The EMBO Journal vol.15 no.15 pp.4016–4029, 1996

Evidence for the presence of a small U5-like RNA in active *trans*-spliceosomes of *Trypanosoma brucei*

Jan M.Dungan, Kenneth P.Watkins and Nina Agabian[1]

Intercampus Program in Molecular Parasitology, University of California, San Francisco and University of California, Berkeley, San Francisco, CA 94143-0422, USA

[1]Corresponding author

The existence of the *Trypanosoma brucei* 5′ splice site on a small RNA of uniform sequence (the spliced leader or SL RNA) has allowed us to characterize the RNAs with which it interacts *in vivo* by psoralen crosslinking treatment. Analysis of the most abundant crosslinks formed by the SL RNA allowed us previously to identify the spliced leader-associated (SLA) RNA. The role of this RNA in *trans*-splicing, as well as the possible existence of an analogous RNA interaction in *cis*-splicing, is unknown. We show here that the 5′ splice site region of the SL RNA is also crosslinked *in vivo* to a second small RNA. Although it is very small and lacks a 5′ trimethylguanosine (TMG) cap, the SLA2 RNA possesses counterparts of the conserved U5 snRNA stem–loop 1 and internal loop 1 sequence elements, as well as a potential trypanosome snRNA core protein binding site; these combined features meet the phylogenetic definition of U5 snRNA. Like U5, the SLA2 RNA forms an RNP complex with the U4 and U6 RNAs, and interacts with the 5′ splice site region via its putative loop 1 sequence. In a final analogy with U5, the SLA2 RNA is found crosslinked to a molecule identical to the free 5′ exon splicing intermediate. These data present a compelling case for the SLA2 RNA not only as an active *trans*-spliceosomal component, but also for its identification as the trypanosome U5 structural homolog. The presence of a U5-like RNA in this ancient eukaryote establishes the universality of the spliceosomal RNA core components.

Keywords: SL RNA/SLA RNA/*trans*-splicing/*Trypanosoma brucei*/U5

Introduction

The two-step transesterification reaction that removes introns from pre-mRNAs takes place in the spliceosome, a dynamic ribonucleoprotein complex which associates with nascent transcripts in eukaryotic cell nuclei (for a review, see Moore *et al.*, 1993). In 'higher' eukaryotes, the spliceosome functions to splice together exons that exist in a precursor mRNA *in cis*. Some of the more primitive eukaryotes use a version of this machinery to splice a leader sequence onto the 5′ ends of mRNAs *in trans*. Nematodes, trematodes and *Euglena* perform pre-mRNA splicing both *in cis* and *in trans*, while trypanosomes and other kinetoplastid organisms appear to splice all pre-mRNAs solely *in trans* (reviewed in Bonen, 1993). Spliceosomal snRNAs, first defined in higher organisms and thought to be at the catalytic heart of the process, have been characterized in nematodes and Kinetoplastida (Tschudi *et al.*, 1986, 1988; Mottram *et al.*, 1989; Thomas *et al.*, 1990; Shambaugh *et al.*, 1994). Nematodes possess homologs for all of the snRNAs used in the *cis*-spliceosome (U1, U2, U4, U5 and U6); it is not clear whether all or just the U2/U4/U6 subset of these participate in *trans*-splicing (reviewed in Nilsen, 1993). Kinetoplastid homologs of U2, U6 and U4 have been characterized, but U1 and U5 counterparts have been difficult to identify. The absence of U1 may be easier to rationalize, since it functions early in *cis*-spliceosome assembly (prior to either of the transesterification reactions) and may be dispensible or, in some cases, even inhibitory (Konforti *et al.*, 1993; Konforti and Konarska, 1995; Crispino *et al.*, 1994; Crispino and Sharp, 1995; Tarn and Steitz, 1995). The lack of a U5 analog has been more perplexing, however, given its apparent central role in the *cis*-spliceosome (reviewed in Moore *et al.*, 1993 and Umen and Guthrie, 1995; MacMillan *et al.*, 1994; Newman *et al.*, 1995). The absence of these two RNAs has led to proposals that sequences or structures present in the 5′ splice site-bearing spliced leader RNA (SL RNA) might provide the *trans*-spliceosome with whatever critical functions U1 and U5 contribute to the *cis*-spliceosome (Bruzik *et al.*, 1988; Steitz, 1992).

Among the spliceosomal snRNAs, U5 snRNA is the least conserved at the primary sequence level. Comparative analyses (Frank *et al.*, 1994, and references therein) have revealed a consensus secondary structure which contains a few conserved sequence elements, including the 11 nucleotide (nt) loop 1 (L1) sequence (5′-YGCCUUUY-AYY-3′), internal loop 1 (IL1) sequences implicated in binding U5-specific proteins, and a binding site for the snRNP core proteins, the Sm antigens. Nucleotides in L1 have been shown to interact directly with the 5′ and 3′ splice sites through both steps of splicing by genetic and physical criteria (Newman and Norman, 1991, 1992; Wassarman and Steitz, 1992; Wyatt *et al.*, 1992; Cortes *et al.*, 1993; Sontheimer and Steitz, 1993; Newman *et al.*, 1995), underscoring the importance of U5 as a fundamental part of the spliceosome. The additional observations that the highly conserved U5-specific protein, PRP8 (Hodges *et al.*, 1995), is intimately associated with both splice sites and the branch point region (reviewed in Umen and Guthrie, 1995; MacMillan *et al.*, 1994) paints a picture of the spliceosome in which U5 snRNP components provide structural continuity for the catalytic machinery.

The apparent lack of U5 in organisms which carry out only *trans*-splicing challenges its seeming central role, and introduces the possibility that the intermolecular

© Oxford University Press

version of splicing either does not require the U5 functions, or that it uses some other means of synapsing and stabilizing substrate RNAs in the spliceosome during the reaction. The search for spliceosomal snRNAs by analysis of TMG-capped RNAs and by homology-based techniques has failed to reveal candidates for either U1 or U5 in kinetoplastid organisms (Mottram *et al.*, 1989; K.P. Watkins, unpublished observations). Using psoralen crosslinking treatment to search for small RNAs that interact with the 5′ splice site *in vivo*, we have previously identified a novel RNA, the spliced leader-associated (SLA) RNA, that contains a conserved potential to base pair with sequences near the 5′ splice site in the kinetoplastid SL RNA (Watkins *et al.*, 1994; Roberts, 1996). The function of this RNA in *trans*-splicing is unknown: it may be a highly diverged form of either U5 or U1 (from the structure of the crosslinks it forms at the 5′ splice site) or possibly an snRNA unique to either *trans*-splicing or trypanosomes (Watkins *et al.*, 1994). Crosslinks of the SLA RNA to the 5′ splice site region account for the majority of the crosslinks created in the SL RNA by psoralen treatment. Analysis of the remaining abundant SL RNA crosslinked species reveals that it consists of an intermolecular crosslink to a second small RNA, the SLA2 RNA, which appears to be a structural homolog of the U5 snRNA. We present a characterization of the SLA2 RNA and its interaction with the SL RNA.

Results

Characterization of a second SL RNA-associated RNA

Procyclic *Trypanosoma brucei* were treated with the psoralen derivative 4′-aminomethyltrioxsalen (AMT) and UV (365 nm) irradiated as described (Watkins *et al.*, 1994). Displayed in Figure 1A is a Northern analysis of the crosslinks formed by the SL RNA, side by side with a panel showing which of these species consist of intermolecular crosslinks to the previously characterized SLA RNA (Watkins *et al.*, 1994). To facilitate the purification of sufficient amounts of the remaining uncharacterized crosslinked species (‘Z’ in Figure 1A), the SL RNA present in several milligrams of crosslinked total RNA was specifically 3′ end-labeled using a procedure developed by Hausner *et al.* (1990). This ‘splint-labeling’ method exploits the ability of a modified version of T7 DNA polymerase to primer-extend an RNA when it is hybridized to an oligodeoxynucleotide that overhangs its 3′ end (see Materials and methods). This allowed direct visualization of the crosslinked SL RNA species by autoradiography during gel purification, instead of requiring blind fractionation of unlabeled crosslinked RNA. The electrophoretic pattern of SL RNA crosslinks observed by this method is essentially the same as that observed by Northern analysis (data not shown). Cross-linked RNAs were separated from co-migrating uncrosslinked RNAs by successive electrophoresis through denaturing gels of differing acrylamide sieves, as described (Watkins *et al.*, 1994). After fractionation through two rounds of preparative 6% (29:1) PAGE, followed by preparative 10% (29:1) PAGE, eluted crosslinked RNA species were 3′ end-labeled using T4 RNA ligase and [5′-^{32}P]cytidine 3′,5′-bis-phosphate

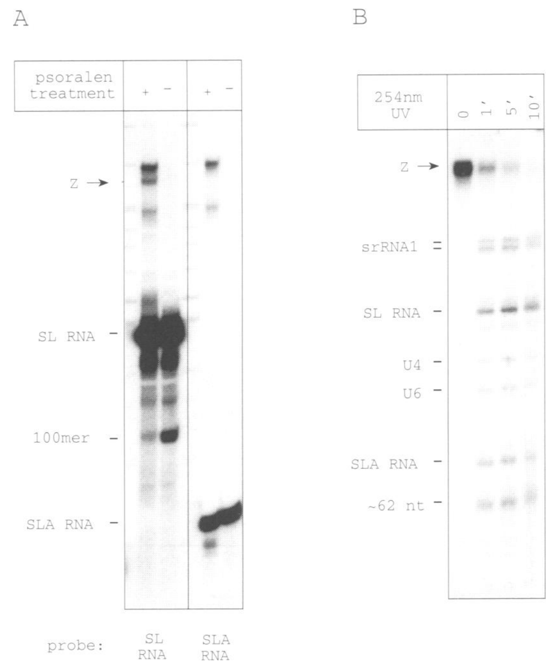


Fig. 1. What is the remaining unidentified SL RNA psoralen crosslink? (**A**) RNA samples isolated from untreated cells and cells UV irradiated in the presence of psoralen were treated with HeLa debranching extracts as per Patzelt *et al.* (1989), and loaded onto a denaturing 6% (29:1) polyacrylamide gel in pairwise repeats, separated by end-labeled pBR-MspI molecular weight markers. After transfer to nylon membrane, filters were UV treated and sliced through the midpoints of the molecular weight marker lanes. Filter strips were probed for the SL RNA and SLA RNA, and reassembled for autoradiography. The unidentified SL RNA crosslinked species is denoted ‘Z’. Debranched SL RNA intron is denoted ‘100mer’. (**B**) The ‘Z’ species crosslink and other co-purifying crosslinked RNAs were 3′ end-labeled with [5′-^{32}P]pCp by T4 RNA ligase and subjected to a shortwave UV photoreversal time course. RNAs released by this treatment were subjected to sequencing analysis; all but one of these RNAs were readily identifiable, as indicated.

([5′-^{32}P]pCp) and re-purified by 10% (29:1) PAGE (data not shown).

End-labeled crosslinked RNAs present in the ‘Z’-containing fraction were revealed upon photoreversal of AMT crosslinks using shortwave UV (254 nm) light, a time course of which is shown in Figure 1B. End-labeled RNAs released by this treatment were again purified by 10% (29:1) PAGE, before being subjected to both chemical and enzymatic sequencing reactions (Donis-Keller *et al.*, 1977; Peattie, 1979; Donis-Keller, 1980). Sequence obtained for the 3′ 25–40 nt of each of these RNAs identified them as two forms of small ribosomal RNA1 (srRNA1), SL RNA, U4, U6, SLA RNA and an unknown RNA doublet of ~62 nt (Figure 1B; sequencing data not shown). Previous characterization of AMT crosslinked RNAs of *T.brucei* has shown that, depending on the sieve of PAGE, the SL RNA crosslinked species have electrophoretic mobilities that overlap with the U4/U6 crosslinked species as well as crosslinked and uncrosslinked forms of srRNA1. However, since RNase H experiments had indicated that ‘Z’ did not contain any of these other co-migrating RNAs (Watkins *et al.*, 1994 and data not shown), we focused our efforts on obtaining the sequence of the unknown 61 and 62 nt RNA(s).

