Nucleic Acids Research, 1996, Vol. 24, No. 3 515–522

Characterization of the highly divergent U2 RNA homolog in the microsporidian *Vairimorpha necatrix*

Peter DiMaria*, Branka Palic, Bettina A. Debrunner-Vossbrinck[1], Julie Lapp and Charles R. Vossbrinck[1]

Department of Chemistry, Delaware State University, Dover, DE 19901, USA and [1]Department of Entomology, 320 Morrill Hall, 505 S. Goodwin Avenue, University of Illinois, Urbana, IL 61801, USA

Received August 31, 1995; Revised and Accepted November 30, 1995

EMBL accession no. Z50072

ABSTRACT

An RNA homologous to U2 RNA and a single copy gene encoding the RNA homolog have been characterized in the microsporidian, *Vairimorpha necatrix.* The RNA which is 165 nucleotides in length possesses significant similarity to U2 RNA, particularly in the 5′ half of the molecule. The U2 homolog contains the highly conserved GUAGUA branch point binding sequence seen in all U2 RNAs except those of the trypanosomes. A U2 RNA sequence element implicated in a U2:U6 RNA intermolecular pairing is also present in the U2 homolog. The *V.necatrix* U2 RNA homolog differs at positions previously found to be invariant in U2 RNAs and appears to lack an Sm binding site sequence. The RNA can be folded into a secondary structure possessing three of the four principal stem–loops proposed for the consensus U2 RNA structure. A cis-diol containing cap structure is present at the 5′ end of the U2 homolog. Unlike the cap structures seen in U-snRNAs and mRNAs it is neither 2,2,7-trimethylguanosine, γ-monomethyl phosphate, nor 7-methylguanosine.

INTRODUCTION

The splicing of nuclear pre-mRNAs is apparently common to all eukaryotic organisms. The splicing of the pre-mRNA substrate takes place in the context of a dynamic supramolecular complex termed the spliceosome (for reviews see 1,2). Spliceosomes, which have been particularly well characterized in mammalian and yeast systems, include a number of small nuclear RNAs (snRNAs). These relatively abundant, uridine rich RNAs are found in small ribonucleoprotein complexes (snRNPs) and appear to be ubiquitous in eukaryotes. A hallmark of the snRNAs (with the exception of the U6 RNA) appears to be the presence of the hypermethylated 2,2,7-trimethylguanosine 5′ cap structure (3). The U1, U2, U5 and U4/U6 snRNPs are present in the spliceosome and have been shown by a variety of means to be necessary for splicing (2). A number of functionally relevant base-pairing interactions among snRNAs and also between snRNAs and the pre-mRNA have been identified. Prominent among these are interactions of U2 RNA with the pre-mRNA and U6 RNA (4).

While molecular genetic and biochemical approaches have been indispensible in achieving an ever increasing understanding of the structure and function of snRNPs, comparative analysis of highly divergent snRNAs is also a valuable tool in this undertaking. Using phylogenetic approaches, considerable insight into the structure of snRNAs can be gained by including extremely primitive (divergent) organisms in comparative molecular biological studies of snRNAs. If structural and functional homologs of snRNAs are present in such divergent organisms, they may also provide a glimpse of the nuclear pre-mRNA splicing apparatus as it first emerged in eukaryotes. Currently there is a paucity of information regarding putative snRNAs in such organisms.

In this report we have identified and characterized an RNA from the microporidian *Vairimorpha necatrix* which is structurally homologous to U2 RNA. Organisms in the phylum Microspora are among the most ancient eukaryotes thus far identified (5). These obligate intracellular parasites lack mitochondria; it has been proposed that the microsporidia diverged prior to the acquisition of this organelle in the eukaryotic line. They have a primitive nuclear division and their ribosomes and rRNAs are of prokaryotic size (5,6). The U2 RNA homolog identified in this organism shows a high degree of divergence from the U2 consensus sequence in that a number of variant nucleotides occur at positions previously found to be invariant. Also, in sharp contrast to other snRNAs it appears to contain a 5′ cap structure which is not 2,2,7-trimethylguanosine, 7-monomethylguanosine, nor γ-monomethyl phosphate.

