCTTATTACAACCTACCAAGAG	Proteasome subunit N3	2×10^{-49}	<i>Oncorhynchus mykiss</i>
9	AK062346	-GCTTATTACAACCTACCAAGAG	60S ribosomal protein L15	5×10^{-68}	<i>Anguilla japonica</i>
10	AK100502	GGCTTATTGCAACTACCAAGAG	Unnamed protein product	9×10^{-42}	<i>Tetraodon nigroviridis</i>

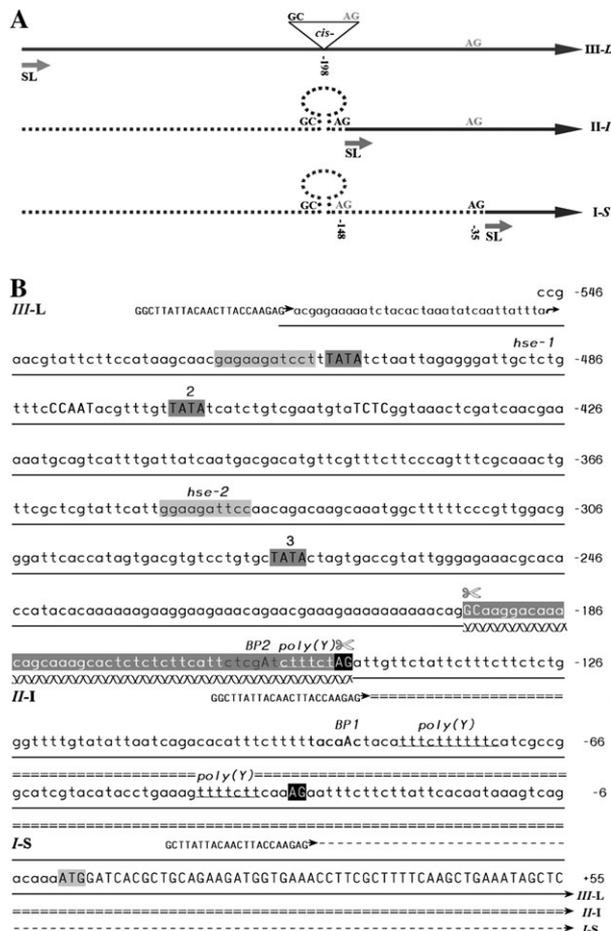


FIG. 2.—Mapping 5'-ends of *Adineta ricciae* *hsp82-1* transcripts. (A) Schematic representation of alternative SL-spliced *hsp82-1* mRNA 5' UTRs, signified by cDNAs I-S (341 bp), II-I (456 bp), and III-L (842 bp). The III-L sequence matches the genomic version exactly, apart from the 48-bp *cis*-spliced intron represented as a triangle. Arrows labeled SL indicate SL positions within transcripts. Splice acceptor dinucleotides (AG) involved in *trans*-splicing are in black; nonparticipating acceptors are dimmed. (B) Alignment of cDNAs against *hsp82-1* 5' UTR (accession number AY823992). Sequence of the *hsp82-1* 5' UTR and part of the coding sequence (upper case) are shown. I-S, II-I, and III-L SLs are presented as SL sequence itself followed by an arrow at the head of lines depicting the extent of transcripts. Both AG acceptor sites are in capital letters and shaded in black; pyrimidine tracks (*poly*(Y)) are underlined; the invariant A nucleotide within branch point sites BP1 and BP2 is in bold upper case. HSE motifs (*hse-1* and *hse-2*) are highlighted in light gray. Both a CCAAT general transcriptional factor motif and a GAGA box/CT-element are in bold upper case. TATA boxes (1, 2, and 3) are highlighted in dark gray. The 48-bp intron within the *hsp82-1* 5' UTR is indicated with a double-helix and symbols and highlighted in gray.

the AY394701 sequence. Resulting PCR products were cloned and four isolates with insert sizes corresponding to amplified fragments were analyzed. One of the four clones represented a *hsp82* gene, termed *hsp82-2*, which differs from the previously reported *hsp82-1* sequence (8 nt substitutions within the coding region). The remaining three clones, however, although different in length, were all derived from *hsp82-1* and contained the SL sequence at their 5' termini. They are named as follows: I-S for the short (341 bp) product, II-I for the intermediate size (456 bp), and

III-L for the longest (842 bp) clone (fig. 2A). All three clones included the previously unreported 5'-end of the *hsp82-1* coding region (48 bp), together with different extents of the upstream UTR (5' UTR); genomic sequence corresponding to most of III-L was also obtained after PCR amplification from *A. ricciae* genomic DNA.

