

Characterization of a novel trypanosomatid small nucleolar RNA

Alexander Levitan, Yu-xin Xu, Claudia Ben-Dov, Herzel Ben-Shlomo, Yafei Zhang and Shulamit Michaeli*

Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

Received November 10, 1997; Revised and Accepted February 18, 1998

DDBJ/EMBL/GenBank accession no. AF050095

ABSTRACT

Trypanosomes possess unique RNA processing mechanisms including *trans*-splicing of pre-mRNA and RNA editing of mitochondrial transcripts. The previous finding of a trimethylguanosine (TMG) capped U3 homologue in trypanosomes suggests that rRNA processing may be related to the processing in other eukaryotes. In this study, we describe the first trypanosomatid snoRNA that belongs to the snoRNAs that were shown to guide ribose methylation of rRNA. The RNA, identified in the monogenetic trypanosomatid *Leptomonas collosoma*, was termed snoRNA-2 and is encoded by a multi-copy gene. SnoRNA-2 is 85 nt long, it lacks a 5' cap and possesses the C and D boxes characteristic to all snoRNAs that bind fibrillarin. Computer analysis indicates a potential for base-pairing between snoRNA-2 and 5.8S rRNA, and 18S rRNA. The putative interaction domains obey the rules suggested for the interaction of guide snoRNA with its rRNA target for directing ribose methylation on the rRNA. However, mapping the methylated sites on the 5.8S rRNA and 18S rRNA indicates that the expected site on the 5.8S is methylated, whereas the site on the 18S is not. The proposed interaction with 5.8S rRNA is further supported by the presence of psoralen cross-link sites on snoRNA-2. GenBank search suggests that snoRNA-2 is not related to any published snoRNAs. Because of the early divergence of the Trypanosomatidae from the eukaryotic lineage, the presence of a methylating snoRNA that is encoded by a multi-copy gene suggests that methylating snoRNAs may have evolved in evolution from self-transcribed genes.

INTRODUCTION

In eukaryotes, ribosomal RNA (rRNA) processing takes place in the nucleolus from a large rRNA precursor to generate the mature 5.8S, 18S and 28S rRNAs. The cleavages that generate the mature rRNAs occur at the 5' external transcribed spacer (5' ETS) and two internal transcribed spacer sequences (ITS1 and ITS2) (1). A number of small nucleolar RNAs (snoRNAs) were shown to be

essential for the rRNA processing events (2). The most abundant of these is the U3 that functions early in the process close to the 5' end of the transcript and its binding commits the rRNA transcript for processing (3,4). U8, the second most abundant snoRNA, is required for cleavage upstream of the 5.8S and at the 3' end of 28S rRNA (5). U22 is essential for processing at both the 5' and 3' ends of 18S rRNA (6), whereas U14 is implicated in the maturation of 18S rRNA (7) and RNase MRP (a relative of the RNaseP ribozyme) is required for correct processing of the 5.8S sequences (8). Common to snoRNAs is the presence of short motifs, called boxes C and D, that bind an abundant nucleolar protein, fibrillarin (2).

In mammalian cells, except from U3, U8 and U13, many snoRNAs are processed from introns of pre-mRNA and therefore possess 5' monophosphate. The C and D boxes are essential for direct or indirect binding of fibrillarin (9) and for the accumulation of snoRNAs. Together with the bound proteins the C and D boxes form part of the processing signal necessary for maturation of these snoRNAs (10). One characteristic feature of these intronic snoRNAs is that the majority contain long stretches that are perfectly complementary to the highly conserved regions of either 18S or 28S rRNAs (11). Direct interaction between the snoRNA and the proposed site was demonstrated for the yeast U14 (12).

Recently it was shown that anti-sense snoRNAs function as guide RNAs in directing site-specific ribose methylation of pre-rRNA (13). The site of methylation is dictated by the distance from the D box and is always the fifth nucleotide (nt) downstream from the position matched to the D box (13). Altering the distance between D box and the snoRNA-rRNA duplex shifted the methylation site and mutating the sequence of the snoRNA in the region involved in duplex formation affected methylation (13,14). The importance of pre-rRNA methylation to ribosome biogenesis is unclear. However, the fact that 2'-O-methylated nucleotides are located in conserved core structures of rRNA, may suggest that the methyl groups contribute to the recognition or folding essential for the function of rRNA (15–17). More recently, snoRNAs carrying the conserved H and ACA boxes were shown to direct site specific synthesis of pseudouridine (ψ) on rRNAs. As for the anti-sense snoRNAs carrying the C/D boxes, strict rules determine the site of modification. The modification site is dictated by base-pairing of the snoRNA with

*To whom correspondence should be addressed. Tel: +972 8 9343626; Fax: +972 8 9468256; Email: bfshula@weizmann.weizmann.ac.il

the complementary rRNA elements flanking the position of the Ψ and by the conserved distance between the 3' ACA box and the Ψ (17–19).

Trypanosomatids are protozoan parasites that diverged early in the evolution of eukaryotes (20) and possess unique RNA processing mechanisms, such as *trans*-splicing (21). Like *cis*-splicing, *trans*-splicing requires the participation of small nuclear RNAs (snRNAs) (22). With no exceptions all the Trypanosomatidae U snRNAs are shorter than their other eukaryotic counterparts and none binds the Sm antigens that bind snRNPs from yeast to man (23,24).

Very little is known about rRNA processing in trypanosomes. The identification of RNA B as the U3 homologue, was the first evidence that trypanosomes possess snoRNAs (25). However, the trypanosome machinery may involve additional snoRNAs as their rRNA processing pathway includes extra steps needed for the generation of small rRNA that are derived from the 28S rRNA (26,27). Studies performed in *Trypanosoma brucei* also indicate the existence of a 35 kDa fibrillar protein (25). Using this anti-fibrillar antibody, several additional snoRNAs ranging in size from 70 to 150 nt were identified, but were not further analyzed (25).

In this study, we cloned and sequenced a novel snoRNA that belongs to the family of anti-sense snoRNAs. Twelve contiguous nucleotides of snoRNA-2 are complementary to the trypanosomatid 18S rRNA near the 5' end, and 10 nt are complementary to the 5.8S rRNA. In addition, snoRNA-2 obeys the rules of a guide-methylating snoRNA, as the bases complementary to the rRNAs lie upstream to either the D or D' boxes. Whereas the expected site on the 5.8S rRNA is methylated, the site on the 18S is not. A single mismatch is located 1 nt downstream from the expected methylated site on the 18S rRNA, which may prevent methylation on the expected site. Using *in vitro* psoralen-induced cross-linking, we demonstrated that the crosslinked sites on snoRNA-2 are consistent with its proposed function in guide-methylating 5.8S rRNA.

