3986–3994 *Nucleic Acids Research, 1999, Vol. 27, No. 20*

Cloning and mutational analysis of the *Leptomonas seymouri* U5 snRNA gene: function of the Sm site in core RNP formation and nuclear localization

Mathias Bell and Albrecht Bindereif*

Institut für Biochemie, Humboldt-Universität/Charité, Monbijoustraße 2, D-10117 Berlin, Germany

Received July 19, 1999; Revised and Accepted August 30, 1999

DDBJ/EMBL/GenBank accession nos AJ243568, AJ243569

ABSTRACT

We have cloned the single-copy gene for the *trans*-spliceosomal U5 snRNA from the trypanosomatid species *Leptomonas seymouri*, using U5 RNA affinity selection and cDNA cloning. Sequence comparison revealed that the *trans*-spliceosomal U5 RNAs from trypanosomatid species share certain characteristic features. Interestingly, the affinity selection procedure yielded—in addition to the bona fide U5 RNA—a closely related small RNA, which can be folded into the same secondary structure, but carries three changes in the loop sequence. This raises the possibility that there may be a larger family of U5-like RNAs in trypanosomes. To study the U5 snRNP assembly and function in trypanosomes we have established a stable expression system in *L.seymouri*. Two cell lines have been generated that express U5 RNAs with mutations in the Sm site, resulting in a defect of core snRNP formation. In addition, the U5 Sm-mutant RNAs behaved differently in cell fractionation, implying a defect in nuclear localization. In sum, this demonstrates for the first time that the Sm site of trypanosome snRNAs contributes an essential element for stable core RNP assembly and may be important for nuclear localization, in analogy to the Sm site function of *cis*-spliceosomal snRNAs in higher eucaryotes.

INTRODUCTION

Trans mRNA splicing has been studied mainly in the trypanosome and nematode systems and follows the same principal two-step mechanism of transesterification reactions as conventional *cis* mRNA splicing. From many studies in the mammalian and yeast systems, we know in considerable detail the *cis*-splicing machinery; in contrast, our knowledge on RNA and protein cofactors as well as on their interactions with the *trans* mRNA substrates is still very limited (for recent reviews, see 1–3). In the last few years, it has become clear, however, that there are homologous protein factors in *cis*- and *trans*-splicing systems: first, a 40 kDa U2 snRNP protein from *Trypanosoma brucei*, which in its N-terminal half is homologous to the mammalian U2 A′ protein (4); second, a 277 kDa U5-specific protein from *T.brucei* with extensive homology to the *cis*-spliceosomal PRP8/p220 factors (5); and third, a set of at least five Sm-analogous proteins common to the *trans*-spliceosomal SL RNP and the U2, U4/U6 and U5 snRNPs (6–8).

In addition to protein factors, the small nuclear RNAs (snRNAs) U2, U4 and U6 are involved in *trans* splicing and are conserved between the trypanosome *trans*- and conventional *cis*-splicing systems; a *trans*-spliceosomal U1 RNA homolog, however, has not been detected in trypanosomes; this probably reflects the *trans*-splicing specific manner, in which the unique 5′ splice site of the SL RNA is recognized and activated. Because sequence conservation of U5 RNAs is generally low, a *trans*-spliceosomal U5 homolog has only recently been detected in the trypanosomatid species *T.brucei* (5,9) and *Leptomonas collosoma* (10). These *trans*-spliceosomal U5 RNAs represent the shortest known U5 RNAs, and conservation is restricted mainly to the 5′ stem–loop. In the *cis*-spliceosome, U5 RNA plays—in close collaboration with the associated PRP8/p220 protein—an essential and central role in splice-site selection and specificity (reviewed in 11,12). A detailed mutational analysis and functional studies of the role of the U5 snRNP in *trans* splicing are therefore of importance and promise more insight into interactions that are distinct between the *cis*- and *trans*-splicing machineries.

We have cloned and analyzed the *Leptomonas seymouri* U5 RNA gene, and established an expression system for this *trans*-spliceosomal U5 RNA. This opens up a detailed mutational analysis of the U5 snRNP assembly and function in trypanosomes. Using this expression system we demonstrate here for the first time a role of the conserved Sm-analogous binding site of *trans*-spliceosomal snRNAs in core snRNP assembly and nuclear localization, in analogy to the Sm site function of *cis*-spliceosomal snRNAs in higher eucaryotes.

MATERIALS AND METHODS

U5 expression and mutant constructs

As basic vectors for the expression of U5 mutant RNAs, pBNeo and pBHyg were constructed. For pBNeo, a 2.3 kb

*To whom correspondence should be addressed at present address: Institut für Biochemie, Fachbereich Biologie, Justus-Liebig-Universität Geißen, H.-Buff-Ring 58, D-35392 Gießen, Germany. Tel: +49 641 35 420; Fax: +49 641 35 419; Email: albrecht.bindereif@charite.de

fragment containing the neomycin resistance gene, the PARP promoter and the PARP 3′ region was isolated from pT13-11 (13) by *Eco*NI cleavage, Klenow fill-in reaction and subsequent *Sal*I cleavage; this fragment was cloned between the *Eco*RV and the *Sal*I restriction sites of pBluescript. pBHyg was constructed in the same way, using the corresponding fragment with the hygromycin gene, which was derived from pND-1 (kindly provided by P. Patnaik).

To introduce the tags and the Sm site mutations into the U5 gene locus a two-step PCR technique was used (14). First, using the genomic clone as a template, 5′ and 3′ fragments were amplified; the internal primers overlapped in a region of ~20 nt and introduced the mutation, and the flanking primers contained terminal *Xho*I restriction sites. Second, the two overlapping fragments were mixed, and the full-length fragment carrying the mutation was amplified with the flanking primers, followed by *Xho*I cleavage and cloning into the *Xho*I site of pBNeo. The U5 loop mutations were introduced by outward PCR, using pBHyg/Ls U5 G46A, which was amplified using kinased tail-to-tail primers carrying the mutation. The PCR product was religated, transformed, and the mutation was confirmed by DNA sequencing.

Transfection of *L.seymouri*

Stably transfected cell lines of *L.seymouri* were obtained by electroporation, using the BTX electroporation system, 0.2 mm gap cuvettes and 20–100 μg plasmid DNA per transfection (3×10^7 cells in a volume of 400 μl ZPFM) (15). Mutant cell lines were selected in the presence of G418 (Geneticin, Gibco BRL) or hygromycin (Boehringer Mannheim), each at a concentration of 100 μg/ml. The constructs with U5 loop mutations were transfected into *L.seymouri* carrying a spliced leader RNA construct, pDH/SL-sub8 (15).

Cell culture and extract preparation

Cultures of *L.seymouri* (15) and of the procyclic form of *T.brucei* strain 427 (4) were grown as described. S100 extracts from *T.brucei* strain 427 and *L.seymouri* were prepared as previously (4). The cell fractionation of *L.seymouri* basically followed the procedure of Field and Field (16).