I-S, II-I, and III-L all share the same sequence (−33 to +1) upstream of the initiation codon (ATG) and the −148 to −34 region is identical in both II-I and III-L (fig. 2B). Upstream of position −33, the gene sequence contains a splice acceptor site (AG) at position −35, a splice intermediate branch point (BP1: TACAAC) at position −89, and two pyrimidine tracts between them, consistent with *trans*-splicing of the SL exon to abut position −33, as seen in cDNA clone I-S. Similarly, the SL sequence of cDNA clone II-I adjoins sequence at position −148 immediately following a splice acceptor site, which is preceded by a pyrimidine tract and near consensus branch point (BP2: CTCGAT), all located within a 48-bp intron of the *hsp82-1* gene. This intron, which has a rare (GC; position −196) splice donor sequence (Burset, Seledtsov, and Solovyev 2001), has been removed in cDNA clone III-L, demonstrating that the splice acceptor at −150 can be used for either *trans*- or *cis*-splicing. As expected, if *trans*-splicing is occurring, the SLs within cDNAs I-S and II-I were not present in the *hsp82-1* genomic sequence. The three forms of *hsp82-1* cDNAs could be due to either alternative *trans*-splicing of the SL exon to different splice acceptors in the same long precursor mRNA or to the use of alternative promoters generating different species of pre-mRNA, or both. Several putative promoter elements were identified within the 5' UTR of *hsp82-1* genomic DNA. These include two heat shock element (HSE) motifs (*hse-1* and *hse-2* in fig. 2), a CCAAT box, a GAGA box/CT-element, and a number of TATA boxes. Mapping of transcription initiation sites will be required to distinguish between these possibilities.

Individual cDNAs corresponding to SL RNAs of both *A. ricciae* and *Philodina* sp. were generated by 3'-end RNA LM PCR. Sequence analysis of independent cloned LM PCR products showed that rotifer SL RNA can be either 105 or 106 nt long and includes the SL exon itself (23 nt), a splice donor dinucleotide (GT), and a putative Sm-binding motif (fig. 3A). Five variants of the Sm-like motif were identified with a consensus AAYUYUGA (where Y is a pyrimidine), similar to the consensus, AGCUUUGG, of the primitive chordate *C. intestinalis* (Vandenbergh et al. 2001). A number of species-specific differences are also discernible within the introns of *A. ricciae* and *Philodina* sp. SL RNAs. For instance, in *Philodina* sequences there were several nucleotide substitutions within the putative Sm-binding site and near the 3'-end of the intron, which were never present at the same positions in *Adineta* cDNAs. Similarly, a major variant seen in several *A. ricciae* SL RNAs was not found in *Philodina* sp. This variant showed an interesting feature, i.e., complementary nucleotide transitions at two independent sites: ATC → GCT and GAT → AGC in the 75–92 nt SL RNA region, together with G → A and C → T at positions 81 and 86. These changes have implications for SL RNA secondary structure (see below).

