

# RNA Trans-splicing in Flatworms

ANALYSIS OF TRANS-SPLICED mRNAs AND GENES IN THE HUMAN PARASITE, *SCHISTOSOMA MANSONI*\*

(Received for publication, April 14, 1995, and in revised form, June 28, 1995)

Richard E. Davis‡, Cara Hardwick, Paul Tavernier§, Scott Hodgson, and Hardeep Singh

From the Department of Biology, San Francisco State University, San Francisco, California 94132 and the

§Department of Chemistry, Bethel College, St. Paul, Minnesota 55112

**Characteristics of trans-splicing in *Schistosoma mansoni* were examined to explore the significance and determinants of spliced leader (SL) addition in flatworms. Only a small subset of mRNAs acquire the SL. Analysis of 30 trans-spliced mRNAs and four genes revealed no discernable patterns or common characteristics in the genes, mRNAs, or their encoded proteins that might explain the functional significance of SL addition. While the mRNA encoding the glycolytic enzyme enolase is trans-spliced, mRNAs encoding four other glycolytic enzymes are not, indicating trans-splicing is not prevalent throughout this metabolic pathway. Although the 3' end of flatworm SLs contribute an AUG to mRNAs, the SL AUG does not typically serve to provide a methionine for translation initiation of reading frames in recipient mRNAs. SL RNA expression exhibits no apparent sex, tissue, or cell specificity. Trans-spliced genes undergo both cis- and trans-splicing, and the sequence contexts for these respective acceptor sites are very similar. These results suggest trans-splicing in flatworms is most likely associated either with some property conferred on recipient mRNAs by SL addition or related to some characteristic of the primary transcripts or transcription of trans-spliced genes.**

Trans-splicing is an RNA processing event that accurately joins sequences derived from independently transcribed RNAs. In one form of trans-splicing, a leader sequence (the spliced leader, SL)<sup>1</sup> is donated from the 5' end of a small, non-polyadenylated RNA (the spliced leader RNA, SL RNA) to pre-mRNAs to form the 5'-terminal exon of mature mRNAs (for recent reviews see Refs. 1-6). This form of RNA maturation was first described in trypanosomes (7, 8) and subsequently in other kinetoplastida and the flagellated protozoan *Euglena* (9). The identification of trans-splicing in two divergent invertebrate phyla, first in nematodes (10) and then in flatworms (11), suggests that this particular form of RNA processing may be an important form of gene expression common in early metazoa.

The general distribution of trans-splicing and its origin in

metazoa is currently not known. Furthermore, both the origin of early metazoan groups and the phylogenetic relationships between flatworms, nematodes, and other early invertebrates have been difficult to delineate (12, 13). Trans-splicing may have arisen independently in several invertebrate lineages (6) and, if true, the characteristics and functional significance of spliced leader addition might also be different in diverse metazoan groups. Trans-splicing is of particular interest in flatworms (Phylum Platyhelminthes) as these metazoa may represent the earliest bilateral animals, and one possible evolutionary tree places a flatworm-like ancestor as the progenitor of a number of other early invertebrate groups (12, 13). We have recently shown that trans-splicing is present in diverse trematode flatworms and in a predominantly free-living group generally considered to represent primitive flatworms (14).<sup>2</sup> This suggests that spliced leader addition may have been present in the flatworm progenitor and in the ancestors of parasitic flatworms.

The primary function(s) of most trans-splicing in metazoa remains unknown. We have analyzed several characteristics of spliced leader addition in the flatworm *Schistosoma mansoni* to explore the biological significance of trans-splicing in flatworms and to provide a comparative metazoan perspective. We previously noted that not all mRNAs acquire the spliced leader in schistosomes (11). In the present study, we identified and partially characterized 30 mRNAs and four genes that are trans-spliced in *S. mansoni* to increase our understanding of the molecular characteristics and general properties of trans-splicing in flatworms. The mRNAs were examined to determine 1) if there are any discernable patterns in the proteins they encode, 2) if mRNAs in a particular pathway are trans-spliced as a group, 3) if any other general characteristics of trans-spliced mRNAs were evident, and 4) if the AUG conserved at the 3' end of all flatworm SLs (11, 14) provides the methionine for translation initiation of recipient mRNAs. Genes coding for trans-spliced mRNAs were analyzed to investigate the general organization of these genes and for conserved elements associated with the trans-splice acceptor sites that might distinguish these sites from cis-splice acceptor sites or facilitate bringing the SL RNA and pre-mRNA substrates together for trans-splicing. Finally, the expression of the SL RNA and several trans-spliced mRNAs was also examined by *in situ* hybridization in adult worms to determine if there is any possible sex, tissue, or cell specificity in trans-splicing.

Our results described herein suggest that the functional significance of flatworm trans-splicing does not appear to be correlated with specific types of mRNAs or the proteins they encode nor with restricted expression of the SL RNA to specific

\* This work was supported by National Institutes of Health Grant AI 32709. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) U30175-U30183, U30258-U30266, and U30291.

‡ To whom correspondence should be addressed: Dept. of Biological Sciences, Fordham University, Bronx, NY 10458. Fax: 718-817-3645; Email: rdavis@murray.fordham.edu.

<sup>1</sup> The abbreviations used are: SL, spliced leader; SL RNA, spliced leader RNA; ORF, open reading frame; PCR, polymerase chain reaction; TMG, trimethylguanosine.

<sup>2</sup> R. E. Davis, C. Botka, J. Villanueva, and C. Hardwick, manuscript in preparation.

cells, tissues, or sex. This suggests that the functional significance of trans-splicing in flatworms is more likely associated either with properties conferred on recipient mRNAs by addition of the spliced leader or related to the characteristics of transcription and the primary transcripts of trans-spliced genes.

#### MATERIALS AND METHODS

**Organisms**—Mice infected with *S. mansoni* and adult worms were kindly provided by Ron Blanton (Department of Geographic Medicine, Case Western Reserve University) and George Newport (University of California at San Francisco).

**Nucleic Acid Isolation and Blot Analyses**—Genomic DNA was isolated from frozen worms powdered on dry ice as described (15). Total RNA was purified either by guanidinium-hot phenol (15) or acid-phenol extraction (16). Poly(A)<sup>+</sup> mRNAs were selected either by oligo(dT)-cellulose (15) or biotinylated oligo(dT) and streptavidin-coated paramagnetic particles (Poly(A)Tract mRNA Isolation, Promega, Madison, WI). Agarose Northern blots, genomic Southern blots, probe preparation, and hybridization and washing conditions were as described (11, 15).