We found that the RNAs in the doublet possess the

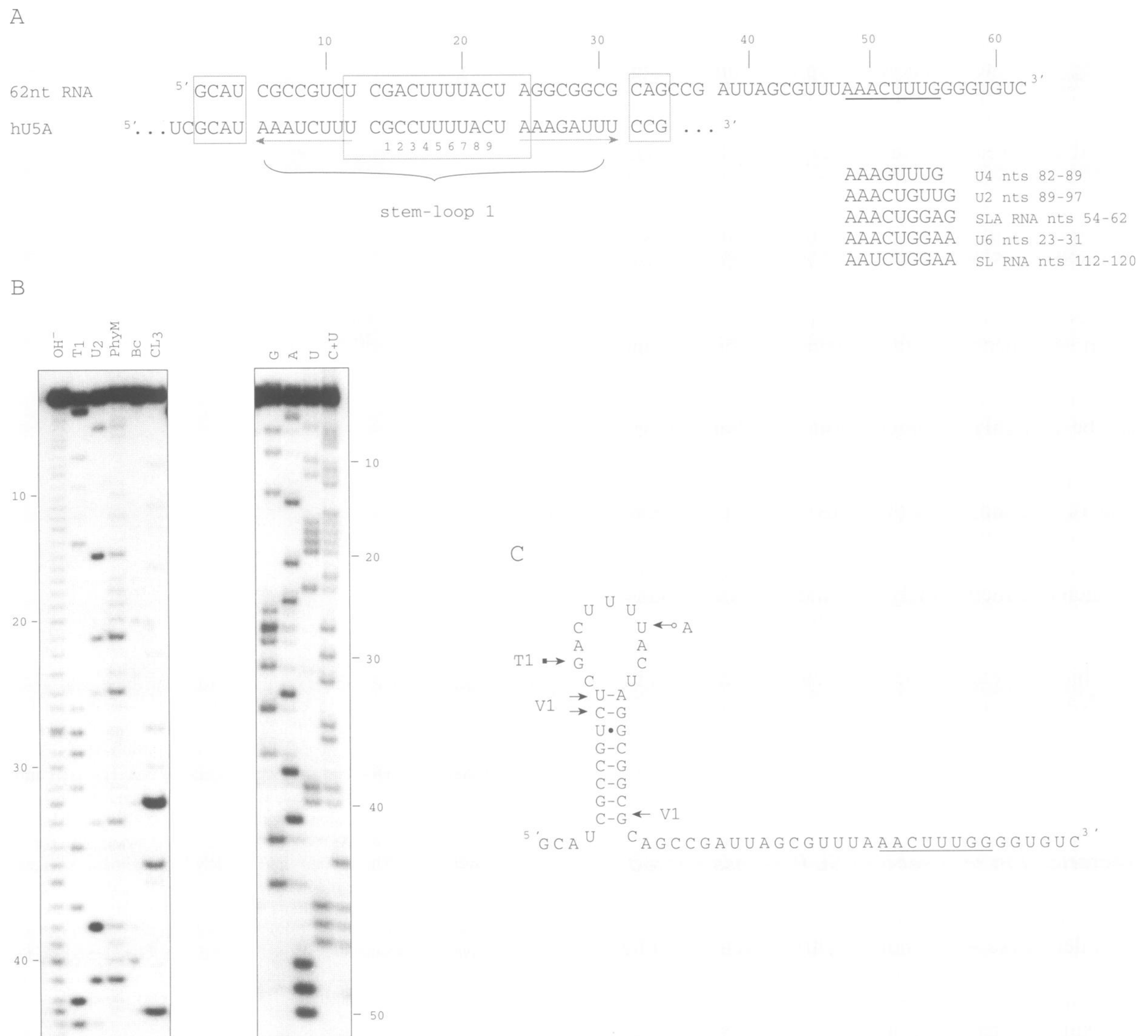


Fig. 2. Sequence analysis of the unknown species reveals it to be a 62 nt RNA with homology to U5 snRNA. (**A**) Sequence of the 62 nt RNA. The 5′ domain of the 62 nt RNA contains a sequence that is the same as the central portion of the conserved stem–loop 1 domain of human U5A at 12 out of 13 residues (boxed). The U5 loop 1 consensus sequence numbering convention is indicated. To the right is a list of sequences present in *T.brucei* spliceosomal RNAs at positions where core binding proteins may bind; the underlined sequence in the 62 nt RNA is most similar to those found in U4 and U2. (**B**) Comparison of alkaline hydrolysis (lane OH^-) with enzymatic (lanes T_1, U_2, PhyM, Bc and CL_3, which show cleavages after G, A, A+U, U+C and C residues, respectively) and chemical (lanes G, A, U and C+U) sequencing reactions on 3′ splint-labeled 62 nt RNA shows an absence of 2′-*O*-methyl modifications and pseudouridines. (**C**) The results of limited nuclease digestion of 3′ splint-labeled 62 nt RNA with nucleases under non-denaturing conditions are consistent with a model of the 5′ half of the 62 nt RNA as a stem–loop structure. Arrows denote sites of the single strand-specific T_1 and A cleavages, and those of nuclease V_1, which preferentially cleaves residues present in base-paired or stacked regions.

same RNA sequence, differing by the presence or absence of a 3′-terminal C residue. The complete sequence of this RNA (Figure 2A) was obtained by a combination of reverse transcriptase sequencing of both the RNA and its complementary DNA (data not shown) in addition to direct chemical and enzymatic sequencing of 3′ splint-labeled RNA (Figure 2B and data not shown; see Materials and methods). All of these methods were required due to extreme compression phenomena in several specific regions of the sequence. The order of a final 2 nt compression was confirmed by restriction enzyme digestion analysis of the cloned gene (W.P.Mitchell, personal communication). At the primary sequence level, this RNA does not bear significant homology to any known RNA, other than a 13 nt stretch (nt 12–24) which matches the central portion of the U5 conserved stem–loop 1 consensus sequence at 12 out of 13 positions (Figure 2A).

When the sequence of this small RNA is subjected to secondary structure modeling (Zuker *et al.*, 1991; MFOLD version 2), several structures are obtained which place the L1-like sequence into the loop of a hairpin. Results of limited digestion using RNases V_1, CL_3 and T_1 are consistent with the folding of the 5′ half of the molecule shown in Figure 2C; the 3′ half is not attacked by these enzymes under single-hit digestion conditions (data not shown). Other notable features include nt 1–4 (5′-GCAU-3′) and 32–34 (5′-CAG-3′), which are similar in sequence and location to the conserved IL1 sequences implicated in binding U5-specific proteins (Black and Pinto, 1989; Bach and Lührmann, 1991; Frank *et al.*, 1994). Although

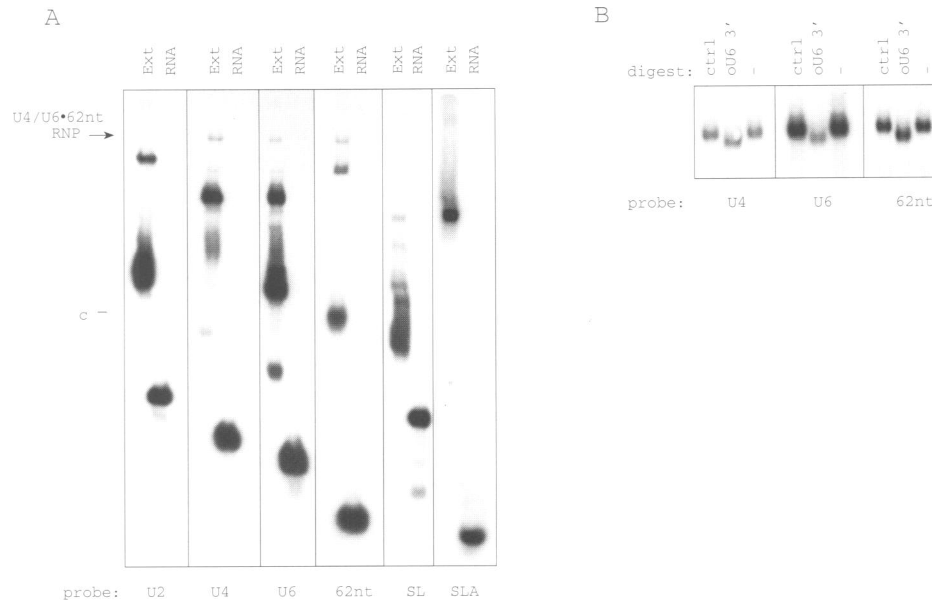


Fig. 3. The 62 nt RNA is present in *T.brucei* extracts in RNP complexes which also contain the U4 and U6 RNAs. (**A**) One of the SLA2 RNP complexes co-migrates with a U4/U6 RNP complex. Crude *T.brucei* extract (lanes Ext) and naked RNA (lanes RNA) samples were loaded in repeated configuration for non-denaturing PAGE, followed by transfer to a nylon filter and probing for the small RNAs of *T.brucei*, as indicated. (**B**) The 62 nt RNA is present in an RNP complex which also contains the U4 and U6 RNAs. Crude *T.brucei* extract was subjected to a mock RNase H digestion with no exogenous oligodeoxynucleotide (lanes –), a control digest with an oligodeoxynucleotide targeting the SL RNA (lanes ctrl), or digestion with an oligodeoxynucleotide targeting the 3′ end of the U6 RNA (lanes oU6 3′). These assays were displayed by non-denaturing PAGE as for (A) and probed as indicated, except that the gel was electrophoresed for twice the volt-hours.

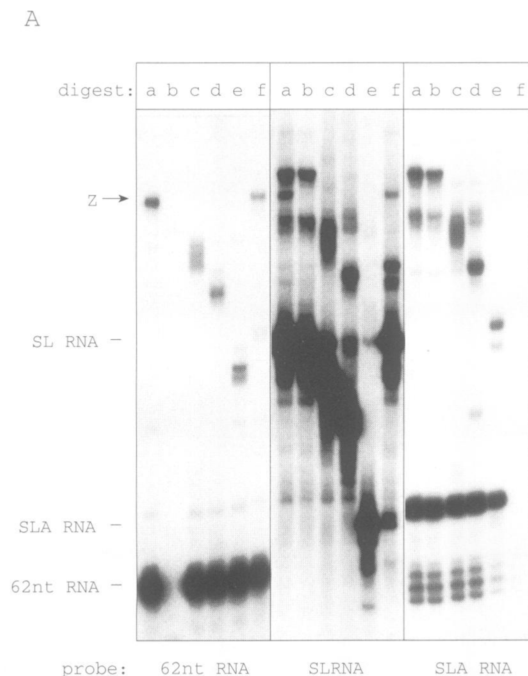
Frank *et al.* (1994) found that IL1 must be closed by a stem in order to complement U5 function in yeast, this sequence in the 62 nt RNA cannot be closed by a hairpin, since the 5′ side of it is at the 5′ terminus of the molecule.

Nucleotides 48–56 (5′-AAACUUUGG-3′) of this small RNA are reminiscent of sequences present in the other *T.brucei* snRNAs (see Figure 2A) at positions where the antigenically divergent *T.brucei* snRNA core proteins may bind (Cross *et al.*, 1991; Palfi *et al.*, 1991; Gunzl *et al.*, 1992; Palfi and Bindereif, 1992; J.M.Dungan, unpublished observations). Non-denaturing gel electrophoresis of crude *T.brucei* extracts shows that the 62 nt RNA is present in several RNP complexes (Figure 3A, panel probed for 62 nt RNA). As is characteristic of snRNP core particles, one of these RNP complexes (noted 'c' in Figure 3A) is stable to high concentrations of salt and sediments in Cs_2SO_4 equilibrium density gradients at a density intermediate (~1.3 g/ml) between those of free RNA (~1.6 g/ml) and free protein (~1.2 g/ml; data not shown). Of particular note, we find that one of the slower, salt-sensitive 62 nt RNP complexes co-migrates with RNP complexes hybridizing to probes for U4 and U6 RNAs, but none of the other *T.brucei* snRNAs (Figure 3A). RNase H targeting of the 3′ end of U6 RNA in these extracts alters the electrophoretic migration of the complex observed with probes to all three RNAs (Figure 3B), showing that all three RNAs are present in this RNP complex.