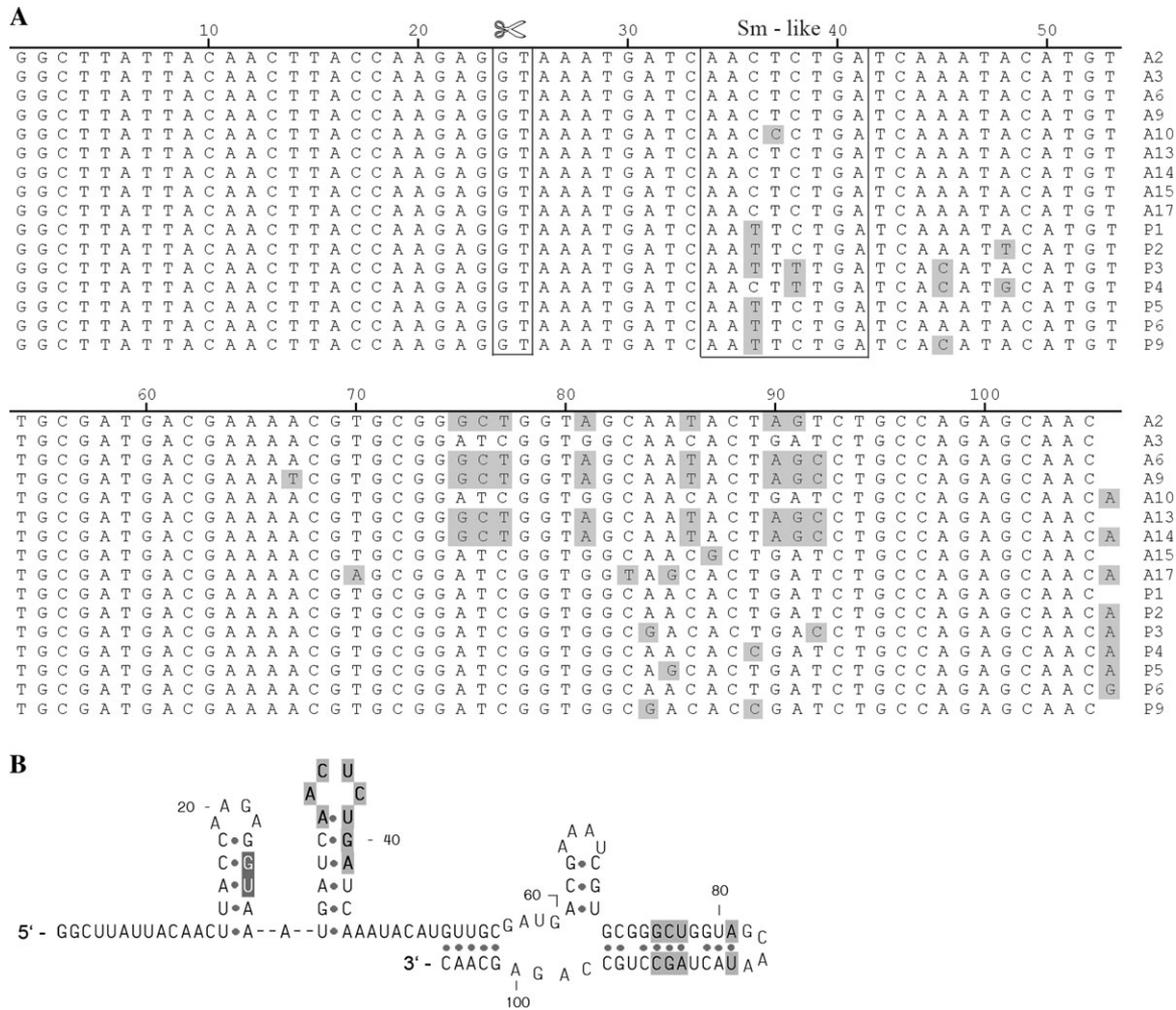


FIG. 3.—Rotifer SL RNA sequences and secondary structure prediction. (A) Independent cDNA clones from *Adineta ricciae* (A series; A9: accession number AY823993) and *Philodina* sp. (P series; P2: accession number AY823994). Features include SL exon (bases 1–23), splice donor dinucleotide GT (bases 24–25, boxed), and putative Sm-binding motif (bases 34–41, boxed). Variant bases are shaded. (B) Rotifer SL RNA secondary structure prediction of *A. ricciae* SL RNA clone A9 performed using Mfold without folding constraints; output is manually edited for presentation. Main hairpins are labeled I, II, and III. Splice donor dinucleotide is shown in white lettering in hairpin I, Sm-like-binding motif is shaded in hairpin II, and complementary sequence variants are shaded in hairpin III.