**Primer Extension Analysis and Rapid Amplification of cDNA Ends (RACE)**—Primer extension was performed as described (15) using end-labeled *HpaII* pBR322 and DNA sequencing reactions as molecular size markers. 5'-RACE was performed with the 5'-RACE System from Life Technologies, Inc. (according to the manufacturer's instructions) and as described previously (17). PCR products were directly sequenced using end-labeled nested primers and the fmol DNA Sequencing System (Promega, Madison, WI). For additional sequence analysis (18), the PCR products were either cloned into a pT7Blue T-vector (19) (Novagen, Madison, WI) or Bluescribe/Bluescript plasmid vectors (Stratagene Cloning Systems, La Jolla, CA). The synaptobrevin cDNA isolated from one of the SL-enriched libraries lacked the 3' end of the mRNA based on comparative analysis with other synaptobrevin mRNAs. To obtain the 3' end of the mRNA, 3'-RACE was performed as described (20) with the modification of Rother (21).

**SL-enriched and 5'-RACE cDNA Library Construction**—SL-enriched libraries were prepared as described (11, 22, 23) with several variations. In general, first strand cDNA was synthesized from 0.2 to 1 µg of poly(A)<sup>+</sup> or 5 µg of total RNA using oligo(dT) primer-adaptors (*XbaI* = GTCGACTCTAGATTTTTTTTTTTTTT, dT57 = AAGGATCCGTCGACATCGATAATACGACTCACTATAAGGGATTTTTTTTTTTTTTTT, or *QtdT* = CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGCTTTTTTTTTTTTTTTTTT) and Superscript reverse transcriptase (Life Technologies, Inc.) using the conditions recommended. cDNAs were amplified using the *XbaI* oligo(dT) primer-adaptor or nested primers in the adaptors (*Ri* = GACATCGATAATACGAC, *Ro* = AAGGATCCGTCGACATC, *Qi* = GAGGACTCGAGCTCAAGC, or *Qo* = CCAGTGAGCAGAGTGACG) and SL primer-adaptors (*BamHI*: CGG-GATCCGAACCGTCACGGTTTACT or CGGGATCCGAACCGT-CACGGTTTTACTCTTG) using the following general conditions: 30 cycles of 1 min denaturation at 94 °C, 1 min annealing at 55–60 °C, and 2.5 min extension at 72 °C. SL-enriched cDNA was also prepared by synthesizing second strand cDNA with the SL primer-adaptors using PFU (Stratagene, La Jolla, CA) at 60 °C without amplification. cDNAs were either directly cloned into Bluescribe or Bluescript vectors (Stratagene Cloning Systems) using restriction sites in the adaptors or they were size fractionated on a 1% agarose gel and products greater than 1000 bases gel purified (Magic PCR Prep, Promega) prior to cloning. 5'-RACE cDNA libraries were constructed as described previously and SL containing clones identified by colony hybridization (14).

**Library Screening, Isolation of λ-DNA, and Genomic Insert Mapping**—An EMBL-3 genomic adult schistosome library was screened and λ-DNA isolated as described previously (15). Genomic inserts were restriction mapped by "hot mapping" as described (15). Relevant regions of genomic λ clones were identified by hybridization and subcloned for further analyses into either Bluescribe or Bluescript vectors.

**In Situ Hybridization**—Adult worms were isolated from the hepatic portal system, washed several times in phosphate-buffered saline, and fixed in 4% paraformaldehyde. 6-µm paraffin sections were hybridized with antisense [<sup>35</sup>S]RNA transcripts, washed, dipped in photographic emulsion, and developed for 0.25–5 days as described (24). After development, sections were stained with hematoxylin and eosin.

**Sequence Analysis**—Plasmid DNA was prepared as described previously (11, 15) or by Magic Plasmid Prep (Phannga). Clones were sequenced by the dideoxynucleotide method on alkali-denatured plasmids using the USB Sequenase Kit (U. S. Biochemical Corp.) as described

(11, 15). Sequencing was facilitated by subcloning and the primer walking strategy as described (25).

**Sequence and RNA Secondary Structure Analysis**—Nucleic acid sequences were compiled and analyzed using AssemblyLIGN and MacVector sequence analysis software (Eastman Kodak). Multiple alignment of sequences was performed using GeneWorks (IntelliGenetics, Mountain View, CA) and Genetics Computer Group (GCG) (Madison, WI) software packages. Protein structure was analyzed by MacVector, GeneWorks, and GCG software packages, and RNA secondary structure was analyzed by MFold in the GCG sequence analysis software package or by MulFOLD (26). Oligonucleotide primers for primer extension of RNA, DNA sequencing, and PCR were designed with the aid of Oligo 4.0 primer design software (NBI, Plymouth, MN). cDNA and protein sequences were compared to sequence data bases using electronic mail servers at the National Center for Biotechnology Information (NCBI) using the BLAST set of programs for protein and nucleotide similarities (27); at the European Bioinformatics Institute (EBI) (28) using FASTA (29), BLITZ (30, 31), and QUICK analyses for protein and nucleotide similarities; and BLOCKS for protein pattern similarities (32).

#### RESULTS

**Isolation and Analysis of SL mRNAs**—Not all mRNAs in *S. mansoni* undergo trans-splicing, and thus only a subset of mRNAs acquire the SL sequence (11). Available evidence suggests that a low percentage of schistosome mRNAs are trans-spliced based on the following: 1) only one of five glycolytic mRNAs we examined is trans-spliced (see below), 2) the frequency of trans-splicing among mRNAs and their genes whose 5' ends have been characterized and reported in the literature or data bases, 3) analysis of several types of schistosome cDNA libraries, and 4) our comparison of SL-enriched and 5'-RACE cDNA libraries constructed for schistosomes with similar libraries constructed for *Fasciola hepatica* (14) and *Ascaris*.<sup>3</sup> In contrast, all mRNAs in trypanosomes (33) and a large percentage, 70–90%, of nematode mRNAs (*C. elegans* ~70% and *Ascaris* ~80–90%) (3, 34) are thought to be trans-spliced.