MATERIALS AND METHODS

Preparation of nucleic acid

DNA isolation. Vairimorpha necatrix spores were opened by germination. Spores were incubated in 50 mM phosphate buffer (pH 11.4) for 45 min followed by centrifugation to remove the incubation buffer. The spore pellet was resuspended in 50 mM borate buffer (pH 9.2), 15 mM KCl. Germination was monitored by phase contrast light microscopy. After ~3 min, when the majority of spores had germinated, the solution was extracted with phenol–chloroform and the nucleic acid was precipitated from the aqueous phase with ethanol. The nucleic acid was treated with 0.2 mg/ml ribonuclease A for 30 min at 37°C, followed by phenol–chloroform extraction and ethanol precipitation.

* To whom correspondence should be addressed

RNA isolation. Nucleic acid was isolated using the glass bead method, as described previously (7). The nucleic acid was treated with 2 µg/ml ribonuclease-free deoxyribonuclease I for 1 h at 37°C, followed by phenol–chloroform extraction and ethanol precipitation.

Lambda library construction and screening

Vairimorpha necatrix genomic DNA was partially digested with *Sau*3A and ligated to bacteriophage lambda L47.1 which had been digested with *Bam*HI. Phages were plated on *Escherichia coli* strain P2392 lysogen. Screening of the library was carried out essentially as previously described utilizing 5′-^{32}P-labeled U2-L15 oligonucleotide (see DNA sequencing) as a hybridization probe (8).

Northern blot analysis

RNAs were resolved on 8% polyacrylamide gels containing 8 M urea or on 5.75% aminophenylboronate gels (see modification of RNAs) and electroblotted onto nylon membrane (Hybond N, Amersham; or Zeta-Probe, BioRad). After fixing the RNA to membranes (UV-crosslinking, or baking for 1 h at 80°C), prehybridization and hybridization were carried out as described above for the lambda library screening.

Southern blot analysis

Vairimorpha necatrix genomic DNA (5 µg) was cut with *Eco*RI, *Hin*dIII or *Bam*HI and resolved by agarose gel electrophoresis and blotted onto nitrocellulose (9). The filters were hybridized to a 120 bp *Sau*3A fragment corresponding to nucleotides 22–142 of the U2 homolog which was ^{32}P-labeled by the random primers method (9).

DNA sequencing

A 2.5 kb *Hin*dIII fragment containing the U2 gene was subcloned into M13 mp18. DNA sequencing was done with T7 DNA polymerase (Sequenase kit, USB), using oligonucleotide primers U2-L15 5′-CAGATACTACACTTG-3′, U2.2 5′-CTCTCAAA-GTTATCGGC-3′, U2.3 5′-TCTTAAGATGCCAGAGA-3′, and U2.4 5′-CAGGTGGACCAATGACAA-3′. U2.2 was constructed based on the sequence obtained with U2-L15, and U2.3 and U2.4 were constructed based on the sequence obtained with U2.2. The primer extension sequencing reactions were carried out using 5′-^{32}P-labeled primers or unlabeled primers in the presence of [α-^{35}S]dATP and were resolved on 8% polyacrylamide gels containing 8 M urea and visualized by autoradiography.

RNA sequencing and end mapping

RNA dideoxy sequencing with reverse transcriptase was done using 5′-^{32}P-labeled U2-L15 primer as described previously (8,10). The 3′ end was mapped by annealing total RNA to an M13 mp18 clone containing the antisense strand of the U2 homolog. The hybrids were digested with mung bean nuclease. The resulting protected DNA fragment was annealed to the U2.2 primer. The primer was extended with T7 DNA polymerase (Sequenase kit, USB) in the presence of dNTPs and [α-^{35}S]dATP. After denaturation, the radiolabeled primer extension product was run on a sequencing gel alongside a sequencing ladder generated using the same primer and the intact M13 mp18 clone. The electrophoretic mobility of the primer extension product in relation to the sequencing ladder allows positioning of the 3′ end of the RNA.

Immunoprecipitation of RNA

For immunoprecipitations, 2.5 mg of Protein A–Sepharose in NET-2 [150 mM NaCl, 50 mM Tris–HCl (pH 7.4), 0.05% (v/v) Nonidet P-40] was incubated with rabbit anti-7-monomethylguanosine serum or with anti-2,2,7-trimethylguanosine mouse monoclonal antibody (K121) plus rabbit anti-mouse IgG (11) with mixing, for 1 h. The Protein A–Sepharose conjugates were washed three times with 1 ml NET-2 and resuspended in a final volume of 300 µl NET-2. Seventy-five microliters of the Protein A–Sepharose conjugate were combined with the RNA antigen for a final reaction volume of 150 µl. The mixture was allowed to incubate at 4°C with gentle shaking for 2 h. The samples were then microfuged briefly and the resulting supernatant was collected and saved. The immunopellets were washed five times with 1 ml aliquots of NET-2, and resuspended in NET-2. Both the immunosupernatants and the resuspended immunopellets were purified by phenol–chloroform extraction and were subsequently ethanol precipitated.