Secondary structure prediction of rotifer SL RNA was performed using Version 3.1 of Mfold without folding constraints (Zuker 2003). Rotifer SL RNA was predicted to fold into three stem loops, two of which were located towards the 5'-end of the molecule (fig. 3B). Part of the SL exon and the splice donor dinucleotide (GU) were located in the double-stranded region of hairpin I, similar to that predicted for all other SL RNAs, while the putative Sm-binding site was found in the loop region of hairpin II. Lowering the temperature from the default 37°C to 26°C, at which rotifers are grown, had no effect on predicted folding. In most SL RNAs described to date, secondary structure predictions locate the possible Sm-binding motif to a single-stranded region, but this feature is imposed in Mfold by the investigator. We chose not to apply such a constraint to rotifer SL RNA because the proposed Sm-binding site is located in a more 5' position compared to analogous motifs of all other SL RNAs with the exception of that of *C. intestinalis*. Forcing the Sm motif region to adopt a single-

stranded structure led to the abolishment of hairpin II and part of hairpin III and resulted in unlikely base pairing between the 5' - and 3'-ends of rotifer SL RNA. The complementary transitions in the variant *A. ricciae* sequence mentioned above are located within hairpin III and are predicted to base pair with each other. The maintenance of base pairing in hairpin III of both major SL RNA variants in *A. ricciae* is consistent with the predicted structural role for these residues.

A noteworthy feature of rotifer SL RNAs is their resemblance to those of nematodes. The SL exon itself, at 23 nt is almost identical in size to the 22 nt of all major nematode SLs, whereas SL sequences from other organisms vary in size from 16 to 51 nt; a three stem-loop structure is also predicted for both rotifer and nematode SL RNAs. Interestingly, their sequences also appear to be related: an alignment of the *A. ricciae* SL RNA sequence with SL1 RNA from *C. elegans* (fig. 4A) showed similarity in the SL exon itself, around the exon-intron junction and



FIG. 4.—Comparison of bdelloid rotifer and nematode SL RNAs. ClustalW alignment of *Adineta ricciae* clone A2 SL RNA (as DNA sequence; abbreviated Ar SL) with (A) *Caenorhabditis elegans* (abbreviated Ce) SL1 and (B) *Oscheius* sp. CEW1 (abbreviated Os) SL1 RNAs. (C) Alignment of *C. elegans* SL1 and SL2 RNAs. Matching residues are shaded gray; the SL exon is boxed.

also throughout the intron, and gave a ClustalW score of 32 (indicating percent sequence identity adjusted for gap penalties; see *Methods*). Alignments of rotifer SL RNA with SL1 RNA from other nematodes gave similar results, e.g., with SL1A RNA of *Oscheius* (formerly *Dolichorhabditis*) sp. CEW1 (Evans et al. 1997; fig. 4B), which gives a score of 26. The similarity between rotifer SL and nematode SL1 RNA sequences is comparable to that between *C. elegans* SL1 and SL2 RNAs, which gives a score of 21 in the alignment of figure 4C and between *C. elegans* and *Oscheius* sp. SL1 RNAs, with a score of 29 (alignment not shown). Although intriguing, these findings do not demonstrate homology between SL RNA genes from bdelloid rotifers and nematodes; it remains possible that the similarity observed is due to chance. If so, we might expect alignments of rotifer SL RNA sequences with those of some nematodes to give a considerably lower percentage identity. Indeed, the *Wuchereria bancrofti* SLG1 RNA (Dassanayake, Chandrasekharan, and Karunanayake 2001) gives a score of only 9 after alignment with *A. ricciae* SL RNA. Comparisons of rotifer SL RNA sequences with those of nonnematode species, e.g., cestode or larvacean, showed no similarity (data not shown).

Discussion

Bdelloid rotifers, one of three classes of the phylum Rotifera, have several features of interest to geneticists and biochemists, including obligate asexual reproduction (Mark Welch and Meselson 2000) and anhydrobiosis, a reversible state of metabolic arrest induced by desiccation (reviewed in Tunnacliffe and Lapinski 2003). During studies on the molecular basis of anhydrobiosis in bdelloid rotifers, we found a number of full-length cDNAs to contain a common sequence at the 5' termini, which is diagnostic of SL RNA-mediated *trans*-splicing. The 23-nt SL sequence was attached to a range of different cDNAs from two bdelloid species grown in the laboratory and probably

one or more other species, which contaminate rice plants. Because at least 50%–65% of cDNA sequences analyzed contained the SL exon, *trans*-splicing is an important mRNA-processing pathway in bdelloid rotifers.