Because only a relatively small subset of mRNAs appears to acquire the spliced leader in schistosomes, we identified and characterized trans-spliced mRNAs and several of their genes as one approach to determine if their type or organization could provide information on the potential function(s) and regulation of trans-splicing in flatworms. We used several approaches to construct cDNA libraries enriched for mRNAs with spliced leaders and isolated and characterized portions of 30 trans-spliced mRNAs (see "Materials and Methods"). These cDNAs were analyzed to determine if there are any discernable patterns in the type of computer-predicted proteins encoded as well as the general sequence or secondary structure characteristics of these mRNAs. cDNAs were also examined to determine if addition of the SL to mRNAs was required to provide the initiator methionine for open reading frames (ORFs) or contributed some other property to the 5' ends of the mRNAs. Representative cDNAs were selected from each of the libraries and analyzed either by primer extension analysis or direct sequencing of 5'-RACE products to provide independent confirmation that the cDNAs represented mRNAs with 5'-terminal SLs. From these analyses we estimated that at least 80% of the clones isolated from the SL-enriched libraries represent mRNAs with 5'-terminal spliced leaders.

**Are There Any Patterns in Trans-spliced mRNAs or Their Encoded Proteins?**—Open reading frames initiated by methionine were identified by computer-assisted translation of the cDNAs. Both the nucleotide and protein sequences were compared with known sequences in data bases using the BLAST (Blastp and Blastx), FASTA, MPsrch, and BLOCKS algorithms to identify significant similarities with known sequences. Sev-

<sup>3</sup> R. E. Davis, unpublished results.

TABLE I  
Spliced leader AUG context in flatworms

Sequences represent the 3'-terminal nucleotides of three flatworm spliced leaders containing a potential initiator AUG. The AUG is in bold and conserved nucleotides upstream from the AUG are underlined. The sequence for *S. mansoni* is from Ref. 11, *Fasciola hepatica* from Ref. 14, and *Stylochus zebra* from (R. E. Davis, manuscript in preparation).

<i>S. mansoni</i>	G AUU UGU <u>UGC</u> <b>AUG</b>
<i>F. hepatica</i>	A UAU UAG <u>UGC</u> <b>AUG</b>
<i>S. zebra</i>	C AAU AAU <u>UGC</u> <b>AUG</b>

eral parameter matrices were used for these analyses as described (35). Eight of the 30 trans-spliced mRNAs encoded proteins which were homologous to protein sequences or conceptual translations in data bases. These significant matches included the glycolytic enzyme enolase, a homolog of the synaptic vesicle protein synaptobrevin, a homolog of the mitochondrial ATPase inhibitor, a member of the alcohol dehydrogenase family (carbonyl reductase - NADPH), cyclophilin, a guanine nucleotide-binding protein (G protein  $\beta$  subunit-like), and an unidentified open reading frame in *C. elegans* and within a bacterial operon. We had identified trans-splicing previously in the *S. mansoni* mRNA encoding HMG-CoA reductase (11, 25), and these sequences were also used in our study. No identifiable patterns are evident in this set of proteins. Additional characterization of all the trans-spliced cDNAs, including predicted protein properties and structure, characteristics of 5'- and 3'-untranslated regions, and RNA secondary structures did not identify any apparent patterns in trans-spliced mRNAs or their encoded proteins.

*Does the Spliced Leader Contribute an Initiator Methionine?*—The 3'-terminal nucleotides of all flatworm SLs<sup>1</sup> constitute a potential translation initiator methionine (Table I). Using 60 non-trans-spliced schistosome mRNAs derived from nucleic acid data bases, we generated a preliminary *S. mansoni* translation initiation consensus Aanna(a/u)Aaa**AUG**ncna described in Table II. Comparison of this initiation consensus with the sequence context of the SL AUG shows that they differ significantly, and that the adenine at the -3 position, known to be important in other organisms, is absent in the SL. The longest ORFs in the trans-spliced *S. mansoni* mRNAs examined are rarely initiated by the SL AUG indicating that trans-splicing does not typically serve to provide an essential AUG. Thus, it seems unlikely that the primary function of spliced leader addition in schistosomes is to provide an initiator methionine for ORFs. However, two mRNAs are predicted to be initiated by the SL AUG based on conceptual translation. In these two mRNAs, SL1-6 (950+ bases) and SL1-17 (1150 bases), the ORFs extend at least 350 bases before the next in-frame AUG is present. One of the conceptual translations of these mRNAs has similarity with a motif in G protein  $\beta$  subunit-like proteins (SL1-17). Demonstration of the existence of proteins initiated by the SL AUGs in schistosomes requires further study. In the enolase mRNA, the SL AUG is in-frame and within 10 nucleotides of a second downstream AUG that exhibits a more typical eukaryotic translation initiation context. In other mRNAs, the SL contributes an upstream and out-of-frame AUG. The mean distance between the SL AUG and the predicted initiator AUG for the dominant ORF was  $50 \pm 50$  (S.D.) nucleotides with a typical range of 6–150 (two mRNAs with 5'-untranslated regions over 500 bases were excluded from this analysis). Finally, computer-generated RNA secondary structure predictions for the 5' ends of trans-spliced mRNAs (5' terminus to 100 bases 3' of the initiator methionine) did not show any consistent or common structural motifs in recipient mRNAs.

*Are Other Glycolytic mRNAs Trans-spliced?*—One of the iso-

lated trans-spliced schistosome mRNAs is predicted to encode the glycolytic enzyme enolase. Schistosomes exhibit an extremely high rate of glycolysis. Their energy metabolism is primarily homolactate fermentation, and the worms can consume glucose equivalent to 20% of their dry weight/h (36, 37). We hypothesized that the high rate of glycolysis might be facilitated by trans-splicing of glycolytic mRNAs as a group. SL addition might then contribute to coordinate expression, enhanced translation, or subcellular localization of glycolytic mRNAs. To explore this hypothesis, we analyzed several other glycolytic mRNAs for the presence of spliced leaders and investigated whether proteins in a common pathway might be derived from trans-spliced mRNAs. We used direct sequencing of 5'-RACE products to characterize the 5'-terminal sequences of the mRNAs coding for four other schistosome glycolytic enzymes (glyceraldehyde 3-phosphate dehydrogenase, triose phosphate isomerase, aldolase, and phosphofructokinase). Northern blot hybridization with probes derived from these 5'-terminal sequences was then used to determine if the mRNAs are trans-spliced. Control experiments on well characterized schistosome mRNAs and previous studies<sup>2</sup> indicate that our 5'-RACE conditions consistently generate products that extend to the 5' termini of mRNAs (14, 17). None of these four other glycolytic enzyme mRNAs exhibited the schistosome spliced leader nor did they have any 5'-terminal sequences in common (the TPI analysis was conducted simultaneously with these mRNAs and described previously (17)). Northern blot hybridizations using antisense oligonucleotides to the 5' termini of glyceraldehyde 3-phosphate dehydrogenase, aldolase, triose phosphate isomerase, and phosphofructokinase demonstrated hybridization only to discrete mRNAs of the predicted size for the corresponding eukaryotic glycolytic mRNA and not to a small RNA or a smear as would be expected if the 5' terminus of the mRNA were a spliced leader. These data indicate that these other glycolytic enzyme mRNAs are not trans-spliced and that glycolytic mRNAs do not appear to be trans-spliced as a group in schistosomes.