Oligonucleotides

The following DNA oligonucleotides were used: U5-Tbs1, 5′-GCA TCG CCG TCT CGA-3′, corresponding to nt 1–15 of *T.brucei* U5 snRNA; U5-Tb4, 5′-GAC ACC CCA AAG TTT AAA CGC-3′, complementary to nt 42–62 of *T.brucei* U5 snRNA; seqA-1, 5′-TCA GTC AAT CCG GTT C-3′, complementary to nt 52–67 of *L.seymouri* U5-relA RNA; seqA-2, 5′-CGC GTG ATG TGT CTA C-3′, corresponding to nt 1–16 of *L.seymouri* U5-relA RNA; Lsd1-1, 5′-GGC TGC AGG TAG CAG AAC-3′, corresponding to nt –92 to –74 upstream of the *L.seymouri* U5 gene; Lsu2-1, 5′-TCC GCG TGA GAT GAT GC-3′, complementary to nt 189–205 downstream of the *L.seymouri* U5 gene; seqC-2, 5′-CCT AGG TAG GCC TCA AAA T-3′, complementary to nt 54–72 of *L.seymouri* U5 snRNA; seqC-6, 5′-ATG ATA TCA AAT TTT GAG GC-3′, corresponding to nt 44–63 of *L.seymouri* U5 snRNA (carrying two mismatches); seqC-7, 5′-TTA GTA AAA GTC GAA GCA G-3′, complementary to nt 11–29 of *L.seymouri* U5 snRNA; LsU5/ΔSm, 5′-CCT AGG TAG GCC TTT GAT A-3′, complementary to nt 48–66 of *L.seymouri* U5 ΔSm RNA; LsU5/sub-Sm, 5′-CCT AGG TAG GCC TCT CCT C-3′, complementary to nt 54–72 of *L.seymouri* U5 sub-Sm RNA; sal-adapt-primer, 5′-TAT ATC GAC CCA CGC GTC CG-3′; and *Not*I poly(dG) primer, 5′-GAC TAG TTC TAG ATC GCG AGC GGC CGC CC$(G)_{15}$-3′.

For U5 snRNP affinity selections a biotinylated antisense 2′-*O*-methyl/2′-*O*-allyl RNA oligonucleotide was used: Tb-U5A, 5′-UXU XUA* GUA* A*A*A* GUC GA*G XUX U-3′ [A*, 2′-*O*-allyl 2-aminoadenosine derivative (17); X, biotin-dT; U, 2′-*O*-methyl U; C, 2′-*O*-methyl C; and G, 2′-*O*-methyl G].

Antisense affinity selection of *L.seymouri* U5 RNA

U5 snRNA selection on a preparative scale was performed as described (6). The efficiency and specificity of affinity selection was determined by 3′ end-labeling of the RNA with [^{32}P]pCp and analysis by denaturing polyacrylamide–urea gel electrophoresis. Prior to cDNA cloning the RNA was size-fractionated by denaturating gel electrophoresis and purified.

cDNA cloning of the U5 RNA gene from *L.seymouri*

cDNA cloning of *L.seymouri* U5 snRNA was performed using the SUPERSCRIPT™ Plasmid System (Gibco BRL). Affinity-selected RNA was poly(C)-tailed by poly(A) polymerase (Gibco BRL) for 2 h at 37°C. First and second strand synthesis as well as *Sal*I adapter ligation were performed as described for the SUPERSCRIPT™ Plasmid System, except that the *Not*I poly(dG) primer adapter was used for first strand synthesis instead of the corresponding poly(dT) DNA oligonucleotide. An aliquot of the cDNA including the terminal *Sal*I adapter was amplified by PCR [using Pwo DNA polymerase (Boehringer Mannheim) and 0.8 μM Sal-adapt-primer]. Products were size-fractionated by agarose gel electrophoresis and gel-purified, followed by ligation into *Eco*RV-cut pBluescript, transformation of *Escherichia coli* JM109, and sequencing of the inserts of isolated plasmids.

Since the expected 5′ loop of U5 was missing in the sequenced products, cDNA cloning of affinity-selected RNA in addition was performed according to the procedure of O'Brien and Wolin (18). Products in the expected size range were purified and re-amplified with the anchored dT primer and either the seqC-1 or seqA-1 primers directed against the sequences obtained from the initial SUPERSCRIPT™ cDNA cloning procedure described above. PCR products were isolated, ligated to *Eco*RV-cut pBluescript, re-amplified with pBluescript-dependent primers, and sequenced, resulting in U5 and U5-relA sequences. In addition, the RNA sequence of U5-relA was confirmed by primer-extension sequencing, using oligonucleotide seqA-1.

Genomic cloning of the *L.seymouri* U5 snRNA gene

To generate a longer probe for genomic cloning of the *L.seymouri* U5 snRNA gene, flanking regions were amplified by inverse PCR, using genomic *Hinc*II and *Pst*I fragments and primers seqC-6 and seqC-7, followed by sequencing of gel-purified PCR products. The U5 snRNA gene was cloned from a *L.seymouri Bgl*II library [ZAP Express™ cDNA Gigapack III Gold Cloning Kit (Stratagene)]; phages were screened by plaque lift assay using the U5 snRNA probe (see the section on non-radioactive hybridization techniques and probes). Putative positive plaques were isolated and re-purified. The genomic inserts were excised *in vivo* as described (Stratagene). Isolated

plasmid DNA was analyzed by restriction digestion and Southern blots.

Genomic cloning of the *T.brucei* U5 snRNA gene

A bacteriophage P1 library of the *T.brucei* strain TREU 927/4 on high-density filters (kindly obtained from Sarah Melville and Vanessa Leech, Cambridge, UK) was screened with a *T.brucei* U5 RNA probe (see below). Plasmid DNA from the positive clone 17D12 was isolated, cut with different restriction enzymes and analyzed by Southern blot for U5 RNA gene containing fragments. A 4.7 kb *Pst*I fragment was isolated, subcloned into pBluescript, and sequenced.

Non-radioactive hybridization techniques and probes

One microgram total RNA was separated by denaturing polyacrylamide–urea gel electrophoresis, using DIG-labeled DNA molecular weight marker VIII (Boehringer Mannheim). The gel was equilibrated in 0.5× TBE and transferred onto a nylon membrane by semi-dry blotting for 1 h at 400 mA (Bio-Rad). RNA was crosslinked by UV light and probed as described below.

For northern blot analysis as well as for P1 library and λ phage screens the digoxigenin-11-dUTP (DIG) system from Boehringer Mannheim was used. Except for the P1 high-density filters in all cases positively charged nylon membranes were taken. DIG-labeled probes were generated by PCR, including the PCR DIG labeling mix as desoxynucleotides and genomic DNA as a template. For generating the *T.brucei* U5 snRNA probe by PCR, primers U5-Tbs1 and U5-Tb4 were used, for the *L.seymouri* U5-relA probe primers seqA-1 and seqA-2, and for the *L.seymouri* U5 snRNA probe primers Lsd1-1 and Lsu2-1. Hybridization was performed in standard buffer (P1 library screen and plaque hybridization) or in Church buffer (northern blot); washes of the blot and probe detection with AP-conjugated anti-DIG-Fab fragments were done as described in the Boehringer manual.

