MOLECULAR AND CELLULAR BIOLOGY, Mar. 1989, p. 1212–1223
0270-7306/89/031212-12$02.00/0

Vol. 9, No. 3

Isolation and Sequence of Four Small Nuclear U RNA Genes of *Trypanosoma brucei* subsp. *brucei*: Identification of the U2, U4, and U6 RNA Analogs

JEREMY MOTTRAM,[1]† KAREN L. PERRY,[1] PAUL M. LIZARDI,[2] REINHARD LÜHRMANN,[3] NINA AGABIAN,[1] AND RICHARD G. NELSON[1]‡*

University of California Intercampus Program in Molecular Parasitology, Schools of Pharmacy and Medicine, San Francisco, and School of Public Health, Berkeley, University of California at San Francisco, Laurel Heights Campus, San Francisco, California 94143[1]; Centro de Investigación sobre Ingeniería Genética y Biotecnología, Universidad Nacional Autónoma de México, 62270 Cuernavaca, Mexico[2]; and Max-Planck-Institut für Molekulare Genetik, D-1000 Berlin 33 (Dahlem), Federal Republic of Germany[3]

Received 6 June 1988/Accepted 22 November 1988

Trypanosomes use *trans* splicing to place a common 39-nucleotide spliced-leader sequence on the 5′ ends of all of their mRNAs. To identify likely participants in this reaction, we used antiserum directed against the characteristic U RNA 2,2,7-trimethylguanosine (TMG) cap to immunoprecipitate six candidate U RNAs from total trypanosome RNA. Genomic Southern analysis using oligonucleotide probes constructed from partial RNA sequence indicated that the four largest RNAs (A through D) are encoded by single-copy genes that are not closely linked to one another. We have cloned and sequenced these genes, mapped the 5′ ends of the encoded RNAs, and identified three of the RNAs as the trypanosome U2, U4, and U6 analogs by virtue of their sequences and structural homologies with the corresponding metazoan U RNAs. The fourth RNA, RNA B (144 nucleotides), was not sufficiently similar to known U RNAs to allow us to propose an identity. Surprisingly, none of these U RNAs contained the consensus Sm antigen-binding site, a feature totally conserved among several classes of U RNAs, including U2 and U4. Similarly, the sequence of the U2 RNA region shown to be involved in pre-mRNA branchpoint recognition in yeast, and exactly conserved in metazoan U2 RNAs, was totally divergent in trypanosomes. Like all other U6 RNAs, trypanosome U6 did not contain a TMG cap and was immunoprecipitated from deproteinized RNA by anti-TMG antibody because of its association with the TMG-capped U4 RNA. These two RNAs contained extensive regions of sequence complementarity which phylogenetically support the secondary-structure model proposed by D. A. Brow and C. Guthrie (Nature [London] 334:213–218, 1988) for the organization of the analogous yeast U4-U6 complex.

Eucaryotic cells contain many small nuclear (sn) U RNAs complexed in vivo with proteins to form ribonucleoprotein (RNP) particles termed snRNPs (3, 9). Individual snRNPs are distinguished by the U snRNA species they contain and by their complement of unique and shared protein antigens (2, 6, 40). Eleven U snRNAs have been detected in mammals; eight of these (U1 through U8) have been isolated and sequenced (21, 41, 42). They are abundant RNAs (10^4 to 10^6 copies per cell) ranging in size from 57 to 216 nucleotides (nt) and are usually encoded by multicopy gene families. With the exception of U6, the U snRNAs are characteristically capped at their 5′ termini by 2,2,7-trimethylguanosine (TMG). In *Saccharomyces cerevisiae*, 24 U snRNA species have been identified; these are approximately 100-fold less abundant than their metazoan counterparts, range in size from 70 to 1,000 nt, and, except for the U3 analog, are encoded by single-copy genes (1, 36, 43, 59). Several of the yeast U RNA genes have been shown genetically to be necessary for cellular viability (3, 8, 22, 38, 50, 51). The indispensable yeast U RNAs (U1, U2, U5, and U4-U6) belong to the subclass whose snRNPs are immunoprecipitable by autoimmune antisera having Sm specificity (44). These Sm snRNPs play essential roles in the processing of nuclear mRNA precursors in both metazoa and yeasts; in vitro, each is required for formation of the 40- to 60S pre-mRNA splicing complex, or spliceosome, which carries out the reactions leading to the excision of introns and ligation of exons (3, 30, 49).

Trypanosome mRNAs differ significantly in formation and structure from those of most other eucaryotes. All trypanosome mRNAs share an identical 39-nt noncoding leader sequence at their 5′ ends termed the spliced leader (SL; 14, 37). The SL is encoded not by these structural genes but rather by a highly reiterated and tandemly organized family of SL genes (13, 33) which are transcribed to produce a primary transcript of 139 nt (10, 20, 31). This SL RNA contains the 7-methylguanosine-capped SL sequence at its 5′ end (39) followed by a consensus 5′ splice site dinucleotide, GU. Although little is known of the detailed structure of structural-gene primary transcripts in trypanosomes, it is clear that they contain a single exon which is immediately preceded at its 5′ end by a consensus 3′ splice site dinucleotide, AG; it is at this location that the SL is found in the mature mRNA. To date, no trypanosome gene that is interrupted by a conventional intervening sequence has been found.

The SL RNA acts as a 5′ exon donor in an intermolecular or *trans*-splicing reaction which places the SL on structural gene transcripts (32, 55). As in *cis* splicing, the first step in

* Corresponding author.
† Present address: Institute of Genetics, University of Glasgow, Glasgow, G11 5JS, Scotland, United Kingdom.
‡ Present address: Departments of Pharmaceutical Chemistry and Medicine, UCSF San Francisco General Hospital, 1001 Potrero Avenue, San Francisco, CA 94110.

the reaction is thought to be cleavage at the 5′ splice site coupled to formation of a 5′-2′ phosphodiester branched intermediate containing the SL RNA intron and structural gene intron-exon. In trypanosomes, this intermediate has a Y rather than a lariat structure, since the intervening sequence separating the SL and structural-gene exons is composed of introns derived from two primary transcripts.

These findings suggest that the basic biochemical reactions involved in *cis* and *trans* splicing are related and are probably mediated by spliceosomes containing similar U snRNPs (48). Formation of a *trans* spliceosome, however, would require the recognition, assembly, and juxtaposition of two distinct substrate RNAs into one complex; these functions might be carried out by a novel U snRNP(s) unique to *trans*-splicing organisms or by a common U snRNP(s) that has developed additional functions. To examine this hypothesis, we have begun to identify and characterize trypanosome U RNAs-RNPs and to define their roles in *trans* splicing. This report describes six candidate TMG-capped U RNA species in *Trypanosoma brucei* and the cloning and sequencing of four of the corresponding single-copy genes. These data definitively identify three of these RNAs as the trypanosome U2, U4, and U6 analogs.

MATERIALS AND METHODS

RNA and DNA preparation. IsTat 1.A or 1.1 bloodstream trypanosomes (53) were prepared as described previously (32). Trypanosomes were collected by centrifugation, suspended in a minimal volume of phosphate-buffered saline-glucose, and lysed by addition of 10 volumes of lysis buffer (120 mM NaCl, 10 mM EDTA, 25 mM Tris hydrochloride [pH 7.5], 2% Sarkosyl [27]) containing 1 mg of proteinase K per ml. The lysate was incubated at 37°C for 90 min and extracted with equal volumes of phenol-chloroform and chloroform; the nucleic acids were then precipitated by the addition of 2.5 volumes of ethanol. After RNase A digestion (0.1 mg/ml, 37°C, 60 min), DNA was extracted and precipitated as described above. Alternatively, RNA was purified in the same manner after digestion with DNase 1 (0.05 mg/ml, 37°C, 90 min) in 20 mM Tris hydrochloride (pH 7.5), 10 mM $CaCl_2$, and 0.5 mg of proteinase K per ml (58). RNA was also prepared by the LiCl-urea and guanidine thiocyanate-CsCl centrifugation procedures as described previously (32); no qualitative differences were noted in the anti-TMG immunoprecipitation patterns obtained with the three types of RNA preparation.

RNA immunoprecipitation, end labeling, and sequencing. Total RNA (2 mg) and anti-TMG immunoglobulin G (0.3 mg) (28) were incubated on ice for 60 min in 1 ml of 50 mM Tris hydrochloride (pH 7.4)–150 mM NaCl–1 mM EDTA–0.05% Nonidet P-40. A 5-mg (0.2-ml) amount of swollen protein A–Sepharose was added, and incubation was continued for 30 min. Immune complexes were pelleted by centrifugation and washed 10 times in the same buffer by centrifugation. The final pellet was digested with proteinase K (2 mg/ml, 37°C, 30 min) in 0.2 ml of lysis buffer; after centrifugation, the supernatant was deproteinized as described above and extracted with diethyl ether; RNA was precipitated with ethanol.

RNA samples were 3′ end labeled with [5′-^{32}P]pCp (3,000 Ci/mmol; Dupont, NEN Research Products, Boston, Mass.) and T4 RNA ligase (Pharmacia Fine Chemicals, Piscataway, N.J.) as described by England et al. (15). Alternatively, after removal of the terminal cap with tobacco acid pyrophosphatase (Promega Biotech), RNA samples were 5′ end labeled with [γ-^{32}P]ATP (3,000 Ci/mmol; Dupont, NEN) and polynucleotide kinase (39). End-labeled RNAs were fractionated on an 8% acrylamide–7 M urea sequencing gel; individual RNAs were excised from the gel and recovered by elution into 0.5 M ammonium acetate, 0.2% sodium dodecyl sulfate, and 0.4 mM EDTA. The filtered eluate was extracted twice with phenol-chloroform and diethyl ether, and the labeled RNA was ethanol precipitated in the presence of 1 μg of tRNA carrier. Equal portions of each RNA were partially digested with base-specific RNases T1 (G), U2 (A), CL3 (C), Phy M (A+U), and *Bacillus cereus* (C+U) as recommended by the supplier (Bethesda Research Laboratories, Inc., Gaithersburg, Md.). The sequencing reactions were resolved on 8 and 16% acrylamide–7 M urea gels and visualized by autoradiography.