U5 snRNAs characterized thus far possess a 5′ trimethylguanosine (TMG) cap and pseudouridine modifications at certain positions in the most highly conserved portion of the molecule, the L1 sequence (Reddy and Busch, 1988; Szkukalek *et al.*, 1995). In contrast to the above similarities of the 62 nt RNA to U5, immunoprecipitation results indicate that this small RNA appears not to possess either a 5′ TMG or a 7-methyl guanosine (7mG) cap (T.G.Roberts and J.M.Dungan, unpublished observations). In addition, the 62 nt RNA 5′ end is not affected by periodate oxidation and β-elimination treatment, nor is it a substrate for guanylyl transferase (data not shown). These results are consistent with a 5′ end bearing a modified di- or tri-phosphate (as is found on U6; Singh and Reddy, 1989), a 5′ hydroxyl or a 5′ monophosphate. Finally, the 62 nt RNA appears to lack the usual internal modifications found in U5 snRNA: none of the U residues are resistant to hydrazine treatment (Figure 2B, lane U), indicating an absence of pseudouridines, and alkaline hydrolysis of 3′ splint-labeled RNA yields an even ladder (Figure 2B, lane OH^-; and data not shown), indicating an absence of sugar modifications at the 2′ position.

The ~62 nt RNA is crosslinked to the SL RNA

To confirm that the 62 nt RNA accounts for the unidentified SL RNA crosslinked species 'Z', we performed oligodeoxynucleotide-directed RNase H analysis of crosslinked RNA (Figure 4A; diagrammed in Figure 4B) targeting the 62 nt RNA (lanes b), the SL RNA (lanes c–e) or the SLA RNA (lanes f). A mock reaction containing no oligodeoxynucleotide was included for comparison (lanes a). Aliquots of RNA from these digests were electrophoresed as a set, loaded in triplicate on the same gel; nucleic acids were electrophoretically transferred to a nylon filter which was then sliced into thirds and probed for the 62 nt RNA, the SL RNA or the SLA RNA (see Materials and methods). Comparing the three RNA probes of the mock digestion reaction (lanes a), the sum of crosslinks visualized with the SLA RNA and 62 nt RNA probes accounts for all of the major crosslinked species seen with the SL RNA probe. Furthermore, neither the



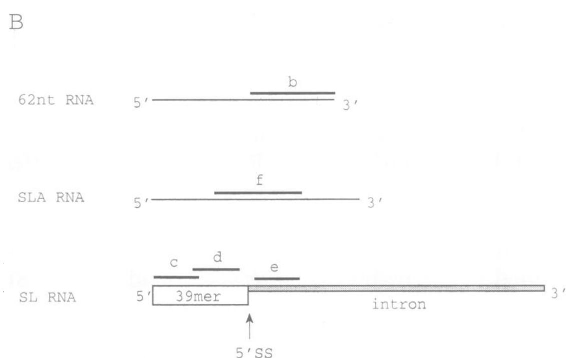


Fig. 4. RNase H targeting shows that the 62 nt RNA accounts for the remaining SL RNA crosslinked species. (**A**) RNA from psoralen-treated cells was incubated with RNase H in the absence (lanes a) or presence of oligodeoxynucleotides targeting the 62 nt RNA (lanes b), SL RNA (lanes c–e) or SLA RNA1 (lanes f), as indicated. Aliquots of the digests were electrophoresed in repeated load configuration and analyzed as in Figure 1. (**B**) Diagram of sites targeted by oligodeoxynucleotides used in (A).

SLA RNA nor the 62 nt RNA has significant crosslinks apart from the set shared with the SL RNA. Thus, the psoralen-entrappable interactions for these two RNAs are with the SL RNA, and vice versa. This is evidence that RNA–RNA crosslinking by psoralen *in vivo* is quite stringent, and does not reflect possible non-specific interactions with, for example, tRNA or rRNA, each of which are several orders of magnitude more abundant than any of the *T.brucei* snRNAs (e.g. see Figure 1 in Watkins and Agabian, 1991). Indeed, the relative abundance of the SL RNA crosslinked species is comparable with the abundance of U4/U6 crosslinks (see Figure 1 in Watkins *et al.*, 1994), an interaction that is stabilized in *T.brucei* by two regions of extensive base-pairing interactions.

The identification of the 62 nt RNA–SL RNA crosslink by co-migration is upheld upon RNase H digestion of the specific subsets of SL RNA crosslinks. In addition to efficiently and specifically targeting the 62 nt RNA, RNase H digestion using oligo b, complementary to the 3′ end of the 62 nt RNA (Figure 4A, lanes b), causes the specific ablation of the ‘Z’ crosslinked species (seen in the panels probed for the SL RNA and the 62 nt RNA), while the remainder of the SL RNA crosslinks are unaffected. Conversely, RNase H targeting of the SL RNA (lanes c–e) specifically alters the mobility of only the crosslinked form of the 62 nt RNA. An oligodeoxynucleotide that efficiently targets the SLA RNA (lanes f), however, has no effect on the mobility of any species visualized with the 62 nt RNA probe or of the SL RNA ‘Z’ crosslinked species. These data prove that the remaining prominent psoralen crosslinked SL RNA species ‘Z’ consists of an intermolecular crosslink between the SL RNA and the 62 nt RNA, which we therefore refer to as the SLA2 RNA.

The SLA2 RNA L1-like sequence is crosslinked to the exon just upstream of the 5′ splice site

RNase H digestions roughly localize the ‘Z’ crosslink positions to the 5′ half of the SLA2 RNA and the splice site region of the SL RNA. This information is embedded in the analysis shown in Figure 4. Similar RNase H analysis performed on gel-purified SL–SLA2 crosslinked species, 3′ end-labeled by T4 RNA ligase with pCp, confirms that the SL RNA crosslink site is 3′ of the oligo d target (nt 16–35), while digestion with oligo e releases the uncrosslinked 3′ end of the SL RNA (data not shown). To refine this map, we analyzed a highly purified fraction of the ‘Z’ crosslinked species both by primer extension and by RNase sequencing reactions. RNA substrate for primer extension analysis of crosslink positions was purified as described for Figure 1, with an additional RNase H digestion targeting the SLA RNA (to eliminate any minor co-migrating non-SLA2–SL RNA crosslinks) prior to a final additional gel purification. This sample was then primer extended with a 5′ end-labeled oligodeoxynucleotide complementary to either the 3′ end of the SL RNA (Figure 5A, lane Z) or the 3′ end of the SLA2 RNA (Figure 5B, lane Z), and electrophoresed next to corresponding primer extension sequencing ladders (lanes C, U, A and G) and control reactions (lanes N, N′) made from total RNA. SL RNA primer extension stops in lane N show obstructions to reverse transcriptase in the SL RNA population present in total RNA, including branched splicing intermediates (stops near G40) and secondary structures; these stops are also apparent in the sequencing ladders made from total RNA (lanes C, U, A, G). The N′ lane shows additional stops in total RNA caused by treatment with psoralen (presumably a combination of monoadducts and crosslinks); the ratio of these stops to full-length reverse transcripts reflects the degree of psoralen adduct formation and the heterogeneity of species present in the RNA sample. SL RNA primer extension of the highly purified crosslinked RNA substrate (lane Z) shows almost complete obstruction to full-length reverse transcription with stops at positions G39, U38, U37 and A36. These data are consistent with psoralen adducts at U38, U37 and U35 (position –2, –3 and –5 relative to the 5′ splice site; see Watkins *et al.*, 1994 for a discussion of the interpretation of primer extension stop sites regarding psoralen adducts). A similar analysis of SLA2 RNA by

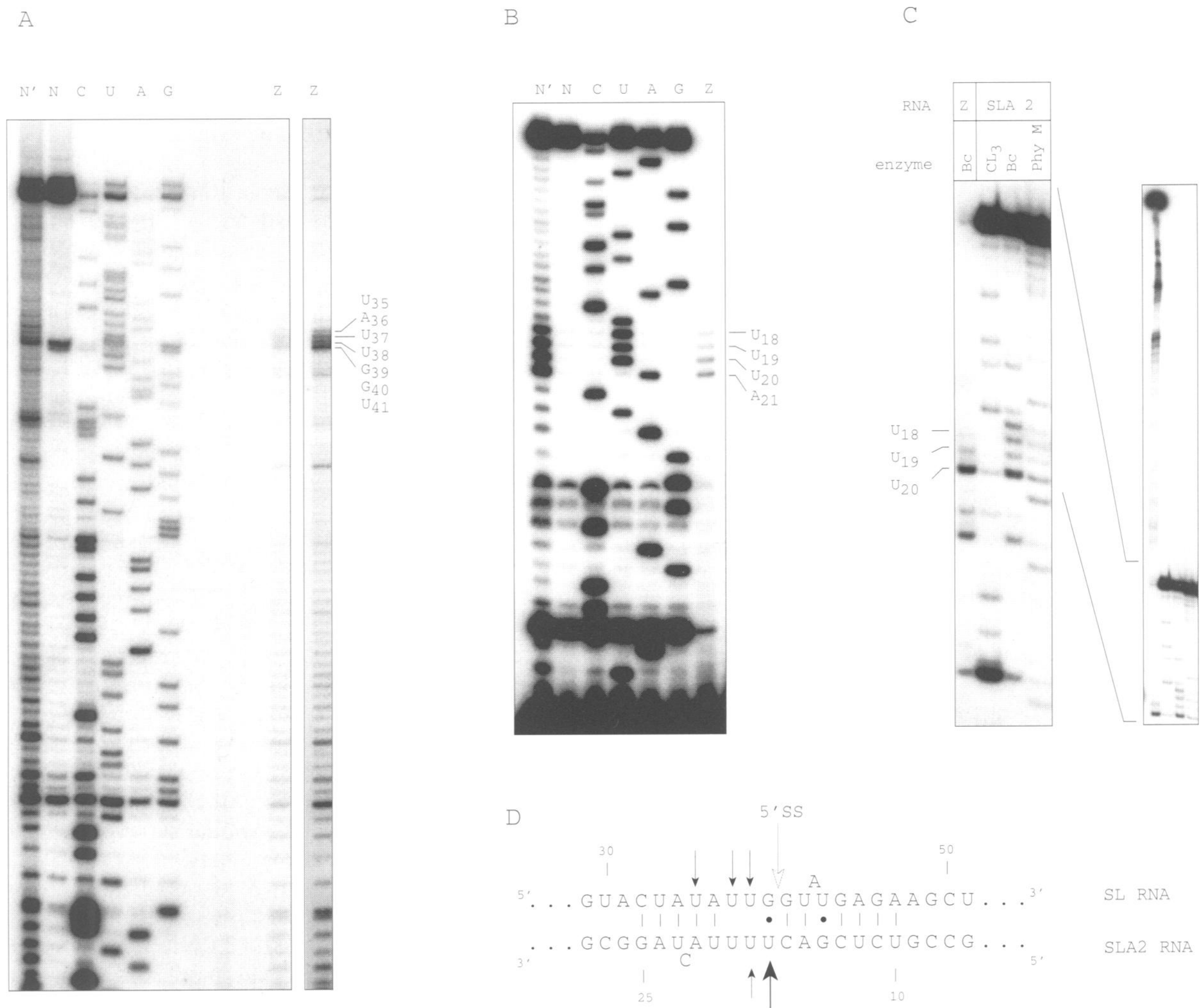


Fig. 5. Primer extension and nuclease analysis defines the SLA2–SL RNA crosslink position. (**A**) SL RNA primer extension analysis. Reference ladders were made by primer extension of total RNA made from untreated cells (lanes N, C, U, A and G) or total RNA from psoralen-treated cells (lanes N′). The 'Z' crosslinked RNA sample was digested with RNase H and oligonucleotide f (see Figure 4), targeting the SLA RNA prior to the final gel purification to avoid reverse transcriptase stops from any possible co-migrating SLA–SL RNA crosslinks. A darker exposure of the 'Z' primer extension reaction is shown in the panel on the right. (**B**) SLA2 RNA primer extension analysis was performed on the same RNA samples as those used in part (A). (**C**) A partial RNase B.c. digest of SLA2-splint-labeled 'Z' crosslinked species was electrophoresed next to uncrosslinked splint-labeled SLA2 RNA reference ladders. The panel on the right is photoreduced to show the relative migration of 'Z' and species cleaved upstream of the crosslink position, which migrate aberrantly since they are still crosslinked to the SL RNA. (**D**) Diagram of possible base-pairing interactions between the SLA2 and SL RNAs consistent with psoralen crosslinks between SL U38–SLA2 U17 and SL U37–SLA2 U18. Psoralen adduct positions are indicated by black arrows.

primer extension shows enhanced stops at A21, U20, U19 and U18, consistent with psoralen adducts at U20, U19, U18 and U17. Again, while these stops are present in both lanes N′ and Z, the level of obstruction to full-length reverse transcription in lane Z is almost complete, reflecting the purity of the crosslinked RNA substrate.