Primer extension

DNA oligonucleotides complementary to the RNA to be detected were 5′-end-labeled with [γ-^{32}P]ATP and polynucleotide kinase. 50 000 c.p.m. of these primers (3.5×10^5 c.p.m./pmol) were annealed to the RNA (10 min at 70°C, then 10 min on ice). For primer extension Expand^{TM} reverse transcriptase (Boehringer Mannheim) was used in the supplied reaction buffer supplemented with 0.5 mM dNTPs (for full-length primer extension) or 0.5 mM ddCTP mix (dATP, dGTP, dTTP, ddCTP; for terminated primer extension). The reaction was performed for 10 min at 30°C followed by 50 min at 42°C. Products were analyzed by denaturing polyacrylamide–urea gel electrophoresis.

CsCl density gradient centrifugation

CsCl density gradient centrifugation was done as described previously (5). The density distribution across the gradient was measured in a parallel gradient run and ranged from 1.31 (lane 1) to 1.47 g/ml (lane 10). RNA prepared from each fraction was divided and analyzed by primer-extension assays with wild-type and mutant-specific primers. The fractionation of wild-type U5 snRNA (shown in Fig. 6A) was determined with wild-type specific primer, using gradient fractions from U5ΔSm extract. The fractionation of U5ΔSm and U5 sub-Sm RNAs was determined with the mutant-specific primers and RNA from the respective mutant cell lines. Similarly, the U5 loop mutations were detected by terminated primer extension, using seqC-2 primer.

RNA secondary structure prediction

To predict the secondary structure of RNAs the program RNAdraw V1.0 (Matzura Multimedia, Sweden) was used. The integrity of the 5′ stem–loop was confirmed for the wild-type and all mutant derivatives of U5. However, the remaining 3′ terminal portion shows some variability in folding for the different RNAs; since we have no experimental data to distinguish between these structures, the 3′ portion of U5 RNAs is drawn as single-stranded in Figure 1.

Nucleotide sequence accession numbers

The nucleotide sequence data reported in this paper will appear in the EMBL/GenBank database under the accession numbers AJ243568 (*T.brucei* U5 snRNA gene) and AJ243569 (*L.seymouri* U5 snRNA gene).

RESULTS AND DISCUSSION

U5 snRNA from *L.seymouri*: antisense affinity selection and cDNA cloning

To study the role of the U5 snRNP in *trans* splicing by mutational analysis we established an expression system for the U5 snRNA, using the trypanosomatid species *L.seymouri*. First we cloned the U5 snRNA gene from *L.seymouri*. Since among the spliceosomal snRNAs U5 is least conserved (19), we devised an antisense affinity selection procedure for the purification of U5 RNA–protein complexes. This approach was based on the assumption that the U5 loop sequence is conserved in the *trans*-spliceosomal U5 RNAs as it is in all known *cis*-spliceosomal U5 RNAs. A biotinylated antisense 2′-*O*-methyl/2′-*O*-allyl RNA oligonucleotide directed against the 11 nt of the U5 loop from *T.brucei* was used to affinity-select U5 RNPs from *L.seymouri* S100 extract by strepavidin agarose. RNA was purified from the beads, 3′-^{32}P-end-labeled and analyzed by gel electrophoresis, revealing a cluster of several bands in the 60–80-nt range; a control experiment showed that the appearance of these bands depended on the presence of the biotinylated oligonucleotide (data not shown). cDNA was generated from the selected RNA and cloned. Sequence analysis of several cDNA clones yielded one sequence that could be folded into an RNA secondary structure containing a stem–loop with an 11-nt loop identical to the *T.brucei* U5 loop sequence and adjacent sequences that could be folded into an 8-bp stem (for a secondary structure model, see Fig. 1A).

Interestingly, one of the other cDNA sequences can be folded into an RNA containing a stem–loop related to that of the bona fide U5 snRNA (U5-relA RNA; see Fig. 1B): as in all other U5 snRNAs its loop would consist of 11 nt that, however, include three base changes (in comparison to the trypanosomatid consensus: G→U at position 2; U→A at position 8; and C→U at position 10 of the loop). The adjacent stem would contain 7 bp including an additional, hypothetical A–C pair (note also the A–C pair at the corresponding position of the *L.collosoma* U5 sequence). Similarly, as for the other trypanosomatid U5 RNA sequences, there is no significant homology outside of the 5′ stem–loop (data not shown).

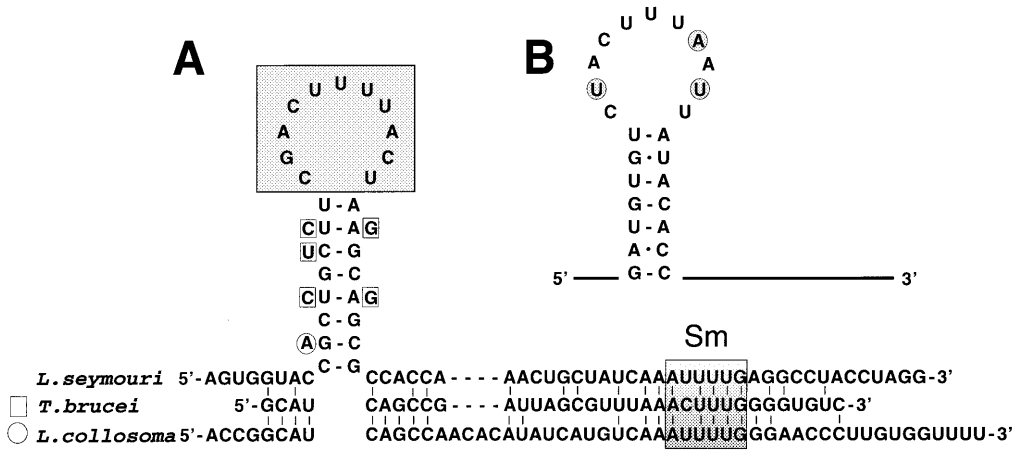

Figure 1. Sequence comparison of the *trans*-spliceosomal U5 snRNAs and U5-relA RNA from trypanosomatid species. (**A**) The U5 RNA sequence from the following three trypanosomatid species are aligned: *L.seymouri* (this study), *T.brucei* (9) and *L.collosoma* (10). The conserved regions 5′ loop and Sm site are boxed. In the 5′ stem region the *L.seymouri* sequence is shown, with deviations from this sequence indicated by boxed (*T.brucei*) and circled positions (*L.collosoma*). (**B**) For the U5-relA RNA from *L.seymouri* only the 5′ stem–loop is represented, with the three deviations from the trypanosomatid consensus in the loop indicated by circles.