Oligonucleotide probes complementary to portions of the enzymatically determined RNA sequences were synthesized by, and were the kind gift of, P. Barr of Chiron Corp. The following oligonucleotides were used: RNA A-1, 5′-CTGA TAAGAACAGTTTAATAAC; RNA B-3, 5′-TTATTTCTC ATTTRARAGGTT; RNA C-2, 5′-AAA(G/T)TTTCCCCGA AGAGTACCGG; and RNA D-1, 5′-TCTTCTCTGTTGAA TTTCC.

Northern and Southern hybridizations. For Northern (RNA) analysis of small RNAs, 10 μg of total RNA was electrophoresed on an 8% polyacrylamide–7 M urea gel and electroblotted to a Nytran membrane (Schleicher & Schuell, Inc., Keene, N.H.). Genomic Southern analysis using 5 μg of DNA per lane was performed as described previously (52). Hybridization probes were prepared by 5′ end labeling 20 pmol of each oligonucleotide with [γ-^{32}P]ATP and polynucleotide kinase (29). Membranes were hybridized at 37°C with 10^7 cpm of probe in 5× SSC (SSC is 0.15 M NaCl plus 0.015 M sodium citrate) (29), 0.25% Sarkosyl, and 100 μg of heparin per ml. After hybridization, the filters were washed three times at room temperature and once at 45°C in 5× SSC–0.1% sodium dodecyl sulfate. Hybridization and wash conditions were the same for both Northern and Southern blots with each of the probes.

Isolation and sequencing of U RNA genomic clones. Recombinants containing the RNA A, B, and D genes were isolated from a *T. brucei* IsTat 1.1 genomic library in λ1059 by plaque hybridization using the corresponding oligonucleotide probes. An RNA C genomic clone was isolated by colony hybridization from a minilibrary constructed in pUC19 from *Hin*dIII-digested genomic DNA which had been size selected to contain the 5.5-kilobase-pair fragment that hybridized most strongly with the RNA C-2 probe (see Fig. 3).

DNA was isolated and purified from bacteriophage and plasmid clones as described elsewhere (29). Small restriction enzyme fragments, 300 to 400 base pairs, containing the U RNA coding region were subcloned in both orientations in M13mp18 or -mp19 and fully sequenced by the dideoxy method of Sanger et al. (47). The fragments that were subcloned and sequenced for each U RNA gene and their designations are as follows: RNA A, 384-nt *Acc*I fragment, M13-AA1 and -AA2; RNA B, 300-nt *Dde*I/*Hae*III fragment, M13-BD2 and -BD4; RNA C, 368-nt *Alu*I fragment, M13-CA22 and -CA11; and RNA D, 342-nt *Hae*III fragment, M13-U619 and -U6H1. In each case, the first M13 clone designated contained the U RNA coding strand.

The 5′ ends of RNAs B through D were mapped by primer extension sequencing from total RNA as described by Murphy et al. (32). The 3′ ends of these RNAs were approximately determined on the basis of the known positions of the 5′ ends and the sizes of the RNAs estimated from a sequenc-

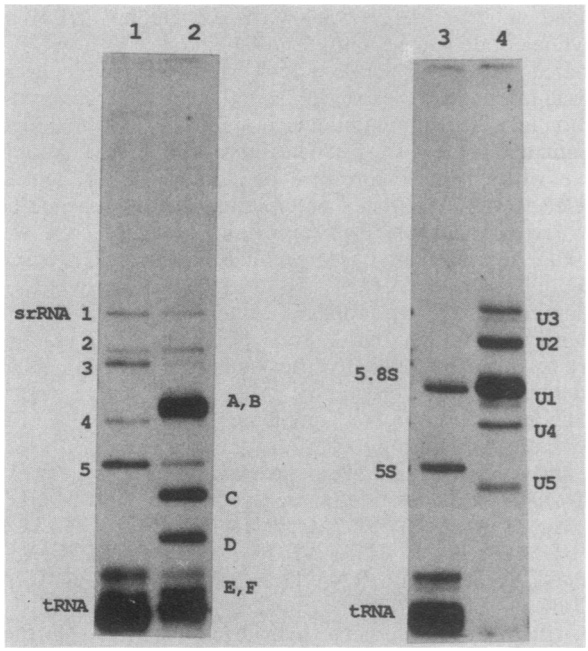

FIG. 1. Identification of candidate trypanosome U RNAs by immunoprecipitation. Trypanosome or mouse NSO myeloma cell total RNA was incubated with anti-TMG serum, and immune complexes were collected with protein A–Sepharose. Total and precipitated RNAs were 3′ ^{32}P end labeled and electrophoresed on an 8% polyacrylamide–7 M urea gel. Lanes: 1, total *T. brucei* RNA; 2, immunoprecipitated *T. brucei* RNA; 3, total mouse myeloma RNA; 4, immunoprecipitated mouse myeloma RNA.

ing gel calibrated with 5S and 5.8S trypanosome rRNAs and *Hpa*II-digested pBR322 DNA size standards. We believe that the positioning of the 3′ ends in this manner is accurate to within a few nucleotides.

RESULTS

Immunoprecipitation of trypanosome RNAs by anti-TMG antiserum. Candidate U RNAs were isolated by immunoprecipitation from total RNA by using polyclonal antiserum directed against the characteristic U snRNA TMG cap (Fig. 1). The autoradiographic pattern of end-labeled total trypanosome RNA before immunoprecipitation (lane 1) was identical to the ethidium bromide staining pattern and revealed a number of putative tRNA species of 70 to 90 nt and the five abundant small ribosomal RNAs unique to kinetoplastid protozoa. Immunoprecipitation with anti-TMG antiserum (lane 2) identified six candidate trypanosome U RNAs (arbitrarily labeled A through F) of approximately 150 nt (A and B), 110 nt (C), 95 nt (D), 85 nt (E), and 75 nt (F). As a positive control for antibody specificity, we also examined the pattern of RNAs immunoprecipitated by this antiserum from mouse myeloma cell total RNA (lane 4). The precipitated RNAs had the sizes expected for the most abundant TMG-capped metazoan U snRNAs: U1, U2, U4, and U5 RNAs as well as the nucleolar U3 RNA. All candidate trypanosome U RNAs differed in size from their metazoan analogs.

Although this work focused on RNAs A through D, the region of RNAs E through F in lane 2 contained at least three RNA species that were not resolved in this gel (data not shown). Furthermore, no RNAs of these exact sizes were observed in end-labeled total RNA. Thus, immunoprecipitation with anti-TMG cap antibody highly enriched for at least seven trypanosome RNA species that are relatively rare in total RNA.

Partial sequencing of candidate U RNAs and isolation of the cognate genes. Total immunoprecipitated RNA was 5′ or 3′ end labeled and fractionated on a preparative sequencing gel. RNAs A through D were eluted from gel slices and subjected to enzymatic RNA sequencing analysis (Fig. 2). Although it was not possible to determine the complete sequence of any of these RNAs because of sequencing ambiguities caused by nucleotide modifications and RNA secondary structure, this analysis did allow us to identify regions of relatively unambiguous sequence of sufficient length to permit the synthesis of complementary oligonucleotide probes (see Materials and Methods). To verify the specificity of these probes, each oligonucleotide was 5′ ^{32}P end labeled and hybridized with a single lane from a Northern blot of total trypanosome RNA. Each probe hybridized specifically with a single RNA species of the expected size (data not shown).

The genes encoding RNAs A through D were identified by genomic Southern analysis using the oligonucleotide probes (Fig. 3). Oligonucleotides A-1, B-3, and D-1 hybridized with a single DNA restriction fragment of variable size in each of five enzymatic digests, which suggested that these RNAs are encoded by single-copy genes that are not closely linked. Although oligonucleotide C-2 hybridized with several restriction fragments in each digest, in every case one fragment hybridized more strongly than did the others. Subsequent Southern analysis using the cloned RNA C gene as a probe detected only the strongly hybridizing fragment in each digest, which indicated that RNA C is also encoded by a single-copy gene (data not shown).

λ1059 bacteriophage clones containing the RNA A, B, or D gene were isolated from a *T. brucei* genomic library by using the oligonucleotide probes. The RNA C gene was isolated from a pUC19 library constructed with the 5- to 6-kilobase fraction of *Hin*dIII-digested genomic DNA by using the RNA C-2 probe. Small restriction fragments (300 to 400 base pairs) containing each candidate U RNA gene were subcloned into M13 vectors, and each was sequenced on both strands (see below).