To distinguish the position of the crosslink site(s) from presumed monoadduct sites in this run of U residues in SLA2, we re-purified the 'Z' crosslinked species as before, except that the starting material was 'splint-labeled' using an oligo overhanging the 3′ end of the SLA2 RNA (see Materials and methods). This material was subjected to partial digestion with *Bacillus cereus* (B.c.) RNase (Figure 5C, 'Z' RNA, lane B.c.), which cleaves after C and U residues. Uncrosslinked, end-labeled SLA2 RNA from psoralen-treated cells was digested with RNases in parallel to generate a set of reference ladders (lanes CL_3, B.c. and Phy M). Although it has been reported that some RNases cleave psoralen-adducted RNAs aberrantly (Garrett-Wheeler *et al.*, 1984), comparison of nuclease digestion patterns of uncrosslinked SLA2 RNA from untreated and psoralen-treated cells showed no difference (Figures 5C versus 2B; and data not shown). The pattern of RNase B.c. cleavage is identical for the crosslinked and uncrosslinked RNA samples, until residue U18, the signal for which is much fainter in the crosslinked RNA sample, while the signal for residue U17 is completely absent. This is consistent with residue U17, and possibly also U18, being sites of crosslinking with the SL RNA, in parallel with the sites of both psoralen- and 4-thiouridine-mediated crosslinking of U5 to the 5′ splice site region in HeLa cell extracts (Wassarman and Steitz, 1992; Wyatt *et al.*, 1992; Sontheimer and Steitz, 1993).

Psoralens are tricyclic compounds known to intercalate

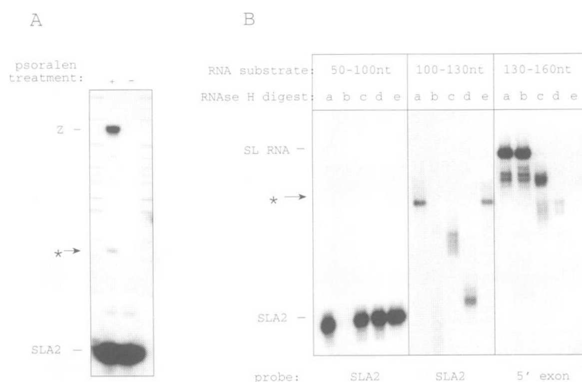


Fig. 6. The SLA2 RNA is crosslinked to an apparent *trans*-splicing intermediate, the free 5′ SL exon. (**A**) Northern analysis of RNA as in Figure 1, probed with SLA2 antisense DNA. The crosslinked species of interest is denoted by an asterisk (*). (**B**) RNA from psoralen-treated cells was size fractionated by gel electrophoresis as described in the text. Size-selected RNA substrate was subjected to RNase H digestion using the same oligodeoxynucleotides (b–e) as in Figure 4, and followed by Northern analysis, as indicated.

double-stranded nucleic acids and form covalent adducts to residues on either side (usually pyrimidines) upon irradiation with longwave UV light (for a review, see Cimino *et al.*, 1985). In the ribosome, psoralen treatment has been shown to crosslink nucleotides of adjacent base pairs in helical regions as well as those which interact over long-range distances, with respect to secondary structure predictions (Thompson and Hearst, 1983). When the regions of the SL RNA and SLA2 RNA that crosslink to each other are inspected, it is obvious that the two RNAs have the potential to form base pairs which might facilitate AMT crosslink formation; this is depicted in Figure 5D. Genetic and phylogenetic analyses will be required to determine any significance of these possible interactions. It is also possible that some other tertiary or protein-mediated interactions stabilize the juxtaposition of these regions of the SL RNA and the SLA2 RNA prior to crosslinking *in vivo*.

Analysis of a crosslink between the SLA2 RNA and a putative trans-splicing intermediate

The potential structural homology of the SLA2 RNA to the U5 snRNA, its presence in an RNP complex with the U4 and U6 RNAs, and the positions of its crosslink(s) to the 5′ splice site region are all indications that the SLA2 RNA is a trypanosome version of U5. Mammalian U5 has been shown to interact physically with the 5′ splice site region in both the precursor mRNA and the splicing intermediate free 5′ exon during splicing *in vitro* (Wyatt *et al.*, 1992; Sontheimer and Steitz, 1993). In the course of analyzing the SLA2 RNA crosslinked species, we noticed a minor species that had the characteristics of a crosslink to the free 39 nt 5′ exon intermediate (see the asterisked species in Figure 6A). With respect to end-labeled DNA molecular weight markers, this species migrated at ~115 nt on a denaturing 6% gel and at ~150 nt on a denaturing 10% gel (data not shown). While RNase H analysis of this species in total RNA was suggestive (data not shown), the presence of the other, more abundant crosslinked species in the same reactions made the interpretation of these data subject to ambiguity. We therefore used size-selected RNA to repeat the type of RNase H analysis shown in Figure 4 on the putative SLA2–5′ exon intermediate crosslink.

Several milligrams of RNA from psoralen-treated cells were size selected from two successive rounds of preparative 6% (29:1) PAGE, using radiolabeled DNA molecular weight markers to estimate RNA size fractions. We used the ~50–100, ~100–130 and ~130–160 nt fractions for RNase H digestions of uncrosslinked SLA2 RNA, putative SLA2–free 39mer crosslink and uncrosslinked SL RNA, respectively. Portions of these fractions were treated with RNase H and oligodeoxynucleotides as previously diagrammed (Figure 4B), targeting the SLA2 RNA (lanes b) or regions of the SL RNA (lanes c–e), or subjected to mock digestion (lanes a), followed by Northern analysis as shown (Figure 6B). As in Figure 4, these treatments efficiently and specifically targeted the SLA2 RNA and the SL RNA (seen in RNA substrate panels 50–100 and 130–160 nt, respectively). The crosslink to the putative *trans*-splicing intermediate (100–130 nt RNA substrate panel) is efficiently targeted by oligodeoxynucleotides complementary to the SLA2 RNA (lanes b), and two regions of the 39 nt spliced leader exon (lanes c and d), but not by an oligo targeting the intron portion of the SL RNA immediately adjacent to the 5′ splice site (lanes e). This result is consistent with the identification of this species as a crosslink between the SLA2 RNA and free 39 nt 5′ exon. Reprobing of the Northern panels in Figure 6B shows that there are a similar set of crosslinks between the SLA RNA previously identified (Watkins *et al.*, 1994) and putative free 39 nt 5′ exon (data not shown). The putative SLA RNA–5′ exon crosslinked species are more heterogeneous than the SLA2 RNA–5′ exon species shown here, likely reflecting the more heterogeneous set of crosslinks formed between the SLA RNA and the SL RNA (Watkins *et al.*, 1994).

If the SLA2–5′ exon species contains authentic *trans*-splicing intermediate, digestion with RNase T1 should yield a 10 nt fragment derived from the 3′ end of the spliced leader which possesses a 3′-OH (5′-UACUAUA-UUG_{OH}-3′) as a result of its cleavage during the first step of splicing. To determine whether this was the case, the remainder of the ~100–130 nt size-selected RNA fraction was 3′ splint labeled using an oligodeoxynucleotide specific for the SLA2 RNA. The splint-labeled, size-selected RNA was re-purified by denaturing 10% (19:1) PAGE, to shift the crosslinked RNA away from the majority of the uncrosslinked RNAs with which it co-migrated during the first gel purifications. This material was then 3′ end-labeled with T4 RNA ligase and [5′-^{32}P]pCp, before (Figure 7A) or after (Figure 7B) crosslink photoreversal treatment. Although there were other RNA species present in this highly purified fraction which were also end-labeled by RNA ligase, RNase H digestion demonstrates that the species containing the SLA2 RNA (asterisked species in Figure 7A, lane a) also possesses all of the 5′ exon-containing material (Figure 7A: compare RNase H digests c and b with the mock digestion, lane a). Psoralen crosslink photoreversal of this material showed that T4 RNA ligase labeled the crosslinked 5′ exon at only ~10% of the efficiency of that achieved when the material was photoreversed prior to the end-labeling reaction (data not shown), consistent with the likely steric

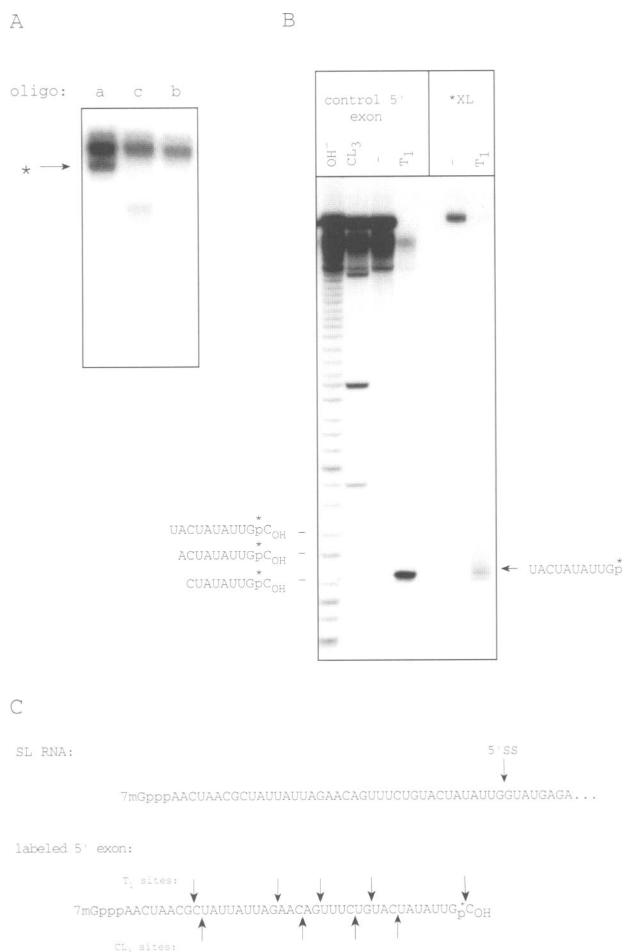


Fig. 7. The SLA2 RNA is crosslinked to authentic 5′ exon splicing intermediate. The putative 39mer–SLA2 RNA crosslinked species was gel purified and 3′ end-labeled as described in the text. (**A**) [5′-^{32}P]pCp-labeled material was subjected to RNase H treatment and electrophoresed on a 10% (29:1) sequencing gel. The faster-migrating band (*) present in mock-digested starting material (lane a) is efficiently targeted by an oligodeoxynucleotide complementary to either the 5′ end of the SL (lane c) or the 3′ end of the SLA2 RNA (lane b). (**B**) The apparent 39mer released by photoreversal treatment (see text) has the same electrophoretic mobility as endogenous free 5′ exon, splint labeled as described (lanes –). Partial degradation of splint-labeled endogenous free 5′ exon by RNase CL_3 and alkaline hydrolysis provides a frame of reference (lanes OH^- and CL_3). Partial digestion lanes for the control 5′ exon each contain ~1000 c.p.m., while complete T_1 digests and undigested *XL lanes each contain only ~100 c.p.m.. Complete digests of photoreleased 39mer and endogenous free 5′ exon RNAs result in T_1 fragments of identical mobility (lanes T_1).

hindrance of RNA ligase by the proximity of the psoralen crosslink site to the 3′ end of the 5′ exon.