In addition, several cDNA sequences were recovered with homologies to tRNAs, probably reflecting some non-specific tRNA binding in the affinity selection with the antisense U5 loop (data not shown).

Both of the U5 and the U5-relA RNAs are expressed in *L.seymouri*, as shown by northern hybridization analysis (Fig. 2) and by primer-extension assays (U5 snRNA: see below and Fig. 5B, lane WT; U5-relA RNA: data not shown). For the U5 snRNA two RNA species of ~72 and ~65 nt, respectively, were detected in total RNA by northern analysis (Fig. 2, lane U5 snRNA); the latter one represents most likely a 3′ shortened derivative, since primer extension resulted only in a single major product. The U5-relA RNA probe detected a single RNA species of ~67 nt (Fig. 2, lane U5-relA RNA).

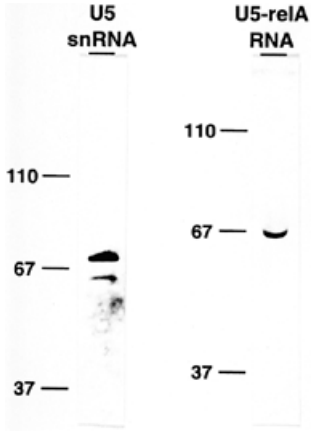

Figure 2. Detection of U5 snRNA and U5-relA RNA in *L.seymouri* by northern hybridization. Total RNA from *L.seymouri* was analyzed by denaturing gel electrophoresis, and U5 snRNA and U5-relA RNA were detected by northern hybridization with specific DIG-labeled probes. Marker sizes are given in nucleotides.

Sequence comparison of trypanosomatid U5 snRNAs

The three *trans*-spliceosomal U5 snRNA sequences from the trypanosomatid species *T.brucei* (62 nt; 9), *L.collosoma* (80 nt; 10) and *L.seymouri* (72 nt; this study) allow a first phylogenetic sequence comparison (Fig. 1A). Sequence homology is clearly concentrated in the 5′ stem–loop portion of U5. The 11-nt loop sequence is identical in the three species, and significantly, all three contain the one deviation from the *cis*-spliceosomal consensus (C→A at the third position in the loop). The loop is closed by a stem of 8 bp, which is well supported by two compensatory changes involving Watson–Crick base pairs (C–G→U–A at the second and the fifth position from the top of the 5′ stem); in addition, the U–G pair (third base pair from the top) is changed into a C–G pair. In contrast, outside of the 5′ stem–loop there is no significant homology between the *T.brucei* and either of the *L.seymouri* or *L.collosoma* sequences, except for a short stretch near the 3′ end (5′-A_3U_4G-3′, positions 52–59 in *L.seymouri*) that resembles the Sm binding site of *cis*-spliceosomal snRNAs and may represent the binding site for the common (Sm-analogous) proteins of trypanosomes (6; see below).

```
-406  AGGGGAAGCG CGATGGGCTC ATAACCCATA GGTCGTTGGA TCGAAACCAA

-356  TCGGTGCTAT AGTTTTTAAC CACAAGGCGC CATCATTCAA TCTCAATGCC

                    tRNA^Leu ──►  box A
-306  CATTTTCCGC ACGAAATCAA GGTGAGATGG TCGAGTGGTC TAAGACGTTG

                                  box B
-256  GATTAAGGTT CCAATCCTTT CGAGGGCGTG GGTTCAAACC CCACTCTCAT

-206  CACTTTTTCC CATTCCCAGC CCTCGGGAAG GGAAAAAAGA GGACCCACCC

      box B                                          box A
-156  GGGTTTGAAC CGGGGACCAT TCGGACTGCA GCCGAATGCT CTACCACTGA

      ◄── tRNA^Cys
-106  GCTATGAATC CAACGGCTGC AGGTAGCAGA ACACATACTC CGCAAACAGT

 -56  TTAATAAGCA TTATCCCAGA CTCATGCAAC ACCACATCCA GTTAGTACCA

           U5 snRNA ──►
  -6  AATGAAAGTG GTACCGCTGC TTCGACTTTT ACTAAGCAGC GCCACCAAAC

   45  TGCTATCAAA TTTTGAGGCC TACCTAGGCC ATAAAGGGAC TTTTTCGTTG

   95  AAAGGCAACA GAAATTACAG CGGCCGCGGT AAAAAGCAAG CACGCCGTCT

  145  GTTTCACTTC TTTGTCGCCT ATAAAAATGT TTCAACGACA ACGAGCATCA

  195  TCTCACGCGG AA
```

Figure 3. Genomic sequence of the U5 snRNA gene locus from *L.seymouri*. The sequence of the 0.6 kb region from the *L.seymouri* U5 snRNA gene locus is given, with the putative A and B boxes indicated. The U5 sequence is marked by a black box; the upstream tRNA sequences are underlined and the arrows indicate their orientation.

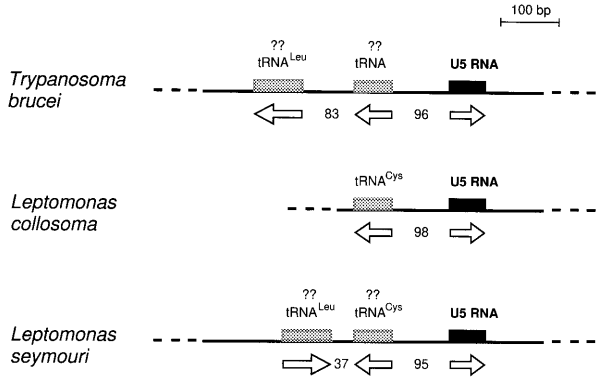

Figure 4. Genomic organization of trypanosomatid U5 snRNA genes from *T.brucei* (this study), *L.collosoma* (10) and *L.seymouri* (this study). The upstream tRNA sequences and their relative orientation are schematically shown; distances between tRNA and U5 snRNA sequences are given in base pairs.

U5 snRNA genes from *L.seymouri* and *T.brucei*: gene structure and comparison with other trypanosomatid snRNA genes

On the basis of the new U5 cDNA sequence from *L.seymouri*, a probe was generated to detect the corresponding gene by Southern blot hybridization (data not shown). For five restriction digests of genomic DNA only a single strong hybridization signal was found, and therefore the U5 RNA gene appears to be present in a single copy in the *L.seymouri* genome.

Using the same probe we then cloned and sequenced a 3.7 kb *Bgl*II fragment containing the *L.seymouri* U5 RNA gene. Figure 3 shows the U5 sequence together with ~540 bp of flanking sequences; a schematic representation of this genomic region is given in Figure 4. Two upstream tRNA sequences containing canonical A and B boxes were found: first, a proximal $tRNA^{Cys}$, in the opposite orientation to the U5 sequence and at a distance of 95 bp; and second, a distal $tRNA^{Leu}$, in the same orientation as U5 and with a 37-bp spacing between the two tRNA sequences (Fig. 4). Both of these tRNA sequences fold into the classical tRNA secondary structure (data not shown); however, we do not know whether they are expressed.