*Isolation, Analysis, and Potential Patterns in the Organization of Trans-spliced Genes*—Genomic clones containing the trans-splice acceptor regions of several mRNAs processed by spliced leader addition were isolated and analyzed. The isolation of genomic clones corresponding to HMG-CoA reductase was described previously (11). Two genes, enolase and L11, were sequenced in their entirety (Fig. 1), whereas only 5' regions of synaptobrevin (exons 2–4 and ~400 bases upstream) and HMG-CoA reductase (exons 2–4 and ~400 bases of upstream) were characterized (Fig. 1). The L11 gene has no significant similarity with current sequences in data bases. All four genes appear to be single copy genes based on analysis of their corresponding genomic clones, Southern blots, and genomic titrations. General characteristics of all four genes include the presence of introns and both variable exon and intron size. In the L11 gene, intron sizes are all quite small including 31, 32, and 34 nucleotide introns, and an exon is present that is only 34 nucleotides. Small exons and introns can also be found in the other trans-spliced genes (Fig. 1) and have previously been described in several nontrans-spliced schistosome genes (15, 38). Exon and intron sizes range from very small to large in schistosome genes and no correlation of exon or intron size or gene organization with trans-splicing is evident.

*Analysis of Trans-splice Acceptor Sites and Upstream Regions for Conserved Elements*—The presence of both trans- and cis-splicing within the same gene raises questions regarding the regulation and discrimination of trans- versus cis-splicing within the primary transcript. In order to compare consensus

TABLE II  
Translation initiation consensus for *S. mansoni* mRNAs

60 mRNA sequences were derived from the NCBI database and the sequence contexts of the translation initiation AUGs used to generate the nucleotide percentages at each position. The consensus at each position was generated using the following criteria: 1) a specific nucleotide is shown when the highest nucleotide percentage is at least 15% greater than other nucleotide percentages present; 2) small letters are used when the highest nucleotide percentage does not exceed 50% and is not at least twice any other nucleotide percentage present; and 3) bold capital letters are used when the nucleotide percentage exceeds 50% and is greater than twice any other nucleotide percentage for that position.

Nucleotide % at each position	Consensus															Nucleotide
	A	a	n	n	a	a/u	a	A	a	a		n	c	n	a	
	-10 <sup>a</sup>	-9	-8	-7	-6	-5	-4	-3	-2	-1	<b>AUG</b>	4	5	6	7	
A	52	43	23	26	49	42	47	71	45	40	<b>AUG</b>	31	20	30	40	A
G	9	11	25	9	15	15	20	18	13	19	<b>AUG</b>	33	18	26	18	G
C	15	18	23	28	10	10	17	3	21	15	<b>AUG</b>	8	39	10	21	C
U	24	28	29	37	26	33	17	8	21	26	<b>AUG</b>	28	23	34	21	U

<sup>a</sup> Position.

TABLE III  
Nucleotide percentages and consensus sequences for cis- and trans-splice acceptor sites in *S. mansoni* genes

Six trans-splice acceptor sites for the four trans-spliced genes analyzed in this study (enolase and L11; HMG-CoA reductase and synaptobrevin each have two trans-splice acceptor sites), 60 cis-splice acceptor sites for 23 other genes derived from the NCBI database, and the 10 cis-splice acceptors in the four trans-spliced genes were used to prepare the consensus shown. Letter style at each position within the consensus was derived using the following criteria: 1) capital letters when the nucleotide percentage is equal to or exceeds 40% and is greater than twice any other nucleotide percentage present; 2) small letters when the highest nucleotide percentage does not exceed 50% and is not at least twice any other nucleotide present; and 3) a specific nucleotide is shown only when the highest nucleotide percentage is at least 15% greater than other nucleotide percentages for that position in the consensus.

		Trans-splice acceptor sites ( <i>N</i> = 6)																												
		-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
	A	50	34	33	12	17	17	0	17	17	0	0	33	0	0	0	17	17	0	0	100	0	83	33	17	33	0	17	16	0
	C	0	33	0	11	16	16	17	0	0	33	0	0	0	33	0	0	17	50	50	0	0	0	17	17	0	33	33	17	50
	G	0	33	0	11	17	17	33	17	33	0	17	17	0	17	33	0	0	17	0	0	100	17	0	16	17	0	33	17	17
	U	50	0	67	66	50	50	50	66	50	67	83	50	100	50	67	83	66	33	50	0	0	0	50	50	50	67	17	50	33
Consensus		w	n	U	U	U	U	u	U	u	U	U	w	U	y	U	U	U	y	y	<u>A</u>	<u>G</u>	A	w	U	w	U	n	U	y

		Cis-splice acceptor sites ( <i>N</i> = 60)																												
		-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
	A	22	22	22	31	32	22	22	19	29	22	33	28	27	25	23	20	11	18	15	100	0	30	31	21	25	27	32	27	36
	C	20	20	22	8	18	20	17	21	16	12	12	24	20	15	13	24	24	15	23	0	0	16	23	11	23	14	20	7	18
	G	10	13	9	11	5	13	7	16	11	11	10	6	7	9	11	15	2	16	3	0	100	34	9	27	25	18	18	25	23
	U	48	45	47	49	45	45	54	44	44	55	45	42	46	51	53	41	63	51	59	0	0	20	37	41	27	41	30	41	23
Consensus		U	U	U	w	w	U	U	U	w	U	w	w	w	U	U	y	U	U	U	<u>A</u>	<u>G</u>	n	n	n	n	w	n	n	n