To map the U RNA coding regions within the sequenced DNA subclones, we analyzed the 5′ ends of RNAs B through D by primer extension dideoxy sequencing, using the respective oligonucleotides to prime synthesis from total RNA. The RNA A 5′ terminus was not mapped since it has been determined for the analogous *Trypanosoma brucei* subsp. *gambiense* RNA (56). Primer extension reactions were analyzed in parallel with sequencing reactions obtained with the same primer and the M13 subclone containing the respective noncoding strand. Both the RNA B and C reverse transcript sequences were colinear with the corresponding DNA sequences and showed strong stops at the last two positions of a CTT triplet (Fig. 4). Therefore, both the RNA B and C 5′ ends begin with the same trinucleotide, AAG, and are positioned within the DNA sequence as indicated in Fig. 5. The RNA D reverse transcript sequence showed strong stops at a location corresponding to a GG dinucleotide in the noncoding-strand sequence. Although the sequences derived from both the RNA and the DNA templates contained identical ambiguities just 3′ of these stops, the DNA subclone sequence immediately 5′ to these stops was clear and allowed us to position the 5′ end of RNA D as shown in Fig. 5.

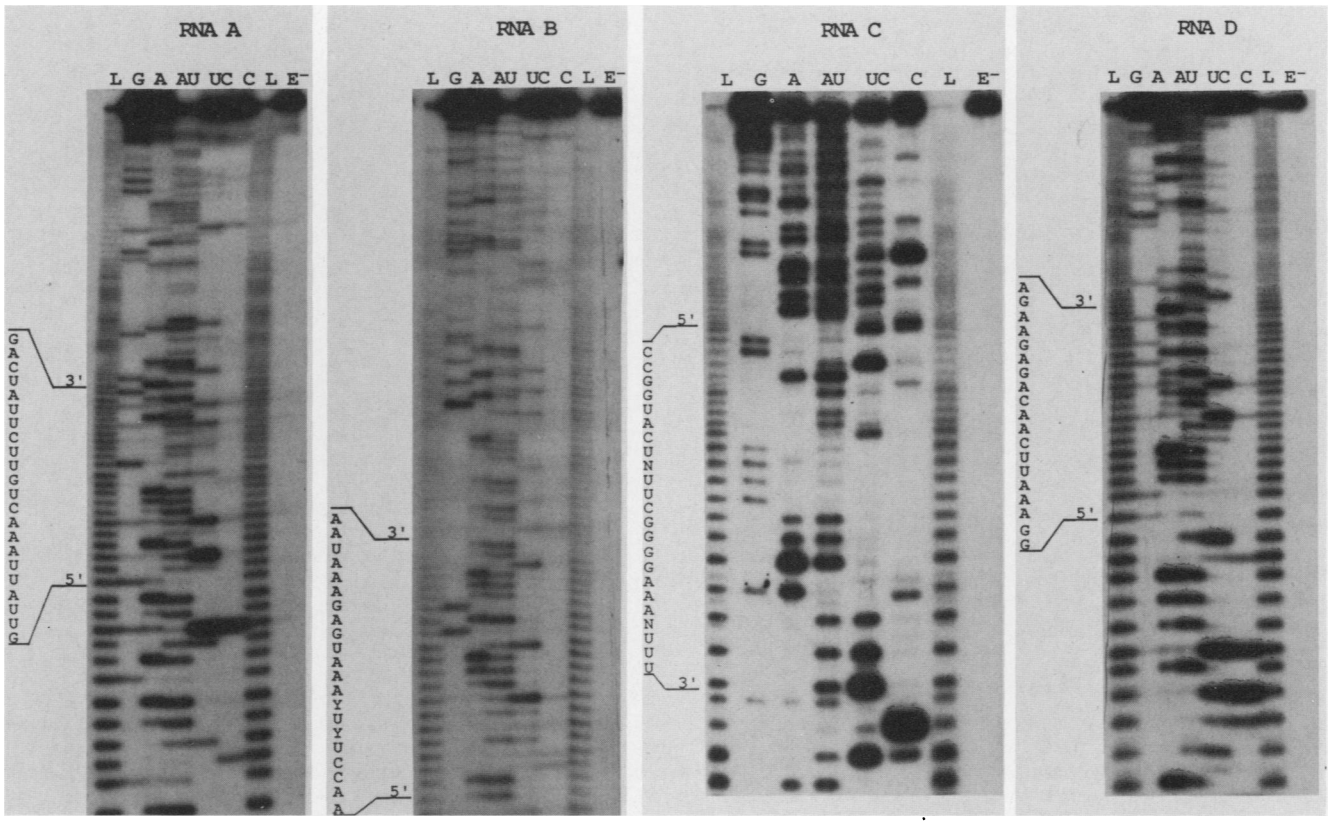

FIG. 2. Partial enzymatic sequence determination of RNAs A through D. Trypanosome RNAs immunoprecipitated by anti-TMG serum were ^{32}P end labeled, separated on a sequencing gel, individually eluted, and subjected to partial digestion with base-specific RNases. RNase digests were resolved by electrophoresis on 8% (RNA B) or 16% (RNAs A, D, and C) polyacrylamide–7 M urea sequencing gels. Shown are autoradiographs of sequencing ladders generated with 5′-^{32}P-end-labeled RNAs A, B, and D and 3′-^{32}P-end-labeled RNA C. Lanes: L, alkaline partial hydrolysis ladder; G, guanine-specific RNase T1 ladder; A, adenine-specific RNase U2 ladder; AU, adenine-uridine-specific RNase Phy M ladder; UC, uridine-cytosine-specific RNase *B. cereus* ladder; C, cytosine-specific RNase CL3 ladder; E^-, no enzyme control. RNA sequences indicated adjacent to each panel are those for which complementary oligonucleotides were synthesized as hybridization probes (see Materials and Methods).

The RNA A, C, and D genes encode the trypanosome U2, U4, and U6 RNA analogs. Comparison of the trypanosome RNA A through D gene sequences with U snRNA gene sequences present in the structural RNA data base of the GenBank Genetic Sequence Data Bank revealed that the RNA A, C, and D coding regions had considerable sequence homology with metazoan U2, U4, and U6 snRNAs, respectively (>50% identity). The sequence of the *T. brucei* subsp. *brucei* RNA A gene (data not shown) was almost identical to the sequence of a small RNA gene isolated from *T. brucei* subsp. *gambiense*, also identified as encoding U2 (56). The sequences differed only by a single C-to-T transition within the coding region at nt 134 and a C-to-G transversion at nt +89 in the 3′-flanking region. The computer-optimized alignments (using the program Malign) of the RNA C and D coding sequences with the corresponding *Drosophila* U4 and U6 RNA genes (12, 46) are shown in Fig. 5B and C, respectively. Search of the data base with the RNA B coding sequence revealed limited sequence homologies with several metazoan U RNA genes; the most statistically significant alignment was obtained with rat U3b RNA (54). Although we do not believe that these data are sufficiently definitive to identify RNA B as the trypanosome U3 analog (see below), we also present this sequence alignment (Fig. 5A).

The sequence of trypanosome U4 RNA (111 nt; Fig. 5B) was the least evolutionarily conserved of the three identified U RNAs. It was 57% identical to the 143-nt *Drosophila* U4 RNA after the introduction of 11 and 5 single-nucleotide gaps, respectively, in the two sequences. There were three blocks of sequence identity between the RNAs: 9 of 10 nt from positions 5 through 15, 8 of 9 nt from positions 33 through 41, and 5 of 5 nt from positions 49 through 53. Each of these conserved regions is hypothesized to be involved in intermolecular base pairing with U6 RNA. Trypanosome U4 was 33 nt smaller than *Drosophila* U4 and is the smallest U4 RNA yet identified. The 5′ ends of the two RNAs aligned reasonably well, and almost the entire size difference is attributed to a 30-nt section missing from the 3′ end of the protozoan RNA (Fig. 5B). This deletion began at a position corresponding to the Sm antigen-binding site (RAU_nGR) in *Drosophila* U4, a sequence that was absent from the trypanosome analog. Interestingly, sequence divergence between the trypanosome and *Drosophila* U2 RNAs also began precisely at the position of the *Drosophila* U2 Sm-binding site; trypanosome U2 RNA did not contain this sequence (data not shown).

Trypanosome U6 RNA (99 nt) was 69% identical to its *Drosophila* analog (106 nt) after allowance for eight single-nucleotide gaps in the trypanosome sequence and one in the *Drosophila* sequence (Fig. 5C). There were three blocks of perfect identity between the RNAs: 6 nt from positions 7 through 12, 16 nt from positions 37 through 52, and 12 nt