RNase protection

The 2.5 kb *Hin*dIII fragment containing the *V.necatrix* U2 gene was subcloned into pTZ19U (USB). The T7 and T3 RNA polymerase promoters flanking the insert were used to generate ^{32}P-uniformly labeled RNAs in both the sense (T7 transcript) and antisense (T3 transcript) orientation in the presence of the appropriate polymerase. These RNA transcripts were hybridized in separate experiments to total *V.necatrix* RNA and the unhybridized RNA was digested with RNase A and T1 RNase. The samples were extracted with phenol–chloroform and precipitated with ethanol. Samples were resolved on 8% polyacrylamide gels containing 8 M urea and visualized by autoradiography.

Modification of RNAs

5′ cap structures were removed by periodate oxidation and β-elimination (12). Total RNA (10–15 µg) was incubated in 50 µl 20 mM NaOAc, 50 mM $NaIO_4$ for 90 min at 0°C in the dark. The reaction was terminated with 50 µl ethylene glycol. The $NaIO_3$ formed was removed by microdialysis for 4 h at 4°C. The sample was ethanol precipitated and resuspended in 75 µl water. Twenty-five microlitres 1 M lysine (pH 9.3) was then added, and the sample was incubated in the dark for 4 h at room temperature. The β-elimination reaction was terminated by the addition of 30 µl glacial acetic acid, and the sample was precipitated by the addition of 2.5 vol ethanol. *In vitro* capping was carried out in 10 µl reactions containing 50 mM Tris–HCl (pH 7.9), 1.2 mM $MgCl_2$, 6 mM KCl, 5 U RNasin, 2.5 mM DTT, 2 mM GTP, 0.3 mM adenosyl-L-methionine (AdoMet) and 0.5 U guanylyltranferase (Gibco BRL). The reactions were stopped with 0.5 µl 100 mM EDTA, purified by phenol–chloroform extraction, and ethanol precipitated.

The RNA samples were resolved by aminophenylboronate gel electrophoresis (12). The 5.75% aminophenylboronate gels were prepared by mixing 15 g urea, 4.3 ml acrylamide (40% w/v of a 19:1 acrylamide/bis mixture in water), 3.2 ml 0.92 M Tris-acetate (pH 9.0), and 90 mg *N*-acryloyl-3-aminophenylboronic acid to a

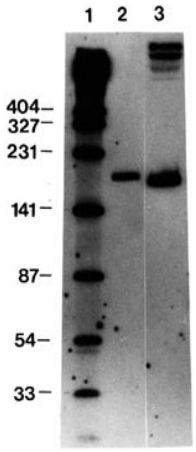

Figure 1. Identification of the *V.necatrix* U2 RNA homolog by Northern blot hybridization analysis. Lane 1, size markers (kinase-labeled *Taq* I fragments of bacteriophage ϕx174; fragment sizes indicated in left margin); lane 2, total rat liver RNA (5 µg); lane 3, total *V.necatrix* RNA (5 µg). RNAs were probed with kinase-labeled U2-L15 oligonucleotide.

final volume of 30 ml. Polymerization was initiated with 450 µl of ammonium persulfate and 18 µl of TEMED.

RESULTS

Demonstration of the presence of a U2 RNA homolog in microsporidia

Our first efforts were focused on the search for a U2 RNA homolog in the microsporidian *V.necatrix*. Comparison of U2 RNAs from a variety of divergent taxa has shown that the sequence extending from nucleotides 20 to 47 is highly conserved (2). This region includes the branch point recognition sequence (GUAGUA), which is present in all known species with the exception of the trypanosomes (13,14). An oligonucleotide probe (U2-L15) complementary to this invariant region (8) was utilized for Northern blot hybridization analysis of microsporidial RNA. As seen in Figure 1, a *V.necatrix* RNA (lane 2) which is signficantly smaller than the rat U2 RNA (lane 3) hybridized to the U2-L15 probe.