The ~39 nt RNA released from the crosslinked material by photoreversal treatment and post-labeled by T4 RNA ligase was purified by denaturing 10% (19:1) PAGE. It was then treated with calf intestinal alkaline phosphatase to allow direct comparison with a standard generated by splint labeling free 5′ exon present in total RNA (see Materials and methods). Figure 7B shows that the phosphatased, [5′-^{32}P]pCp-labeled RNA released by photoreversal treatment does indeed co-migrate with the splint-labeled free 39mer standard (compare lanes –). These two RNAs should differ only by the presence of a 3′-terminal ribose (on the [5′-^{32}P]pCp-labeled RNA) versus a 3′-terminal deoxyribose (on the splint-labeled standard). The end-labeling of the photo-released 39mer RNA species by T4 RNA ligase and [5′-^{32}P]pCp demonstrates that it originally possessed a 3′-OH.

Partial digestion of the standard RNA using RNase CL_3 (which cleaves after C residues) and alkaline hydrolysis (Figure 7B, lanes CL_3 and OH^-) defines the reference register for the size and sequence of the end-labeled spliced leader fragments. Complete digestion of the two RNAs with RNase T_1 results in 10 nt fragments of identical electrophoretic mobility. Since the complete digests vary from the partial digests by both the loss of the 3′-terminal C residue and the presence of the 3′-terminal phosphate left by RNase T_1 (the position of the radiolabel in both cases), the resulting fragment is expected to migrate as it does, out of register with the partial alkaline hydrolysis ladder (which should possess the 3′-terminal OH group of the 3′ end label). These results demonstrate that the SLA2 RNA is indeed crosslinked *in vivo* to an RNA which is equivalent to the free 5′ exon *trans*-splicing intermediate. The formal possibility exists that the free 5′ exon we observe is not the free 5′ exon splicing intermediate, *per se*. However, consistent with the identification of this species as a crosslink to a true splicing intermediate, inhibition of *trans*-splicing *in vivo* with the drug sinefungin (McNally and Agabian, 1992) also results in the specific elimination of this crosslink (K.P.Watkins, J.M.Dungan and N.Agabian, manuscript in preparation). We present the characterization of the SLA2 RNA crosslink to the free 5′ exon as further strong support of the identification of the SLA2 RNA as the trypanosome counterpart of the U5 snRNA.

Discussion

We have used psoralen crosslinking treatment to search in *T.brucei* for RNAs involved in *trans*-splicing *in vivo*. Analysis of crosslinked species formed by the 5′ splice site-bearing SL RNA allowed us previously to identify the novel 72 nt SLA RNA, which accounts for most of the RNA–RNA interactions that can be trapped in this manner (Watkins *et al.*, 1994). The SLA RNA contained no significant homologies to any known snRNA, other than a sequence (5′-CUUUUA-3′) near the 5′ end which is also found in the conserved loop of U5 snRNA. Because of this, we offered several hypotheses as to its possible role in the *trans*-spliceosome: that it may be a highly diverged form of either U5 or U1 (from the structure of the crosslinks it formed at the 5′ splice site) or possibly an snRNA unique to either *trans*-splicing or trypanosomes. Although the 5′-CUUUUA-3′ sequence has proven not to be a conserved feature of the SLA RNA (Palfi *et al.*, 1994), this does not rule out the possibility that the SLA RNA may perform some U5 function(s). Comparative analysis shows that there is a conserved complementarity between the SLA RNA and a sequence just upstream of the 5′ splice site across a range of distantly related species of kinetoplastidae (Roberts, 1996), consistent with one of the models of interaction presented to account for psoralen crosslinking *in vivo* (see Watkins *et al.*, 1994; Figure 7A). In this work, we have characterized a second spliced leader-associated RNA which accounts for the remaining

abundant SL RNA crosslinked species and which is likely to be the trypanosome version of U5 snRNA. Because it is so different from U5 found in other organisms, we use four separate criteria to reach this conclusion: (i) it appears to be structurally homologous to the minimally essential portion of U5; (ii) it crosslinks to 5′ exon sequences in a manner exactly analogous to the *cis*-splicing U5–5′ splice site interaction; (iii) it is present in a tri-snRNP complex with the U4/U6 RNAs, as is U5 in other systems; (iv) it can be crosslinked to a molecule that is identical to the free 5′ exon *trans*-splicing reaction intermediate, as U5 can be in *cis*-splicing. These data implicate the SLA2 RNA not only as an active participant in *trans*-splicing, but also as the trypanosome version of U5.

Interactions between the SLA2 RNA and the 5′ splice site region

Oligodeoxynucleotide-directed RNase H digestion, primer extension and direct ribonuclease analyses indicate that the SL RNA crosslink to the SLA2 RNA consists mainly of diadducts between exon positions –2, –3 and –5 and SLA2 RNA residues U17 and possibly U18 (corresponding to U5 loop 1 consensus residues 4 and 5). Psoralens generally mediate crosslinking of pyrimidines on opposing strands after intercalation into a double-helical region, although RNA tertiary interactions are also known to be entrapped by psoralen crosslinking (Cimino *et al.*, 1985). If the occurrence of these psoralen crosslinks is interpreted to substantiate possible base pairing between the 5′ splice site and the SLA2 RNA loop 1 sequence, the alignment of these two regions about the crosslink positions shows they do indeed have the potential to form a set of base pairs compatible with the formation of crosslinks between residues U17/exon position –2, and U18/exon position –3 (see Figure 5D). However, these potential base pairs with the SLA2 RNA encompass regions of the SL RNA implicated in interactions with the SLA RNA (Roberts, 1996) and U6 RNA. It is therefore possible that the SLA2–SL RNA interaction is mutually exclusive with that of either or both of these other RNAs. However, as previously noted, U6 RNA can be psoralen crosslinked to this region of the intron (Watkins *et al.*, 1994), and we also find a very low-abundance psoralen crosslinked species which contains both the SLA2 and U6 RNAs crosslinked to the SL RNA (data not shown). Characterization of SLA2 RNA homologs in other kinetoplastid organisms will allow us to determine whether these potential SLA2–SL RNA interactions are conserved. It may be more likely that the observed crosslinking arises from juxtaposition of these two regions via some other tertiary or protein-mediated interaction.

Genetic studies show that the sequence of the U5 conserved loop can profoundly affect the choice of 5′ splice site cleavage in both yeast and mammalian cells (Newman and Norman, 1991, 1992; Cortes *et al.*, 1993). These studies imply that U5 loop sequences may be involved in some sort of base-pairing interactions with exon nucleotides just upstream of the 5′ splice site. The lack of conserved sequence in *cis*-spliced 5′ exons likely prevents this interaction from being a rigidly conserved set of Watson–Crick base pairs, and it has been postulated that the ability of U residues to base pair 'promiscuously' may account for the absolute conservation of the stretch of uridines in the U5 loop 1 (Newman and Norman, 1992). There is no similar constraint in Kinetoplastida, where all 5′ exons are *trans*-spliced and identical within a species. The conservation of U5 loop I and the homologous region of the *T.brucei* SLA2 RNA hints that there must be some other purpose for this stretch of sequence. We note that the base pairing suggested by the location of the psoralen crosslinks aligns the SLA2 RNA loop 1 sequence with the SL RNA in a manner exactly analogous to that proposed for U5 with the 5′ splice site by Newman and Norman (1992), positioning SLA2 RNA residue U17 (corresponding to U5 loop 1 residue 4) opposite the spliced leader exon position –1, and residue C16 (corresponding to loop 1 position 3) opposite intron position G1.

Although both the genetic data and the psoralen crosslinks suggest that base pairing may be an important feature of the interaction of U5 and the 5′ splice site (Newman and Norman, 1991, 1992; Wassarman and Steitz, 1992; Cortes *et al.*, 1993), it is likely that, rather than consisting of a static structure, these interactions occur over time. Indeed, physical studies of U5 in *cis*-splicing extracts using 4-thiouridine (4SU) to crosslink factors interacting with specific positions of the 5′ exon indicate that there is a succession of interactions with this region. Loop residues 7 and 5 juxtapose to exon residue –2 early in the process (Wyatt *et al.*, 1992), while loop residues 5 and 4 are close to exon position –1 later, during the transesterification reactions (Sontheimer and Steitz, 1993; Newman *et al.*, 1995). If the psoralen crosslinks we have characterized reflect a similar set of interactions in *trans*-splicing, perhaps only a subset proximal to the 5′ splice site is present in the crosslink to the free 39mer exon. Although 4SU crosslinking reflects proximity rather than indicating a base-pairing interaction, it is notable that both the observed 4SU crosslinks to U5 loop positions (Wyatt *et al.*, 1992; Sontheimer and Steitz, 1993; Newman *et al.*, 1995) and our SLA2 RNA psoralen crosslinks are compatible with the Newman model.

The existence of the trypanosome 5′ splice site on a small RNA of uniform sequence has facilitated our characterization of the RNAs with which it interacts *in vivo*. The remaining U5 RNA–RNA interactions occur at the 3′ splice sites, which are diverse in *T.brucei* as in other organisms. Because of this and the lack of an *in vitro* splicing system from kinetoplastid extracts, detection and analysis of snRNA–3′ splice site interactions require either scaling up of experimental starting material by three to four orders of magnitude, or a genetic approach. Since the SLA2 RNA is present as a single-copy gene in *T.brucei* (W.P.Mitchell, J.Holmes, J.M.Dungan and N.Agabian, unpublished observations), we should be able to determine whether it is essential and to dissect its role(s) in *trans*-splicing genetically. A recently developed extrachromosomally maintained vector for the transformation of *T.brucei* (Metzenberg and Agabian, 1994) will allow us to examine the effects of mutating both the SLA2 RNA gene and a 3′ splice site reporter construct.

SLA2 RNP complexes

The identification of the SLA2 RNA as a U5 homolog predicts its existence in trypanosome versions of a 10S core RNP particle, a 20S RNP complex containing additional U5-specific binding proteins, and a 25S tri-snRNP

complex with the U4/U6 snRNP, in addition to spliceosomal complexes (reviewed in Will *et al.*, 1993 and Moore *et al.*, 1993). The snRNA core binding proteins of 'higher' eukaryotes have recently been characterized to be members of a family of RNA binding proteins which contain an evolutionarily conserved pair of sequence motifs that likely affect protein folding and are involved in protein–protein interactions (Cooper *et al.*, 1995; Hermann *et al.*, 1995; Séraphin, 1995). This group of proteins was initially characterized by their common recognition by anti-Sm antisera, but not all members of this newly expanded family are apparently precipitable with anti-Sm antibodies. The RNA binding requirements for this core of proteins are somewhat complex, consisting of a single-stranded consensus Sm site (PuA$(U)_{3-6}$NUGPu), whose binding affinity is influenced by surrounding sequences and/or structures (Jarmolowski and Mattaj, 1993). Although *T.brucei* contains neither proteins that react with anti-Sm antibodies (Michaeli *et al.*, 1990; Palfi and Bindereif, 1992) nor a good match to the Sm consensus binding site in its snRNAs, it does have a set of proteins common to U2, U4/U6 and SL RNP particles (Palfi *et al.*, 1991), and inspection of protected regions of RNA in these RNPs (Cross *et al.*, 1991; Gunzl *et al.*, 1992; J.M.Dungan, unpublished observations) reveals a common sequence that varies by simple one or two nucleotide changes present in each of the *T.brucei* snRNAs, including both of the SLA RNAs (see Figure 2A). We have shown by non-denaturing gel electrophoresis that the SLA2 RNA is present in *T.brucei* extracts in several RNP complexes (Figure 3), one of which is resistant to high-salt isopycnic centrifugation conditions, as is characteristic of the core snRNPs (Lelay-Taha *et al.*, 1986). The first SLA RNA we identified was found in a salt-resistant core RNP and was precipitable with antibodies that recognize *T.brucei* snRNA core binding proteins (Palfi *et al.*, 1994; Roberts, 1996); we predict that the Cs_2SO_4-stable SLA2 RNP particles we observe also contain these core binding proteins.