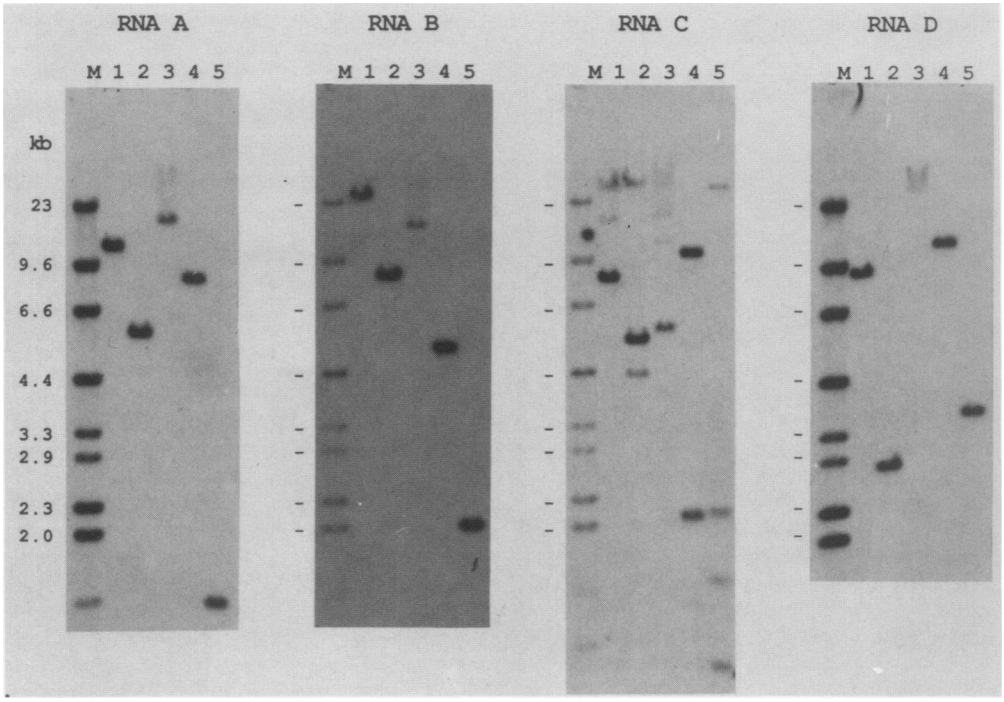

FIG. 3. Identification of genes encoding RNAs A through D. Trypanosome genomic DNA (5 μg) was digested with five different restriction endonucleases, and the digests were electrophoresed on a 0.8% agarose gel, transferred to a Nytran membrane, and hybridized with 5′-^{32}P-end-labeled oligonucleotide probes A-1, B-3, C-2, and D-1 in panels RNA A through D, respectively. Lanes: M, ^{32}P-labeled DNA size markers (lambda DNA digested with *Hin*dIII and pBR322 DNA digested with *Eco*RI-*Ava*I and *Bam*HI-*Ava*I); 1 through 5, trypanosome DNA digested with *Eco*RI, *Hin*dIII, *Bam*HI, *Pst*I, and *Hin*cII, respectively.

from positions 57 through 68. The extensive sequence homology (~90% identity) in the central region (nt 37 through 68) of these molecules is also shared with all other U6 RNAs that have been sequenced and has been hypothesized to reflect the presence of a site of U4-U6 intermolecular base pairing.

The trypanosome RNA B gene encodes an unidentified U RNA. The 144-nt RNA B was not sufficiently similar to any of the major U RNAs whose trypanosome counterparts have not been identified (U1, U3, U5, U7, and U8 RNAs) or to four sequenced nonessential yeast U RNAs (snR3, snR5, snR8, and snR9; 36) to allow assignment of its identity. Although the RNA B sequence could be aligned at 63% of its positions with identical residues in the 215-nt rat U3b sequence (Fig. 5A), most of these residues occurred in blocks of three or fewer contiguous nucleotides which were often separated by large gaps. Most relevant to the proposition that RNA B may be the trypanosome U3 equivalent are the results of sequence and secondary-structure comparisons of the U3 RNAs from a number of eucaryotes (17, 34). These phylogenetic comparisons have shown that although the generic U3 sequence is poorly conserved in its entirety, four blocks of primary sequence (boxed in Fig. 5A), and the overall secondary structure, are much more strongly conserved. The RNA B sequence contained a good match to only one of the four regions of conserved sequence (7 of 9 nt in box C), and it could not be folded into the consensus U3 secondary structure. Therefore, if RNA B is the trypanosome U3 analog, it has diverged much further from its metazoan and yeast counterparts than have the other trypanosome U RNAs identified in this study.

As described for U3 RNA, comparisons of the metazoan and yeast U1-snR19 (22, 51) and U5-snR7 (38) RNAs have shown that although the generic U1 and U5 sequences are only poorly conserved, certain domains of strict primary-sequence and secondary-structure conservation are present. RNA B contained none of the conserved primary- or secondary-structure motifs revealed in these comparisons, nor did it contain a consensus Sm antigen-binding-site sequence. However, at least as judged by the criterion of possessing a TMG cap, RNA B is a legitimate U RNA (see below).

Trypanosome U6 RNA does not contain a TMG cap. Among the U snRNAs analyzed to date, U6 has a unique set of properties: it does not contain the characteristic TMG cap (7), it is transcribed by RNA polymerase III (23), it is the only U RNA known to be essential for pre-mRNA processing that lacks an Sm antigen-binding site (4), and it is found in vivo both as a free snRNP and as an intermolecularly base paired component of the U4-U6 snRNP (5, 8, 16, 50). It is because of its association with U4 RNA, which contains both an Sm antigen-binding site and a TMG cap, that metazoan and yeast U6 RNAs can be immunoprecipitated from nuclear extracts by anti-Sm and anti-TMG sera and from deproteinized RNA by anti-TMG serum.

To determine whether trypanosome U6 RNA also lacks a TMG cap and to verify that the trypanosome U4 and B RNAs contain this cap, we examined the anti-TMG immunoprecipitability of the individual RNAs isolated from total RNA by hybrid selection. Trypanosome U2 RNA has been shown to contain a TMG cap (56) and was not included in this analysis. Figure 6 shows each isolated RNA fraction before incubation with antibody (initial) as well as the RNAs immunoprecipitated by the anti-TMG serum (pellet) and those remaining in the supernatant. Both the hybrid-selected B and U4 RNAs were found predominantly in the pellet after immunoprecipitation, whereas U6 RNA was found exclu-

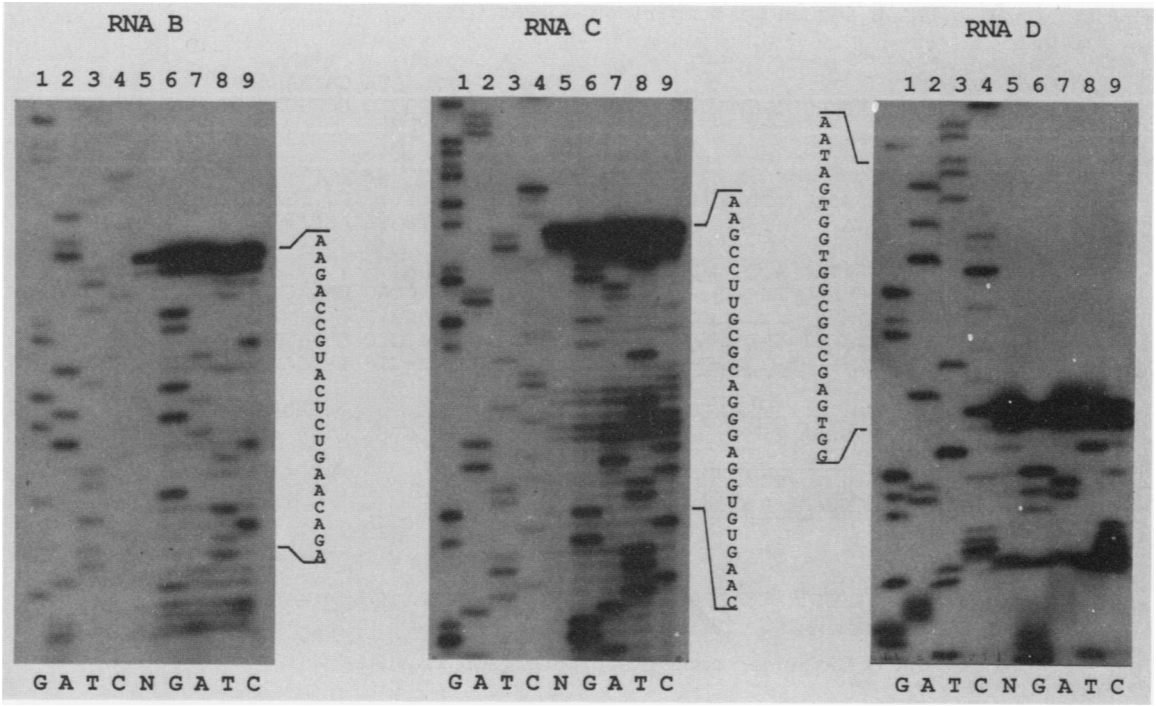

FIG. 4. Primer extension mapping of the 5′ ends of RNAs B, C, and D. Trypanosome total RNA (25 μg) was annealed with 5′-^{32}P-end-labeled oligonucleotides B-3, C-2, and D-1 (panels RNA B through D, respectively), divided into five equal portions, and extended by avian myeloblastosis virus reverse transcriptase in the absence (lane 5) or presence (lanes 6 through 9) of dideoxynucleotide GTP, ATP, TTP, and CTP. The same primers were annealed with M13 subclones containing the respective noncoding DNA strand and extended by Klenow polymerase in the presence of the same dideoxynucleotide triphosphates (lanes 1 through 4). The RNA sequence predicted by the cDNA sequence is shown on the right of panels RNA B and C, and the predicted noncoding-strand DNA sequence extending upstream from position −2 is shown to the left of panel RNA D.

sively in the supernatant. The trypanosome B and U4 RNAs must, therefore, contain a TMG cap. Conversely, these data and those of Fig. 1 (lane 2) indicate that although U6 RNA can be precipitated from total trypanosome RNA by anti-TMG antiserum, the isolated RNA itself is not precipitable. Therefore, like all other U6 RNAs examined, the trypanosome analog is not capped with TMG and hence must associate, presumably through intermolecular base pairing, with another U RNA which is. In other organisms this RNA is, of course, U4. The extensive sequence complementarity between the trypanosome U4 and U6 RNAs (see Discussion) suggests that this is also the case in trypanosomes.

DISCUSSION

The experiments described here were undertaken to identify molecular components likely to be involved in the formation and functioning of the *trans*-splicing apparatus in trypanosomes. Since *trans* and *cis* splicing appear to be mechanistically related, it is probable that U RNAs essential for *cis* splicing might also be used in *trans* splicing. We have therefore identified U RNA species in *T. brucei* by immunoprecipitation with antiserum specific for the U RNA TMG cap. The six RNAs identified in these experiments represent the minimum number of immunoprecipitable RNAs, since longer autoradiographic exposures often reveal additional species. However, given the limitations inherent in estimating mass and stoichiometry by 3′ end labeling with RNA ligase, these appear to be the most abundant TMG-capped trypanosome RNAs.