The gene encoding the U2 RNA homolog

To characterize further this U2 RNA homolog, we sought to isolate the gene encoding this RNA. A *V.necatrix* genomic library screened with the U2-L15 probe showed several positive plaques. These positive plaques were subjected to further analysis and the putative U2 gene was localized to a 2.5 kb *Hin*dIII fragment by restriction mapping. To ensure that this fragment contained sequence coding for the U2 RNA homolog rather than a pseudogene, plasmid subclone constructs were used to generate radioactive RNA probes in both the sense and antisense orientation. An RNase protection product corresponding in size to the full length *V.necatrix* U2 RNA homolog was observed when the antisense RNA probe was hybridized to total *V.necatrix* RNA, as would be expected for a true gene (data not shown).

Vairimorpha necatrix genomic DNA cut with a variety of restriction endonucleases was analyzed by Southern blot analysis. In all cases a single band was observed to hybridize to a *V.necatrix* U2 RNA homolog probe, indicating that the U2 homolog is coded for by a single copy gene (data not shown). This result also provides further evidence that the genomic clone is not a pseudogene, but codes for the U2 RNA homolog.

Figure 2 shows the DNA sequence of the U2 RNA homolog and flanking regions. The coding region is aligned with human and consensus U2 RNA sequences (2). The 5′ end of the RNA coding region (designated +1) was positioned by comparing the *V.necatrix* RNA sequence, determined separately by RNA dideoxy sequencing, with the genomic DNA sequence (data not shown). The 3′ end of the coding region was mapped (as described in Materials and Methods) to nucleotide +165 of the genomic DNA sequence (data not shown). The only known U2 RNAs which are smaller than the 165 nucleotide *V.necatrix* U2 homolog are those of the trypanosomes (13,14).

Primary and secondary structure of the microsporidian U2 RNA homolog

Examination of the aligned sequences in Figure 2 shows that significant homology is present particularly in the 5′ region between the *V.necatrix* U2 RNA homolog and the human and consensus U2 RNAs. The *V.necatrix* sequence from nucleotides 21 to 43 with the exception of T21, exactly matches the highly invariant nucleotide 24–46 region of the consensus U2 RNA (2). This region includes the GTAGTA (nucleotides 32–38; consensus sequence) branch point recognition sequence found in all U2 RNAs except those of trypanosomes (13,14). The underlined nucleotides (nucleotides 7, 9, 10, 15 and 44) in the 5′ region represent divergence at positions in the *V.necatrix* U2 homolog previously found to be invariant in U2 RNAs (indicated by upper case letters in the consensus sequence). U2 RNAs exhibit a greater degree of divergence in the 3′ half of the molecule than the 5′ half. The *V.necatrix* and human sequences show very limited similarity in the 3′ region. Interestingly, the alignment introduces a gap between nucleotides 94 and 95 of the *V.necatrix* sequence. This corresponds to the location (nucleotides 99–105; human and consensus) of the highly conserved Sm binding sequence ($A(U)_{4-6}G$) found in all U2 RNAs (17) except those of trypanosomes which lack the Sm site (13,14). However, it should be noted that the 3′ region of the *V.necatrix* sequence does contain an ATTTTTTA sequence at nucleotides 132–139. While this sequence is consistent with an Sm binding site, the location of this sequence is quite different from that of the consensus Sm binding site and may be involved in a secondary structure pairing. A conserved sequence element which does appear to be present in the 3′ region is GCTTGC, nucleotides 127–132. This sequence encompasses a putative tetraloop sequence (2) in the stem III region of the consensus U2 RNA (Fig. 3B).

Using computer-assisted free energy minimalization, the *V.necatrix* U2 RNA homolog can be folded into the structure pictured in Figure 3A, possessing a free energy of –29.7 kcal/mol (18–20). The consensus U2 RNA secondary structure is shown in Figure 3B with the characteristic stem–loop structures labeled (I, IIa, IIb, III and IV) (2). The stem–loop structures in Figure 3A which resemble the consensus stem–loop structures are correspondingly labeled (eg. I, IIa, IIb, III). Two other stem–loops in the *V.necatrix* U2 homolog without obvious correspondence to the stem–loops of the consensus model are labeled x and y. Comparison of the stem–loop I of the homolog and the consensus structures shows the former to lack the internal loop and to have an AUC in the apical loop region rather than the uuuu found in the