To determine whether this gene organization is conserved in trypanosomatid U5 snRNA genes and as a basis for U5 snRNA expression and mutational studies, we have also cloned the *T.brucei* U5 snRNA gene, using the published U5 snRNA sequence (9; see Materials and Methods). Sequence analysis of the upstream region of the *T.brucei* U5 snRNA gene showed that it contains two tRNA sequences. First, a proximal tRNA-like sequence in the opposite orientation as U5 and at a distance of 96 bp; this tRNA sequence carries A and B boxes and shows—restricted to the 3′ half—78% identity to a *T.brucei* $tRNA^{Asn}$ sequence (accession number Z11884). Second, a distal $tRNA^{Leu}$ sequence, which can be folded into the canonical tRNA structure as supported by seven compensatory base changes and an 81% identity to a known *T.brucei* $tRNA^{Leu}$ gene (accession number X13750).

This gene organization is very reminiscent of other trypanosomatid snRNA genes (see, for example, 20,21) and suggests that U5 RNA transcription in *Leptomonas*—as snRNA transcription in other trypanosomatids—is directed by a bidirectional RNA polymerase III promoter in close linkage with upstream tRNA sequences (reviewed in 22). Most likely it is only the proximal A and B boxes that are relevant to U5 RNA expression: their distance relative to the U5 RNA is conserved, and in the case of the *T.brucei* U5 snRNA gene, the A/B boxes are intact, even though the context cannot be folded into a canonical tRNA secondary structure. In contrast, the orientation of the distal tRNA sequence is not conserved.

Establishing an expression system for the *L.seymouri* U5 snRNA

For the expression and mutational analysis of U5 snRNA assembly and function in the trypanosome system, we cloned the U5 snRNA gene from *L.seymouri* into a pBluescript derivative, pBNeo. pBNeo is stably maintained as an episome in *Leptomonas* and contains the G418 (neomycin) resistance gene as a selectable marker. In order to distinguish expression of the transfected gene from that of the endogenous U5 RNA gene, a single point mutation, G46A, was introduced in a region without apparent sequence conservation (pBNeo/Ls U5 G46A; Fig. 5A). As an alternative marker, we inserted 5 nt (5′-GGUUU-3′)

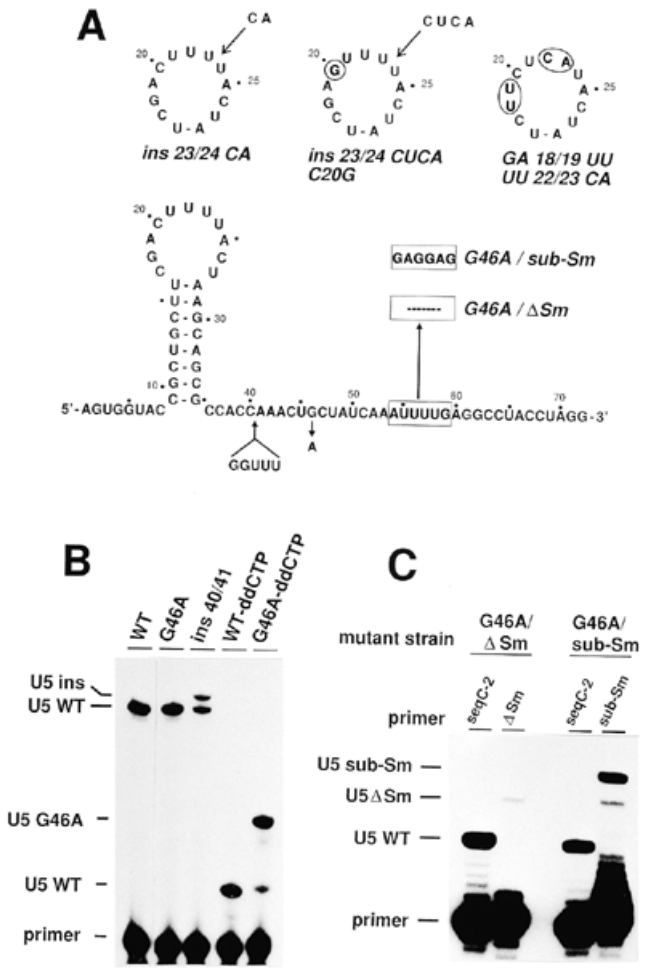

Figure 5. *In vivo* expression of mutant derivatives of the *trans*-spliceosomal *L.seymouri* U5 snRNA. (**A**) Schematic representation of *L.seymouri* U5 snRNA mutant derivatives. The following two tags were introduced into the U5 sequence within an episomal expression construct: an insertion of 5 nt (5′-GGUUU-3′) between positions 40 and 41, U5 ins 40/41, and a single point mutation, G46A. In addition, mutations were made in two highly conserved regions: first, a substitution (Ls U5 G46A/sub-Sm) or a deletion (Ls U5 G46A/ΔSm) were introduced in the Sm-like sequence of *L.seymouri* U5 RNA (5′-AUUUUG-3′, positions 54–59). Second, the loop sequence was altered by three mutations shown above: ins 23/24 CA; ins 23/24 CUCA and C20G; GA 18/19 UU and UU 22/23 CA (substituted positions circled). (**B**) The two tagged U5 expression constructs (Ls U5 G46A and ins 40/41) were stably transfected into *L.seymouri*, total RNA was prepared, and U5 RNA expression was analyzed by primer extension; RNA from untransfected wild-type cells served as a control. The expression of the wild-type U5 RNA and U5 G46A RNA was detected by full-length and ddCTP-terminated primer extension, respectively, using oligonucleotide seqC-2 (lanes WT, G46A, WT-ddCTP and G46A-ddCTP). U5 RNA with the insertion was detected by full-length primer extension (lane ins 40/41). For both mutations primer extension allowed the simultaneous detection of endogenous U5 RNA and U5 RNA from the transfected gene (positions of primer-extension products marked on the side). (**C**) The two constructs with Sm site mutations were stably transfected into *L.seymouri*, total RNA was prepared, and U5 RNA expression was analyzed by ddCTP-terminated primer extension (lanes G46A/ΔSm and G46A/sub-Sm). Wild-type U5 and mutant Sm U5 RNAs were detected in two separate primer-extension reactions, using seqC-2 and Sm-mutant specific primers, respectively (bands identified on the side).

between positions 40 and 41 (pBNeo/Ls U5 ins 40/41). Using these two constructs stable cell lines were established, and U5 RNA expression was analyzed by primer extension. Figure 5B shows that either of the tagged U5 derivatives was expressed efficiently and at levels comparable to those of the endogenous wild-type gene; expression of the U5 snRNA with the 5 nt insertion was detected by full-length primer extension (compare lanes WT and ins 40/41), U5 G46A RNA expression by terminated primer-extension reactions (compare lanes WT-ddCTP and G46A-ddCTP); either assay of the expressed U5 RNAs resulted in two primer-extension products, which quantitatively reflect the endogenous U5 RNA and U5 RNA from the transfected gene, respectively. In the U5 RNA with the single point mutation G46A the same 5′ end was found as in wild-type U5 RNA (see full-length primer extension in lane G46A); northern analysis showed no effect on 3′ end formation (data not shown). For introducing additional mutations into U5, we decided to use G46A as a marker, which is expected to interfere least with the U5 snRNA structure and function.