In further support of the SLA2 RNA as a U5 homolog, we find that one of the complexes present in crude extracts also contains the U4 and U6 RNAs, and may therefore be a trypanosome version of the U4/U6·U5 tri-snRNP complex. The entry of the tri-snRNP complex is a critical step in the formation of the *cis*-spliceosome; the presence of a similar complex in *T.brucei* suggests that this aspect of the spliceosome assembly pathway may also be conserved. As with the U4/U6·U5 tri-snRNP complex, the U4/U6·SLA2 complex appears to be disrupted by treatment with high salt (data not shown). It will be of interest to determine whether the two slower migrating SLA2 RNP complexes contain trypanosome versions of the U5-specific proteins, especially the highly conserved PRP8, present in the U5 20S and 25S complexes (reviewed in Will *et al.*, 1993).

Upon reflection, the SLA RNAs might appear to constitute a separate category of spliceosomal RNP which lacks an inverted nucleoside cap yet contains core binding proteins. Studies of the biogenesis of small RNAs in other eukaryotes show that there is a relationship between the transcribing RNA polymerase, the acquisition of an inverted nucleoside cap and, particularly, the binding of core proteins and hypermethylation of the inverted cap to TMG (reviewed in Mattaj, 1988; Izaurralde and Mattaj, 1992; see also Terns *et al.*, 1995). In trypanosomes, these relationships appear to be at least partially uncoupled. The *T.brucei* U2 snRNA is an apparent RNA polymerase III transcript (Ullu and Tschudi, 1990; Fantoni *et al.*, 1994), and yet it acquires an inverted TMG nucleoside cap. The SL RNA is bound by core proteins, but its inverted cap dinucleotide is left as a 7mG, while the next four nucleotides are hypermethylated (Perry *et al.*, 1987; Bangs *et al.*, 1992). The SLA and SLA2 RNAs both apparently lack inverted nucleoside 5′ caps, but may possess some other 5′ modification similar to that of U6 (Roberts, 1996 and data not shown). In recent work which may help resolve the relationship of 5′ cap modification to the binding of snRNP core proteins, newly identified members of the expanded Sm motif family have been shown to bind U6 (Cooper *et al.*, 1995; Séraphin, 1995). We note with interest that the *T.brucei* U6 RNA contains a match to the putative *T.brucei* snRNP core protein binding site noted above (Figure 2A). It is also possible that the SLA RNAs' lack of an inverted cap structure might be the default result of 5′ end biogenesis from a short-lived precursor. Consistent with this, long exposures of Northerns show there are indeed two longer forms of the SLA2 RNA (~73 and ~90 nt) which are not present when transcription is inhibited *in vivo* (Figure 6A; and data not shown).

Finally, although all U5 RNAs thus far characterized possess both a TMG cap and pseudouridine modifications in the conserved loop 1 (Szkukalek *et al.*, 1995), it has been found that neither the specific TMG cap nor the pseudouridine modifications are necessary for *cis*-splicing in HeLa extracts, while reconstitution with snRNP core proteins is required for the reaction (Ségault *et al.*, 1995). Thus, it seems likely that the lack of a TMG cap and the minimal size of the SLA2 RNA are consequences of the divergence of trypanosome RNA maturation pathways, reducing the evolutionary constraints on U5 sequence close to the absolute minimum required for splicing.

The trypanosome trans-spliceosome

The criteria that distinguish the SLA2 RNA as a U5 homolog are not so clearly applicable to the first SLA RNA we characterized, which may have a function unique to *trans*-splicing (Watkins *et al.*, 1994; Roberts, 1996). If the psoralen crosslinking treatment that trapped interactions of the SLA RNAs with the SL RNA did so because they both occur via base pairing with the exon regions noted (Figure 5D; Watkins *et al.*, 1994; T.G.Roberts *et al.*, submitted), then it would seem that the interactions of these two RNAs with the 5′ splice site might be mutually exclusive, barring the formation of a triple helix. The apparent overlapping complementarity of the SLA RNAs with the 5′ exon is reminiscent of the interactions of U1 and U6 with the 5′ end of the intron in the *cis*-spliceosome (reviewed in Moore *et al.*, 1993). If the 5′ splice site interactions with the SLA RNA and the SLA2 RNA complex are mutually exclusive, then the SLA RNA–5′ splice site interaction might be analogous in function (though not structure) to the U1–5′ splice site interaction, which appears to be displaced by the U4/U6·U5 tri-snRNP during spliceosome assembly (Konforti *et al.*, 1993; Moore *et al.*, 1993). Confounding the clear exclusion of the SLA RNA as a form of U5 is limited evidence suggesting that

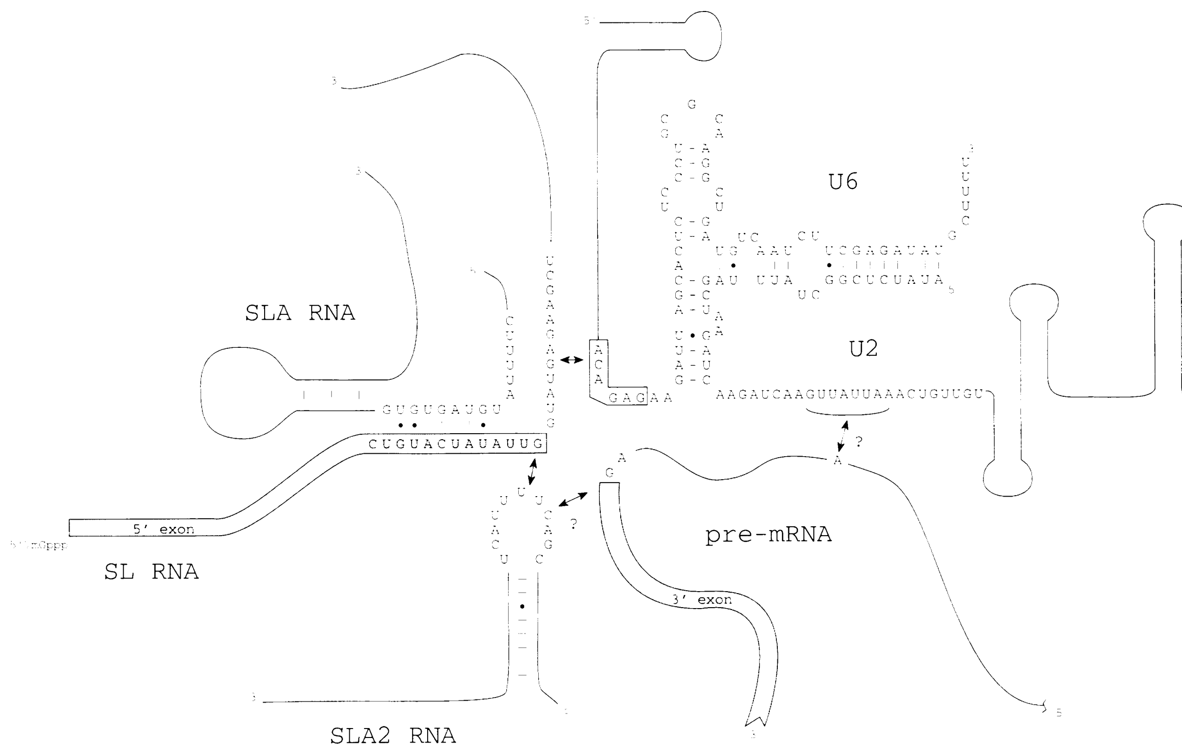


Fig. 8. Model of the trypanosome *trans*-spliceosome: the SLA2 RNA may not displace the SLA RNA during *trans*-splicing. The SLA RNA is shown base paired to the SL exon via sequences which are conserved in Kinetoplastida (Roberts, 1996). U6/U2 base pairings are adapted from Watkins and Agabian (1991) and Madhani and Guthrie (1992). Interactions between the small RNAs and substrates are indicated by arrows (see the text); question marks denote putative interactions for which there are as yet no data.

it too may be crosslinked to the free 5′ 39 nt exon (discussed above). Furthermore, part of the potential base-pairing interaction of the SLA2 RNA with the 5′ splice site involves the very intron sequences for which U1 and U6 seem to compete in the *cis*-spliceosome. It seems most likely that the SLA2 RNA–5′ splice site interaction is not exclusively dependent on complementarity, but is facilitated by other factors during spliceosome assembly. In this case, the SLA RNA base paired to the exon upstream may not be displaced during *trans*-spliceosome assembly, and the three RNAs may be simultaneously associated with the 5′ splice site region (see Figure 8). This might explain the existence of psoralen crosslinks of both of the SLA RNAs to the free 5′ exon, as well as the observation, discussed above, of a crosslinked species containing the SLA2, U6 and SL RNAs.

Although it is not known whether *trans*-splicing in Kinetoplastida is directly analogous to *trans*-splicing in the distantly related nematodes, the process in both organisms has been shown to require the U2, U4 and U6 snRNAs by oligodeoxynucleotide-directed RNase H knockout experiments in permeabilized trypanosome cells (Tschudi and Ullu, 1990) and in splicing extracts made from *Ascaris* embryos (Hannon *et al.*, 1991). The involvement of U5 in this process was previously debatable, since there was no clear candidate homolog for U5 in trypanosomes and the *Ascaris* U5 RNP was found to be refractory to RNase H targeting. U2, U6 and U5 RNAs appear to comprise the RNA heart of the *cis*-spliceosome; the apparent absence of U5 from this intricate network of RNA interactions in trypanosome *trans*-splicing was therefore conspicuous. The data presented here strongly support a model of the *trans*-spliceosome in which a trypanosome U5 structural homolog is juxtaposed to the 5′ splice site in the catalytic center during the trans-esterification reactions. Although the conserved ability of the spliced leader RNA to switch between two secondary structures may reflect some *trans*-splicing-specific function(s) (LeCuyer and Crothers, 1993), this study shows that U5 is not replaced by intramolecular interactions within the SL RNA, as has been proposed (Steitz, 1992). The presence of a U5-like RNA in *T.brucei* demonstrates that the fundamental components of the spliceosome's RNA core are phylogenetically conserved from the very ancient protozoa through mammals.

Materials and methods

Cell growth, RNA isolation and extract preparation

Growth of *T.brucei brucei* IsTat 1.1 procyclics, *in vivo* psoralen crosslinking, *in vitro* crosslink photoreversal treatments and RNA isolation were performed as described previously (Watkins *et al.*, 1994). Extracts used for non-denaturing gel electrophoresis and Cs_2SO_4 gradients were prepared as follows. Procyclic *T.brucei brucei* IsTat 1.1 were grown to a density of ~8×10^6 cells/ml, harvested and washed twice in buffer A [20 mM Tris–HCl pH 7.5, 1 mM EDTA, 1 mM EGTA, 1 mM spermidine, 0.3 mM spermine, 5 mM β-mercaptoethanol, 10 μg/ml leupeptin, 84 μg/ml phenylmethylsulfonyl fluoride (PMSF)] before freezing in liquid nitrogen. A cell pellet containing ~4×10^{10} trypanosomes was thawed and resuspended by Dounce in a final volume of 8 ml buffer A. The resulting lysate was centrifuged for 20 min at 11 500 r.p.m. in a type 70.1 Ti rotor. The pellet fraction was resuspended in an equal volume of buffer B (buffer A plus 0.4 M potassium glutamate and 4 mM magnesium acetate). Following incubation with agitation for 20 min at 4°C, this suspension was centrifuged for 30 min at 20 000 r.p.m. in a type 70.1 Ti rotor. The supernatant was aliquoted and flash frozen in liquid nitrogen. Samples for non-denaturing gel electrophoresis were

thawed, diluted with four parts of buffer C (20 mM HEPES–KOH pH 7.9, 2 mM $MgCl_2$, 5 mM β-mercaptoethanol) and incubated in the presence or absence of 1 μM of the indicated oligodeoxynucleotide for 1 h at 30°C. Samples were diluted with an equal volume of buffer D (buffer C plus 50 mM potassium glutamate, 10% glycerol) containing 0.1 μg/ml bromophenol blue and xylene cyanol before non-denaturing gel electrophoresis. Cs_2SO_4 gradient centrifugation was performed as previously described (Michaeli *et al.*, 1990), followed by dialysis of fractions against buffer D before non-denaturing gel electrophoresis.