Evidence for the complexity of *trans*-splicing in bdelloid rotifers was obtained through analysis of *A. ricciae hsp82-1* cDNAs. The *hsp82* gene has been used in rotifers for studies on phylogenetic relationships and asexual reproduction: in four bdelloid genomes examined, two or more copies of *hsp82* have been identified (Mark Welch and Meselson 2000). Cytogenetic analysis of *Philodina roseola* revealed the presence of four *hsp82* genes with each copy on a different chromosome (Mark Welch, Mark Welch, and Meselson 2004). These can be grouped as two gene pairs according to sequence similarity: copies 1 and 2 differ by 3.5% and copies 3 and 4 by 6%, but the former differ from the latter by ~47%. To date, only one *hsp82* sequence from *A. ricciae* has been deposited in databases, but cDNAs from this gene, *hsp82-1*, revealed features characteristic of SL *trans*-splicing and showed that the SL exon could be spliced to three different positions in the *hsp82-1* 5' UTR. Differential *trans*-splicing patterns have also been reported in trypanosomes (Mair et al. 2000; Nepomuceno-Silva et al. 2001; Manning-Cela, Gonzalez, and Swindle 2002) and can be explained either by alternative *trans*-splicing of the same pre-mRNA or by use of alternative promoters. Distinguishing between these possibilities will require accurate promoter mapping experiments, but whichever is correct, the long *hsp82-1* (III-L) transcript identified must be a product of both conventional *cis*-splicing and SL-dependent *trans*-splicing. The former relies on the recruitment of a rare GC donor site, which shows a near perfect match with the consensus exon-intron sequence MAGGCAAGT found in mammalian GC-AG type introns (Burset, Seledtsov, and Solovyev 2001). The splice acceptor in this intron can also participate in *trans*-splicing which gives rise to II-I type transcripts. Such “dual-purpose” splice acceptors have been observed in the *Trypanosoma brucei* PAP (poly(A) polymerase) gene (Mair et al. 2000) and a trehalose synthase gene of the anhydrobiotic nematode, *Aphelenchus avenae* (Goyal et al. 2005). Interestingly, a second copy of the *hsp82* gene in *A. ricciae*, *hsp82-2*, which is very similar to *hsp82-1* in the 5' UTR, does not contain a splice donor site in the equivalent position, where its sequence is AC instead of GC; *hsp82-2* also lacks the most downstream splice acceptor dinucleotide (deletion of AG at position -35; fig. 2) necessary for formation of a transcript like I-S. This suggests that fewer variant transcripts will result from *trans*-splicing of *hsp82-2* pre-mRNAs and, indeed, only cDNAs equivalent to II-I of *hsp82-1* have been obtained to date (unpublished data).

The discovery of *trans*-splicing in Rotifera extends the known phylogenetic range of this phenomenon to five metazoan phyla. In itself, this increases the likelihood of an ancient origin for *trans*-splicing, coupled with its loss in several lineages, because it is arguably less stringent to lose a biological function than to gain one, as the opposing hypothesis demands. The apparent similarity between rotifer and nematode SL RNAs is intriguing; one explanation might be that the sequences are homologous and that therefore

trans-splicing is ancestral to the divergence between their respective evolutionary lineages. However, previously, SL RNAs from different phylogenetic groups have shown no relatedness (Nilsen 2001) and, indeed, considerable diversity of SL RNA sequences can exist even within a single phylum, i.e., Platyhelminthes (Davis 1997), assuming this group is monophyletic. Given this degree of interphyletic and intraphyletic diversity, it seems unlikely that SL RNA sequence would be conserved between rotifers and nematodes. More likely is that the sequence similarity arose by chance. Furthermore, apart from patches of matching sequence in the 5'-end of the SL exon and around the splice donor site, which might be due to functional constraints, the remaining identities are scattered throughout the intron region in the rotifer-nematode SL RNA alignments and often occur at different positions in the *C. elegans* and *Oscheius* sp. CEW1 sequences (figs. 4A,B). Further difficulties become apparent when current interpretations of animal phylogeny are taken into account, because rotifers and nematodes are placed in separate taxonomic groups of Bilateria, i.e., Lophotrochozoa and Ecdysozoa, respectively (Halanych 2004). Rotifera and Nematoda are widely separated in evolution, therefore, and if their SL RNA genes were homologous, we might expect to see a comparable or greater degree of identity between corresponding sequences of rotifers and a nearer evolutionary neighbor. However, although flatworms (Platyhelminthes) are grouped with rotifers in Platyzoa, Lophotrochozoa, there is no similarity between their respective SL RNA sequences.