Cloning and sequencing of the single-copy genes encoding the four largest immunoprecipitated RNAs (A through D) definitively identified RNAs A and D as the trypanosome U2 and U6 analogs, respectively. The RNA C sequence is less homologous to its proposed metazoan U4 RNA counterpart; however, the fact that the sequence is extensively complementary to the trypanosome U6 RNA sequence strongly supports this assignment. The RNA B sequence is not sufficiently similar to any of the other known U RNA sequences to permit identification of a metazoan or yeast analog. Whether RNA B represents a very divergent example of a major U RNA species, a U RNA whose analog has not been identified in other organisms, or a U RNA specific to trypanosomes remains to be determined.

Trypanosome U RNAs do not contain a consensus Sm antigen-binding site. In mammals and yeasts, the subset of U snRNPs which are essential for pre-mRNA splicing contain proteins carrying the Sm antigenic determinant (30, 44). Each of these Sm RNPs is associated with a U RNA containing the Sm antigen-binding sequence $RAU_{3-6}GR$. This sequence typically occurs in a single-stranded region of U RNA secondary structure, and its location is phylogenetically conserved within each of the Sm U RNAs (3, 4). None of the trypanosome U RNAs we have sequenced contain the consensus Sm-binding sequence, nor do they contain a shared sequence motif that might have the same function as the Sm-binding site has in other eucaryotes. These observations are consistent with preliminary experiments suggesting that *T. brucei* U snRNPs are not precipitable from whole-cell extracts with a number of anti-Sm (and anti-RNP) sera (R. G. Nelson, unpublished observations; S. Michaeli, personal communication). Nevertheless, it is clear that each of the trypanosome U RNAs (and the SL RNA) is associated

A. RNA B sequence; comparison to Rat U3b.

```
- 73                                CTG AGGTTGAATA CATAAGAAAC
- 50 TGTTTATTAC CGATATTAGA CTATCGCGAT AATTTTGGAC GCCGGATGTT


  1 AAGACCGTAC TCTGAACAGA .ATCGTTT.T AT..... .... GAGTAC RNA B
  1 -----TA--- -T-...---G G---A---C- --AGTTCGTT ACTA---A-G R U3b

 37 AACCTCTTAA A.TG.AGAAA TAACCAACAA CCA..AA... .........
 48 TTT----G-C TG--T---GC ..---CGA-- ---CG-GGAC GAGACATAGC

 70 ......TCCT GATGATGAAG GTTGCT...G ..GCTCCGCC CGGAGAGCTG
 96 GTCCCC---- --GCG----- CCG---CTA- GT---G-TT- T-TGC-----

109 ..TTC..... ..GATGAACG GCAGGAGCCG GTTCC..... ...AGAAGGA
146 CC-CTTGCCA TT-----T-- TTCTTCT-TC C---GGGGGG GTA---G---

142 .......... TCC.....
196 GGGAACGCAG --TGAGTGGA


+  1 TTTTTTTTGT TCCACCTTGG CTCGCTGCAA CGGGAACTGC CTGGCTCGCT
+ 51 TCTTTTGAGG CC
```

B. RNA C sequence; comparison to Drosophila U4.

```
-129                          AGCTACGAA GGCAGTTGCA GAAACCGACG
-100 AGGACGGCGT TTGTTTATCT TGTTTTGATT CCTAAAAATA AAAACGTTAT
- 50 TTTGTTAGAC TATAGTCCGT TTCTATGGGC ATTTTGGCTT TCCCCGGCGC


  1 AAGCCTT.GC GCAGGGAGGT GTG...AACG CAAGAT.... .CCTCAGGTG  RNA C
  1 -.-.---A-- ----T-GCAA TACCGT---C A-T--AGCCT C---G-----  D U4

 42 A..TTGTTCA CTAGTGCAAT ACTATATCCG GTACTCCTTC GGGGAAAGTT
 49 CGG--A--G. -----TG--A ---T--A--A ACC-A-GCCA T----C.--G

 90 TGCTACCCAC CACGGGTGGG A......... .......... ........
 97 AAA----GT- ---TAC.--C -ATTTTTGGA AGCCCTTACG AGGGCTAA


+  1 GGTCCTTTTT CGTCTGCGGA CAGATGCAGT TTCTTTTTTC TTACGGGCTG
+ 51 CCTTGTGACT GCTGCGGCCC TCAAAGGCGA GTGACGCATA TGGCCATTAA
+101 CTGGACAAAG TGTAAGAGCT
```

C. RNA D sequence; comparison to Drosophila U6

```
-101 GGCCACGCCG AAGCAGGGAT ATTTTTAGTG TTTTGTTTTT ATTTAGAAAT
- 51 GACTTTTTCC ATAAAATGTT AGAGAAGGAA CAATAGTGGT GGCGCCGAGT


  1 GGAGCC..CT TCGG..GGAC ATCCAC.AAA CTGGAAATTC AACAGAGAAG  RNA D
  1 .-TT-TTG-- ----CA-A-- --AT--T--- A-T-G--CGA T---------  D U6

 46 ATTAGCACTC TCCCTGCGCA AGGCTGATGT .CAA..TCTT CGAGAGATAT
 50 -------TGG C--------- ---A---CAC G---AA--G- GA--C-T-CC

 93 AGCTTTT.
100 -CA----T


+  1 CGGTTGATGC AGTCATTCGT CTTTAATAAT ATGTCTGATC TTTTTTCTTC
+ 51 GCCGAAGTAG AGTCGGAGTC GGAGTTGGGG GGGTTTCCGT CCACATTTGT
+101 CCCCATCGGC AGAGAGTTAC CTGCCGCCCT GTTCCGGCC
```

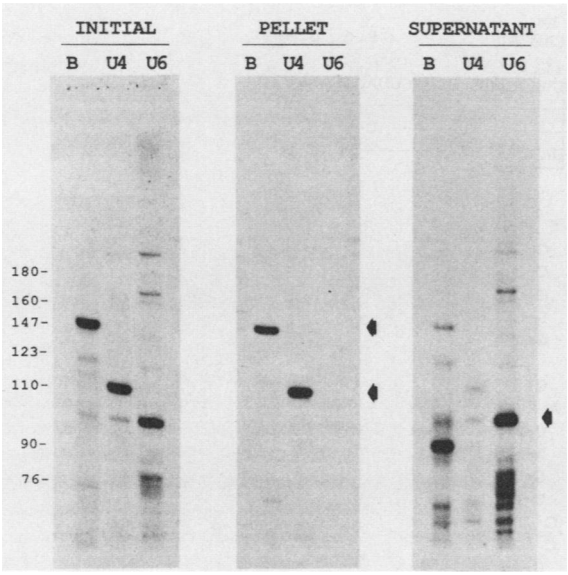

FIG. 6. Trypanosome U6 RNA does not contain a TMG cap. RNA B, U4 RNA, and U6 RNA were hybrid selected from total trypanosome RNA with the complementary M13 clones. RNAs were 3′ ^{32}P end labeled and incubated with anti-TMG serum, and immune complexes were collected with protein A–Sepharose. The panel labeled INITIAL shows each hybrid-selected RNA fraction before exposure to antiserum; the other two panels show RNAs present in the pellet and supernatant fractions after immunoprecipitation. Lanes: B, RNA B; U4, U4 RNA; U6, U6 RNA. Arrows indicate locations of the RNAs after immunoprecipitation; i.e., RNA B and U4 RNA are present in the pellet, and U6 RNA is present in the supernatant. The migration of denatured DNA size markers (pBR322 digested with *Hpa*II) is shown on the left. The origin and identity of the ca. 85-nt RNA in lane B of the supernatant fraction is unknown.

with several proteins in vivo and that these RNPs have properties characteristic of metazoan U snRNPs (S. Michaeli et al., manuscript in preparation). Further experiments are required to determine whether any of these trypanosome proteins are structurally or functionally analogous to metazoan and yeast proteins that contain the Sm epitope.

The putative pre-mRNA branchpoint recognition sequence of trypanosome U2 RNA is divergent. The sequences of the *T. brucei* subsp. *brucei* (RNA A; data not shown) and *T. brucei* subsp. *gambiense* (56) U2 RNA genes are almost identical; the coding sequences differ by a single C-to-T transition at position 134. As pointed out by Tschudi et al. (56), trypanosome U2 is extensively homologous with both metazoan and yeast U2 RNAs over the first 100 nt. This 5′ or U2 domain region is highly conserved among all known U2 RNAs and is thought to contain sequence or structural determinants that are required for the function of the U2 snRNP in pre-mRNA processing. A model of yeast U2 RNA–pre-mRNA interaction based on in vivo evidence suggests that base pairing between the two RNAs leads to recognition of the adenosine residue selected for 2′-5′ branching (35). Metazoan and yeast U2 RNA sequences are exactly conserved in the putative branchpoint-binding region; pre-mRNA branchpoint sequences are conserved exactly in *S. cerevisiae* (26) and less so metazoa (18, 19). In Fig. 7A we compare the sequences of the region of U2 RNA surrounding the putative branchpoint recognition sequence from trypanosomes, rats, *Drosophila melanogaster*, and *S. cerevisiae*. The RNA sequences are very similar with a single exception: the trypanosome sequence is totally divergent over the position corresponding to the putative branchpoint recognition sequence of the other U2 RNAs. An examination of sequences upstream of the acceptor splice site of many trypanosome structural-gene pre-mRNAs reveals no conserved, complementary sequence that could bind this region of U2 RNA in a way analogous to that proposed for the metazoan or yeast U2–pre-mRNA interaction (18, 19, 35). Therefore, if this region of U2 sequence is involved in selection or recognition of trypanosome pre-mRNA branchpoints, the latter sequences must be both degenerate and different from those used in metazoan and yeast pre-mRNAs.