Oligodeoxynucleotides and probes

Oligodeoxynucleotides used for 3′ end splint labeling of RNAs: oJD41 (5′-GGGACACCCCAAAGTTTAAACGC-3′), complementary to the 3′ end of the SLA2 RNA; and oSL3′ (5′-GAGTGGAGGTCATCCGAC-CCC-3′), complementary to the 3′ end of the SL RNA; oJD45 (5′-GCAATATAGTACAGAAACTGTTCTAATAATAGCGTT-3′), complementary to the SL RNA free 5′ exon. Oligodeoxynucleotides used in RNase H digestions: oligo b is oJD41, oligo c is complementary to SL RNA nt 1–20, oligo d is complementary to SL RNA nt 16–35, oligo e is complementary to SL RNA nt 37–57, and oligo f is complementary to nt 23–44 of the SLA RNA (Watkins *et al.*, 1994). Oligodeoxynucleotides used for primer extension reactions: oJD39 (5′-CGGCCCCAGGG-TTTAGGCGC-3′), oJD41 and oSL3′. oSLA2(+) is a sense oligodeoxynucleotide of the same sequence as SLA2 RNA nt 3–22 (5′-ATCGCCGT-CTGGACTTTTAC-3′).

Antisense RNA probes for U2, U4, U6 and SL RNAs were prepared as described previously (Watkins and Agabian, 1991). SLA RNA was probed with oligodeoxynucleotide oKW74 (Watkins *et al.*, 1994), 5′ end-labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP. Antisense DNA probes for the SLA2 RNA and SL RNA 5′ end were made by asymmetric PCR with *Thermus aquaticus* (*Taq*) DNA polymerase. Reactions contained ~2 ng DNA template, 1 μM oligodeoxynucleotide and 10 mM Tris pH 9.0 at 25°C, 50 mM KCl, 3 mM $MgCl_2$, 0.1% Triton X-100, 3 μM each dATP, dGTP and dTTP, 1.3 μM (4 mCi/ml) [α-^{32}P]dCTP and 1 U *Taq* DNA polymerase. Reactions were incubated for 10–25 cycles of 30 s intervals at 94, 50 and 72°C, followed by phenol–chloroform extraction and P6-DG (Bio-Rad) spin column desalting to separate probe from unincorporated nucleotides. Asymmetric PCR for generating antisense SL RNA 5′ end probe used gel-purified *Sal*I–*Eco*RI insert from pSPSL2 (Michaeli *et al.*, 1990) as template and oJD9 (complementary to SL RNA nt 31–50) as primer. SLA2 RNA template: AMV reverse transcriptase was used with oJD41 and total *T.brucei* RNA to generate SLA2 RNA cDNA as described (Watkins *et al.*, 1994). cDNA was amplified by PCR using oligos oSLA2(+) and oJD41. PCR conditions used were the same as those for generating probe, except that all four dNTPs were non-radioactive and present at 0.2 mM. Amplified PCR product was gel purified before use as probe template. Asymmetric PCR for generating antisense SLA2 RNA used oJD41 as primer.

Gel electrophoresis and Northern analysis

Non-denaturing gel electrophoresis was performed as per Konarska (1989) in 4% (80:1) acrylamide:bis-acrylamide gels buffered with 50 mM Tris base–50 mM glycine. Denaturing PAGE was performed in 8 M urea and 1× TBE (89 mM Tris, 89 mM borate, 2.5 mM EDTA). For Northern analysis, samples were electrophoretically transferred to magna NT nylon filters (MSI) in 0.25× TAE (10 mM Tris–acetate pH 8.0, 0.25 mM EDTA) at 15 V for 12 h in the cold. Transferred nucleic acids were UV crosslinked to filters prior to probing. Antisense RNA probes were hybridized at 55°C in 60% formamide, 5× SSPE [0.9 M NaCl, 40 mM sodium phosphate pH 7.4, 5 mM EDTA], 5× Denhardt's [0.1% each ficoll, polyvinylpyrrolidone and bovine serum albumin (BSA)], 0.5% SDS, 0.1 mg/ml *Escherichia coli* tRNA, and washed three times at 55°C in 0.1× SSPE, 0.1% SDS. Antisense DNA probes were hybridized at 45°C in 5× SSPE, 5× Denhardt's, 0.5% Sarkosyl, 0.1 mg/ml *E.coli* tRNA, and washed three times at 45°C in 1× SSPE, 0.5% Sarkosyl (filters probed for free 5′ exon and SLA2 RNA).

Enzymatic manipulations and RNA sequencing

3′ End-labeling reactions. T7 DNA polymerase (Sequenase version 2.0, US Biochemicals) and oligodeoxynucleotides oJD41, oSL3′ or oJD46 (listed above) were used to selectively 3′ end-label the SLA2 RNA, the SL RNA or the SL RNA free 5′ exon, respectively, using the splint-labeling technique developed by Hausner *et al.* (1990). Prior to end-labeling, the total RNA sample was digested with 0.05 U/μl RQ1 DNase I (Promega) at 37°C for 2 h in 20 mM HEPES–KOH pH 7.5, 0.1 M NaCl, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 0.05 mg/ml BSA, 0.1 U/μl RNasin (Promega). DNase I digestion was stopped by phenol–chloroform extraction and nucleic acids were ethanol precipitated after adjusting the aqueous phase to 2 M NH_4OAc. The sample was then reprecipitated from 0.3 M NaOAc, washed with 75% ethanol and dried. DNase-treated RNA sample (~100 μg) was annealed with the splint oligodeoxynucleotide (2–20 μM) in 0.1 M NaCl, 2 mM EDTA, 40 mM Tris pH 7.9, by boiling for 1 min, followed by slow cooling to room temperature. An equal volume of a mixture containing 60 mM Tris pH 7.9, 20 mM $MgCl_2$, 10% glycerol, 2 mM DTT, 2 U/μl RNasin, 200 μCi [α-^{32}P]dCTP (3000 Ci/mmol) and 2 U/μl Sequenase version 2.0 was then added to the annealed RNA–oligo sample; the total reaction volume was 10 μl. Reactions to splint label the SL RNA additionally included 50 μM dATP and dTTP, to fill in its heterogeneous 3′ end (K.P.Watkins, unpublished observations). The splint reaction was incubated for 20 min at room temperature, stopped by the addition of EDTA to 25 mM, and phenol–chloroform extracted. RNA was separated from unincorporated nucleotides by two rounds of ethanol precipitation from 2 M NH_4OAc, as above. After gel purification, the yield of a particular splint-labeled RNA varied, depending on the RNA target, its relative abundance, and the length and T_M of the oligodeoxynucleotide chosen. Omission of the DNase I digestion step and/or inclusion of all four dNTPs leads to an increased background of random incorporation. Gel-purified RNA samples were 3′ end-labeled with [5′-^{32}P]pCp and T4 RNA ligase as per England *et al.* (1980).

Sequencing analysis. AMV reverse transcriptase reactions were performed as described (Watkins *et al.*, 1994), using either total RNA with 5′ end-labeled oJD41, or SLA2 RNA cDNA with 5′ end-labeled oSLA2(+) as templates and primers. oJD39 is partially complementary to the 3′ end of the SLA2 RNA; it was used to obtain sequence of the 5′ end of the SLA2 RNA. Sequence of the 5′ end was used to make oSLA2(+), which was then used to sequence cDNA of poly(A)-tailed SLA2 RNA, as described (Watkins *et al.*, 1994). This information was used to synthesize oJD41. Psoralen adduct positions in the SL RNA were determined by primer extension of the RNA sample with 3′ end-labeled oSL3′. Enzymatic RNA sequencing reactions were performed as per Donis-Keller (Donis-Keller *et al.*, 1977; Donis-Keller, 1980); chemical sequencing reactions were as per Peattie (1979), except that only 200–600 c.p.m. 3′ end-labeled RNA was used in each reaction containing 1 μg (for nuclease reactions) or 10 μg (for chemical reactions) carrier *E.coli* tRNA. Periodate and β-elimination treatment of 3′ splint-labeled SLA2 RNA was performed as per Greenburg and Burn (1988). Guanylyl transferase reactions were carried out at 37°C for 1 h in 50 mM HEPES–KOH (pH 7.9), 1.2 mM $MgCl_2$, 2.5 mM DTT, 125 μM S-adenosyl methionine, 0.1 mM GTP, 0.1 U/μl RNasin and 0.2 U/μl guanylyl transferase (Gibco). To assess possible RNA secondary structure, gel-purified 3′ splint-labeled SLA2 RNA was resuspended in 20 mM HEPES–KOH (pH 7.9), 0.1 M KCl, 5 mM $MgCl_2$, 0.5 mg/ml carrier *E.coli* tRNA, incubated for 2 min at 50°C, and slow-cooled to room temperature prior to partial digestion with RNases V_1, T_1 and A; 'single hit' digestion conditions were determined by titration and time course.

To verify the identity of the 5′ exon–SLA2 RNA crosslinked species, gel-purified [5′-^{32}P]pCp-labeled 39mer photo-released from the (*) SLA2 RNA crosslink was first treated with 0.4 U calf intestinal alkaline phosphatase (CIAP) in the presence of 2 μg carrier *E.coli* tRNA to remove the 3′-terminal phosphate in 4 μl 10 mM Tris pH 7.5, 1 mM EDTA at 37°C. CIAP was heat inactivated at 70°C for 5 min, and the sample was split between two tubes. One of these was digested with 1 U T1 RNase for 30 min at 37°C, followed by 15 min at 50°C. Both were then diluted with 2 vols 10 M urea, 0.25× TBE loading buffer, boiled for 1 min and electrophoresed on a 10% (19:1) sequencing gel next to similar T1 treatments and reference ladders of the standard (endogenous SL RNA free 5′ exon, which had been splint labeled using the oligo oJD46 described above).

RNA used in RNase H reactions was first treated with DNase I as outlined for the splint-labeling protocol above, then hybridized with the oligodeoxynucleotide of choice at 10 μM in 0.1 M HEPES–KOH pH 7.5, 0.1 M NaCl, 1 mM EDTA by boiling for 1 min and slow cooling to room temperature. To this was added an equal volume of 1 mg/ml carrier *E.coli* tRNA, 0.5 mg/ml BSA, 20 mM $MgCl_2$, 2 mM DTT, 0.1 U/μl RNasin, 0.1 U/μl RNase H; the 10 μl reaction was incubated at 37°C for 30 min, followed by phenol–chloroform extraction and ethanol precipitation.

Acknowledgements

Dedicated to the memory of William M.Dungan, D.V.M. We would like to thank Cammy Diges for her excellent technical assistance, and T.Guy

Roberts, Wayne P.Mitchell, Jeff Holmes, Ted White and members of the Agabian lab for their interest, comments and helpful discussions. This work was supported by grants to N.A. from the John D. and Catherine T.MacArthur Foundation and the NIH.

References

Bach,M. and Lührmann,R. (1991) Protein-RNA interactions in 20S U5 snRNPs. *Biochim. Biophys. Acta*, **1088**, 139–143.

Bangs,J.D., Crain,P.F., Hashizume,T., McCloskey,J.A. and Boothroyd, J.C. (1992) Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J. Biol. Chem.*, **267**, 9805–9815.

Black,D.L. and Pinto,A.L. (1989) U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. *Mol. Cell. Biol.*, **9**, 3350–3359.

Bonen,L. (1993) *Trans*-splicing of pre-mRNA in plants, animals and protists. *FASEB J.*, **7**, 40–46.

Bruzik,J.P., Van,D.K., Hirsh,D. and Steitz,J.A. (1988) Trans splicing involves a novel form of small nuclear ribonucleoprotein particles. *Nature*, **335**, 559–562.

Cimino,G.D., Gamper,H.B., Isaacs,S.T. and Hearst,J.E. (1985) Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Annu. Rev. Biochem.*, **54**, 1151–1193.

Cooper,M., Johnston,L.H. and Beggs,J.D. (1995) Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J.*, **14**, 2066–2075.

Cortes,J.J., Sontheimer,E.J., Seiwert,S.D. and Steitz,J.A. (1993) Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5′ splice sites *in vivo*. *EMBO J.*, **12**, 5181–5189.

Crispino,J.D. and Sharp,P.A. (1995) A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes Dev.*, **9**, 2314–2323.

Crispino,J.D., Blencowe,B.J. and Sharp,P.A. (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science*, **265**, 1866–1869.

Cross,M., Gunzl,A., Palfi,Z. and Bindereif,A. (1991) Analysis of small nuclear ribonucleoproteins (RNPs) in *Trypanosoma brucei*: structural organization and protein components of the spliced leader RNP. *Mol. Cell. Biol.*, **11**, 5516–5526.

Donis-Keller,H. (1980) Phy M: an RNase activity specific for U and A residues useful in RNA sequence analysis. *Nucleic Acids Res.*, **8**, 3133–3142.

Donis-Keller,H., Maxam,A.M. and Gilbert,W. (1977) Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res.*, **4**, 2527–2538.

England,T.E., Bruce,A.G. and Uhlenbeck,O.C. (1980) Specific labeling of 3′ termini of RNA with T4 RNA ligase. *Methods Enzymol.*, **65**, 65–74.

Fantoni,A., Dare,A.O. and Tschudi,C. (1994) RNA polymerase III-mediated transcription of the trypanosome U2 small nuclear RNA gene is controlled by both intragenic and extragenic regulatory elements. *Mol. Cell. Biol.*, **14**, 2021–2028.

Frank,D., Roiha,H. and Guthrie,C. (1994) Architecture of the U5 small nuclear RNA. *Mol. Cell. Biol.*, **14**, 2180–2190.

Garrett-Wheeler,E., Lockard,R.E. and Kumar,A. (1984) Mapping of psoralen crosslinked nucleotides in RNA. *Nucleic Acids Res.*, **12**, 3405–3423.

Greenberg,J.R. and Burn,V.E. (1988) Proteins associated with rabbit reticulocyte mRNA caps during translation as investigated by photocrosslinking. *Nucleic Acids Res.*, **16**, 3437–3454.

Gunzl,A., Cross,M. and Bindereif,A. (1992) Domain structure of U2 and U4/U6 small nuclear ribonucleoprotein particles from *Trypanosoma brucei*: identification of *trans*-spliceosomal specific RNA–protein interactions. *Mol. Cell. Biol.*, **12**, 468–479.

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

Hausner,T.P., Giglio,L.M. and Weiner,A.M. (1990) Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev.*, **4**, 2146–2156.

Hermann,H., Fabrizio,P., Raker,V.A., Foulaki,K., Hornig,H., Brahms,H. and Lührmann,R. (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein–protein interactions. *EMBO J.*, **14**, 2076–2088.

Hodges,P.E., Jackson,S.P., Brown,J.D. and Beggs,J.D. (1995) Extraordinary sequence conservation of the PRP8 splicing factor. *Yeast*, **11**, 337–342.

Izaurralde,E. and Mattaj,I.W. (1992) Transport of RNA between nucleus and cytoplasm. *Semin. Cell Biol.*, **3**, 279–288.

Jarmolowski,A. and Mattaj,I.W. (1993) The determinants for Sm protein binding to Xenopus U1 and U5 snRNAs are complex and non-identical. *EMBO J.*, **12**, 223–232.

Konarska,M.M. (1989) Analysis of splicing complexes and small nuclear ribonucleoprotein particles by native gel electrophoresis. *Methods Enzymol.*, **180**, 442–453.

Konforti,B.B. and Konarska,M.M. (1995) A short 5′ splice site RNA oligo can participate in both steps of splicing in mammalian extracts. *RNA*, **1**, 815–827.

Konforti,B.B., Koziolkiewicz,M.J. and Konarska,M.M. (1993) Disruption of base pairing between the 5′ splice site and the 5′ end of U1 snRNA is required for spliceosome assembly. *Cell*, **75**, 863–873.

LeCuyer,K.A. and Crothers,D. (1993) The *Leptomonas collosoma* spliced leader RNA can switch between two alternate structural forms. *Biochemistry*, **32**, 5301–5311.

Lelay-Taha,M.N., Reveillaud,I., Sri-Wadada,J., Brunel,C. and Jeanteur,P. (1986) RNA–protein organization of U1, U5, and U4-U6 small nuclear ribonucleoproteins in HeLa cells. *J. Mol. Biol.*, **189**, 519–532.

MacMillan,A.M., Query,C.C., Allerson,C.R., Chen,S., Verdine,G.L. and Sharp,P.A. (1994) Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev.*, **8**, 3008–3020.

Madhani,H.D. and Guthrie,C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*, **71**, 803–817.

Mattaj,I.W. (1988) U snRNP assembly and transport. In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, pp. 100–114.

McNally,K.P. and Agabian,N. (1992) *Trypanosoma brucei* spliced-leader RNA methylations are required for *trans*-splicing *in vivo*. *Mol. Cell. Biol.*, **12**, 4844–4851.

Metzenberg,S. and Agabian,N. (1994) Mitochondrial minicircle DNA supports plasmid replication and maintenance in nuclei of *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA*, **91**, 5962–5966.

Michaeli,S., Roberts,T.G., Watkins,K.P. and Agabian,N. (1990) Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of *Trypanosoma brucei*. *J. Biol. Chem.*, **265**, 10582–10588.

Moore,M.J., Query,C.C. and Sharp,P.A. (1993) Splicing of precursors to mRNAs by the spliceosome. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–358.

Mottram,J., Perry,K.L., Lizardi,P.M., Lührmann,R., Agabian,N. and Nelson,R.G. (1989) Isolation and sequence of four small nuclear U RNA genes of *Trypanosoma brucei* subsp. *brucei*: Identification of the U2, U4, and U6 RNA analogs. *Mol. Cell. Biol.*, **9**, 1212–1223.

Newman,A. and Norman,C. (1991) Mutations in yeast U5 snRNA alter the specificity of 5′ splice-site cleavage. *Cell*, **65**, 115–123.

Newman,A.J. and Norman,C. (1992) U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell*, **68**, 743–754.

Newman,A.J., Teigelkamp,S. and Beggs,J.D. (1995) snRNA interactions at 5′ and 3′ splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *RNA*, **1**, 968–980.

Nilsen,T.W. (1993) *Trans*-splicing of nematode premessenger RNA. *Annu. Rev. Microbiol.*, **47**, 413–440.

Palfi,Z. and Bindereif,A. (1992) Immunological characterization and intracellular localization of *trans*-spliceosomal small nuclear ribonucleoproteins in *Trypanosoma brucei*. *J. Biol. Chem.*, **267**, 20159–20163.

Palfi,Z., Gunzl,A., Cross,M. and Bindereif,A. (1991) Affinity purification of *Trypanosoma brucei* small nuclear ribonucleoproteins reveals common and specific protein components. *Proc. Natl Acad. Sci. USA*, **88**, 9097–9101.

Palfi,Z., Xu,G.L. and Bindereif,A. (1994) Spliced leader-associated RNA of trypanosomes. Sequence conservation and association with protein components common to *trans*-spliceosomal ribonucleoproteins. *J. Biol. Chem.*, **269**, 30620–30625.

Patzelt,E., Perry,K.L. and Agabian,N. (1989) Mapping of branch sites in *trans*-spliced pre-mRNAs of *Trypanosoma brucei*. *Mol. Cell. Biol.*, **9**, 4291–4297.

Peattie,D.A. (1979) Direct chemical method for sequencing RNA. *Proc. Natl Acad. Sci. USA*, **76**, 1760–1764.

Perry,K.L., Watkins,K.P. and Agabian,N. (1987) Trypanosome mRNAs have unusual 'cap 4' structures acquired by addition of a spliced leader. *Proc. Natl Acad. Sci. USA*, **84**, 8190–8194.

Reddy,R. and Busch,H. (1988) Small nuclear RNAs: RNA sequences, structure, and modifications. In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, pp. 1–37.

Roberts,T.G. (1996) *Characterization of components of the trans spliceosome in Trypanosoma brucei brucei*. Ph.D. thesis. University of California, Berkeley, USA.

Ségault,V., Will,C.L., Sproat,B.S. and Lührmann,R. (1995) *In vitro* reconstitution of mammalian U2 and U5 snRNPs active in splicing: Sm proteins are functionally interchangeable and are essential for the formation of functional U2 and U5 snRNPs. *EMBO J.*, **14**, 4010–4021.

Séraphin,B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.*, **14**, 2089–2098.

Shambaugh,J.D., Hannon,G.E. and Nilsen,T.W. (1994) The spliceosomal U small nuclear RNAs of *Ascaris lumbricoides*. *Mol. Biochem. Parasitol.*, **64**, 349–352.

Singh,R. and Reddy,R. (1989) Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc. Natl Acad. Sci. USA*, **86**, 8280–8283.

Sontheimer,E.J. and Steitz,J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**, 1989–1996.

Steitz,J.A. (1992) Splicing takes a Holliday. *Science*, **257**, 888–889.

Szkukalek,A., Myslinski,E., Mougin,A., Lührmann,R. and Branlant,C. (1995) Phylogenetic conservation of modified nucleotides in the terminal loop 1 of the spliceosomal U5 snRNA. *Biochimie*, **77**, 16–21.

Tarn,W.-Y. and Steitz,J.A. (1995) SR proteins can compensate for the loss of U1 snRNP functions *in vitro*. *Genes Dev.*, **8**, 2704–2717.

Terns,M.P., Grimm,C., Lund,E. and Dahlberg,J.E. (1995) A common maturation pathway for small nucleolar RNAs. *EMBO J.*, **14**, 4860–4871.

Thomas,J., Lea,K., Zucker,A.E. and Blumenthal,T. (1990) The spliceosomal snRNAs of *Caenorhabditis elegans*. *Nucleic Acids Res.*, **18**, 2633–2642.

Thompson,J.F. and Hearst,J.E. (1983) Structure of *E. coli* 16S RNA elucidated by psoralen crosslinking. *Cell*, **32**, 1355–1365.

Tschudi,C. and Ullu,E. (1990) Destruction of U2, U4, or U6 small nuclear RNA blocks trans splicing in trypanosome cells. *Cell*, **61**, 459–466.

Tschudi,C., Richards,F.F. and Ullu,E. (1986) The U2 RNA analogue of *T. brucei gambiense*: implications for a splicing mechanism in trypanosomes. *Nucleic Acids Res.*, **14**, 8893–8903.

Tschudi,C., Krainer,A.R. and Ullu,E. (1988) The U6 small nuclear RNA from *Trypanosoma brucei*. *Nucleic Acids Res.*, **16**, 11375.

Ullu,E. and Tschudi,C. (1990) Permeable trypanosome cells as a model system for transcription and *trans*-splicing. *Nucleic Acids Res.*, **18**, 3319–3326.

Umen,J.G. and Guthrie,C. (1995) The second catalytic step of pre-mRNA splicing. *RNA*, **1**, 869–885.

Wassarman,D.A. and Steitz,J.A. (1992) Interactions of small nuclear RNAs with precursor messenger RNA during *in vitro* splicing. *Science*, **257**, 1918–1925.

Watkins,K.P. and Agabian,N. (1991) *In vivo* UV crosslinking of U snRNAs that participate in trypanosome *trans*-splicing. *Genes Dev.*, **5**, 1859–1869.

Watkins,K.P., Dungan,J.M. and Agabian,N. (1994) Identification of a small RNA that interacts with the 5′ splice site of the *Trypanosoma brucei* spliced leader RNA *in vivo*. *Cell*, **76**, 171–182.

Will,C.L., Behrens,S.E. and Lührmann,R. (1993) Protein composition of mammalian spliceosomal snRNPs. *Mol. Biol. Rep.*, **18**, 121–126.

Wyatt,J.R., Sontheimer,E.J. and Steitz,J.A. (1992) Site-specific crosslinking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. *Genes Dev.*, **6**, 2542–2553.

Zuker,M., Jaeger,J.A. and Turner,D.H. (1991) A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. *Nucleic Acids Res.*, **19**, 2707–2714.

Received on February 9, 1996; revised on March 15, 1996