At present, therefore, the significance of the apparent similarity between rotifer and nematode SL RNA sequences is difficult to assess and clearly its interpretation would be better informed with more data on the phylogenetic distribution of *trans*-splicing, as called for by Nilsen (2001). In the context of bdelloid rotifers, it would be of immediate interest to look for *trans*-splicing in the other main group of rotifers, the monogononts, and in the closely related acanthocephalans (Garey et al. 1996; Mark Welch 2001). A systematic approach of this kind should result in more rapid progress of our understanding of *trans*-splicing evolution.

Acknowledgments

We would like to thank Claudia Ricci (University of Milan) for providing *Adineta ricciae*. This work was funded by BBSRC grant 8/S19912. A.T. is the Anglian Water Fellow in Biotechnology of Pembroke College, Cambridge.

Literature Cited

- Bektesh, S. L., and D. I. Hirsh. 1988. *C. elegans* mRNAs acquire a spliced leader through a *trans*-splicing mechanism. *Nucleic Acids Res.* **16**:5692.
- Blaxter, M., and L. Liu. 1996. Nematode spliced leaders-ubiquity, evolution and utility. *Int. J. Parasitol.* **26**:1025–1033.
- Blumenthal, T. 1995. *Trans*-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet.* **11**:132–136.
- Burset, M., I. A. Seledtsov, and V. V. Solovyev. 2001. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res.* **29**:255–259.
- Campbell, D. A., N. R. Sturm, and M. C. Yu. 2000. Transcription of the kinetoplastid spliced leader RNA gene. *Parasitol. Today* **16**:78–82.
- Dassanayake, R. S., N. V. Chandrasekharan, and E. H. Karunanayake. 2001. *Trans*-spliced leader RNA, 5S-rRNA genes and novel variant orphan spliced-leader of the lymphatic filarial nematode *Wuchereria bancrofti*, and a sensitive polymerase chain reaction based detection assay. *Gene* **269**:185–193.
- Davis, R. E. 1997. Surprising diversity and distribution of spliced leader RNAs in flatworms. *Mol. Biochem. Parasitol.* **87**:29–48.
- Davis, R. E., H. Singh, C. Botka, C. Hardwick, M. Ashraf el Meanawy, and J. Villanueva. 1994. RNA *trans*-splicing in *Fasciola hepatica*. Identification of a spliced leader (SL) RNA and SL sequences on mRNAs. *J. Biol. Chem.* **269**:20026–20030.
- Denker, J. A., P. A. Maroney, Y. T. Yu, R. A. Kanost, and T. W. Nilsen. 1996. Multiple requirements for nematode spliced leader RNP function in *trans*-splicing RNA. *RNA* **2**:746–755.
- Elbashir, S. M., W. Lendeckel, and T. Tuschl. 2001. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**:188–200.
- Evans, D., D. Zorio, M. MacMorris, C. E. Winter, K. Lea, and T. Blumenthal. 1997. Operons and SL2 *trans*-splicing exist in nematodes outside the genus *Caenorhabditis*. *Proc. Natl. Acad. Sci. USA* **94**:9751–9756.
- Ferguson, K. C., P. J. Heid, and J. H. Rothman. 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.
- Ganot, P., T. Kallesoe, R. Reinhardt, D. Chourrout, and E. M. Thompson. 2004. Spliced-leader RNA *trans*-splicing in a chordate, *Oikopleura dioica*, with a compact genome. *Mol. Cell. Biol.* **24**:7795–7805.
- Garey, J. R., T. J. Near, M. R. Nonnemacher, and S. A. Nadler. 1996. Molecular evidence for Acanthocephala as a sub-taxon of Rotifera. *J. Mol. Evol.* **43**:287–292.
- Goyal, K., J. A. Browne, A. M. Burnell, and A. Tunnacliffe. 2005. Dehydration-induced *tps* gene transcripts from an anhydrobiotic nematode contain novel spliced leaders and encode atypical GT-20 family proteins. *Biochimie* (in press).
- Greenbaum, N. L., I. Radhakrishnan, D. J. Patel, and D. Hirsh. 1996. Solution structure of the donor site of a *trans*-splicing RNA. *Structure* **4**:725–733.
- Halanych, K. M. 2004. The new view of animal phylogeny. *Annu. Rev. Ecol. Syst.* **35**:229–256.
- Hitchcock, R. A., G. M. Zeiner, N. R. Sturm, and D. A. Campbell. 2004. The 3' termini of small RNAs in *Trypanosoma brucei*. *FEMS Microbiol. Lett.* **236**:73–78.
- Kikuchi, S., K. Satoh, T. Nagata et al. (70 co-authors). 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science* **301**:376–379.
- Krause, M., and D. Hirsh. 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49**:753–761.
- Lapinski, J., and A. Tunnacliffe. 2003. Anhydrobiosis without trehalose in bdelloid rotifers. *FEBS Lett.* **553**:387–390.
- Mair, G., H. Shi, H. Li et al. (14 co-authors). 2000. A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA. *RNA* **6**:163–169.
- Manning-Cela, R., A. Gonzalez, and J. Swindle. 2002. Alternative splicing of LYT1 transcripts in *Trypanosoma cruzi*. *Infect. Immun.* **70**:4726–4728.
- Mark Welch, D. B. 2001. Early contributions of molecular phylogenetics to understanding the evolution of Rotifera. *Hydrobiologia* **446/447**:315–322.