		Cis-splice acceptor sites in trans-spliced genes ( <i>N</i> = 10)																												
		-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
	A	60	50	50	30	30	30	30	10	20	10	20	40	10	40	10	0	0	0	0	100	0	40	20	20	20	40	30	30	10
	C	10	10	0	30	0	0	20	30	10	50	0	10	0	10	0	20	10	30	60	0	0	0	0	20	10	10	20	10	30
	G	0	0	10	0	30	30	10	20	20	10	0	0	0	0	10	10	10	10	0	0	100	30	20	20	20	20	10	20	30
	U	30	40	40	40	40	40	40	40	50	30	80	50	90	50	80	70	80	60	40	0	0	30	60	40	50	30	40	40	30
Consensus		A	w	w	n	n	n	n	n	U	Y	U	w	U	w	U	U	U	U	Y	<u>A</u>	<u>G</u>	n	U	n	U	n	n	n	n

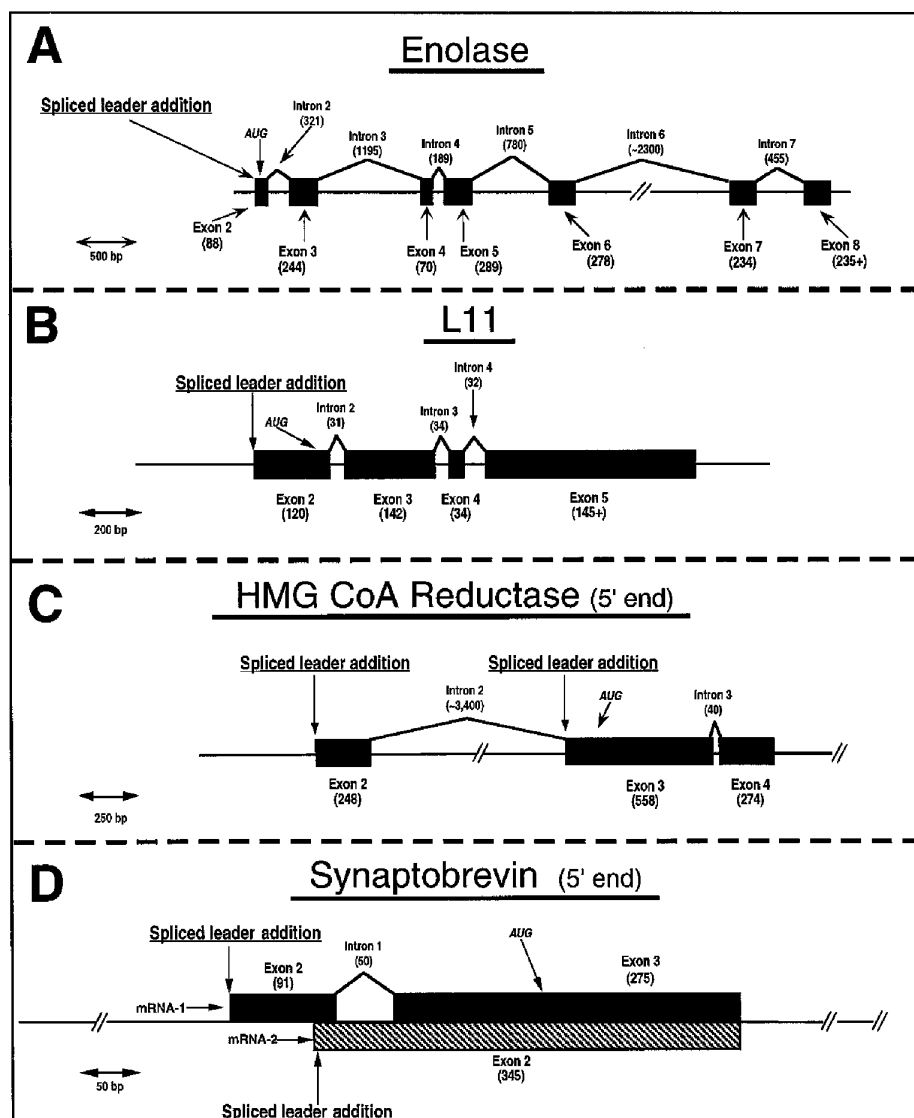
sequences for trans- versus cis-splice acceptor sites, we compiled sequences for trans-splice acceptor sites (6 = both HMG-CoA and synaptobrevin genes express two trans-spliced mRNAs and thus have two distinct trans-splice acceptor sites), cis-splice acceptor sites in trans-spliced genes (10 sites), and cis-splice sites in other schistosome genes (over 60 sites derived from 23 genes in nucleic acid data bases) (Table III). This sequence comparison showed few differences between these three types of acceptor sites. From this small sampling, the trans-splice acceptor site exhibits a preference for an adenine as the first nucleotide in the exon that acquires the SL, an absolutely conserved U at the -7 position in the intron, and a slightly more pronounced polypyrimidine tract compared with other acceptor sites (Table III).

Secondary structure and base pairing interactions have been implicated as phylogenetically conserved elements associated with self-splicing and snRNA-mediated cis- and trans-splicing. We examined the regions adjacent to the trans-splice acceptor sites in the four genes for homologous sequences or potential secondary structures that might be involved in facilitating the interaction of the two RNA substrates and/or the specificity of the trans-splicing reaction. Conserved elements were not ob-

served in the trans-spliced genes.

*Is There Sex, Tissue, or Cell Specificity in the Generation of the SL RNA or Trans-spliced mRNAs?*—We used *in situ* hybridization to determine if SL RNA expression was present only in particular cells or tissues. Restricted expression of the SL RNA might contribute to differential expression of genes requiring trans-splicing. *In situ* hybridization of adult worms using an antisense SL RNA hybridization probe (Fig. 2), however, showed that the SL RNA was expressed in both males and females and in almost all tissues and cells. Localization of the SL RNA was greatest in tissues with large numbers of nuclei and the grains localized in highest concentration over the nuclei (Fig. 2B, D-F). Notably, although almost all nuclei show expression of the SL RNA, all nuclei do not exhibit the same levels of expression. Although there are several possible explanations for this observation, one consistent with a short SL RNA half-life, such as that observed in trypanosomes (~6 min) (39), is that the expression of the SL RNA might be cell-cycle-regulated. Analysis of several trans-spliced mRNAs (data not shown) did not demonstrate any unusual tissue or cellular localization.