Mutational analysis of the *L.seymouri* U5 snRNA: the Sm site, but not the conserved loop sequence, is required for assembly of a stable U5 core RNP

Phylogenetic sequence comparison of the three trypanosomatid U5 snRNAs had revealed only two conserved elements: the 5′ loop sequence (positions 17–27 of the *L.seymouri* U5 RNA) and the putative Sm region (5′-AU_4G-3′, positions 54–59; see above and Fig. 1A; Fig. 5). The latter element resembles the Sm site of *cis*-spliceosomal snRNAs U1, U2, U4 and U5 (consensus RAU_{3-4}NUGR; 23,24), which usually resides in a single-stranded stretch between two stem–loops; in the corresponding regions of the *trans*-spliceosomal snRNAs (SL, U2, U4 and U5) a similar, although more degenerate sequence element can be found (data not shown). Binding of core protein components occurs in the Sm region of the SL RNA, as shown by RNase H protection and oligonucleotide masking experiments (4,25), in analogy to Sm protein binding to the Sm site of snRNAs from higher eucaryotes (26, and references therein). Consistent with such a conserved role of the Sm site, mammalian Sm proteins have been shown to bind trypanosomal snRNAs *in vitro* (27). As established in higher eucaryotes, Sm protein binding is required for cap trimethylation of snRNAs (28); the Sm core together with the m_3G cap structure constitute a bipartite signal important for nuclear snRNP translocation (29,30). Therefore it was of general interest to determine whether trypanosomal snRNPs follow a similar biogenesis pathway. An additional relevant aspect in this context concerns the different 5′ ends of the trypanosome snRNAs: in contrast to U5 (9) and U6 snRNAs, U2 and U4 snRNAs carry the canonical m_3G cap, as shown for the *T.brucei* snRNAs (31), whereas the SL RNA contains a unique, so-called cap4 structure (32).

As a first step to elucidate the role of the Sm site in trypanosomes, we have introduced two mutations into the *L.seymouri* U5 snRNA (Fig. 5A): in Ls U5 G46A/ΔSm 6 nt are deleted (AU_4G; positions 54–59), and in Ls U5 G46A/sub-Sm the AU_4 sequence of the Sm site (positions 54–58) is replaced by GAGGA. These constructs were transfected into *L.seymouri*, stable cell lines were established, and U5 RNA expression was tested, using wild-type and mutant-specific primers (Fig. 5C). The specificity of these primers was confirmed in control reactions with RNA from untransfected cells and from a

U5 G46A-expressing cell line (data not shown). In conclusion, both U5 RNA derivatives with the Sm site either deleted or substituted were expressed: U5 G46A/sub-Sm at levels similar to wild-type (compare lanes seqC and sub-Sm in the mutant strain G46A/sub-Sm), and U5 G46A/ΔSm at comparatively low levels (compare lanes seqC-2 and ΔSm in the mutant cell line G46A/ΔSm). Southern blot analysis of genomic DNA revealed no significant difference of the episome copy number in the two mutant cell lines (data not shown). Most likely the relatively low level of the Sm deletion derivative (U5 G46A/ΔSm) is caused by a stability defect of this mutant snRNA and its inability to stably associate with common proteins (see below). We did not detect any effect of these Sm mutations in 5′ and 3′ end formation: using full-length primer extensions and northern hybridization analysis, no apparent difference was observed between RNA from wild-type cells and mutant Sm cell lines (data not shown).

On the basis of these two cell lines expressing Sm-mutant derivatives of U5, we analyzed U5 core RNP assembly and stability by CsCl density gradient centrifugation. Under the highly stringent conditions of CsCl gradients, a core complex of the *trans*-spliceosomal U5 snRNP is stable, as shown previously for the *T.brucei* U5 core RNP (5) as well as for the SL, SLA, U2 and U4/U6 RNPs (4,33–35). Nuclear extracts from the two *Leptomonas* cell lines were fractionated through CsCl gradients, and RNA was prepared from each gradient fraction for primer extension analysis. As Figure 6A clearly shows, the endogenous U5 RNA peaked exclusively around fraction 6, corresponding to a stable U5 core RNP with a density of 1.36 g/ml, which is consistent with our previous study of the *T.brucei* U5 snRNP (5). The distribution of U5 G46A RNA–protein complexes across the gradient was identical to that observed for the endogenous U5 snRNP (data not shown); this single point mutation therefore does not affect core snRNP assembly.

In contrast, the two U5 Sm-mutant RNAs exhibited a drastically different distribution after CsCl gradient fractionation, with a peak in fractions 9 and 10 corresponding to free RNA (U5 G46A/ΔSm RNA, Fig. 6B; U5 G46A/sub-Sm RNA, Fig. 6C). This demonstrates that either deleting or substituting the Sm site of U5 strongly interferes with formation of a stable core RNP. Our analysis cannot distinguish whether the Sm-mutant RNAs do not bind core proteins anymore or only lack the characteristic, high stability of the core RNP. However, we can conclude that the Sm site contributes an essential element for the formation of a stable U5 core RNP.

In addition to the Sm sequence, we tested the other conserved element of U5, the 5′ loop, for a function in core snRNP assembly. Three mutations were introduced: an insertion (ins 23/24 CA), an insertion/substitution (ins 23/24 CUCA and C20G) and a substitution (GA 18/19 UU and UU 22/23 CA) (Fig. 5A; the particular choice of these mutations was originally based on experiments addressing a potential U5 loop/SL RNA interaction). Cell lines of *L.seymouri* were created that were stably transfected with these mutant U5 snRNA constructs. The expression of the mutant U5 snRNAs was assayed by primer extension: in comparison to the endogenous wild-type U5 snRNA, the expression of the insertion (ins 23/24 CA) and insertion/substitution mutants (ins 23/24 CUCA and C20G) was low, while the substitution mutant (GA 18/19 UU and UU 22/23 CA) was expressed at comparable levels (Fig. 7, input lanes T; data not shown). Nuclear extracts were prepared from

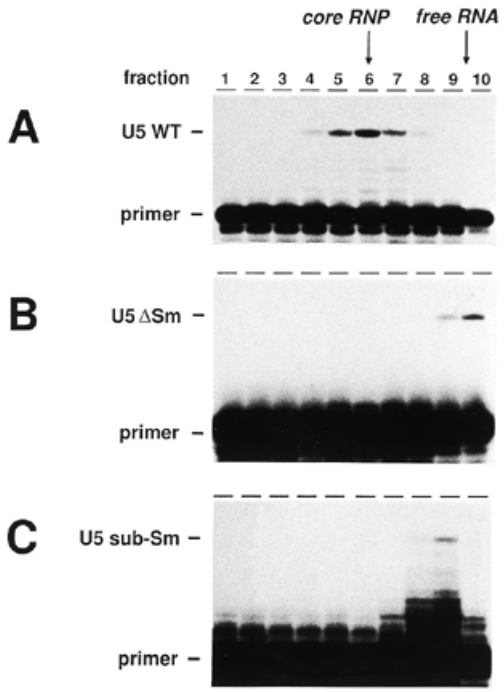

Figure 6. U5 RNAs with Sm mutations are defective in core snRNP formation. Nuclear extract was prepared from cell lines of *L.seymouri* expressing U5 RNAs with the sub-Sm and ΔSm mutations. RNA–protein complexes and free RNA were fractionated by CsCl gradient centrifugation, and RNA from fractions 1–10 (top to bottom) was analyzed by primer extension for the distribution of wild-type (WT) U5 RNA (**A**), U5 ΔSm (**B**) and U5 sub-Sm (**C**). The positions of primer-extension products and primers are marked on the side. The distribution of U5 core RNPs (fractions 5–7) and of free RNA (fractions 9 and 10) is indicated above.

these mutant cell lines, and U5 core snRNP assembly was characterized by CsCl density gradient centrifugation (Fig. 7). Primer-extension assays revealed the distribution of endogenous and mutant U5 snRNAs across the CsCl gradient. Each of the U5 snRNAs with loop mutations behaved as wild-type U5 snRNA and was detected in the form of core complexes with the characteristic density (Fig. 7A–C; see fractions 4 and 5). We conclude that—in contrast to the Sm sequence—the conserved loop of U5 does not contribute an essential element for core snRNP assembly.

Function of the Sm site in nuclear localization of the U5 RNP

Since in higher eucaryotes the Sm core functions as a nuclear localization signal during snRNP biogenesis, we next wanted to assess whether it plays a similar role in the trypanosome system. We used the two *L.seymouri* cell lines expressing U5 RNAs with defective Sm sequences, U5 G46A/ΔSm and U5 G46A/sub-Sm, and for comparison cells that express U5 G46A. The intracellular U5 RNA distribution was determined by cell fractionation, comparing with each other an aliquot of the total fraction, the cytoplasmic fraction, and the nuclear

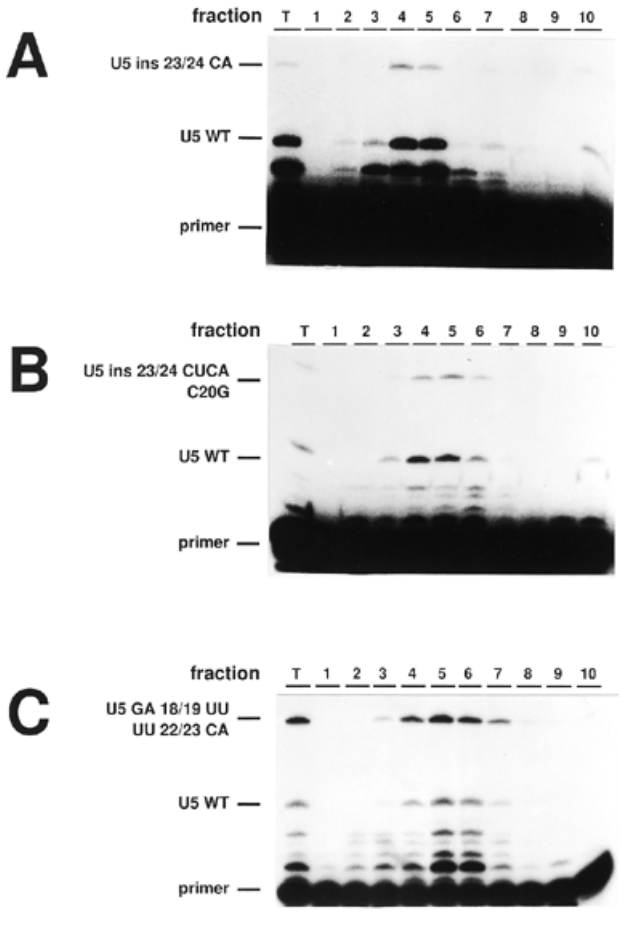

Figure 7. Core snRNP formation of U5 RNAs with loop mutations. Nuclear extract was prepared from cell lines of *L.seymouri* expressing U5 RNAs with loop mutations as indicated. RNA–protein complexes and free RNA were fractionated by CsCl gradient centrifugation, and RNA from fractions 1–10 (top to bottom) was analyzed by primer extension for the distribution of wild-type (WT) and U5 mutant RNA ins 23/24 CA (**A**), ins 23/24 CUCA and C20G (**B**) and GA 18/19 UU and UU 22/23 CA (**C**). Note that the density distribution in these gradients was slightly different from that in the gradients shown in Figure 6; therefore the peak of the U5 core snRNP is shifted here to fractions 4 and 5. The positions of primer-extension products and primers are marked on the side.

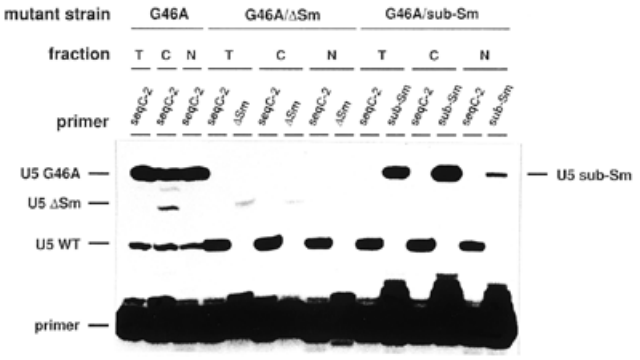

Figure 8. Trypanosome U5 RNAs with mutated Sm sites are affected in nuclear localization. *L.seymouri* cells expressing U5 RNAs with the sub-Sm and ΔSm mutations were lysed by Dounce homogenization. Cell fractionation resulted in cytoplasmic (C) and nuclear (N) extracts. RNA was prepared from each fraction (lanes C and N) as well as from whole cells (lanes T); the distribution of wild-type (internal control) and Sm-mutant U5 RNAs was analyzed by primer extension (positions of primer-extension products indicated on the side; the band between the U5 G46A and the WT primer-extension signals was observed only in some cytoplasmic preparations and may be caused by RNA degradation).

extract. RNA was prepared from each fraction and analyzed by primer extension. Nuclear leakage cannot be completely avoided in this approach (see below and previous studies, for example 36). Therefore results from the cellular fractionation cannot be taken as a quantitative measure to reflect precisely the intracellular RNA localization.