- Mark Welch, D. B., and M. Meselson. 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* **288**:1211–1215.
- Mark Welch, J. L., D. B. Mark Welch, and M. Meselson. 2004. Cytogenetic evidence for asexual evolution of bdelloid rotifers. *Proc. Natl. Acad. Sci. USA* **101**:1618–1621.
- Maroney, P. A., J. A. Denker, E. Darzynkiewicz, R. Laneve, and T. W. Nilsen. 1995. Most mRNAs in the nematode *Ascaris lumbricoides* are trans-spliced: a role for spliced leader addition in translational efficiency. *RNA* **1**:714–723.
- Murphy, W. J., K. P. Watkins, and N. Agabian. 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans-splicing. *Cell* **47**:517–525.
- Nepomuceno-Silva, J. L., K. Yokoyama, L. D. de Mello et al. (11 co-authors). 2001. TcRho1, a farnesylated Rho family homologue from *Trypanosoma cruzi*. *J. Biol. Chem.* **276**:29711–29718.
- Nilsen, T. W. 2001. Evolutionary origin of SL-addition trans-splicing: still an enigma. *Trends Genet.* **17**:678–680.
- Rajkovic, A., R. E. Davis, J. N. Simonsen, and F. M. Rottman. 1990. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc. Natl. Acad. Sci. USA* **87**:8879–8883.
- Stover, N. A., and R. E. Steele. 2001. Trans-spliced leader addition to mRNAs in a cnidarian. *Proc. Natl. Acad. Sci. USA* **98**:5693–5698.
- Sturm, N. R., M. C. Yu, and D. A. Campbell. 1999. Transcription termination and 3'-end processing of the spliced leader RNA in kinetoplastids. *Mol. Cell. Biol.* **19**:1595–1604.
- Sutton, R. E., and J. C. Boothroyd. 1986. Evidence for trans-splicing in trypanosomes. *Cell* **47**:527–535.
- Tessier, L. H., M. Keller, R. L. Chan, R. Fournier, J. H. Weil, and P. Imbault. 1991. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in *Euglena*. *EMBO J.* **10**:2621–2625.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- Tunnacliffe, A., and Lapinski, J. 2003. Resurrecting Van Leeuwenhoek's rotifers: a reappraisal of the role of disaccharides in anhydrobiosis. *Philos. Trans. R. Soc. Lond. B* **358**:1755–1771.
- Vandenbergh, A. E., T. H. Meedel, and K. E. M. Hastings. 2001. mRNA 5'-leader trans-splicing in the chordates. *Genes Dev.* **15**:294–303.
- Zeiner, G. M., N. R. Sturm, and D. A. Campbell. 2003. The *Leishmania tarentolae* spliced leader contains determinants for association with polysomes. *J. Biol. Chem.* **278**:38269–38275.
- Zorio, D. A. R., N. N. Cheng, T. Blumenthal, and J. Spieth. 1994. Operons as a common form of chromosomal organization in *C. elegans*. *Nature* **372**:270–272.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–3415.

Billie Swalla, Associate Editor

Accepted March 16, 2005