FIG. 1. *Schistosoma mansoni* trans-spliced gene organization. Schematics illustrate the exon (boxes) and intron organization for each gene, the location of the trans-splice acceptor site(s), and the location of the translation initiation site (AUG). The horizontal lines represent the extent of sequence generated for each locus. Note that the scales for each schematic vary. A, enolase gene (5,050 nucleotides). The discontinuity between exon 6 and 7 indicates that the entire sequence of the intron was not determined. B, L11 gene (1,020 nucleotides). C, 5' end of the HMG-CoA reductase gene (2,285 nucleotides). The discontinuity between exons 2 and 3 illustrates that the entire sequence of the intron was not determined. D, 5' end of the synaptobrevin gene (1060 nucleotides). The discontinuities (—//—) in the sequence are present to keep the figure to scale. The upstream region corresponds to 400 bases and the downstream region to 300 bases of nucleotide sequence.



## DISCUSSION

No apparent common motifs or patterns were observed in our sampling of 30 trans-spliced schistosome mRNAs and their encoded proteins. Similarly, analysis of the trans-spliced genes did not reveal any unique or inherent characteristics when compared with non-trans-spliced schistosome genes. Although the glycolytic enzyme enolase is derived from a trans-spliced mRNA, four other glycolytic enzymes are not, indicating that trans-splicing of mRNAs does not appear common to this particular metabolic pathway. Furthermore, *in situ* hybridization analysis of adult schistosomes indicates that the SL RNA exhibits no gross sex, tissue, or cell specificity. An AUG is absolutely conserved at the 3' terminus of all flatworm spliced leaders. We found, however, that addition of the spliced leader AUG is not typically required to initiate computer-predicted ORFs in trans-spliced schistosome mRNAs. Together, these observations suggest that the significance of trans-splicing in flatworms is more likely to be correlated either with other properties conferred by the SL on recipient mRNAs or related to some characteristic of the primary transcripts or transcription of trans-spliced genes.

Analysis of *C. elegans* and *Ascaris* mRNAs which acquire spliced leaders (3, 22, 23) and the current data base of trans-spliced nematode mRNAs indicates that it is also unlikely that trans-splicing is related to particular types of pathways, en-

coded proteins, or restricted to particular cells or tissues in nematodes (6, 34, 40). Furthermore, there is no general conservation of particular genes that are trans-spliced in metazoa, since for example, glyceraldehyde 3-phosphate dehydrogenase is trans-spliced in *Caenorhabditis spp.*, but not in schistosomes, and the homolog of the mitochondrial ATPase inhibitor in *Caenorhabditis spp.* is not trans-spliced, while the analogous mRNA in schistosomes acquires an SL (41).

The 5' ends of both nematode (42–44) and flatworm (11, 14) SL RNAs have a trimethylguanosine (TMG) cap. This cap is transferred to nematode actin mRNAs during the trans-splicing reaction (45, 46). Transfer of the TMG cap to mRNAs presumably also occurs in schistosomes. Capping of mRNAs by spliced leader addition appears essential for mRNA stability in trypanosomes (47, 48), and the TMG cap or the SL sequence itself might also affect schistosome mRNA stability, translation, transport, cytoplasmic localization, cis-splicing, or other processing of precursor mRNAs.

Two spliced leaders are present in the nematode *C. elegans*, SL1 and SL2. Although SL1 trans-splicing constitutes the majority of trans-splicing in both *C. elegans* and *Ascaris*, its function remains largely unknown. In trypanosomes, trans-splicing plays a role in resolving polycistronic transcription units into individual mRNAs (49–52). These individual mRNAs are generated by 5' processing through trans-splicing of the SL and

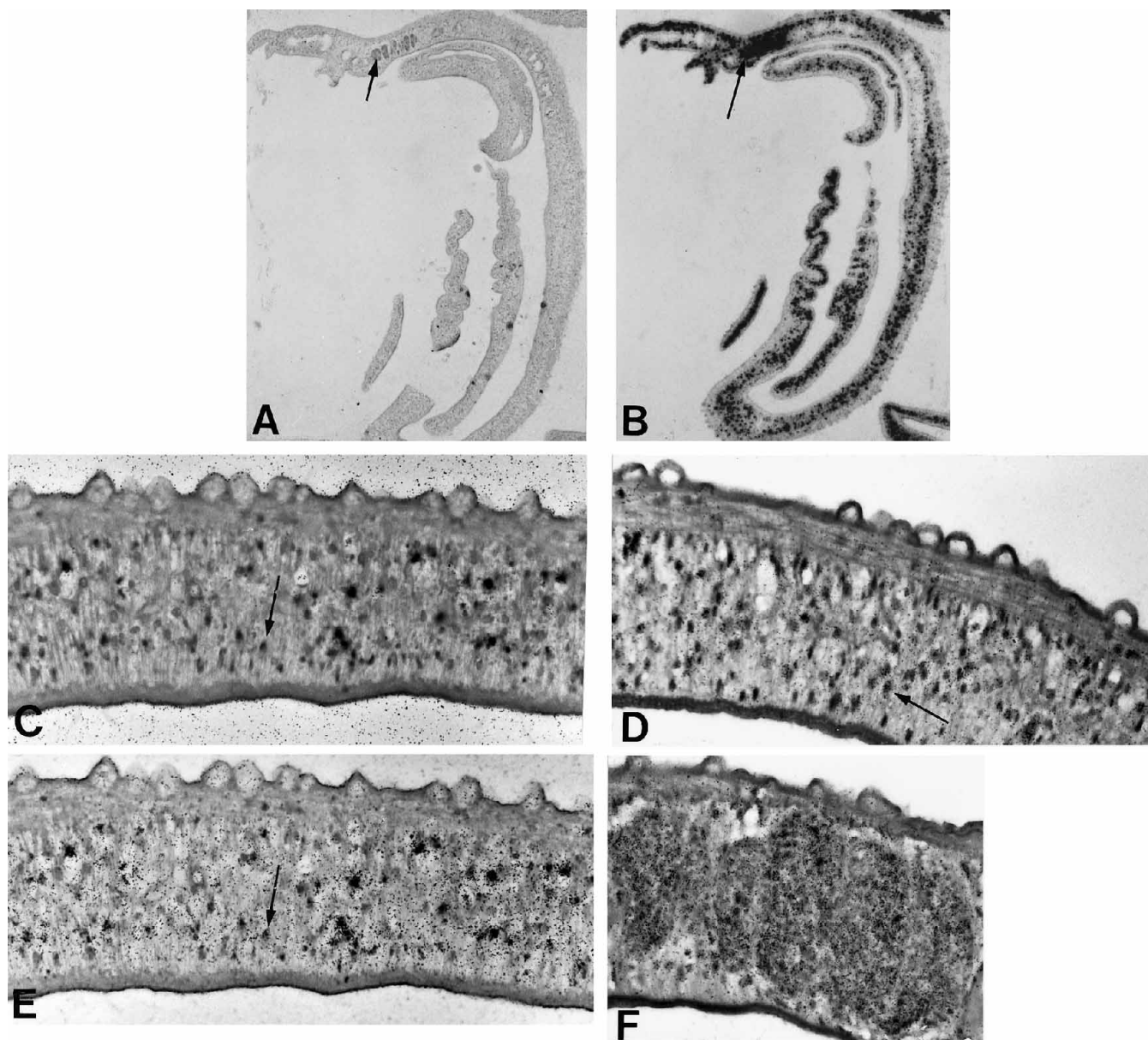


FIG. 2. SL RNA expression in adult *Schistosoma mansoni*. *In situ* hybridization on paraformaldehyde fixed paraffin sections of adult *S. mansoni* was performed with sense or antisense  $^{35}\text{S}$ -labeled SL RNA probes. Control hybridization using a sense RNA corresponding to the SL RNA sequence represents background (A and C). SL RNA expression is shown using an antisense SL RNA probe (B and D-F). Grains associated with the antisense SL RNA probe were absent when sections were pretreated prior to hybridization with RNase A, but were not effected when pre-treated with DNase I (not shown). The arrows in A and B denote one of the five adjacent testes present in the males. F, represents the grains over nuclei in the testes at higher magnification. The arrows in C-E denote nuclei. The nuclei marked in C (SL RNA probe) and E (anti-SL RNA probe) are from adjacent sections. Exposure times for A and B were five times longer than C-E. Magnification: A and B,  $\times \sim 20$ ; C-F,  $\times \sim 200$ .

3' processing via cleavage and polyadenylation. Recently, Blumenthal and colleagues (34, 41) have shown that the subset of trans-spliced *C. elegans* mRNAs acquiring SL2 are processed from internal genes within operons transcribed as polycistronic transcripts. SL2 appears specialized for processing of genes located within these operons in *C. elegans* (6). Except for one unusual case (53), SL1 is not known to be associated with the resolution of polycistronic transcripts in *C. elegans*. It will be of interest to determine if regions upstream or downstream from trans-spliced schistosome genes express detectable mature mRNAs (derived from the same DNA coding strand) to explore the possibility for polycistronic transcription across these loci.

Trans-splicing could be functionally associated with transcription initiation. Transcription initiation sites for these genes might be located significantly upstream of the trans-splice acceptor site or be unusually heterogeneous (1) produc-

ing long 5'-untranslated regions or ones of highly mixed lengths. Trans-splicing might then function to trim the mRNAs and generate shorter, uniform 5' ends. Although inherently difficult in trans-spliced genes, it will be of interest to attempt to identify and characterize transcription initiation sites in the genes described here to investigate this potential function for spliced leader addition in schistosomes.

All four trans-spliced schistosome genes we characterized undergo cis-splicing. Similarly, all nematode genes which undergo trans-splicing almost invariably exhibit cis-splicing. The presence of cis- and trans-splicing within the same primary transcript would ostensibly require the splicing machinery to discriminate between these sites for accurate RNA processing and generation of functionally mature mRNAs. Our comparison of a small sampling of trans-splice and cis-splice acceptor site sequences and their contexts indicates that the two types of

schistosome splice acceptor sites are similar. In nematodes, significant differences between cis- and trans-splice acceptor site sequences have also not been observed (6, 40, 54). The consensus for cis-splicing in nematodes (UUUC/AGG) is similar to that which we describe here in schistosomes (U<sub>2</sub>YU<sub>3</sub>/AGR), although the polypyrimidine tract upstream from the acceptor site in schistosomes is more pronounced than in nematodes. Both nematodes and flatworms have higher A/U content within introns than within exons. The transition in A/U content between introns and exons is significantly greater in nematodes (54) and is a determinant in splice site recognition (55).

Detailed studies using a hybrid gene in transgenic nematodes suggest that when the 5' most splice acceptor site within a primary transcript is not preceded by an upstream splice donor, that these elements are sufficient to identify a transcript as an appropriate SL1 trans-splice acceptor substrate (56). Addition of a 5' splice site upstream of a trans-splice acceptor site in this paradigm alters the splicing exclusively to cis-splicing (57). Thus, a 5' unpaired splice-acceptor site appears necessary and sufficient to direct SL1 trans-splicing to an appropriate site. Similar 5' unpaired splice-acceptor sites may direct trans-splicing to appropriate acceptor sites in schistosomes. In the *S. mansoni* HMG-CoA and the synaptobrevin genes, two distinct trans-spliced mRNAs are produced (11). Whether the two distinct trans-spliced mRNAs from these two genes are derived by alternative trans-splicing within the same primary transcript, distinct transcription initiation sites for the mRNAs, or if inefficient cis-splicing is responsible for the generation of these different mRNAs is currently not known. Analysis of transcription initiation sites and the primary transcription units for schistosome genes will be necessary to provide a better understanding of the substrates, splice acceptor site choices, and processing of trans-spliced genes in schistosomes.

**Acknowledgments**—We thank Lee A. Niswander for the *in situ* hybridization analysis and Ron Blanton and George Newport for adult schistosomes. Unpublished cDNA sequence and DNA primers for aldolase and 3-phosphoglyceraldehyde dehydrogenase were kindly provided by George Newport and phosphofructokinase by Tag Mansour and John Ding. We also thank the genetic engineering class at San Francisco State University for help in screening, isolating, and mapping genomic clones corresponding to trans-spliced genes and J. Villanueva and S. Koepf for plasmid preparation and sequencing.

# REFERENCES

- Huang, X.-Y., and Hirsh, D. (1992) *Genet. Eng.* **14**, 211–229
- Nilsen, T. W. (1992) *Infectious Agents Disease* **1**, 212–218
- Nilsen, T. W. (1993) *Annu. Rev. Microbiol.* **47**, 385–411
- Nilsen, T. M. (1994) *Science* **264**, 1868–1869
- Davis, R. E. (1995) in *Molecular Approaches to Parasitology* (Boothroyd, J. C., and Komuniecki, R., eds) pp. 299–320, Wiley-Liss, New York
- Blumenthal, T. (1995) *Trends Genet.* **11**, 132–136
- Boothroyd, J., and Cross, G. (1982) *Gene (Amst.)* **20**, 279–287
- Van der Ploeg, L., Liu, A., Michels, P., De Lange, T., Borst, P., Majumder, H., Weber, H., Veeneman, G., and Van Boom, J. (1982) *Nucleic Acids Res.* **10**, 3591–3604
- Tessier, L.-H., Keller, M., Chan, R., Fournier, R., Weil, J. H., and Imbault, P. (1991) *EMBO J.* **10**, 2621–2625
- Krause, M., and Hirsh, D. (1987) *Cell* **49**, 753–761
- Rajkovic, A., Davis, R. E., Simonsen, J. N., and Rottman, F. M. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8879–8883
- Raff, R. A., Marshall, C. R., and Turbeville, J. M. (1994) *Annu. Rev. Ecol. Syst.* **25**, 351–375
- Willmer, P. (1990) *Invertebrate Relationships: Patterns in Animal Evolution*, Cambridge University Press, New York
- Davis, R. E., Singh, H., Botka, C., Hardwick, C., el Meanawy, M. A., and Villanueva, J. (1994) *J. Biol. Chem.* **269**, 20026–20031
- Davis, R. E., Davis, A. H., Carroll, S. M., Rajkovic, A., and Rottman, F. M. (1988) *Mol. Cell. Biol.* **8**, 4745–4755
- Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Reis dos, M. G., Davis, R. E., Singh, H., Skelly, P. J., and Shoemaker, C. B. (1993) *Mol. Biochem. Parasitol.* **59**, 235–242
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
- Marchuk, D., Drumm, M., Saulino, A., and Collins, F. S. (1990) *Nucleic Acids Res.* **19**, 1154
- Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8998–9002
- Rother, R. P. (1992) *BioTechniques* **13**, 524–527
- Bektesh, S., Van Doren, K., and Hirsh, D. (1988) *Genes & Dev.* **2**, 1277–1283
- Hannon, G. J., Maroney, P. A., Denker, J. A., and Nilsen, T. W. (1990) *Cell* **61**, 1247–1255
- Frohman, M. A., Boyle, M., and Martin, G. R. (1990) *Development* **110**, 589–607
- Rajkovic, A., Simonsen, J. N., Davis, R. E., and Rottman, F. M. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 8217–8221
- Zuker, M. (1989) *Science* **244**, 48–52
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
- Emmert, D. B., Stoehr, P. J., Stoesser, G., and Cameron, N. (1994) *Nucleic Acids Res.* **22**, 3445–3449
- Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2444–2448
- Smith, T. F., and Waterman, M. S. (1981) *J. Mol. Biol.* **147**, 195–197
- Sturrock, S. S., and Collins, J. F. (1993) Biocomputing Research Unit, University of Edinburgh, U.K.
- Henikoff, S., and Henikoff, J. G. (1994) *Genomics* **19**, 97–107
- Walder, J. A., Eder, P. S., Engman, D. M., Brentano, S. T., Walder, R. Y., Knutzen, D. S., Dorfman, D. M., and Donelson, J. E. (1986) *Science* **233**, 569–571
- Zorio, D. A. R., Cheng, N. N., Blumenthal, T., and Spieth, J. (1994) *Nature* **372**, 270–272
- Altschul, S. F., Boguski, M. S., Gish, W., and Wootton, J. C. (1994) *Nature Genet.* **6**, 119–129
- Bueding, E. (1950) *J. Gen. Physiol.* **33**, 475–495
- Tielens, A. G. M. (1994) *Parasitol. Today* **10**, 346–352
- Craig, S. P., Muralidhar, M. G., McKerrow, J. H., and Wang, C. C. (1989) *Nucleic Acids Res.* **17**, 1635–1647
- Laird, P. W., Zomerdijk, J. C. B. M., de Korte, D., and Borst, P. (1987) *EMBO J.* **6**, 1055–1062
- Blumenthal, T., and Thomas, J. (1988) *Trends Genet.* **4**, 305–308
- Spieth, J., Brooke, G., Kuersten, S., Lea, K., and Blumenthal, T. (1993) *Cell* **73**, 521–532
- Van Doren, K., and Hirsh, D. (1988) *Nature* **335**, 556–559
- Thomas, J. D., Conrad, R. C., and Blumenthal, T. (1988) *Cell* **54**, 533–539
- Maroney, P. A., Hannon, G. J., and Nilsen, T. W. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 709–713
- Liou, R.-F., and Blumenthal, T. (1990) *Mol. Cell. Biol.* **10**, 1764–1768
- Van Doren, K., and Hirsh, D. (1990) *Mol. Cell. Biol.* **10**, 1769–1772
- Tschudi, C., and Ullu, E. (1990) *Cell* **61**, 459–466
- Huang, J., and Van der Ploeg, L. H. T. (1991) *Mol. Cell. Biol.* **11**, 3180–3190
- Agabian, N. (1990) *Cell* **61**, 1157–1160
- Clayton, C. E. (1992) *Prog. Nucleic Acids Res. Mol. Biol.* **43**, 37–66
- Pays, E., Vanhamme, L., and Berberof, M. (1994) *Annu. Rev. Microbiol.* **48**, 25–52
- Tschudi, C. (1995) in *Molecular Approaches to Parasitology* (Boothroyd, J. C., and Komuniecki, R., eds) pp. 255–268, Wiley-Liss, New York
- Hengartner, M. O., and Horvitz, R. (1994) *Cell* **76**, 665–676
- Fields, C. (1990) *Nucleic Acids Res.* **18**, 1509–1512
- Conrad, R., Liou, R. F., and Blumenthal, T. (1993) *Nucleic Acids Res.* **21**, 913–919
- Conrad, R., Thomas, J., Spieth, J., and Blumenthal, T. (1991) *Mol. Cell. Biol.* **11**, 1921–1926
- Conrad, R., Liou, R. F., and Blumenthal, T. (1993) *EMBO J.* **12**, 1249–1255