MATERIALS AND METHODS

Primers

[1] 16865-5'-ATCAGATGCCGGTAGTC-3' complementary to nt 72–86 of snoRNA-2. [2] 17224-5'-CAAATGCAGTGATTCT-3' complementary to nt 61–77 18S rRNA. [3] 19796-5'-CGCCGT-TTGGTTCA-3' complementary to nt 101–115 of the 5.8S rRNA. [4] 5'-GATCCTTCTGGAAC-3' complementary to nt 129–143 of the *Leptomonas collosoma* U3 RNA. [5] 5'-CGGGATC-CGCGTCGACGTGCAGTTGT-3' sense oligonucleotide carrying information from –47 to –64 carrying a *Bam*HI site at the 5' end. [6] 5'-CGGGATCCACAACAACAGGTGCT-3' anti-sense oligonucleotide complementary to positions 84–97 carrying a *Bam*HI site at the 5' end. [7] 5'-CGGGATCCAGCCGGAGCTTGCTC-3' sense to positions 1–16 of the 7SL RNA carrying a *Bam*HI site (28).

Extract preparation

Leptomonas collosoma growth and extract preparation was as previously described (29). Post-ribosomal supernatant (PRS) preparation and DEAE-chromatography were as previously described (24).

Cloning and sequencing

The source of snoRNA-2 was a DEAE flow-through fraction. RNPs enriched in the flow-through fraction were sedimentated on sucrose gradients and the RNA obtained from 10–20S particles was extracted and separated on a preparative denaturing gel. RNA of the 85 nt region was end-labeled at the 5' end after removing the phosphate termini with alkaline phosphatase (30). Labeling was with [γ -³²P]ATP (3000 Ci/mmol) and polynucleotide kinase. The labeled RNA (200 000 c.p.m.) was used to screen a phage genomic library (29,31). Eight independent clones were isolated and a *Sau*3A I 700 nt fragment that hybridized with the RNA was subcloned to pBluescript SK plasmid. The gene was sequenced using the Sanger chain termination method using T3 and T7 primers. The plasmid carrying the *Sau*3A I 700 nt fragment was designated pC-1. Plasmid pC-2, carrying the coding region for snoRNA-2 was constructed for the synthesis of anti-sense snoRNA-2 by PCR using oligonucleotides 5 and 6.

Determination of 2'-O-methylated nucleotides

For partial alkaline hydrolysis, 10 μ g of total RNA was resuspended in 10 mM NaOH and 0.2 mM EDTA, and the samples were boiled for 30 s, 1, 2, 5, 10 and 30 min. The RNA samples were mixed and precipitated with ethanol. One tenth of the RNA sample was subjected to primer extension as previously described (32).

Non-denaturing sucrose gradients

Leptomonas collosoma extracts were treated with 2% SDS and 2 mg/ml proteinase K and incubated at 14°C for 15 min. The lysate was cleared at 50 000 g for 30 min, and RNA was sedimented for 16 h at 22 000 r.p.m. in SW41 rotor as was previously described (25,33). To examine the distribution of snoRNA after heat treatment, the deproteinized extracts were incubated at 65°C for 3 min and then chilled on ice before separating on sucrose gradients. RNA was extracted from the gradient fractions and subjected to northern analysis with oligonucleotide complementary to snoRNA-2 (oligo 1).

Tagging the RNPs with anti-sense oligonucleotide

Aliquots from sucrose gradients were mixed with 50 000 c.p.m. (0.5–1 pmol) of end-labeled oligonucleotide in binding buffer containing 20 mM HEPES–KOH (pH 7.9), 2 mM MgCl₂, 5 mM β -mercaptoethanol and incubated for 1 h at 30°C. Glycerol and bromophenol were added to the samples to a final concentration of 10 and 0.005%, respectively, and the samples were fractionated on a non-denaturing gel. The gels used in this study were 5% acrylamide:bis acrylamide (30:0.8) buffered with 50 mM Tris–glycine (50 mM Trizma base, 50 mM glycine pH 8.8). To examine the specificity of interaction between the labeled oligo and the RNP, competition was performed in the presence of 20 pmol of unlabeled specific or non-specific oligonucleotides under the same binding conditions as described above.

UV cross-linking experiments

Whole-cell extracts were prepared as described above. Proteinase K (2 mg/ml), 2% SDS were added to the extracts, and the samples were incubated at 18°C for 20 min. The lysate was cleared by centrifugation and AMT–psoralen was added to the deproteinized

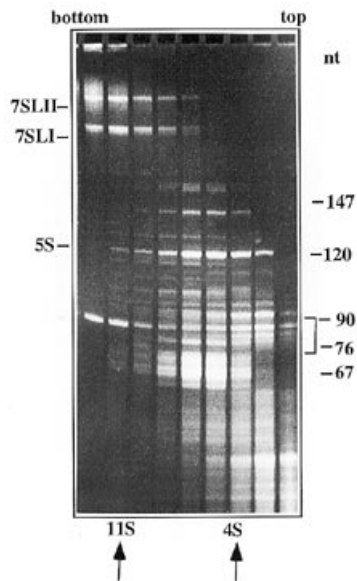


Figure 1. RNA profile of RNPs enriched by column chromatography and sucrose gradient fractionation. Whole cell extracts were prepared from 5×10^{10} cells and the supernatant obtained after centrifugation at 150 000 g (S150) was further fractionated on a DEAE-Sephacel column as previously described (24). The column flow-through fraction was further concentrated and layered on a continuous 10–30% (w/v) sucrose gradient. Gradients were centrifuged at 4°C for 22 h at 35 000 r.p.m. in a Beckman SW41 rotor. The RNA profile derived from deproteinized fractions that was separated on a 10% denaturing gel and stained with EtBr is presented. S value markers were tRNA (4S) and the enzyme catalase (11S). The identity of RNAs is indicated. Marker was pBR322 *Hpa*II digest.

extract at 100 µg/ml. After 1 h incubation on ice, half of the extract was irradiated at 4°C with a UV light for 30 min, and the remainder was kept on ice in the dark as was previously described (34). Irradiation was performed with UV light source at 365 nm with an intensity of 40 mW/cm². Total RNAs extracted from the irradiated and non-irradiated samples were resuspended in buffer containing 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 2 mM EDTA and 0.5% SDS. The samples were heat-denatured for 2 min at 100°C and loaded on a 10–30% linear sucrose gradient. The gradients were centrifuged at 4°C for 20 h at 22 000 r.p.m. in a SW41 rotor. RNA was extracted from the sucrose gradient fractions, separated on a 4% denaturing gel and subjected to northern analysis with an anti-snoRNA-2 RNA probe.

RESULTS

Identification of a novel ~85 nt RNA that is encoded by a multi-copy gene

Analysis of *L.collosoma* RNA enriched in extracts depleted of ribosomes indicates that the abundant small RNAs have already been identified, including the 7SL RNA (28), U3 homologue (25), U2, U4, U6 (23) and SL RNA (29). In addition, a group of abundant small RNAs ranging in size from 70 to 95 nt have also been observed, but, because of their co-migration with tRNAs, they have not yet been identified. Fractionation of extracts on DEAE-Sephacel column revealed the presence of distinct RNAs that did not bind to the column and were found in the flow-through fraction. Further separation of this column fraction on sucrose gradients indicated that the most abundant RNAs are enriched in the 11S peak which includes the 7SL RNA and sRNA-85. sRNA-85 was recently

identified as a tRNA-like molecule, which we propose may be part of the trypanosomatid signal recognition particle (28). Nevertheless, additional, as yet unidentified, small RNAs, ranging in size from 76 to 90 nt were also observed (Fig. 1).

We began to systematically analyze the small RNAs enriched in the 11S peak by direct RNA sequencing and gene cloning. RNAs in the size range of 76–90 nt were excised from the gel and labeled at the 5' end after removing the phosphate termini. Labeled RNA in the size of 85 nt was used to screen a genomic library. Eight identical positive clones were isolated, each of which carried an ~700 nt *Sau*3A I fragment that was further subcloned and sequenced. The sequence of the 678 nt *Sau*3A I fragment is presented in Figure 2A. Further mapping of the plasmid indicated that the RNA is encoded within a 290 bp *Acc*I-*Bst*BI fragment. Inspection of the DNA sequence of this fragment for consensus binding domains indicated the presence of C, D and D' boxes in the sequence that are characteristic to snoRNAs (2) (Fig. 2A). To examine whether the sequence containing the boxes encodes for a small RNA, anti-sense oligonucleotide to the proposed coding region (oligo 1) was used in northern analysis and primer extension experiments. The results presented in Figure 2B indicate that the size of the RNA is ~85 nt. The exact +1 position of the RNA was determined by primer extension (Fig. 2C) and is indicated in Figure 3A. The ability to label the RNA after removing the phosphate termini suggests that the RNA is not capped. A GenBank search, using the entire sequence presented in Figure 2 or only the coding region of the RNA, did not reveal any significant homology to known genes.

The genomic organization of this locus was examined by Southern analysis. The results presented in Figure 3 indicate that the gene is multimeric, since partial digestion of genomic DNA with *Sau*3A I produces a ladder of fragments differing in 700 nt (Fig. 3B), representing multimers of one repeat unit. The data indicate that the gene is repeated at least five times. Minor hybridization fragments were also observed (Fig. 3A), suggesting the presence of sequences in the repeat unit that may exist elsewhere in the genome. A similar genomic organization was reported for the SL RNA gene that was shown to be located both in cluster(s) of repeat units and in unclustered loci which are dispersed from the cluster(s) (35).

The 85 nt RNA is an snoRNA

To examine whether this RNA exists in a ribonucleoprotein complex, whole cell extracts were fractionated on sucrose gradients and the fractions were examined by northern analysis with anti-sense oligonucleotide. The results (Fig. 4A) indicate that this novel RNA, like U3 snoRNA, was found at the top of the gradient as well as associated with high S value particles. Interestingly, tagging the RNPs with anti-sense oligonucleotide (oligo 1) indicates that the RNA was found in two distinct particles (designated I and II in Fig. 4B). The specificity of the interaction with the particle, was examined by competition using the same unlabeled oligonucleotide and a non-specific oligonucleotide as a control (Fig. 4C). The results indicated that only the unlabeled specific oligonucleotide competed with the labeled oligo for binding to the RNP (Fig. 4C, lane b).

To examine whether the novel RNA is a snoRNA, the interaction of the small RNA with rRNA was examined. Most of the snoRNAs interact with rRNA through hydrogen bonding. It

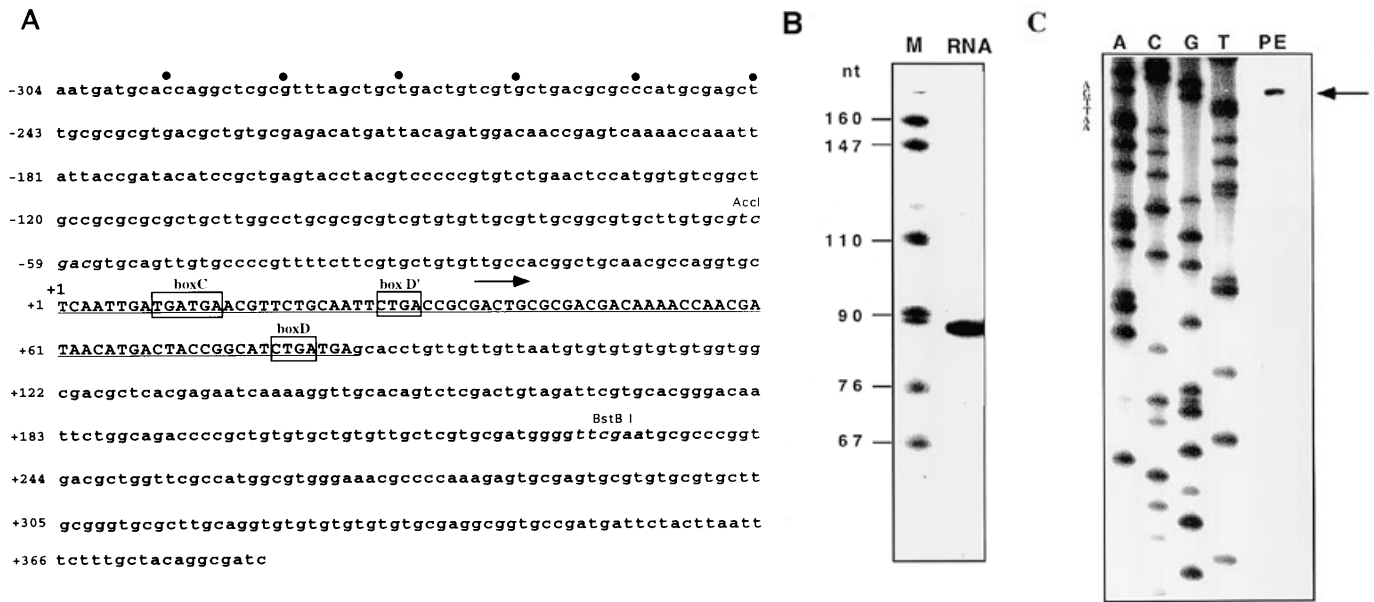


Figure 2. Sequence analysis of the snoRNA-2 locus. (A) The sequence of 678 bp repeat is presented. The coding region of the snoRNA-2 is indicated by upper case lettering and the sequence of snoRNA-2 is underlined. The boxes C, D' and D are indicated. (B) Northern analysis of snoRNA-2. Total *L.collosoma* RNA (10 µg) was separated on a denaturing gel as described in Figure 1. The RNA was subjected to northern analysis with oligo 1. Marker was pBR322 *Hpa*II digest. (C) Primer extension of snoRNA-2. Primer extension analysis was performed using end-labeled oligonucleotide oligo 1. The products of sequencing reaction of the snoRNA-2 clone (pC-1) were used as a reference. The sequence of the cDNA is indicated and the +1 position is marked with an arrow.

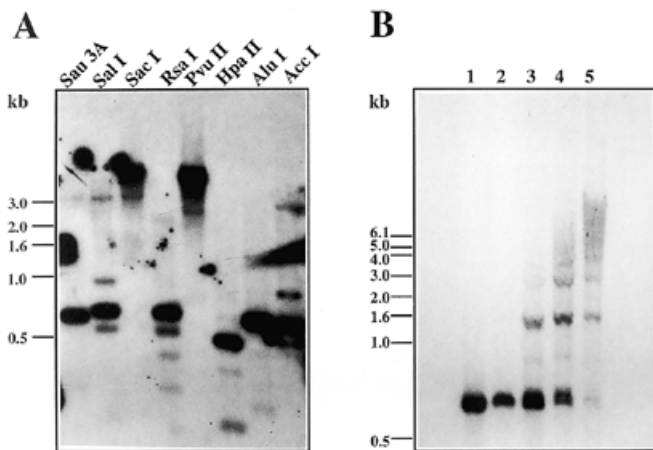


Figure 3. (A) Genomic organization of the snoRNA-2 gene. Southern analysis using different restriction endonucleases. DNA (10 µg) was digested with different restriction enzymes as indicated. After transferring, the membrane was hybridized with the random-primed labeled snoRNA clone carrying the 680 nt repeat unit (pC-1). λ Phage DNA digested with *Hind*III was used as a marker and the sizes of the fragments are indicated. (B) Southern analysis of DNA partially digested with *Sau*3A I. DNA (10 µg) was digested with 4 U of the enzyme. At various time points (lane 1, 1 h; lane 2, 30 min; lane 3, 15 min; lane 4, 5 min; lane 5, 30 s) aliquots were removed and the reaction was terminated by addition of 50 mM EDTA.

was previously demonstrated that yeast snoRNA-rRNA hybrids can be enriched after gently deproteinizing and fractionation on sucrose gradients (25,33).

Leptomonas collosoma RNA was prepared from deproteinized extracts and fractionated on sucrose gradient alongside an RNA sample that was heated to 65°C for 30 min as was previously described (25,33). Northern analysis of the sucrose gradient

fractions from an unheated RNA sample is presented in Figure 5A. The results indicate that the novel RNA, like U3, was found in two peaks: one corresponding to free RNA (fraction 6) and the second that co-migrated with high S value RNAs (fraction 16). Heat treatment of the RNA sample resulted in the dissociation of snoRNA-2 from the high S value RNA and its sedimentation in the region of low S value RNAs molecules. Fractionating the RNA on a 1% formaldehyde gel indicated the presence of 28S α and β in fractions 14–16, whereas the 18S was found mostly in fractions 8–10 (data not shown) as previously observed (25). The new snoRNA was termed snoRNA-2, since the first identified *T.brucei* snoRNA is U3 (25). The different location of snoRNA-2 in the gradient compared with U3 may suggest that either these RNAs do not interact with the same portion of the rRNA precursor or interact with the rRNA precursor at different stages of the processing.

snoRNA-2 is a guide RNA that may direct methylation on 5.8S but not on 18S rRNAs

As a first step towards elucidating the function of snoRNA-2, we examined whether snoRNA-2 belongs to the recently described snoRNAs that serve as guide RNAs for directing site-specific ribose methylation (13–16). A computer search to identify rRNA sequences that may interact with the snoRNA by base-pairing was performed with 18S and 28S rRNA sequences from all members of the Trypanosomatidae. An extensive putative base-pair interaction was revealed between snoRNA-2 and the 5' end of 18S rRNA, as shown in Figure 6A. Interestingly, the base-pair interaction obeys the rules recently suggested for the guide-methylating snoRNAs. snoRNA-2 possesses 12 nt that are complementary to a universal core region of the mature 18S rRNA that is 100% identical in all Trypanosomatidae. In fact, the mammalian snoRNA U27 interacts with sequences that partially overlap with the snoRNA-2-18S

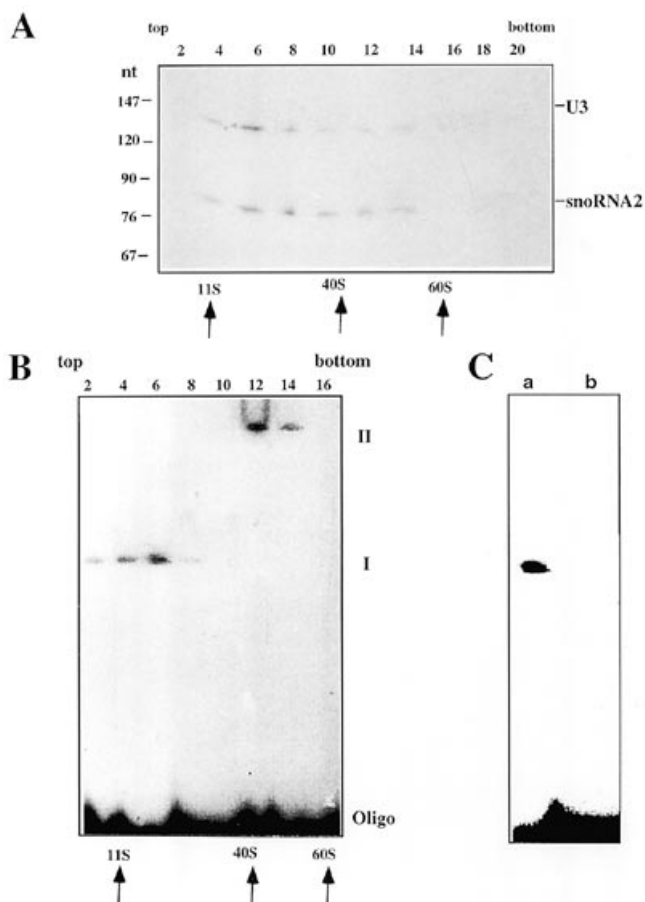


Figure 4. Fractionation of RNPs on sucrose gradients. Low-salt whole-cell extracts were layered on a continuous 10–30% (w/v) sucrose gradient in buffer A containing 100 mM KCl. Gradients were centrifuged for 3 h at 35 000 r.p.m. in a Beckman SW41 rotor. S values were determined using standard 40S and 60S ribosomes from HeLa cell S100 extract and the enzyme catalase (11S). (A) RNA was extracted from the gradient fractions, separated on a 6% denaturing gel and subjected to northern analysis with oligos 1 and 4 anti-sense to snoRNA-2 and U3, respectively. The marker was pBR322 *Hpa*II digest. (B) Tagging of the RNPs with oligo 1. RNPs from the sucrose gradient fractions were incubated with end-labeled oligo 1 and the RNPs were separated on a native gel. The positions of the two RNP complexes I and II are indicated. (C) RNPs from fraction 2 was incubated with end-labeled oligo 1 (lane a) in the presence of 20 pmol of oligo 7 (lane b) in the presence of 20 pmol of oligo 1. The RNPs were separated as in (B).

rRNA interaction domain (13). In all snoRNAs that serve as guide-methylating snoRNAs, the nts that are complementary to the rRNA sequences are located immediately or 1 nt upstream from the D box. The snoRNA-2–18S rRNA interaction presented in Figure 6A obeys these rules. The study on the guide-methylating snoRNAs also suggests that the fifth nucleotide upstream from the D box is always facing a methylated nucleotide in the rRNA sequence. If these rules are applicable to trypanosomatids and if snoRNA-2 is a guide-methylating snoRNA, we would expect that the C at position 18 indicated in Figure 6A would be methylated. To examine whether the C is indeed methylated, the location of the ribose-methylated sites near the proposed interaction domain was determined. Primer extension (using oligo 2) was performed on partially alkali hydrolyzed RNA. The presence of a 2'-O-methyl-nucleotide confers resistance to alkali and therefore a gap is created in the ladder where a methylated nucleotide exists. The

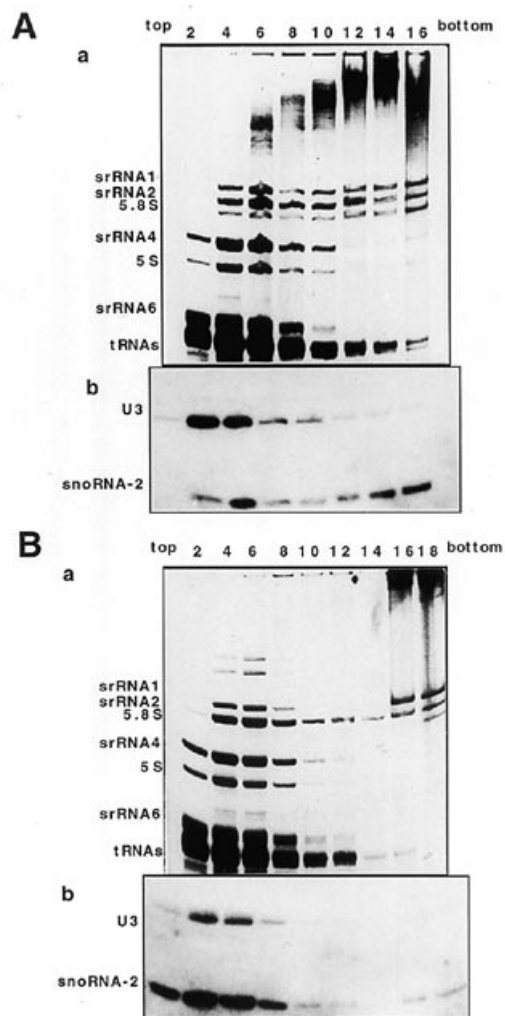


Figure 5. Distribution of deproteinized RNAs on a non-denaturing sucrose gradient. Total RNA was deproteinized under non-denaturing conditions and the RNA was either layered directly on a 10–30% (w/v) sucrose gradient or treated at 65°C for 10 min before loading on the sucrose gradient. Centrifugation was performed at 4°C for 16 h at 22 000 r.p.m. in a Beckman SW41 rotor. RNA isolated from gradient fractions was separated on a 10% denaturing gel, stained with EtBr and subjected to northern analysis with oligo 1 and 4 anti-sense to snoRNA-2 and U3, respectively. (A) Deproteinized RNA was treated at 14°C. (a) Profile of the RNA across the gradients. (b) Northern analysis with oligo 1 and anti-U3 oligo. (B) Deproteinized RNA was treated at 65°C. (a) and (b) are the same as in (A). The identity of the RNAs is indicated.

results presented in Figure 6B suggest that the expected site is not methylated. Interestingly, the site that is methylated in the human 18S rRNA (Fig. 6B, open arrow) is not methylated in the trypanosome 18S rRNA. However, a different site in position +40 (Fig. 6B, closed arrow) is methylated. This information can be further used to predict the sequence of the snoRNA that may govern the methylation at this site. As can be seen only a single mismatch is present in the putative interaction domain, exactly 1 nt downstream from the expected methylated nucleotide. This mismatch is not due to sequence mistakes in either the snoRNA-2 or the 18S rRNA. As mentioned above, the 18S rRNA sequence in this region is 100% identical in all the Trypanosomatidae.

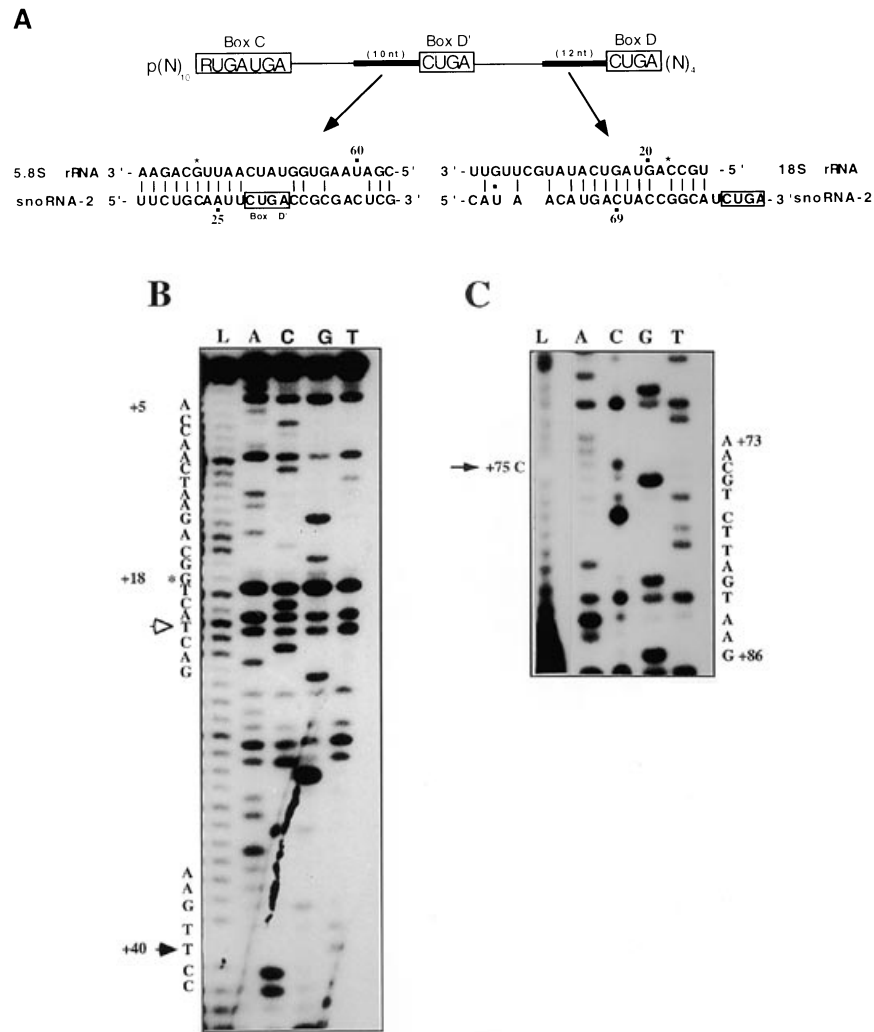


Figure 6. Potential for base-pair interaction between snoRNA-2 and 5.8S, and 18S rRNAs, and the mapping of ribose methylated nucleotides on the rRNA target sites. (A) Schematic presentation of the snoRNA-2 and its potential to interact with rRNA sequences. Regions that are complementary to rRNA are indicated by thick lines. The positions of the predicted methylated nucleotides are indicated with asterisks. (B) Mapping of the methylated nucleotides on the 18S rRNA by primer extension. RNA (10 µg) was subjected to partial alkaline hydrolysis. End labeled oligonucleotide (oligo 2) was used in primer extension of the partially hydrolyzed RNA. The primer extension products were separated on a 6% denaturing gel next to the RNA sequencing ladder generated by primer extension sequencing using the same primer. The position of the expected methylated nucleotide based on the putative interaction with 18S rRNA is indicated with an asterisk. The position of the methylated site on human RNA is indicated with an open arrow and methylated nucleotide on the *L.collsoma* RNA is indicated with a black arrow. The sequence of the cDNA is indicated. (C) As in (B) but with anti-sense oligonucleotide to the 5.8S rRNA. The position of the methylated nucleotide is indicated.

The other predicted site for interaction was found between 5.8S rRNA and snoRNA-2. Ten contiguous nucleotides complementary to 5.8S rRNA exactly upstream to the D' box were observed (Fig. 6A). This interaction also obeys the rules suggested for the guide-methylating snoRNAs (13–16). To examine whether 5.8S rRNA is methylated at this position, the same experiment as presented in Figure 6B was performed but with an anti-sense oligonucleotide to 5.8S rRNA (oligo 3). The results presented in Figure 6C indicate that the predicted site is indeed methylated since a gap was observed that corresponds to methylation at position G75.

To map the site of interaction of snoRNA-2 with pre-rRNA, we used the AMT-psoralen cross-linking approach (34). Studies performed on yeast U3-rRNA interactions using cross-linking indicate that *in vivo* or *in vitro* cross-linking gave essentially the

same results but *in vitro* reactions were more efficient (34). For this reason we performed the experiments *in vitro*. Extracts were gently deproteinized and incubated with psoralen, and irradiated with UV light. To distinguish between intramolecular and intermolecular cross-linking (34), the RNA was boiled, fractionated on 10–30% sucrose gradients and further analyzed on a 4% denaturing polyacrylamide gel. As a control, non-irradiated RNA was prepared and fractionated in parallel with the cross-linked RNA. The results are presented in Figure 7. Comparison between the RNA profiles in Figure 7A indicates that the cross-linking was very efficient, as almost all the small ribosomal RNAs, notably sRNA1 as well as 5.8S rRNA, disappeared from their expected position in the gel apparently due to efficient cross-linking to the 28S rRNA. Hybridization of the blots with anti-snoRNA-2 probe

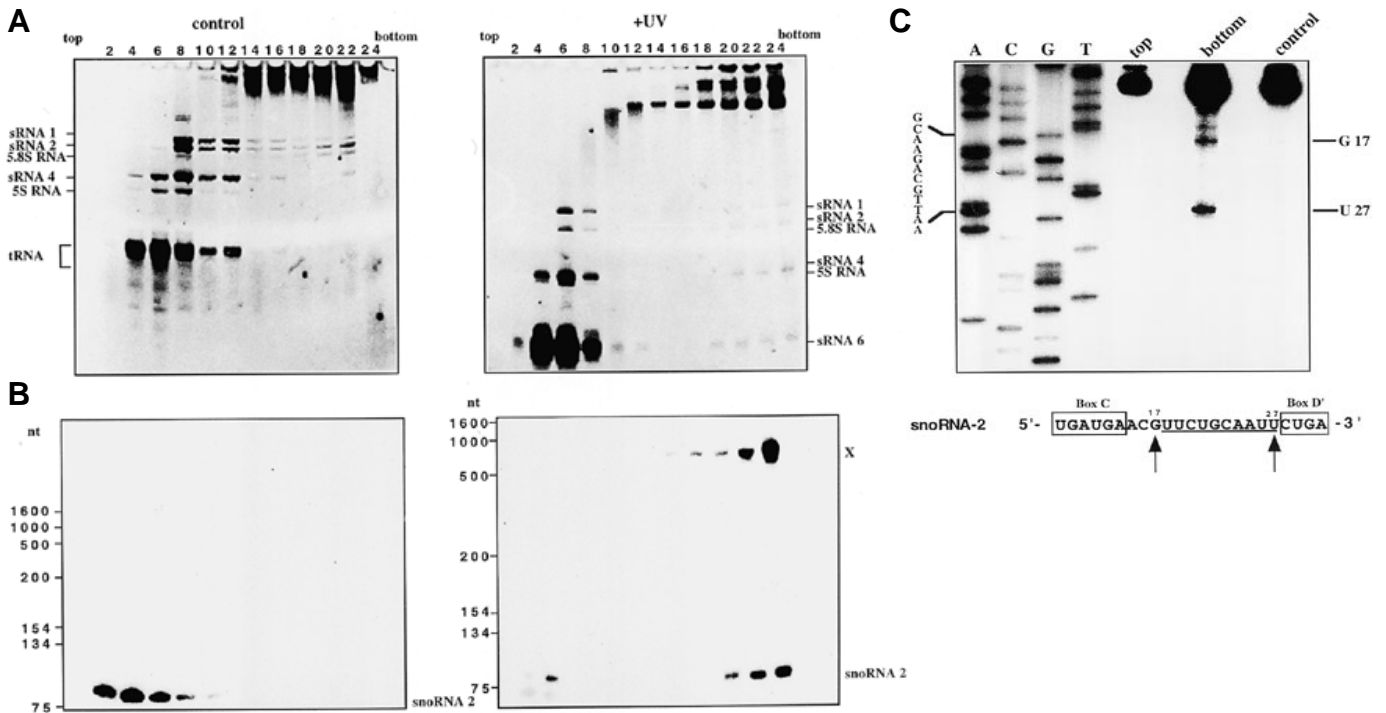


Figure 7. *In vitro* cross-linking of snoRNA-2 with pre-rRNA. Total RNA was extracted from crude deproteinized lysates treated with AMT-psoralen and UV-irradiated, and from the non-irradiated control. The RNA was heat-denatured and run on sucrose gradients. The RNA from sucrose gradient fractions was separated on a 4% denaturing gel. (A) Demonstrates the pattern of the RNA extracted from the sucrose gradients stained with EtBr. The identity of the RNAs is indicated. (B) The gels were subjected to northern analysis with anti-snoRNA-2 RNA probe. The position of snoRNA-2 is indicated and the cross-linked species containing snoRNA-2 is designated X. (C) Location of the cross-linking sites on snoRNA-2 by primer extension. Primer extension reactions were performed on snoRNA-2 from the top fractions (top), from the bottom of the gradient (bottom) and from non-irradiated control. The products of the sequencing reaction of pC-1 were used as a reference. The positions of the major stops are indicated and marked with arrows on the snoRNA-2 sequence. The region proposed for interaction with 5.8S rRNA is underlined.

indicates a band of ~800 nt that hybridized with the snoRNA-2 probe and appeared only after UV irradiation (designated X).

Primer extension was used to further map the cross-linked sites on snoRNA-2. Previous studies have indicated that the reverse transcriptase stops can reach up to and including sometimes the cross-linked nucleotide, but in most cases a single stop is usually observed 1 nt 3' from the cross-linked residue (36). The results presented in Figure 7C indicate the presence of two major stops located in positions U27 and G17, suggesting that the crosslinked nucleotides are U26 and C16. U27 is located within the duplex formed by the 5.8S rRNA and snoRNA-2 (underlined), whereas C16 is located 2 nt upstream to the duplex. The results obtained in this study are in accord with those shown previously for the U3 snoRNA. In this context, the location of cross-linked nucleotides in the U3-ETS region were positioned 1 nt upstream to the RNA duplex, and to the first nucleotide of the duplex (34). The location of the cross-linked site within and immediately outside the proposed duplex between 5.8S rRNA and snoRNA-2 thus supports the proposed base-pair interaction expected for directing methylation on G75 of 5.8S rRNA.

DISCUSSION

In this study we describe the second trypanosomatid snoRNA. snoRNA-2 has the potential to guide ribose methylation on both 5.8S and 18S rRNAs, but may direct methylation only on 5.8S rRNA. snoRNA-2 is encoded by a multi-copy gene and does not possess a 5' cap.

At present, two types of snoRNAs containing C and D boxes that bind fibrillarin have been described. The first group, represented by U3, U8, U14 and U22, participates in rRNA processing (2), whereas the second group, to which most of the recently described snoRNAs belong, functions as guide RNAs for rRNA methylation (13–16). snoRNA-2 belongs to the second group. However, it is premature to exclude the possibility that snoRNA-2 may also be involved in processing. Most of the guide-methylating snoRNAs are encoded by introns of genes and possess impressive complementarity with either 18S or 28S rRNAs (13–16). Most of the intronic snoRNAs are located in parent genes encoding for proteins that are involved in nucleolar function, ribosome structure and protein synthesis (11). With the identification of snoRNA homologues in different eukaryotes it became evident that the same snoRNA can be encoded by different genes, for example, the U14 snoRNA is intron-encoded in mammals (37), but is non-intronic in *Saccharomyces cerevisiae* (38). In yeast, intronic snoRNAs exist, but the majority are encoded by single-copy genes (2). Interestingly, U14 is tandemly arranged in maize (39). This raises questions regarding the evolution of snoRNA genes and their coordinate regulation with respect to other ribosomal components.

In the absence of conventional introns in the Trypanosomatidae genes, it was not surprising to find that snoRNA-2 is encoded by an independent gene. Since Trypanosomatidae diverged very early in the eukaryotic lineage (20), the genomic organization of snoRNA-2 may suggest that the first methylating snoRNA genes

were independently transcribed genes, and that the intronic snoRNAs emerged only later in evolution, starting from yeast.

The finding that snoRNA-2 is encoded by a multi-copy gene places it with two other trypanosomatid small RNAs, the SL (40) and SLA RNAs (41) that are encoded by multimeric genes of 200 and 20 copies, respectively. The other genes coding for small RNAs such as the U2, U4, U6 (23), U5 (42, 43), U3 (25) and 7SL RNA (28) are single-copy genes that are transcribed by RNA polIII (44). All these genes, except U2, carry extragenic regulatory elements (A and B boxes) located in tRNA genes that are oriented divergently with respect to the small RNA genes. Inspecting the sequence upstream to snoRNA-2 both in the same direction of the gene and in the opposite orientation failed to detect such boxes. The data suggest that snoRNA-2 may be similar to SL RNA containing extragenic control elements that are not of the tRNA type (45–47). A possibility also exists that the elements controlling the transcription of the gene are located upstream to the first copy. Transcription studies are required to determine which polymerase transcribes snoRNA-2 and whether the extragenic transcription elements are located in each repeat unit.

The absence of a cap on snoRNA-2 is not surprising, since U3 was shown to be the only *T. brucei* snoRNA that was immunoprecipitated with both anti-TMG and anti-fibrillarin antibodies (25). Most of the snoRNAs containing C and D boxes in other eukaryotes also lack a cap and possess a phosphate at their 5' ends but these snoRNAs are usually processed from intronic RNA.

snoRNAs were reported to exist in two classes of snoRNA–protein complexes, a small size RNP of 10–15S and a larger particle of 60–90S (48). The smaller particles correspond to monomeric particles, whereas the larger complexes are believed to be snoRNAs that are associated with nascent pre-rRNA transcripts which undergo processing (49). The larger complexes observed in this study (complex II) may represent either particles containing snoRNA-2 that are bound to pre-rRNA, or snoRNA-2 associated with other snoRNAs. The finding that some cleavage sites are affected by more than one snoRNA, like in the case of U3, U14 and U22, suggests that such RNAs may exist as part of a multi-snoRNP complex (2). In addition, guide-methylating snoRNAs function in clustered regions of the rRNA and may therefore join the 'processome' as a multi-snoRNA complex. It will be interesting to further explore whether other small RNAs together with snoRNA-2 form complex II.

The ability of snoRNA-2 to fold into a stem–loop structure was examined using the MFOLD program (50), and the results indicate that snoRNA-2 does not exhibit a defined structure apart from the boxes C and D that are brought in close proximity by pairing of the 5' and the 3'-terminal nucleotides. Indeed, the first and last 3 nt of snoRNA-2 are complementary.

snoRNA-2 resembles U3 and U14 in that they all have the potential to interact with two distinct regions on the pre-rRNA. U3 interacts and functions in mediating the cleavage within the 5' ETS and the 5' boundary of 18S rRNA (3) whereas snoRNA-2 has the potential to interact with the 5.8S and 18S RNAs. More recently, a group of anti-sense snoRNAs (U32 and U36) were shown to carry the potential for interaction with both 18S and 28S rRNAs (51).

Despite the extensive complementarity of snoRNA-2 with an appropriate site on 18S rRNA, and despite the fact that snoRNA-2 obeys the structural rules for a guide-methylating snoRNA, the expected site is not methylated. The presence of a mismatch between

the rRNA and the snoRNA-2 one nucleotide downstream from the expected methylated nucleotide may affect the recognition of this site. It was previously reported that one mismatch in the contiguous base-pairing between rRNA and a snoRNA did not interfere with the ability of either U14 or U62 to direct methylation (52,53). However, using tailored U20 snoRNA mutants it was demonstrated that creating a mismatch in the duplex dramatically reduced the level of methylation (14).

The finding that snoRNA-2 does not direct methylation at the expected site does not preclude the possibility that it interacts with 18S rRNA at the site. Further mapping of the cross-linked species of snoRNA-2 with pre-rRNA is needed to examine whether it also interacts with the 18S rRNA. It has been already noticed that in the case of U14 extensive complementarity exists between the snoRNA and the 18S rRNA, but none of the bases in the 18S target site is modified, suggesting that U14 may function in modulating rRNA folding (51). This is also true for U13 that has the potential to interact with two consecutive regions of the 3' terminal domain of 18S that also lack ribose-methylated nucleotides (11).

Because of the early divergence of the Trypanosomatidae from the eukaryotic lineage, studying the Trypanosomatidae repertoire of snoRNAs should contribute to our understanding of how snoRNAs and their coding genes developed in evolution. Our study suggests that it may not be possible to predict the repertoire of the Trypanosomatidae snoRNAs based on the positions of methylated ribose on human rRNAs, because these positions may differ between human and trypanosomes. However, the conserved methylation position on ribose 75 of 5.8S RNA in human and *Xenopus laevis* suggests that homologues to snoRNA-2 may exist in vertebrates but not in *Saccharomyces* that lacks modification on this site (54).

ACKNOWLEDGEMENTS

This work was supported by the MINERVA foundation, Munich/Germany, by the Leo and Julia Forchheimer Center for Molecular Genetics of the Weizmann Institute and by a research grant from the Cemach and Anna Oiserman Research Fund. We wish to thank Ora Asher for her excellent technical assistance.

REFERENCES

- Sollner-Webb, B., Tycowski, K. T. and Steitz, J. A. (1996) In Zimmermann, R. A. and Dahlberg, J. E. (eds), *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis*. CRC Press, Boca Raton, Florida, pp. 469–490.
- Maxwell, E. S. and Fournier, M. J. (1995) *Annu. Rev. Biochem.*, **35**, 897–934.
- Huges, J., M. X. and Ares, M. J. (1991) *EMBO J.*, **10**, 4231–4239.
- Beltrame, M. and Tollervey, D. (1995) *EMBO J.*, **14**, 4350–4356.
- Peculis, B. A. and Steitz, J. A. (1993) *Cell*, **73**, 1233–1245.
- Tycowski, K.T., Shu, M. D. and Steitz, J. A. (1994) *Science*, **266**, 1558–1561.
- Li, H. D., Zagorski, J. and Fournier, M. J. (1990) *Mol. Cell. Biol.*, **10**, 1145–1152.
- Schmitt, M. E. and Clayton, D. A. (1993) *Mol. Cell. Biol.*, **13**, 7935–7941.
- Baserga, S. J., Yang, X. W. and Steitz, J. A. (1991) *EMBO J.*, **10**, 2645–2651.
- Huang, G. M., Jarmolowski, A. J. C. R. and Fournier, M. J. (1992) *Mol. Cell. Biol.*, **12**, 4456–4463.
- Bachelierie, J. P., Michot, B., Nicoloso, M., Balakin, A., Ni, J. and Fournier, M. J. (1995) *Trends Biochem. Sci.*, **20**, 261–264.
- Liang, W. Q. and Fournier, M. J. (1995) *Genes Dev.*, **9**, 2433–2443.
- Kiss-Laszlo, S., Henry, Y., Bachelierie, J. P., Caizergues-Ferrer, M. and Kiss, T. (1996) *Cell*, **85**, 1077–1088.
- Cavaillé, J., Nicoloso, M. and Bachelierie, J. P. (1996) *Nature*, **383**, 732–735.

- 15 Maden, T. (1996) *Nature*, **383**, 675–676.
- 16 Bachellerie, J. P. and Cavallé, J. (1997) *Trends Biochem. Sci.*, **22**, 257–261.
- 17 Smith, C. M. and Steitz, J. A. (1997) *Cell*, **89**, 669–672.
- 18 Ni, J., Tien, A. L. and Fournier, M. J. (1997) *Cell*, **89**, 565–573.
- 19 Ganot, P., Bortolin, M. L. and Kiss, T. (1997) *Cell*, **89**, 799–809.
- 20 Sogin, M. L., Elwood, H. J. and Gunderson, J. H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1383–1387.
- 21 Agabian, N. (1990) *Cell*, **61**, 1157–1160.
- 22 Ullu, E. and Tschudi, C. (1990) *Cell*, **61**, 459–466.
- 23 Mottram, J., Perry, K. L., Lizardi, P. M., Lührmann, R., Agabian, N. and Nelson, R. G. (1989) *Mol. Cell. Biol.*, **9**, 1212–1223.
- 24 Michaeli, S., Roberts, T. G., Watkins, K. P. and Agabian, N. (1990) *J. Biol. Chem.*, **265**, 10582–10588.
- 25 Hartshorne, T. and Agabian, N. (1993) *Mol. Cell. Biol.*, **13**, 144–154.
- 26 White, T. C., Rudenko, G. and Borst, P. (1986) *Nucleic Acids Res.*, **14**, 9471–9489.
- 27 Campbell, D. A., Kubo, K., Clark, G. and Boothroyd, J. C. (1987) *J. Mol. Biol.*, **196**, 113–124.
- 28 Ben-Shlomo, H., Levitan, A., Béjà, O. and Michaeli, S. (1997) *Nucleic Acids Res.*, **25**, 4977–4984.
- 29 Goldring, A., Karchi, M. and Michaeli, S. (1995) *Exp. Parasitol.*, **80**, 333–338.
- 30 Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 31 Goldring, A. and Michaeli, S. (1995) *Gene*, **156**, 139–144.
- 32 Patzelt, E., Perry, K. L. and Agabian, N. (1989) *Mol. Cell. Biol.*, **9**, 4291–4297.
- 33 Tollervey, D. (1987) *EMBO J.*, **6**, 4169–4175.
- 34 Beltrame, M. and Tollervey, D. (1992) *EMBO J.*, **11**, 1531–1542.
- 35 Nelson, R. G., Parsons, M., Barr, P. J., Stuart, K., Selkirk, M. and Agabian, N. (1983) *Cell*, **34**, 901–909.
- 36 Ericson, G. and Wollenzien, P. (1988) *Anal. Biochem.*, **174**, 215–223.
- 37 Liu, J. and Maxwell, E. S. (1990) *Nucleic Acids Res.*, **18**, 6565–6571.
- 38 Balakin, A. G., Lempicki, R. A., Huang, G. M. and Fournier, M. J. (1994) *J. Biol. Chem.*, **269**, 739–746.
- 39 Leader, D. J., Sanders, J. F., Waugh, R., Shaw, P. and Brow, W. S. (1994) *Nucleic Acids Res.*, **22**, 5196–5203.
- 40 Nelson, R. G., Parsons, M., Selkirk, M., Newport, G., Barr, P. and Agabian, N. (1984) *Nature*, **308**, 665–667.
- 41 Roberts, T. G., Dungan, J. M., Watkins, K. P. and Agabian, N. (1996) *Mol. Biochem. Parasitol.*, **83**, 163–174.
- 42 Dungan, J. D., Watkins, K. P. and Agabian, N. (1996) *EMBO J.*, **15**, 4016–4029.
- 43 Xu, Y. X., Ben-Shlomo, H. and Michaeli, S. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 8473–8478.
- 44 Nakaar, V., Dare, A. O., Hong, D., Ullu, E. and Tschudi, C. (1994) *Mol. Cell. Biol.*, **14**, 6736–6742.
- 45 Agami, R., Aly, R., Halman, S. and Shapira, M. (1994) *Nucleic Acids Res.*, **22**, 1959–1965.
- 46 Saito, R. M., Elgort, M. G. and Campbell, D. A. (1994) *EMBO J.*, **13**, 5460–5469.
- 47 Hartree, D. and Bellofatto, V. (1995) *Mol. Biochem. Parasitol.*, **71**, 27–39.
- 48 Epstein, P., Reddy, R. and Busch, R. (1984) *Biochemistry*, **23**, 5421–5425.
- 49 Kass, S. and Sollner-Webb, B. (1990) *Mol. Cell. Biol.*, **10**, 4920–4931.
- 50 Zuker, M., Jaeger, J. A. and Turner, D. H. (1991) *Nucleic Acids Res.*, **19**, 2707–2714.
- 51 Nicoloso, M., Qu, L. H., Michot, B. and Bachellerie, J. P. (1996) *J. Mol. Biol.*, **260**, 178–195.
- 52 Liang, W. Q. and Fournier, M. J. (1995) *Genes Dev.*, **9**, 2433–2443.
- 53 Tycowski, K. T., Smith, C. M., Shu, M. D. and Steitz, J. A. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 14480–14485.
- 54 Maden, B. E. H. (1990) *Progr. Nucleic Acids Res. Mol. Biol.*, **39**, 241–303.