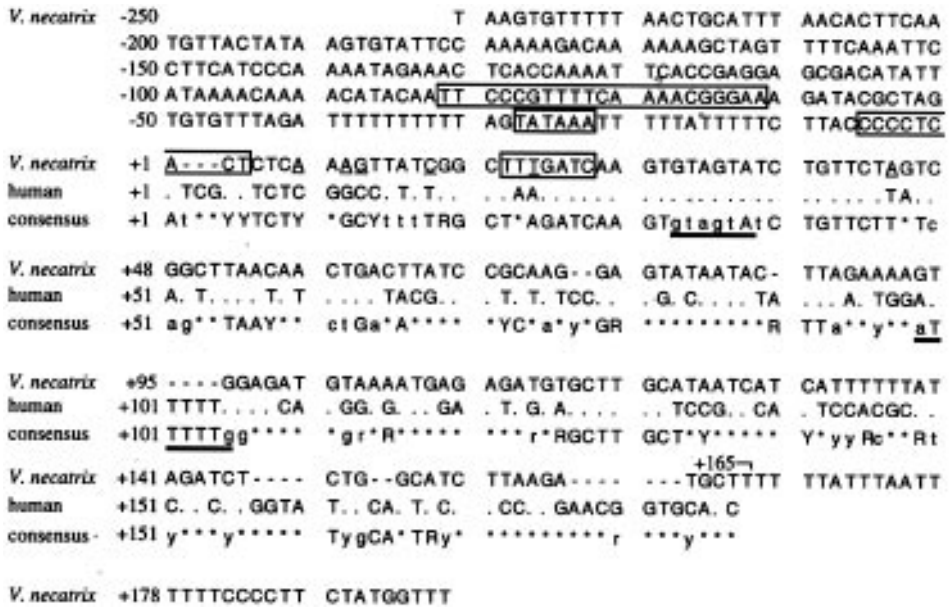

Figure 2. Sequence of the gene encoding the *V.necatrix* U2 homolog and alignment of the *V.necatrix* RNA coding region (+1 to +165) with human and consensus U2 RNA sequences. A 2.5 kb *Hin*dIII fragment of a bacteriophage λ clone containing the RNA coding region was subcloned into M13mp18 and sequenced with oligonucleotide primers U2-L15, U2.2, U2.3 and U2.4. The strand corresponding to the RNA transcript is shown. Dashes (–) represent gaps introduced into the *V.necatrix* coding region to promote alignment with the human and consensus U2 RNAs. Nucleotides in the first 50 bases of the *V.necatrix* coding region which differ from invariant nucleotides of the consensus sequence are underlined. In the human sequence nucleotides which are identical to those of *V.necatrix* are indicated by a period (.). The consensus sequence is from ref. 2 (based on 12 U2 RNAs) Invariant residues are indicated by uppercase letters. Lower case letters represent positions which are identical in all but one species. Asterisks (*) in the consensus sequence represent positions that are variable (including possible deletions or insertions). The branch point recognition sequence (nucleotides 32–38) and Sm binding site (nucleotides 99–105) are bold underlined in the consensus sequence. Potential transcriptional promoter sequence elements are boxed: a TATA box-like sequence centered at –25, a sequence centered at –71 TTCCCGTTTTCAAAACGGGAA which shows perfect dyad symmetry. The 3′ end of the coding region falls within a run of six thymines which is indicative of an RNA polymerase III termination signal. The reverse complements of two sequence elements: –7 to +3 and +19 to +25 (both boxed) are consistent with the Box A and Box B tRNA promoter elements (15,16). The +3 to –7 element, 5′-AGTGAGGGG-3′, corresponds to the box A consensus, 5′-RGYNNRR-GG-3′; and the +25 to +19 element, 5′- GATCAAG-3′, corresponds to the box B consensus, 5′-G(A/T)TCRAN-3′.

consensus structure. The covariant base pair A7/U21 in the *V.necatrix* U2 RNA homolog provides phylogenetic support for the pairing near the base of stem I. The nucleotide corresponding to the *V.necatrix* U21 nucleotide is an adenosine in all other U2 RNAs. As will be discussed later, this phylogenetic substitution has implications regarding U2 RNA interactions with U6 RNA. Structures corresponding to the consensus stem–loops IIa and IIb are present in the homolog. The highly conserved IIa loop region (Fig. 3B) has been proposed to be capable of pseudoknotting by base pairing with a downstream sequence element (21). The IIa loop sequence in the *V.necatrix* model (Fig. 3A) includes the sequence UUAC (nucleotides 52–55) which is consistent with the consensus sequence (UAAY) and could pair with nucleotides 85–88 (UUAG) to form the four base-pair helix, UUAA/UUAG.

The stem–loop III structure of the homolog possesses the highly conserved CUUG loop structure of the consensus structure. Stem–loop y, while corresponding positionally to stem–loop IV, (i.e. 3′ to stem–loop III) does not appear to correspond to the consensus stem–loop IV in terms of its structure. