A common core structure for U3 small nucleolar RNAs

Toinette Hartshorne* and Nina Agabian

Intercampus Program in Molecular Parasitology, School of Pharmacy, University of California, San Francisco, CA 94143-1204, USA

Received May 20, 1994; Revised and Accepted July 15, 1994

GenBank accession nos L32919 and L32920

ABSTRACT

U3 nucleolar small RNA (snRNA) is involved in early processing of the primary rRNA transcript. A secondary structure model for the unusually small *Trypanosoma brucei* U3 snRNA was deduced by comparative analysis of U3 snRNA sequences and by chemical modification and enzymatic cleavage of U3 snRNA in deproteinized and ribonucleoprotein (RNP) forms. Comprehensive alignment of U3 snRNAs from vertebrate, plant, fungal and protozoan species clearly delineated conserved and divergent features. The 5' domain of the *T.brucei* U3 snRNA appears to form one small, flexible 5' stem loop structure followed by a long single-stranded region; this model is a variation on 5' domain structures proposed for other U3 snRNAs which do not conform to a single model. The 3' domain of *T.brucei* U3 snRNA contains four single-stranded sequences conserved between U3 snRNAs. Of these, structural probing determined that the configurations of GAU region and box B and C sequences are altered by protein interactions in U3 snRNP. Conspicuously, the 3' domains of trypanosomal U3 snRNAs lack stem loops II and III, indicating that these structures are not required for conserved U3 snRNA functions.

INTRODUCTION

Eukaryotic small subunit (SSU), 5.8S and large subunit (LSU) rRNAs are transcribed by RNA polymerase I in a single large precursor rRNA (pre-rRNA) which is punctuated by external (ETS) and internal transcribed spacer (ITS) sequences (1). Mature rRNA species are processed from the primary transcript by unknown mechanisms which require multiple nucleolar snRNAs and non-ribosomal proteins (1–2). The ubiquitous U3 snRNA, shown to be essential in yeast (4), is the most abundant of the nucleolar snRNAs found in association with the nucleolar antigen, fibrillarin. The participation of U3 snRNA in the early phases of pre-rRNA processing has been demonstrated by several criteria. *In vivo* crosslinking of RNAs in mammalian and yeast cells showed that U3 snRNA is in close contact with the 5' ETS (5–7) where the first cleavage of the mammalian primary rRNA transcript occurs. U3 snRNA-dependent cleavage of the initial

processing site has been reproduced in mouse cell extracts (8); related sequences in the *Xenopus* ETS are cleaved in a U3-dependent manner in *Xenopus* and mammalian cell extracts, implying close evolutionary relatedness of processing complex constituents between these organisms (9). U3 snRNA is also implicated in other processing events. Disruption of U3 snRNA in *Xenopus* oocytes reduced cleavage at the ITS1/5.8S boundary (10), and depletion of U3 snRNA in yeast cells diminished cleavage within the ETS, at the ETS/18S rRNA boundary, and within ITS1, culminating in a deficit of mature 18S rRNA (11).

Little is known as to how U3 snRNA effects pre-rRNA processing. It is essential to understand the structure of U3 snRNAs to accommodate models of RNA–RNA interactions, as well as of RNA–protein interactions within U3 snRNPs. Prototypical secondary structure models for U3 snRNAs predict a two domain structure. The 5' domains of vertebrate U3 snRNAs seem to form a single large stem loop structure (12–13) whereas fungal and plant U3 snRNAs may form two small stem loop structures within the same region (14–16). The 3' domains of all U3 snRNAs approximate a Y-shaped structure consisting of two stem loops and a central stem structure; the *Saccharomyces cerevisiae* U3 snRNA has an additional 3' stem loop structure. Small regions of sequence conservation, most notably boxes A–D, (4,12,15), are found in U3 snRNAs from vertebrate, plant and fungal organisms; sequences related to boxes C and D homologies are also found in other fibrillarin-associated snRNAs (1–3).

T.brucei is an evolutionarily ancient protozoan organism in which there exist multiple variations on the RNA metabolic pathways of higher eukaryotes, including pre-mRNA *trans* splicing and mitochondrial RNA editing (17). Trypanosomal large subunit (LSU) rRNAs are fragmented into six pieces after removal of ITS sequences present within LSU coding sequences (18,19), however, the early steps of the pre-rRNA processing pathway of *T.brucei* appear similar to those of higher eukaryotes (19,20) and are likely to involve conserved U3 snRNA functions. The initial identification of the *T.brucei* small, TMG-capped RNA, RNA B (21), as a U3 snRNA homolog was difficult due to its small size and divergent sequence and relied on its abundance, nucleolar localization, and isolation in association with fibrillarin and pre-rRNAs (20). The U3-specific box A

*To whom correspondence should be addressed

Figure 1. Phylogenetic comparison of trypanosomal U3 snRNA sequences. The *T. brucei* U3 snRNA sequence is shown (21). Residue differences in the *T. cruzi* U3 snRNA sequence are in shaded boxes and differences in the *L. collosoma* U3 snRNA sequence are in plain boxes. Blocks of strong sequence homology with other U3 snRNAs are outlined (also see figure 7). A region of *T. brucei* U3 snRNA which shares sequence complementarity with residues adjacent to the ETS pre-rRNA primary cleavage site (20) is underlined. The *L. collosoma* and *T. cruzi* U3 snRNA gene sequences have been submitted to GenBank and assigned the accession numbers L32919 and L32920, respectively.

sequence was readily identified in the *T. brucei* U3 snRNA as well as two candidates for the box C homolog; box B and D homologs were not obvious. Additional structural information was needed to determine the degree of similarity between *T. brucei* U3 snRNA and other U3 snRNAs.

This manuscript describes the secondary structure of the *T. brucei* U3 snRNA deduced by a combination of phylogenetic analysis of U3 snRNA sequences from other trypanosomatids and from vertebrates, fungi, plants and protozoa, and by chemical modification and ribonuclease cleavage of *T. brucei* U3 snRNA in naked cellular and RNP forms. We report that the structure of *T. brucei* U3 snRNA implies a minimal functional U3 snRNA which shares structural similarities with other known U3 snRNAs yet is distinguished by the comparative absence of helical structures. Conservation of putative protein and pre-rRNA binding domains were observed between *T. brucei* U3 snRNA and other U3 snRNAs.

MATERIALS AND METHODS

Chemical modifications

Procyclic forms of *T. brucei* strain 427 were grown in BSM (22) supplemented with 5% heat-inactivated fetal calf serum (GIBCO-BRL) to late log density of 10^7 cells/ml. For modification of RNAs with dimethyl sulfate (DMS; Fluka), cell pellets were washed twice with TBS (10 mM Tris-HCl, pH 8.0; 150 mM NaCl) and suspended at 3.5×10^8 cells/ml in CMK buffer (80 mM Na-cacodylate, pH 7.2; 0.1 M KCl; 5 mM MgCl₂).

Suspended cells were broken by sonication (20) and chromatin was pelleted at $15,000 \times g$. RNA was purified from one half of the supernatant by extraction with guanidine HCl and phenol/chloroform/isoamyl alcohol (PCA)(23), ethanol precipitated, then suspended in the original volume of CMK buffer. 0.3 ml aliquots (10^8 cell equivalents) of whole cells, sonic extracts, and deproteinized cellular RNA were incubated with 0, 0.5, 1.0, 1.5, and 3.0 μ l DMS for 15 mins at 25°C or 30 mins at 0°C, then reactions were quenched by addition of 150 μ l ice cold DMS stop buffer (1.5 M NaOAc, pH 5.2; 1 M β -mercaptoethanol; 1 M Tris-HCl, pH 7.5; 0.1 mM EDTA); 3.0 μ l of DMS was added to stop control samples following quenching. RNAs were precipitated with ethanol, then purified by guanidine HCl and PCA extraction followed by ethanol precipitation. Precipitates were suspended in 30 μ l diethyl pyrocarbonate (DEPC)-treated H₂O; 3 μ l were used in primer extension assays.

Modification of RNAs by 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide-metho-p-toluene sulfonate (CMCT; Aldrich Chemical Company) was accomplished as for DMS modification with the following exceptions. Cells were suspended in BMK buffer (80 mM K-borate, pH 8.1; 100 mM KCl; 5 mM MgCl₂) to 5×10^8 cells/ml. 200 μ l aliquots of cell sonicate or deproteinized RNA was incubated with 0, 50, 100, 150, or 200 μ l CMCT (42 mg/ml in BMK) in a total volume of 400 μ l BMK. Reactions were stopped with 500 μ l of 0.5 M K-borate, pH 6.1; 200 μ l of CMCT was added to control samples following quenching.

Figure 2. Chemical modification of U3 snRNA in naked and RNP forms. (A) DMS was incubated with deproteinized RNA (RNA), cell sonicates (RNP), and whole cells (WC) at 25°C for 15 mins. Lanes 1–4 contain samples incubated with 0.16, 0.33, 0.5 and 1.0% DMS, respectively. No DMS was added in Lane O and 1.0% DMS was added to stop control samples in Lane +. (B) CMCT was incubated with deproteinized RNA (RNA) or cell sonicates (RNP) at 25°C for 15 mins. Lanes 1–4 contain samples incubated with 5.25, 10.5, 15.75 and 21.0 µg/ml CMCT, respectively. No CMCT was added in Lane O and 21.0 µg/ml CMCT was added to stop control samples in Lane +. Primer extension analysis of purified RNAs is shown using the oligonucleotide primer, cU3-35; dideoxy sequence lanes (A,T,G,C) and a primer extension lane (N) using total *T.brucei* RNA are shown in parallel. Some non-paired residues may not be detected in these experiments because DMS and CMCT are less reactive to C and G residues, respectively. Modification of residues A69, G84, U89–C92 and G103 could not be examined, and residues A52–A69 were often difficult to discern, due to natural reverse transcription stops and/or gel compressions.

Enzymatic cleavages

Cell sonicates for enzymatic analysis were prepared from 5×10^{10} cells/ml in TMK buffer (80 mM Tris-HCl, pH 7.5; 100 mM KCl; 5 mM MgCl₂) plus protease inhibitors (10 µg/ml pepstatin; 5 µg/ml leupeptin; 20 µg/ml phenylmethylsulfonyl fluoride). Ribonuclease cleavage reactions contained either 200 µl of extract, or 10 µg of deproteinized cellular RNA plus 90 µg of *E.coli* tRNA in 200 µl TMK buffer, and the appropriate ribonuclease. Partial digestion of RNA by RNase A (Boehringer Mannheim), RNase T1 (Boehringer Mannheim), and cobra venom RNase V1 (USB) was achieved by titration of each ribonuclease over a wide range of dilutions, then repeated with a focused range of dilutions. Reactions were incubated at 25°C for 15 mins or 0°C for 30 mins and stopped by PCA extraction.

Primer extension analysis

Chemically modified or enzymatically cleaved U3 snRNAs were analyzed by primer extension analysis (24,25), as previously described (26). Oligonucleotide primers were: cU3-31, 5' CC-TTCATCATCAGGATTTGG 3' (complementary to residues 65–84) and cU3-35, 5' GGATCCTTCTGGAACCGGCT 3' (complementary to residues 125–144). RNA sequence could be read starting 5–8 nts upstream of the primer.

(complementary to residues 125–144). RNA sequence could be read starting 5–8 nts upstream of the primer.

Enzymatic analysis of 3' end-labelled U3 RNA

U3 snRNA was radiolabelled at the 3' end by the oligonucleotide splint labelling technique (27). 50 pmoles of oligonucleotide cU3-3': 5' AAAGGATCCTTCTGGAACCGGCTCCTGC 3' was annealed to U3 snRNA present in 50 µg of *T.brucei* RNA. U3 snRNAs were extended by incubation first with Sequenase, version 1 (USB) in the presence of [α -³²P]dCTP and 5 mM dTTP, then with Klenow fragment (NE Biolabs) and 5 mM dCTP. Approximately 10⁴ cpm of gel purified, labelled U3 snRNA and 10 µg of *E.coli* tRNA carrier in 19 µl of TMK buffer were heated to 50°C for 3 mins, cooled slowly to room temperature, and placed on ice. A 1 µl dilution of ribonuclease was added and incubated for 30 mins. Reactions were terminated by addition of 75 µl 4 mM EDTA, 10 µg tRNA, 5 µl 10% sarkosyl, and 100 µl PCA, vortexed immediately, and extracted twice with PCA. Samples were subsequently extracted with 1-butanol, dried, then suspended in 5 µl of DEPC-treated H₂O and 10 µl urea buffer (9.8 M urea, 1.5 mM EDTA, 0.05% xylene

A

B

C

cyanole). Samples were separated on 15% and 6% polyacrylamide sequencing gels.

RESULTS

Comparative analysis of trypanosomal U3 snRNAs

Phylogenetic comparison of functionally homologous RNAs from diverse organisms has been established as a major discriminator for assignment of higher order RNA structure (28,29). Covariation of paired residues between homologous RNAs strongly supports the existence of a putative helical structure. This method depends on a reliable primary sequence alignment. Due to the small size and divergent sequence of the *T. brucei* U3 snRNA, initially it was not possible to align it with U3 snRNA sequences from higher organisms. Therefore, the *T. brucei* U3 snRNA was first compared to U3 snRNA sequences deduced for this study from the trypanosomatid species, *Trypanosoma cruzi* and *Leptomonas collosoma*. U3 snRNAs of approximately 143 residues were identified from each organism (data not shown); RNA sequences were inferred from the sequences of single copy U3 snRNA genes (figure 1). The *T. cruzi* U3 snRNA sequence shares 92% identity with the *T. brucei* sequence and 90% with the *L. collosoma* sequence; 83% identity exists between the *T. brucei* and *L. collosoma* sequences. The first 35 nucleotides are identical in each species; this region contains the well conserved box A homology of U3 snRNAs. Figure 1 shows a secondary structure consistent with the phylogenetic data derived from the trypanosomatid U3 snRNAs. While other models could be derived with these data (20), this model is accordant with chemical modification and enzymatic cleavage studies and phylogenetic comparison with other U3 snRNAs, presented below.

Chemical modification and enzymatic cleavage of U3 snRNAs

The secondary structure of *T. brucei* U3 snRNA was examined by limited chemical modification and ribonuclease cleavage analysis using either deproteinized cellular RNA or RNA complexed with proteins (24,25). The structure of the RNA in RNP forms is assumed to approximate closely the *in vivo* structure. Comparison of structural information derived from naked RNA and RNP should provide information regarding protein-RNA interactions. In the extracts used in these studies, all detectable U3 snRNA was in RNP forms as analyzed by native gel electrophoresis and Northern hybridization (data not shown). The chemical reagents DMS and CMCT were used to modify residues not involved in Watson-Crick base pairing. Naked RNAs and RNPs in cell free extracts were incubated with DMS and CMCT; additionally, living cells were incubated with DMS as it readily crosses cell membranes. Identical data sets were derived by DMS treatment of RNPs in extracts and in whole cells indicating that RNPs probed in cell sonicates were representative

of biologically relevant forms. DMS methylations of N-1-A and N-3-C (A modified more readily than C) and CMCT methylations of N-3-U and N-1-G (U modified more readily than G) were detected by primer extension analysis using AMV reverse transcriptase (figure 2) (24,25). Reverse transcriptase stops at the nucleotide preceding a modified residue because the altered base cannot base pair. Naked U3 snRNA and RNP forms were also treated with ribonucleases which cleave single-stranded regions with base specificity: RNase A cleaves after pyrimidines and RNase T1 cleaves after guanines (24,25). Cobra venom RNase V1 was used to detect helical regions without base specificity; this enzyme also recognizes some tertiary residue interactions not in Watson-Crick base pairs (30). Cleavages in RNAs were identified by primer extension analysis (figure 3). Primers used for reverse transcription were complementary to U3 residues 125-142 and 65-84; representative experiments using the 3'-most primer are shown in Figures 2 and 3.

For both chemical modification and enzymatic cleavage experiments, reactions were performed with increasing concentrations of chemical probes or ribonucleases. Structural information was recorded from reactions in which fewer than 30% of U3 residues were affected to increase the probability that molecules with only one modification or cleavage were examined and that consequent structural alterations were minimized. Control lanes contained unmodified or uncleaved RNAs to determine natural reverse transcription stops; in chemical analyses, stop control samples were included wherein the maximum amount of chemical was added following the addition of stop solution to ensure that modifications did not occur during RNA purification. Reactions were routinely conducted at 25°C and were done at least three times. Decreasing the reaction incubation temperature to 0°C to potentially stabilize RNA-RNA or RNA-protein interactions did not alter the pattern of modifications detected, although some differences in availability of sequences to ribonuclease digestion were noted (see below). No differences in modification or cleavage patterns were detected between RNAs which were probed directly following suspension into reaction buffers, or which were first heated to 55°C then slowly cooled to room temperature. Relative intensities of bands between naked and RNP form RNAs were compared.

Limited ribonuclease digestion of 3' end labelled U3 snRNA

The structure of naked *T. brucei* U3 snRNA was also examined directly by enzymatic analysis. In comparison to the indirect primer extension method, direct enzymatic analysis allowed examination of the entire U3 snRNA molecule, including 3' end structures. U3 snRNA present in total cellular RNA was uniquely labelled at the 3' end by the splint oligonucleotide method (27) then subjected to enzymatic digestion under non-denaturing conditions at 0°C. This analysis could not be done on RNP form U3 snRNA as it could not be 3' end-labelled while associated

Figure 3. Ribonuclease cleavage of U3 snRNA in naked and RNP forms. Deproteinized RNA (RNA) and RNA in cell sonicates (RNP) were treated at 25°C with: (A) RNase A in lanes 1-6 at concentrations of: 0, 1, 2, 4, 8, 16 μ U/ml for RNA; and in lanes 7-12 at concentrations of: 0, 10, 20, 40, 80, 160 μ U/ml for RNP. (B) RNase T1 in lanes 1-6 at concentrations of 0, 0.1, 0.2, 0.4, 0.8, 1.6 mU/ml for RNA; and in lanes 7-12 at concentrations of 0, 1, 2, 4, 8, 16 mU/ml for RNP; (C) Cobra venom RNase V1 at 0°C in lanes 1-6 and 7-12 at concentrations of: 0, 0.1, 0.2, 0.4, 0.8, 1.6 U/ml for RNA and RNP. Purified RNAs were primer extended using AMV reverse transcriptase and the U3-35 primer; parallel lanes contained dideoxy sequence reactions (A,T,C,G) and a primer extension reaction (N) using total *T. brucei* RNA. Enzymatic cleavage after residues G84, G90-C92, and G105 could not be examined, and cleavages after residues U70-A82 were often difficult to discern, due to natural reverse transcription stops and/or gel compressions.

Figure 4. Ribonuclease analysis of 3' end labelled U3 snRNA. U3 snRNA present in total cellular RNA was specifically labelled at the 3' end, then gel purified and mixed with carrier tRNA. Limited ribonuclease digestions were carried out in identical buffers for each enzyme: lane 0, no enzyme; RNase A, in lanes 1–4, at concentrations of 0.8, 4, 20, 100 μ U/ml; RNase T1, in lanes 1–4, at concentrations of 0.08, 0.4, 2, 10 mU/ml; RNase V1, in lanes 1–4, at concentrations of 0.008, 0.04, 0.2, 1 U/ml; lane 5 contains 1 U/ml RNase V1 and 2 μ g tRNA versus 10 μ g tRNA for all other reactions. The reactions in lanes 4 and 5 were too overdigested for structural analysis purposes. Sequence comparison lanes are 3' end labelled U3 snRNA partially digested by RNases under denaturing conditions, or treated by alkaline hydrolysis: G, RNase T1; A, RNase U2; Y = C,U RNase B. cereus; and H, alkaline hydrolysis lane. Indicated cleavages by RNase V1 are distinguished from cleavages by single-strand-specific enzymes by denoting them as two bases separated by a slash (/). RNase V1 cleavages observed within residues A117–C121 coincided with a region of gel compression of RNAs and these cleavages were not included in figure 5A.

with proteins. Cleaved RNAs were compared to ribonuclease sequencing reactions on both 15% (figure 4) and 6% polyacrylamide sequencing gels (not shown). Whereas RNase A and T1 cleavages leave a 5' hydroxyl, as does alkaline hydrolysis, cobra venom RNase V1 cleavages leave a 5' phosphate; thus, RNase V1 fragments migrate approximately one-half nucleotide faster than hydrolyzed RNAs. A modification of the C88 residue, most likely a ribose modification, was detected as a gap in the alkaline hydrolysis ladder. The direct RNA analysis was more sensitive than the indirect primer extension method, providing structural information not only for the 3'-most

residues but also for residues not discerned due to the relatively high signal-to-background ratio inherent in the reverse transcribed RNAs.

Compilation of U3 snRNA secondary structure data

Structural data derived for naked U3 snRNA by primer extension of chemically modified or ribonuclease cleaved RNAs and by ribonuclease cleavage of 3' end labelled U3 snRNA is compiled in Figure 5A. Structural information derived from chemical and enzymatic probing of U3 snRNAs in protein-complexed forms is compiled in Figure 5B. Differences in structural data sets for RNA and RNP detected by primer extension indicated protein-induced conformations in U3 snRNP RNA. Overall, chemical modification experiments yielded more data on non-base-paired regions of RNA than ribonuclease analysis, due to the relative small sizes and base specificities of the chemical probes. Ribonuclease studies provided secondary structure information, but did not reveal substantial protein protections in U3 snRNPs. Ribonuclease cleavages at residues C5, C10, C17, C58, C59, C62 were seen only by primer extension analysis which is consistent with the possibility that cytosines were more efficiently cleaved in U3 snRNAs present in total RNA, than in gel-purified U3 snRNAs plus tRNA carrier. Cleavages at residues U22, G52, G57, U76, U79, G80, G101, G103, G108, and G112 were relatively more pronounced in direct RNA than indirect cDNA analysis, whereas cleavages at residues U28, U70, G83, G84, U89–U94, U107, A122, and residues 3' to 113 were seen only in direct RNA analysis.

U3 snRNA secondary structure determined by chemical, enzymatic and phylogenetic methods

Structural probing of *T.brucei* U3 snRNA (figure 5) and comparative analysis of U3 snRNA sequences (figure 1 and figure 6) strongly support the secondary structure presented in Figures 1, 5 and 6. In this model, residues within regions A1–C6, A18–C23, C37–A68, G75–U86, C99–C117 are single-stranded. A 5' domain stem loop I is formed by residues G7–C36, and 3' domain helices are formed between residues A69–U73 and A138–U142, and G87–C98 and G119–U131. With this structural information in hand, trypanosomatid U3 snRNA sequences were aligned with representative, complete U3 snRNA sequences known to date from vertebrates, yeasts, plants and protozoa (figure 6). In several instances, this alignment differs from the original structure proposed for a given U3 snRNA to retain maximal structural similarity; for example, *S.cerevisiae*, *S.pombe*, tomato and arabidopsis U3 snRNAs are realigned to decrease the size of stem loop Ib and to maintain a single-stranded 'hinge' region, as well as a 3' terminal stem structure and the GAU region. Strikingly, stem loops II and III, found in all other U3 snRNAs, are clearly deleted in trypanosomatid U3 snRNAs, and sequences 3' to the box A homology can not be aligned with stem loop Ib structures proposed for U3 snRNAs of other unicellular organisms.

The 5' domain of *T.brucei* U3 snRNA includes a single small stem loop structure

A single stem loop I structure appears to form within the 5' region of *T.brucei* U3 snRNA, analogous in size and position to stem loop Ia proposed for fungal and plant U3 snRNAs (15,31,32). Loop residues A18–C23 are highly available to chemical probes and single strand-specific ribonucleases. RNase V1 cleaves after residues C10, U11, U13, and G31 in naked RNA and RNP, and

Figure 5. (A). Secondary structure analysis of deproteinized *T.brucei* U3 snRNA. (B) Secondary structure analysis of ribonucleoprotein form *T.brucei* U3 snRNA. Chemically modified residues are circled; single-strand-specific ribonuclease cleavages are indicated by an arrow with a feathered tail; ribonuclease V1 cleavages are indicated by an arrow with a circular tail. Extent of modification or digestion is denoted by open circles or arrows for weak reactivity, shaded for moderate reactivity and black for strong reactivity. In (A), ribonuclease cleavages which were detected only by indirect analysis are marked by a small open circle on the arrow; those which were detected only in direct RNA analysis are marked by a small solid circle; and those which were stronger in direct than indirect RNA analysis are marked by an asterisk.

additionally at C12 in RNP. These data support helix formation in this region. Contrary to these data, putative stem residues, especially those within the energetically weaker, upper portion

of the structure are accessible to chemical modification (weak base pairs at helix termini or adjacent to bulges are often modified); a weak cleavage by RNase A is also seen in naked

<

Figure 6. Comprehensive phylogenetic comparison of U3 snRNA sequences. Primary U3 snRNA sequences derive from H.s., Human placental cells (47); R.n., Rat U3B, Novikoff heptoma cells (48); M.m., Mouse U3B (49); X.l., *Xenopus laevis* (12); S.c., *Saccharomyces cerevisiae* snR17A (4,16,50); S.p., *Schizosaccharomyces pombe* (15); D.d., *Dictyostelium discoideum* (51); T.a., *Triticum aestivum* (32); L.e., *Lycopersicon esculentum* (14); A.t., *Arabidopsis thaliana* (31) T.t., *Tetrahymena thermophila* (34); T.b., *T. brucei* (21); T.c., *T. cruzi* (this paper); L.c., *L. collosoma* (this paper). Regions of primary sequence identity are indicated by name and by consensus sequence. Putative helical features are indicated. The approximate position of putative stem loop I structures are overlined by dots and arrows followed by an unstructured 'hinge' region; stem loops Ia and Ib are positioned according to wheat U3 snRNA (32) and generally indicate putative helical regions in other plant, fungal and protozoan U3 snRNAs; stem loop Ib structures are variable in sequence and position.

RNA at U28. Within the RNP, 12 of the 22 residues in the helix are accessible to modification, and the subset of these which are modified in the naked U3 snRNA are modified to a higher degree in the RNP. Thus, it appears that a more open conformation occurs in the cellular, proteinated form of the U3 snRNA than in naked RNA. Taken together, the structure of the 5' region appears to oscillate between a stem loop form and a more open conformation; this domain structure may be less frequently paired in native form than in deproteinized U3 snRNA.

The long single-stranded region, C37–A68, shows some evidence of altered structure between deproteinized and native U3 snRNAs. RNase A cleavages are more intense in naked RNA relative to RNP at U44, U48, and U55; moderate chemical enhancements are seen in RNP at U43 and A53. RNase V1 cleaves after U41 and C58 in both forms of the U3 snRNA, and after C59 and A61 in naked RNA; these residues are also recognized by single-strand probes. As no base pairing interactions could be postulated to correspond to RNase V1 cleavages (except for the seemingly single-stranded regions: A1–A4 with U41–U44) it is likely that the enzyme is recognizing non-base-paired tertiary interactions.

Conserved single-stranded regions in the 3' domain of *T.brucei* U3 snRNA

Two single-stranded regions of *T.brucei* U3 snRNA which share identity with conserved sequences of other U3 snRNAs were found within the regions encompassed by residues G74–U86 and C99–C118. In each of these regions, several differences were noted in structural probing results between naked RNA and RNP which were consistent with protein interactions. Within residues G74–U86, which correspond to the previously defined GAU region (15), A81 is modified in naked RNA but not RNP, whereas A75 modification is decreased and A78 modification is enhanced in RNP. Moderate cleavages at U76 and U79 were weaker in RNP versus naked RNA. Moderate T1 cleavages at G74, G77 and G80 were seen in both naked and RNP RNAs, however, G77 and G80 cleavages were less pronounced in reactions performed at 0°C versus 25°C in RNP (data not shown). Within residues C99–C118, which share sequence identity with both box B and C homologies, residues G100, A102, A104, U110, G112, and A113 are accessible to modification in naked RNA but not in RNP, and C99, U109 and U114 modifications are enhanced in naked RNA versus RNP. Conversely, enhancement of RNA modification in RNP is seen for residues C99, U107 and G108. Minor cleavages by RNase T1 at G100 and G112 and by RNase V1 at C106 are seen only in naked RNA, whereas a RNase T1 cleavage at G108 is enhanced in RNP. Multiple protection and enhancement effects seen by chemical and enzymatic probing of protein complexed versus naked U3 snRNA is consistent with protein interactions in these single-stranded regions (24,33). Protected residues may be directly involved in protein interactions. Furthermore, protein binding may alter RNA conformation such that RNA residues become more or less available to structural probes.

Direct RNA analysis provided additional information for naked RNA structure in the regions encompassing residues G74–U86 and C99–C118. RNase A or T1 cleavages at U73, G83 and G84 indicated that these residues are not involved in base pairing. Moderate cleavages by RNase V1 were detected after G105 and C106 and weak cleavages after U107 and U110. The RNase V1 cleavages did not indicate regions which could form detectable, conserved base pairs with another region of the U3 snRNA and

are more likely to be recognizing some other form of higher order structure within the large C99–C118 loop region. Similarly, in structural studies of naked *S.cerevisiae* U3 snRNA in solution, RNase V1 cleavages were also detected within the single-stranded region containing box C sequences (16).

A third 3' single-stranded region of *T.brucei* U3 snRNA is likely to occur within residues U132–A137 which share 3 of 6 residues in common with the box D sequences of mammalian, plant and yeast cell U3 snRNAs. Phylogenetic comparison strongly indicates that this region should remain single-stranded in all U3 snRNAs (figure 6). However, residues within this sequence in the *T.brucei* U3 snRNA were digested by both single strand-specific ribonucleases and RNase V1; the analogous sequence was likewise digested by RNase V1 in human U3 snRNA (13). The RNase V1 cleavages potentially indicate residue stacking in this 3' sequence.

Only two helical structures are found in the 3' domain of *T.brucei* U3 snRNA

A strong, central stem structure forms between residues G87–C98 and G119–U131 in *T.brucei* U3 snRNA. This stem is strongly supported by phylogenetic differences in base pairs between trypanosomatid (figure 1) and all other U3 snRNAs (figure 6). Biochemical evidence of this structure in *T.brucei* RNA was obtained from direct RNA analysis wherein several cleavages by RNase V1 occurred on both sides of the helix (figure 4). Experimental evidence for or against the formation of three strong base pairs in the upper portion of the central stem (G96–C98 and G119–C121) were not provided by these studies.

Formation of a short, 4 to 6 base pair, 3' terminal stem in all U3 snRNAs is well supported by phylogenetic covariation of base pairs (figure 6). Experimental support for this helix in *T.brucei* U3 snRNA came from analysis of 3' end-labelled RNA; two RNase V1 cleavages were seen in the 5' half of the helix between residues U70/C71 and C73/U74 and one in the 3' half of the helix between residues G140/A141. A weak RNase A cleavage at C72 seen in indirect analysis of RNP and modifications of U70 and U73 may indicate breathing of helical ends.

DISCUSSION

The trypanosomal U3 snRNA secondary structure model presented in this study is consistent with phylogenetic comparisons between U3 snRNA sequences and with chemical and enzymatic probing of the U3 snRNA in naked and protein-complexed cellular forms. Structural information obtained for the *T.brucei* U3 snRNA aided alignment of its sequence with U3 snRNA sequences from vertebrates, plants, fungi and protozoa and to clearly indicate conserved and divergent structural features. The 5' domain of trypanosomal U3 snRNAs, which contains a moderately conserved box A homology, can form only a single small stem loop I structure as compared to two structures proposed for other unicellular organisms. The 3' domain of *T.brucei* U3 snRNA contains conserved single-stranded regions which apparently interact with proteins. Whereas we previously reported two box C-like sequences within the *T.brucei* U3 snRNA (20), it is now apparent that the 5'-most of these corresponds to the conserved GAU region and the 3' most sequence is the true box C homolog which abuts box B-like sequences in a large single-stranded loop. Stem-loop II, which separates the box B and C sequences in all other U3 snRNAs, and stem loop III which

usually follows box C sequences, are missing in trypanosomal U3 snRNAs. This structural analysis of *T. brucei* U3 snRNA has provided a framework for examination of the minimal U3 snRNA sequences required for conserved functions in rRNA processing.

The secondary structure of the 5' end of U3 snRNAs remains a conundrum. The 5' sequences of all U3 snRNAs are similar in length, from 68 residues in trypanosomes to 78 residues in plants, and the approximate first 32 residues are related in sequence, yet comparative analysis eliminates the possibility of conserved helices within the 5' domain. One large stem loop structure (residues 5–64) may form in vertebrate U3 snRNAs, followed by a hinge region of 10 residues which is accessible to oligonucleotide binding and subsequent RNase H digestion (10,12,13). That two stem loops may form in fungal and plant U3 snRNAs (15,31,32) was supported by a solution study of the *S. cerevisiae* U3 snRNA structure (16). However, in that model, the proposed stem loop Ib contained sequences which are conserved in the GAU region and the terminal stem residues of all U3 snRNAs. In figure 6, the yeast stem loop Ib has been repositioned to include residues U46 to A64; this alteration is consistent with the experimental evidence and frees residues for participation in conserved 3' domain structures. Even so, proposed stem loop Ib structures vary in position and sequence and are not well supported by comparative analysis. The single small 5' stem loop structure of *T. brucei* U3 snRNA is similar to stem loop Ia of fungal and plant U3 snRNAs, and the positioning of the highly conserved box A homology within the loop and 3' helix residues is alike; the stem loop structure is followed by a single-stranded region which shows some evidence of protein-induced conformation in RNP, but no ability to form a stem loop structure. The sequence of another protozoan U3 snRNA, from *Tetrahymena*, has been reported (34). In this report, no helical structures were suggested for the 5' domain, yet conceivably at least a stem loop Ia could form (figure 6); no convincing stem loop Ib configurations are found. The caveat exists that structures proposed for the 5' domains of various U3 snRNAs are not conserved across a broad phylogenetic range. That any of the proposed helical structures have a functional role in U3 snRNP maturation or function remains to be tested by genetic analysis.

Evidence suggests that the 5' domain of U3 snRNAs may exist in more than one conformation. *In vivo* probing of *Xenopus* U3 snRNP indicates that 5' residues are highly accessible to chemical modification, suggestive of an open structure (12). Residues of the upper 5' stem loop I structure in human U3 snRNA were also available to single-strand-specific probes, although a subset of these were protected in RNP; RNase V1 cleavages of the stem indicated helical potential in both naked RNA and RNP (13). Similarly, the 5' helical structure proposed here for *T. brucei* U3 snRNA was subject to probing by both single-strand and double-strand-specific probes, though stem residues were more readily modified in RNP than naked U3 snRNA. In spite of some incongruities between these studies, together they suggest that the 5' regions of U3 snRNAs may adopt both helical and open conformations. The coexistence of single-stranded conformations and species-specific helical structures may be a conserved feature of the 5' domains of U3 snRNAs.

Flexibility of the 5' domain structure of U3 snRNAs is consistent with the notion that U3 snRNA sequences may be single-stranded for interactions with pre-rRNA sequences, possibly to initiate preribosomal complex assembly. Sequences close to and within the box A region have been indicated in close

associations with 5' ETS sequences of pre-rRNAs following psoralen crosslinking of RNAs in mammalian and yeast systems (5,7,35). In yeast, crosslinks were formed between U3 snRNA and two regions of the ETS (5); one site of U3 crosslinking in the ETS was close to a U3-dependent cleavage and the second site was mapped to an upstream region in the ETS, required for 18S rRNA maturation, which shared base complementarity with the U3 snRNA in a sequence 3' adjacent to stem loop Ia. That particular U3 snRNA–ETS interactions may involve base pairing (5,35) is consistent with previous reports in which U3 snRNA and pre-rRNAs were coisolated from deproteinized, nondenatured RNA preparations from human, yeast and trypanosome cells (20,36–38). The *T. brucei* U3 snRNA contains a sequence 3' to stem loop I which shares complementarity to ETS sequences 3' adjacent to the primary cleavage site (20). By analogy to yeast, these sequences in *T. brucei* may closely contact one another. Preliminary studies indicate that psoralen crosslinks are formed *in vivo* between *T. brucei* U3 snRNA and ETS sequences; specific sites of interaction are currently under investigation (unpublished data).

Single stranded sequences homologous to GAU region and box B, C and D sequences were identified in the 3' domain of *T. brucei* U3 snRNA. Comparative analysis of representative U3 snRNA sequences from ancient protozoans to humans show a high degree of conservation (figure 6). The indicated consensus sequences vary somewhat from those proposed in earlier alignments with fewer U3 snRNA sequences (12,15,39). The two best conserved 3' sequences are within the GAU region (GA/UG/UGAUGA) and box C (UNGAUGAU/ACG); the GAU sequence has also been named box A° in plant U3 snRNAs (32) and has recently been labelled box C' due to its sequence similarity to the box C homology (40). Box B and D sequences are less conserved. Trypanosomal U3 snRNAs share identity with 6 of 9 residues, and the *Tetrahymena* U3 snRNA shares 5 of 9 residues, the box B consensus (GAGCGUGAA/U) of non-protozoan U3 snRNAs. Box D-like sequences in trypanosome, frog, and mold U3s share 4 of 6 identities with the box D 'consensus' (G/UUCUGA). Possible base pairing interactions which have been proposed between the GAU region and box D residues (13), and between box C and D sequences (16) are not conserved in trypanosomal and other U3 snRNAs.

Evidence indicates that the GAU region and box B, C and D sequences interact with proteins. In *T. brucei* U3 snRNA, the major single-stranded regions containing the GAU region and the box B and C sequences show differences in protection and enhancement patterns by structural probes between RNA in naked and protein-complexed forms, suggestive of protein interactions. This study extends conclusions drawn from analysis of the HeLa U3 snRNP in which ribonuclease protected fragments containing the GAU region and box B, C and D sequences were immunoprecipitated by antifibrillarin antibodies; six candidate U3 snRNP proteins were also precipitated (13). Furthermore, mutational analysis has indicated that box C, though not box D, residues are required for reconstitution of synthetic U3 snRNAs with fibrillarin in HeLa cell extracts (41). Mutations of residues G159 and G162, which abolished reconstitution (41,42), are invariant between organisms (figure 6). Fibrillarin associates with a number of nucleolar snRNAs, possibly within a large, heterogeneous complex; direct binding of fibrillarin to U3 snRNA or other nucleolar snRNA has not been established. The protein constituents which directly interact with the conserved protein-protected regions of U3 snRNA remain to be identified.

In higher organisms, box B and C sequences are looped out between stem loop II, and the central stem and stem loop III. The arrangement of this domain is interesting in trypanosomal U3 snRNP as stem loops II and III do not occur and box B and C sequences lie adjacent to one another in a large loop structure. Helical structures have been proposed to stabilize snRNP protein-RNA interactions in U3 snRNPs (41), as well as in other RNPs, but obviously are not necessary in the trypanosomal U3 snRNP. A domain including GAU region and box D sequences between the 3' terminal stem and the central stem structure is perfectly conserved in trypanosomal and other U3 snRNAs. A similar structure may occur in other fibrillarin-associated snRNAs including the U14 snRNA described in mammals and yeast and the recently described mammalian U16 snRNA (40). The 3' terminal stem appears to be required for optimal nuclear import and trimethylation of U3 snRNA in *Xenopus* oocytes (42) and for stability of U14 snRNA in yeast (39).

The dramatic difference in size between trypanosomal and other U3 snRNAs is accounted for by the lack of trypanosomal stem loop II and III structures. In the largest known U3 snRNA, from *S.cerevisiae*, structural variation also occurs in the 3' domain wherein an extra stem loop IV is positioned adjacent to stem loop II. Differences between trypanosomal, yeast and other U2 snRNAs previously helped to define minimal, functional U2 snRNA sequences in mRNA splicing. The unusually small U2 snRNA of *T.brucei* does not contain a stem loop III (21,43) but was otherwise conserved in structure with other U2 snRNAs (26). Yeast U2 snRNA contains an additional 1000 residues within the stem loop III region which are not required for U2 function (44,45). Mutant *Xenopus* U2 snRNAs which lack stem loop III can stably bind U2 snRNP proteins and function in splicing, albeit, with lower efficiency than wild type U2s (46). By analogy with the U2 snRNA data, we hypothesize that the stem loop II and III structures missing in trypanosomal U3 snRNAs will not be required for absolute function of U3 snRNPs in other organisms, though they may have some nonconserved roles not required for U3 snRNP function in trypanosomes.

ACKNOWLEDGEMENTS

We thank members of the Agabian lab, especially J.Dungan and K.Watkins, for helpful discussions and experimental advice, and S.Metzenberg and S.Datta for commenting on the manuscript. This work was supported by grants to N.A. from the John D. and Catherine T. MacArthur Foundation and by National Institutes of Health grant A121975.

REFERENCES

- Sollner-Webb, B., Tyc, K. and Steitz, J. (1993) In Zimmerman, R. and Dahlberg, A. (ed.), *Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Synthesis*. CRC Press, New York.
- Filipowicz, W. and Kiss, T. (1993) *Mol. Biol. Rep.*, 18, 149–156.
- Fournier, M.J. and Maxwell, E.S. (1993) *TIBS*, 18, 131–135.
- Hughes, J.M.X., Konings, D.A.M. and Cesareni, G. (1987) *EMBO J.*, 6, 2145–2155.
- Beltrame, M., and Tollervey, D. (1992) *EMBL J.*, 11, 1531–1542.
- Masser, R.L. and Calvet, J.P. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 6523–6527.
- Stroke, I.L., and Weiner, A.M. (1989) *J. Mol. Biol.*, 210, 497–512.
- Kass, S., Tyc, K., Steitz, J.A. and Sollner-Webb, B. (1990) *Cell*, 60, 897–908.
- Mougey, E.B., Pape, L.K. and Sollner-Webb, B. (1993) *Mol. Cell. Biol.*, 13, 5990–5998.
- Savino, R. and Gerbi, S.A. (1990) *EMBO J.*, 7, 2299–2308.
- Hughes, J.M. and Ares, Jr., M. (1991) *EMBL J.*, 10, 4231–4239.
- Jeppesen, C., Stebbins-Boaz, B. and Gerbi, S.A. (1988) *Nucleic Acids Res.*, 16, 2127–2147.
- Parker, K.A. and Steitz, J.A. (1987) *Mol. Cell. Biol.*, 7, 2899–2913.
- Kiss, T. and Solymosy, F. (1990) *Nucleic Acids Res.*, 18, 1941–1949.
- Porter, G.L., Brennwald, P.J., Holm, K.A. and Wise, J. (1988) *Nucleic Acids Res.*, 16, 10131–10152.
- Segault, V., Mougin, A., Gregoire, A., Banroques, J. and Branlant, C. (1992) *Nucleic Acids Res.*, 20, 3443–3451.
- Perry, K. and Agabian, N. (1991) *Experientia*, 47, 118–128.
- Campbell, D.A., Kubo, K., Clark, C.G. and Boothroyd, J.C. (1987) *J. Mol. Biol.*, 196, 113–124.
- White, T., Rudenko, G. and Borst, P. (1986) *Nucleic Acids Res.*, 14, 9471–9489.
- Hartshorne, T. and Agabian, N. (1993) *Mol. Cell. Biol.*, 13, 144–154.
- Mottram, J., Perry, K.L., Lizardi, P.M., Lüthmann, R., Agabian, N. and Nelson, R. (1989) *Mol. Cell. Biol.*, 9, 1212–1223.
- Biene, E.J., Hammodi, E. and Hill, G.C. (1981) *Exp. Parasitol.*, 51, 408–417.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Cristiansen, J. and Garrett, R. (1988) *Meth. Enzymol.*, 164, 456–468.
- Krol, A. and Carbon, P. (1989) *Meth. Enzymol.*, 180, 212–227.
- Hartshorne, T. and Agabian, N. (1990) *Genes Dev.*, 4, 2121–2131.
- Hausner, T.-P., Giglio, L.M. and Weiner, A.M. (1990) *Genes Dev.*, 4, 2146–2156.
- Noller, H.F. and Woese, C.R. (1981) *Science*, 212, 403–411.
- James, B.D., Olsen, G.J. and Pace, N.R. (1989) *Method. Enzymol.*, 189, 227–239.
- Lowman, H.B., and Draper, D.E. (1986) *J. Biol. Chem.*, 261, 5396–5403.
- Marshall, C., Kiss, T. and Filipowicz, W. (1990) *Nucleic Acids Res.*, 18, 3459–3466.
- Marshall, C., Connelly, S. and Filipowicz, W. (1992) *Plant Mol. Biol.*, 19, 973–983.
- Stern, S., Powers, T., Changchien, L. and Noller, H.F. (1989) *Science*, 244, 783–790.
- Orum, H., Neilsen, H. and Engberg, J. (1993) *Nucleic Acids Res.*, 21, 2511.
- Tyc, K. and Steitz, J.A. (1993) *Nucleic Acids Res.*, 20, 5375–5382.
- Epstein, P., Reddy, R. and Busch, H. (1984) *Biochemistry*, 23, 5421–5425.
- Tollervey, D. (1987) *EMBO J.*, 6, 4169–4175.
- Zagorski, J., Tollervey, D. and Fournier, M.L. (1988) *Mol. Cell. Biol.*, 8, 3282–3290.
- Huang, G.M., Jarmolowski, A., Struck, J.C.R. and Fournier, M.J. (1992) *Mol. Cell. Biol.*, 12, 4456–4463.
- Tycowski, K.T., Shu, M. and Steitz, J.A. (1993) *Genes Dev.*, 7, 1176–1190.
- Baserga, S.J., Yang, X.W. and Steitz, J.A. (1991) *EMBO J.*, 10, 2645–2651.
- Baserga, S.J., Gilmore-Herbert, M. and Yang, X.W. (1992) *Genes Dev.*, 6, 1120–1130.
- Tschudi, C., Richards, F.F. and Ullu, E. (1986) *Nucleic Acids Res.*, 14, 8893–8903.
- Igel, A.H. and Ares, Jr., M. (1988) *Nature*, 334, 450–453.
- Shuster, E.O. and Guthrie, C. (1988) *Cell*, 55, 41–48.
- Hamm, J., Dathan, N.A. and Mattaj, I.W. (1989) *Cell*, 59, 159–169.
- Suh, D., Busch, H. and Reddy, R. (1986) *Biochem. Biophys. Res. Commun.*, 137, 3667–3680.
- Reddy, R., Henning, D. and Busch, H. (1985) *J. Biol. Chem.*, 260, 5715–5719.
- Mazan, S. and Bachellerie, J.P. (1988) *J. Biol. Chem.*, 263, 19461–19467.
- Myslinski, E., Segault, V. and Branlant, C. (1990) *Science*, 247, 1213–1216.
- Wise, J.A., and Weiner, A.M. (1980) *Cell*, 22, 109–118.