U2-SL RNA sequence complementarity. An unusual feature of the trypanosome U2 RNA sequence is its extensive complementarity to the SL RNA. In Fig. 7B, we have arranged the 5′ ends of the U2 and SL RNAs to maximize potential intermolecular base pairing; portions of this model have also been suggested by Tschudi et al. in their discussion of the *T. brucei* subsp. *gambiense* U2 RNA sequence (56). The largest block of uninterrupted base pairing occurs between the central portion of the SL and a 10-nt region of the U2 RNA which overlaps (2 nt) the putative branchpoint recognition sequence (8 of these 10 nt are directly repeated downstream in the stem of the large intramolecular U2 hairpin). A smaller, 7-nt block of uninterrupted complementarity occurs between the SL intron sequence adjacent to the donor splice site and residues 6 through 12 near the 5′ end of the U2 RNA. The remainder of the intermolecular pairing shown in Fig. 7B is less compelling, since several loops and bulges must be introduced into both sequences to optimize the interaction.

Experimental data bearing on the validity of this particular model or on the existence of an intermolecular interaction between the U2 and SL RNAs-RNPs are lacking. Unlike the case with U6 RNA, which also does not contain a TMG cap, we have never observed coprecipitation of the SL RNA from total RNA with anti-TMG serum. This suggests that if the U2 and SL RNAs interact directly by base pairing in vivo, this pairing is either less extensive or more labile than that between the U4 and U6 RNAs. Given the short half-life of the SL RNA (24), any intermolecular pairing between the SL

FIG. 5. Nucleotide sequences of the genes encoding RNAs B through D; comparison with metazoan U RNA sequences. (A through C) Nucleotide sequences of the trypanosome RNA B through D genes, respectively. In the central region of each panel, the trypanosome DNA sequence corresponding to the encoded U RNA is aligned with the DNA sequence of its closest metazoan U RNA counterpart. Numbering begins at +1, which reflects the experimentally determined 5′ end of each trypanosome U RNA. (A) RNA B gene sequence compared with the rat U3b RNA gene sequence (54, 76). Boxed regions correspond to sequences that are phylogenetically conserved among U3 RNAs (17, 34); only one of these is also conserved in the RNA B sequence. (B) RNA C gene sequence compared with the *Drosophila* U4 RNA gene sequence (46). (C) RNA D gene sequence compared with the *Drosophila* U6 RNA gene sequence (12). Symbols: -, identity between sequences; ., gaps inserted to optimize alignment. A hexanucleotide sequence (TTTTGG) that is well conserved in location relative to the transcription start sites of RNA B, U2 RNA, U4 RNA, and the SL RNA is underlined in the 5′-flanking sequences of the RNA B and U4 RNA genes.

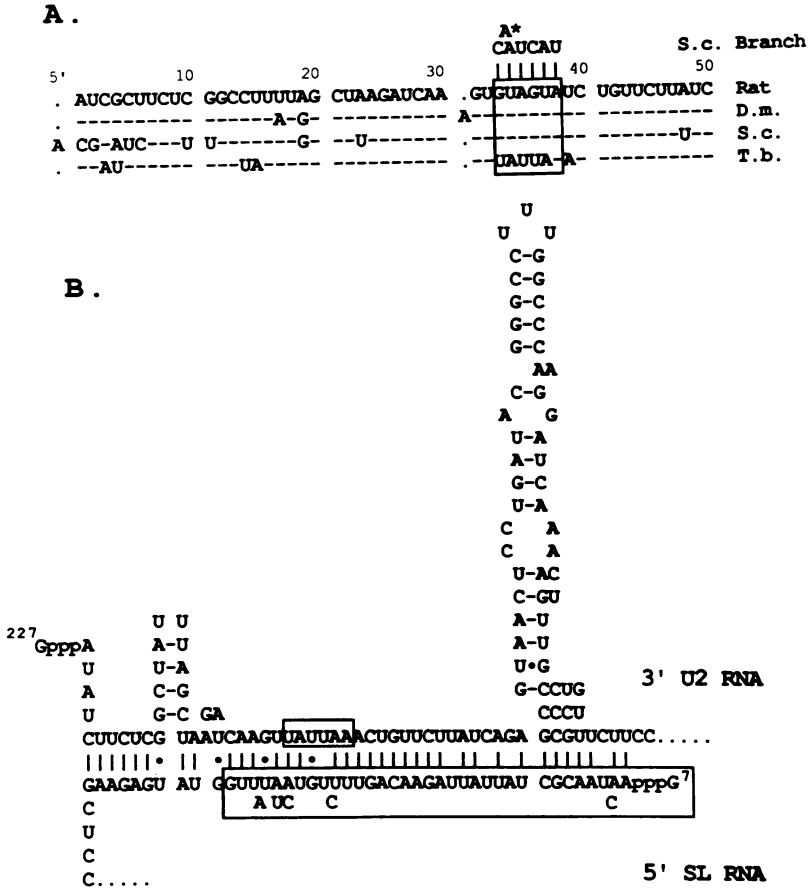

FIG. 7. (A) Comparison of U2 RNA sequences in the putative pre-mRNA branchpoint recognition region. The 5'-terminal 50 nt of the U2 RNA sequences from *S. cerevisiae* (S.c.), *D. melanogaster* (D.m.), and *T. brucei* (T.b.) are compared with the sequence of rat U2 RNA over the same region. Symbols: -, sequence identity with the rat U2 sequence; ., nucleotides not present in a particular sequence. Above the rat sequence, the absolutely conserved yeast pre-mRNA branchpoint sequence (5'-UACUAAC) is shown paired to the putative branchpoint recognition sequence of U2 RNA (pairing with rat U2 is shown, but the yeast sequence is identical in this region). At least for yeasts, there is very good evidence that these two molecules do, in fact, base pair over at least a portion of this region (35). This pairing is hypothesized to result in a bulged adenosine residue (A*) at the site of branch formation (35). Note that although these U2 domain sequences are very similar, the trypanosome sequence diverges from the others over the region thought to be involved in pre-mRNA branchpoint recognition. (B) Potential base pairing between the trypanosome U2 and SL RNAs. Hypothetical model arranged to maximize potential intermolecular base pairing between the 5' domain of U2 RNA and the 5' end of the SL RNA. The region of trypanosome U2 corresponding to the branchpoint recognition sequence of yeast U2 is enclosed in a small box. The SL sequence, enclosed in a large box, is immediately followed by the donor splice site. Vertical lines indicate A · T and G · C pairing; central dots indicate G · U pairing. This model is an extension of one proposed by Tschudi et al. for *T. brucei* subsp. *gambiense* U2 RNA (56).

and U2 RNAs must be transient and perhaps occurs only during the *trans*-splicing reaction.

U4-U6 RNA sequence complementarity. A number of investigators have shown that U4 and U6 RNAs are complexed within a single snRNP in vivo and have suggested that the two RNAs are complexed within the particle by intermolecular hydrogen bonding (5, 8, 12, 16, 50). Consistent with these results, extensive intermolecular complementarity exists in every case in which both the U4 and U6 RNAs have been sequenced (8, 12, 41, 46, 50). This is also true for the trypanosome U4 and U6 sequences. In Fig. 8, the trypanosome RNAs have been arranged in a hypothetical secondary structure based on the model proposed by Brow and Guthrie for the interaction of the analogous yeast RNAs (8). In fitting the trypanosome sequences to this model, the central three-stemmed core and the U6 5'-terminal hairpin have been retained, but we have changed a few areas of U4 and U6 intramolecular structure to better fit the protozoan data. For example, although neither the U4 nor U6 3'-terminal hairpin shown in Fig. 8 can be formed in yeasts, both can be formed in metazoa as well as trypanosomes.

The proposed model contains two main regions of U4-U6 base pairing (indicated in Fig. 8 by I and II) separated by a U4 hairpin loop. Stem I, which occurs between residues 2 and 17 of U4 RNA and residues 55 through 70 of U6 RNA, makes up the longest uninterrupted block of potential intermolecular base pairing. Within this 16-nt stem, trypanosome and yeast U6 RNAs differ in sequence at five positions and the trypanosome and mammalian (or *Drosophila*) U6 sequences differ at three. Each of the differences between the trypanosome and yeast U6 RNAs, and two of the three differences between the trypanosome and mammalian (or *Drosophila*) U6 RNAs, is accompanied by compensatory changes in the respective U4 sequences which maintain base pairing in the stem (see legend to Fig. 8 for locations). Therefore, the available phylogenetic data strongly support the existence of stem I.

The second region of proposed U4-U6 base pairing (stem

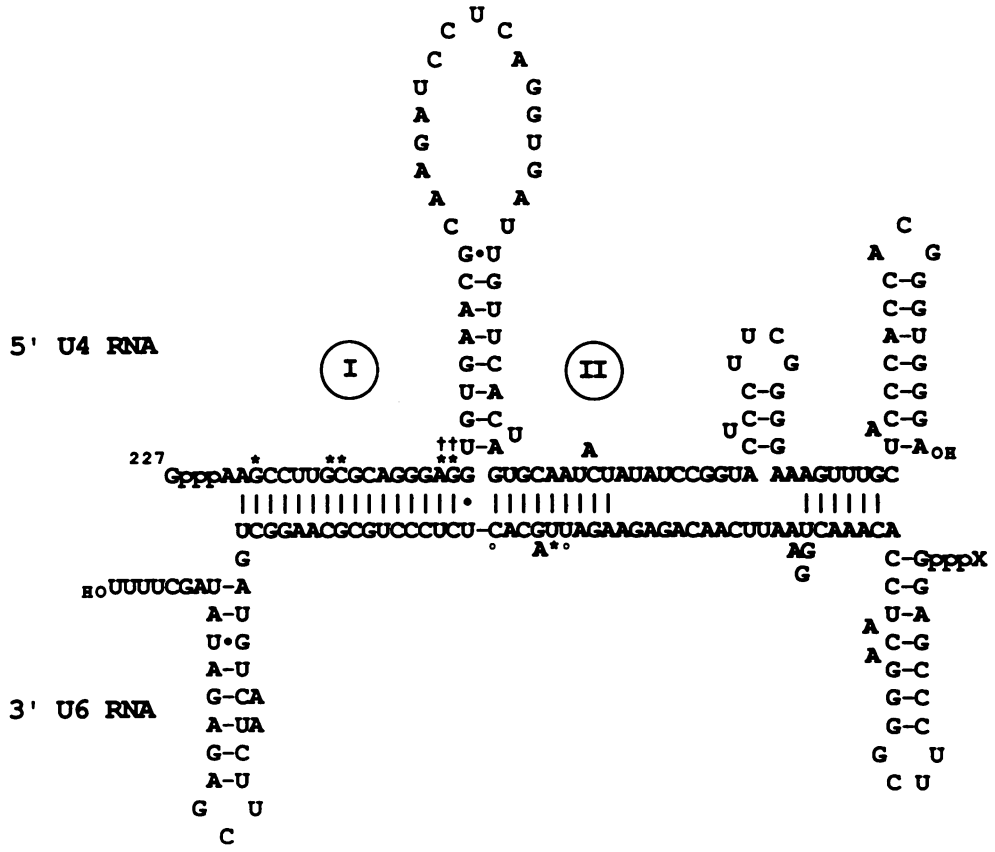

FIG. 8. Hypothetical secondary structure of the U4-U6 RNA complex. The trypanosome U4 and U6 RNA sequences are fit to the model proposed by Brow and Guthrie (8) for the secondary structure of the yeast U4-U6 complex. Most of the intermolecular base pairing shown is similar to that proposed for the yeast RNAs; however, several features of U4 and U6 intramolecular structure have been altered to better fit the protozoan data (see text). I and II, above the intermolecularly hydrogen bonded region of the U4 sequence, refer to the two stems that make up the interaction. Symbols: *, compensatory base changes between yeast and trypanosome U4; +, compensatory changes between mammal-fly and trypanosome U4; °, substitutions or alterations that increase the comparative length or stability of stem II.

II) is supported by photochemical cross-linking data obtained with intact snRNPs in HeLa cell extracts (45). This study suggested that the 8-nt U6 sequence 3'-ACGAUUAG-5' is base paired to a complementary U4 sequence, 5'-UGCUAGUU-3', within the intact HeLa U4-U6 snRNP. The U6 octanucleotide sequence is exactly conserved in trypanosome U6, and although the trypanosome U4 octanucleotide (5'-UGCAAUAC3') differs from its HeLa counterpart, the two sequences can base pair at six of eight positions by bulging the underlined U6 adenosine residue.

Two additional features of the model deserve comment. First, the trypanosome data provide strong phylogenetic support for the existence of the U4 hairpin loop separating the two base-paired U4-U6 stems. As for yeast and mammals, this structure is composed of an 8-base-pair stem and a 17-nt loop. Comparison of the yeast and trypanosome U4 sequences reveals that six of the eight nucleotide pairs in this stem reflect compensatory base changes in the trypanosome U4 sequence. Second, as is the case for the mammalian, *Drosophila*, and yeast U6 RNAs, trypanosome U6 RNA also has the potential to form a 5'-terminal hairpin loop. This structural feature is conserved between trypanosomes and metazoa and consists of a 7-base-pair stem with 5'-CG closure preceding a conserved 5'-UUCG loop. Such CU-UCGG hairpins have been shown to form unusually stable secondary structures in which the loop sequence makes a significant contribution to the overall free energy of the structure (57). Although the larger yeast U6 hairpin loop contains two UUCG sequences, neither is accompanied by CG closure and both reside within the stem of the structure (8). Perhaps what is important about this region of U6 is not the loop sequence itself but the overall stability of the hairpin, which may be provided in *S. cerevisiae* by the presence of a longer stem.

Implications for trypanosome *trans* splicing. Our results indicate that trypanosomes contain at least three (U2, U4, and U6) of the five U RNAs that are known to be required for pre-mRNA processing in mammalian and yeast cells. Do individual trypanosome U RNAs-RNPs play the same role in *trans* splicing as their respective yeast and metazoan analogs play in *cis* splicing? A definitive answer to this question will require development of an in vitro *trans*-splicing system; however, some differences between the two systems might be expected. In trypanosomes, for example, every 5' exon, i.e., the SL, is identical and the small SL RNA containing this sequence and donor splice site is itself complexed with proteins to form a particle with snRNP-like properties (Michaeli et al., manuscript in preparation). Therefore, it is possible that trypanosomes do not require a separate U1 snRNP-based mechanism for the recognition and discrimination of 5' splice sites; the SL RNP may provide both functions. With respect to this hypothesis, it is of some note that no U1 RNA analog was detected among the four largest trypanosome U RNAs examined here, nor was a U1 candi-

date identified by Northern analysis using oligonucleotide probes complementary to phylogenetically conserved regions of U1 RNA sequence (K. P. Watkins, personal communication).

Second, unlike other U2 RNAs, portions of the trypanosome U2 sequence are complementary to the donor splice site and SL sequence, which may be thought of in this context as the pre-mRNA 5′ exon. In other organisms it is U1, not U2, RNA that interacts with the 5′ splice site by base pairing (60). If trypanosome U2 RNA is able to simultaneously base pair with both the SL RNA and the pre-mRNA 3′ branchpoint sequence, it might function to juxtapose the two pre-mRNAs in an appropriate geometry for initiation of the *trans*-splicing reaction. After cleavage at the 5′ splice site, U2 RNA bound at the branchpoint might also serve to retain the SL sequence within the splicing complex in preparation for ligation to the 3′ exon.

Finally, it appears possible that the function or mechanistic role of the trypanosome U4-U6 snRNP might differ, perhaps subtly, from that of its metazoan and yeast counterparts. The protozoan complex appears considerably more stable than the analogous yeast and mammalian structures; stem I of the trypanosome complex contains the largest number of G · C pairs and the longest potential run of uninterrupted and perfect base pairing among known U4-U6 RNAs. Since U4 RNA dissociates from the U4-U6 complex, and from the spliceosome, before the first pre-mRNA cleavage reaction of *cis* splicing (11, 25), the greater (apparent) thermodynamic stability of the trypanosome complex might indicate an altered role of this snRNP in *trans* splicing.

ACKNOWLEDGMENTS

We are especially grateful to David Brow and Christine Guthrie for sharing unpublished data and for their generous help in fitting the trypanosome U4 and U6 RNA sequences to their secondary-structure model. We also thank Phil Barr of Chiron Corp. for synthesis of the oligonucleotides used in this work and Shulamit Michaeli and Kenneth P. Watkins for sharing unpublished data, computer assistance, critical discussion, and encouragement throughout this work.

This work was supported by Office of Naval Research contract N00014-81-C-0570 and in part by Public Health Service grant AI-21975 from the National Institutes of Health to N.A. and the John D. and Catherine T. MacArthur Foundation Consortium in Parasite Biology. K.L.P. is a National Institutes of Health postdoctoral fellow, and N.A. is a Burroughs Wellcome Scholar of Molecular Parasitology.