Significantly, cellular fractionation revealed clear differences between wild-type, which served as an internal standard, and Sm-mutant U5 RNAs (Fig. 8): whereas wild-type U5 RNA and U5 RNA G46A were found at similar levels in the cytoplasmic and nuclear fractions, U5 G46A/sub-Sm RNA was enriched in the cytoplasmic fraction; conversely, the abundance in the nuclear fraction was reduced to ~30%. This suggests that the mutant U5 RNA with an Sm site substitution (U5 G46A/sub-Sm) was defective or inefficient in nuclear translocation. Consistent with this result, the Sm deletion derivative (U5 G46A/ΔSm) could not be detected in the nuclear fraction; the relative abundance in the cytoplasmic fraction, however, was similar to wild-type. Therefore both a substitution and the deletion mutation in the U5 snRNA Sm site appeared to result in a nuclear localization defect.

In sum and combined with the data on core complexes (Fig. 6), these results support a model whereby U5 snRNA binds common proteins during a transient cytoplasmic phase, before the U5 snRNP is translocated into the nucleus. Such a model of snRNP biogenesis in trypanosomes would be directly comparable to the snRNP maturation pathway, as it has been characterized in higher eucaryotes. Extending mutational analyses such as the one presented here for the U5 snRNA to the other *trans*-spliceosomal snRNAs may give more insight into these diverse processes of snRNP maturation.

ACKNOWLEDGEMENTS

We gratefully acknowledge the help of Sarah Melville and Vanessa Leech (*T.brucei* Genome Project, Cambridge, UK) with high-density filters and P1 clones, as well as the advice of Jürgen Brosius in cDNA cloning of small RNAs. We thank Pradeep Patnaik for plasmid constructs, Zsofia Palfi for essential help with U5 snRNP affinity selections and for reading the manuscript, and Shula Michaeli (Weizmann Institute, Israel) for discussions. This investigation received financial support from the Deutsche Forschungsgemeinschaft and the German–Israeli Foundation.

REFERENCES

1. Nilsen,T.W. (1993) *Annu. Rev. Microbiol.*, **47**, 413–440.
2. Nilsen,T.W. (1995) *Mol. Biochem. Parasitol.*, **73**, 1–6.
3. Ullu,E., Tschudi,C. and Günzl,A. (1996) In Smith,D.F. and Parson,M. (eds), *Molecular Biology of Parasitic Protozoa*. IRL Press, Oxford, pp. 115–133.
4. Cross,M., Günzl,A., Palfi,Z. and Bindereif,A. (1991) *Mol. Cell. Biol.*, **11**, 5516–5526.
5. Lücke,S., Klöckner,T., Palfi,Z., Boshart,M. and Bindereif,A. (1997) *EMBO J.*, **16**, 4433–4440.
6. Palfi,Z., Günzl,A., Cross,M. and Bindereif,A. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 9097–9101.
7. Palfi,Z. and Bindereif,A. (1992) *J. Biol. Chem.*, **262**, 20159–20163.
8. Goncharov,I., Palfi,Z., Bindereif,A. and Michaeli,S. (1999) *J. Biol. Chem.*, **274**, 12217–12221.
9. Dungan,J.M., Watkins,K.P. and Agabian,N. (1996) *EMBO J.*, **15**, 4016–4029.
10. Xu,Y.-X., Ben-Shlomo,H. and Michaeli,S. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 8473–8478.
11. Beggs,J.D., Teigelkamp,S. and Newman,A.J. (1995) *J. Cell Sci.*, **19** (Suppl.), 101–105.
12. Newman,A.J. (1997) *EMBO J.*, **16**, 5797–5800.
13. Patnaik,P.K., Kulkarni,S.K. and Cross,G.A. (1993) *EMBO J.*, **12**, 2529–2538.
14. Yon,J. and Fried,M. (1989) *Nucleic Acids Res.*, **17**, 4895.
15. Lücke,S., Xu,G.-L., Palfi,Z., Cross,M., Bellofatto,V. and Bindereif,A. (1996) *EMBO J.*, **15**, 4380–4391.
16. Field,H. and Field,M.C. (1996) *Exp. Parasitol.*, **83**, 155–158.
17. Lamm,G.M., Blencowe,B.J., Sproat,B.S., Iribarren,A.M., Ryder,U. and Lamond,A.I. (1991) *Nucleic Acids Res.*, **19**, 3193–3198.
18. O'Brien,C.A. and Wolin,S.L. (1994) *Genes Dev.*, **8**, 2891–2903.
19. Guthrie,C. and Patterson,B. (1988) *Annu. Rev. Genet.*, **22**, 387–419.
20. Nakaar,V., Dare,A.O., Hong,D., Ullu,E. and Tschudi,C. (1994) *Mol. Cell. Biol.*, **14**, 6736–6742.
21. Xu,G.-L., Wieland,B. and Bindereif,A. (1994) *Mol. Cell. Biol.*, **14**, 4565–4570.
22. Nakaar,V., Tschudi,C. and Ullu,E. (1995) *Parasitol. Today*, **11**, 225–228.
23. Jones Haltiner,M. and Guthrie,C. (1990) *EMBO J.*, **9**, 2555–2561.
24. Branlant,C., Krol,A., Ebel,J.P., Lazar,E., Haendler,B. and Jacob,M. (1982) *EMBO J.*, **1**, 1259–1265.
25. Ullu,E. and Tschudi,C. (1993) *J. Biol. Chem.*, **268**, 13068–13073.
26. Raker,V.A., Plessel,G. and Lührmann,R. (1996) *EMBO J.*, **15**, 2256–2269.
27. Bruzik,J.P., Van Doren,K., Hirsh,D. and Steitz,J.A. (1988) *Nature*, **335**, 559–562.
28. Mattaj,I.W. (1986) *Cell*, **46**, 905–911.
29. Fischer,U., Sumpter,V., Sekine,M., Satoh,T. and Lührmann,R. (1993) *EMBO J.*, **12**, 573–583.
30. Fischer,U. and Lührmann,R. (1990) *Science*, **249**, 786–790.
31. Mottram,J., Perry,K.L., Lizardi,P.M., Lührmann,R., Agabian,N. and Nelson,R.G. (1989) *Mol. Cell. Biol.*, **9**, 1212–1223.
32. Bangs,J.D., Crain,P.F., Hashizume,T., McCloskey,J.A. and Boothroyd,J.C. (1992) *J. Biol. Chem.*, **267**, 9805–9815.
33. Günzl,A., Cross,M. and Bindereif,A. (1992) *Mol. Cell. Biol.*, **12**, 468–479.
34. Michaeli,S., Roberts,T.G., Watkins,K.P. and Agabian,N. (1990) *J. Biol. Chem.*, **265**, 10582–10588.
35. Palfi,Z., Xu,G.-L. and Bindereif,A. (1994) *J. Biol. Chem.*, **269**, 30620–30625.
36. Roberts,T.G., Sturm,N.R., Yee,B.K., Yu,M.C., Hartshorne,T., Agabian,N. and Campbell,D.A. (1998) *Mol. Cell. Biol.*, **18**, 4409–4417.