LITERATURE CITED

1. **Ares, M., Jr.** 1986. U2 RNA from yeast is unexpectedly large and contains homology to vertebrate U4, U5, and U6 small nuclear RNAs. Cell **47**:49–59.
2. **Billings, P. B., and S. O. Hoch.** 1984. Characterization of U small nuclear RNA-associated proteins. J. Biol. Chem. **259**: 12850–12856.
3. **Birnstiel, M. L. (ed.).** 1988. Small nuclear ribonucleoprotein particles. Springer-Verlag KG, Heidelberg.
4. **Branlant, C., A. Krol, J.-P. Ebel, E. Lazar, B. Haendler, and M. Jacob.** 1982. U2 RNA shares a structural domain with U1, U4, and U5 RNAs. EMBO J. **1**:1259–1265.
5. **Bringmann, P., B. Appel, J. Rinke, R. Reuter, H. Theissen, and R. Lührmann.** 1984. Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. EMBO J. **3**:1357–1363.
6. **Bringmann, P., and R. Lührmann.** 1986. Purification of the individual snRNPs U1, U2, U5, and U4/U6 from HeLa cells and characterization of their protein constituents. EMBO J. **13**: 3509–3516.
7. **Bringmann, P., J. Rinke, B. Appel, R. Reuter, and R. Lührmann.** 1983. Purification of snRNPs U1, U2, U4, U5, and U6 with 2,2,7-trimethylguanosine-specific antibody and definition of their constituent proteins reacting with anti-Sm and anti-(U1)RNP antisera. EMBO J. **2**:1129–1135.
8. **Brow, D. A., and C. Guthrie.** 1988. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. Nature (London) **334**:213–218.
9. **Busch, H., R. Reddy, L. Rothblum, and Y. C. Choi.** 1982. SnRNAs, snRNPs, and RNA processing. Annu. Rev. Biochem. **51**:617–654.
10. **Campbell, D. A., D. A. Thornton, and J. C. Boothroyd.** 1984. Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature (London) **311**:350–355.
11. **Cheng, S.-C., and J. Abelson.** 1987. Spliceosome assembly in yeast. Genes Dev. **1**:1014–1027.
12. **Das, G., D. Henning, and R. Reddy.** 1987. Structure, organization, and transcription of *Drosophila* U6 small nuclear RNA genes. J. Biol. Chem. **262**:1187–1193.
13. **De Lange, T., A. Y. C. Liu, L. H. T. Van der Ploeg, P. Borst, M. C. Tromp, and J. H. Van Boom.** 1983. Tandem repetition of the 5′ mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? Cell **34**:891–900.
14. **De Lange, T., P. A. M. Michels, H. J. G. Veerman, A. W. C. A. Cornelissen, and P. Borst.** 1984. Many trypanosome mRNAs share a common 5′ terminal sequence. Nucleic Acids Res. **12**:3777–3790.
15. **England, T. E., A. G. Bruce, and O. C. Uhlenbeck.** 1980. Specific labeling of 3′ termini of RNA with T4 RNA ligase. Methods Enzymol. **65**:65–74.
16. **Hashimoto, C., and J. A. Steitz.** 1984. U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. Nucleic Acids Res. **12**:3283–3293.
17. **Hughes, J. M. X., D. A. M. Konings, and G. Cesareni.** 1987. The yeast homologue of U3 snRNA. Embo J. **6**:2145–2155.
18. **Keller, E. B., and W. A. Noon.** 1984. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. Proc. Natl. Acad. Sci. USA **81**:7417–7420.
19. **Keller, E. B., and W. A. Noon.** 1985. Intron splicing: a conserved internal signal in introns of *Drosophila* pre-mRNAs. Nucleic Acids Res. **13**:4971–4981.
20. **Kooter, J. M., T. De Lange, and P. Borst.** 1984. Discontinuous synthesis of mRNA in trypanosomes. EMBO J. **3**:2387–2392.
21. **Kramer, A.** 1987. Fractionation of HeLa cell nuclear extracts reveals minor small nuclear ribonucloprotein particles. Proc. Natl. Acad. Sci. USA **84**:8408–8412.
22. **Kretzner, L., B. C. Rymond, and M. Rosbash.** 1987. *S. cerevisiae* U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA. Cell **50**:593–602.
23. **Kunkel, G. R., R. L. Maser, J. P. Calvet, and T. Pederson.** 1986. U6 small nuclear RNA is transcribed by RNA polymerase III. Proc. Natl. Acad. Sci. USA **83**:8575–8579.
24. **Laird, P. W., J. C. B. M. Zomerdijk, D. de Korte, and P. Borst.** 1987. *In vivo* labelling of intermediates in the discontinuous synthesis of mRNA in *Trypanosoma brucei*. EMBO J. **6**: 1055–1062.
25. **Lamond, A. I., M. M. Konarska, P. J. Grabowski, and P. A. Sharp.** 1988. Spliceosome assembly involves binding and release of U4 small nuclear ribonucleoprotein. Proc. Natl. Acad. Sci. USA **85**:411–415.
26. **Langford, C. J., and D. Gallwitz.** 1983. Evidence for intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell **33**:519–527.
27. **Lizardi, P.** 1983. Methods for the preparation of messenger RNA. Methods Enzymol. **96**:24–38.
28. **Lührmann, R., B. Appel, P. Bringmann, J. Rinke, R. Reuter, and S. Rothe.** 1982. Isolation and characterization of rabbit anti-2,2,7-trimethylguanosine antibodies. Nucleic Acids Res. **10**:7103–7113.
29. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Maniatis, T., and R. Reed.** 1987. The role of small nuclear

ribonucleoprotein particles in pre-mRNA splicing. Nature (London) **325**:673–678.
31. **Milhausen, M., R. G. Nelson, S. Sather, M. Selkirk, and N. Agabian.** 1984. Identification of a small RNA containing the trypanosome spliced leader: a donor of the shared 5′ sequences of trypanosomatid mRNAs? Cell **38**:721–729.
32. **Murphy, W. J., K. P. Watkins, and N. Agabian.** 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. Cell **47**:517–525.
33. **Nelson, R. G., M. Parsons, P. J. Barr, K. Stuart, M. Selkirk, and N. Agabian.** 1983. Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in *Trypanosoma brucei*. Cell **34**:901–909.
34. **Parker, K., and J. A. Steitz.** 1987. Analyses of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA. Mol. Cell. Biol. **7**:2899–2913.
35. **Parker, R., P.G. Siliciano, and C. Guthrie.** 1987. Recognition of the TACTAAC box during mRNA splicing is yeast involves base pairing to the U2-like snRNA. Cell **49**:229–239.
36. **Parker, R., T. Simmons, E. O. Shuster, P. G. Siliciano, and C. Guthrie.** 1988. Genetic analysis of small nuclear RNAs in *Saccharomyces cerevisiae*: viable sextuple mutant. Mol. Cell. Biol. **8**:3150–3159.
37. **Parsons, M., R. G. Nelson, K. P. Watkins, and N. Agabian.** 1984. Trypanosome mRNAs share a common 5′ spliced leader sequence. Cell **38**:309–316.
38. **Patterson, B., and C. Guthrie.** 1987. An essential yeast snRNA with a U5-like domain is required for splicing *in vivo*. Cell **49**:613–624.
39. **Perry, K., K. P. Watkins, and N. Agabian.** 1987. Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. Proc. Natl. Acad. Sci. USA **84**:8190–8194.
40. **Pettersson, I., M. Hinterberger, T. Mimori, E. Gottlieb, and J. A. Steitz.** 1984. The structure of mammalian small nuclear ribonucleoproteins. J. Biol. Chem. **259**:5907–5914.
41. **Reddy, R.** 1986. Compilation of small RNA sequences. Nucleic Acids Res. **14**(Suppl.):r61–r72.
42. **Reddy, R., D. Henning, and H. Busch.** 1985. Primary and secondary structure of U8 small nuclear RNA. J. Biol. Chem. **260**:10930–10935.
43. **Riedel, N., J. Wise, H. Swerdlow, A. Mak, and C. Guthrie.** 1986. Small nuclear RNAs from *Saccharomyces cerevisiae*: unexpected diversity in abundance, size, and molecular complexity. Proc. Natl. Acad. Sci. USA **83**:8097–8101.
44. **Riedel, N., S. Wolin, and C. Guthrie.** 1987. A subset of yeast snRNAs contains functional binding sites for the highly conserved Sm antigen. Science **235**:328–331.
45. **Rinke, J., B. Appel, M. Digweed, and R. Lührmann.** 1985. Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen crosslinking. J. Mol. Biol. **185**:721–731.
46. **Saba, J. A., H. Busch, D. Wright, and R. Reddy.** 1986. Isolation and characterization of two putative full-length *Drosophila* U4 small nuclear RNA genes. J. Biol. Chem. **261**:8750–8753.
47. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
48. **Sharp, P. A.** 1987. Trans splicing: variation on a familiar theme? Cell **50**:147–148.
49. **Sharp, P. A.** 1987. Splicing of messenger RNA precursors. Science **235**:766–771.
50. **Siliciano, P. G., D. A. Brow, H. Roiha, and C. Guthrie.** 1987. An essential snRNA from *S. cerevisiae* has properties predicted for U4, including interaction with a U6-like snRNA. Cell **50**:585–592.
51. **Siliciano, P. G., M. H. Jones, and C. Guthrie.** 1987. *Saccharomyces cerevisiae* has a U1-like small nuclear RNA with unexpected properties. Science **237**:1484–1487.
52. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
53. **Stuart, K., E. Gobright, L. Jenni, M. Milhausen, L. Thomashow, and N. Agabian.** 1984. The IsTat 1 serodeme of *Trypanosoma brucei*: development of a new serodeme. J. Parasitol. **70**:747–754.
54. **Suh, D., H. Busch, and R. Reddy.** 1986. Isolation and characterization of a human U3 small nucleolar RNA gene. Biochem. Biophys. Res. Commun. **137**:1133–1140.
55. **Sutton, R. E., and J. C. Boothroyd.** 1986. Evidence for trans splicing in trypanosomes. Cell **47**:527–535.
56. **Tschudi, C., F. F. Richards, and E. Ullu.** 1986. The U2 RNA analogue of *T. brucei gambiense*: implications for a splicing mechanism in trypanosomes. Nucleic Acids Res. **14**:8893–8903.
57. **Tuerk, C., P. Gauss, C. Thermes, D. R. Groebe, M. Gayle, N. Guild, G. Stormo, Y. D'Aubenton-Carafa, O. C. Uhlenbeck, I. Tinoco, Jr., E. N. Brody, and L. Gold.** 1988. CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. Proc. Natl. Acad. Sci. USA **85**:1364–1368.
58. **Tullis, R. H., and H. Rubin.** 1980. Calcium protects DNase 1 from proteinase K: a new method for the removal of contaminating RNase from DNase 1. Anal. Biochem. **107**:260–264.
59. **Wise, J., D. Tollervey, D. Maloney, H. Swerdlow, E. J. Dunn, and C. Guthrie.** 1983. Yeast contains small nuclear RNAs encoded by single copy genes. Cell **35**:743–751.
60. **Zhuang, Y., and A. M. Weiner.** 1986. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. Cell **46**:827–835.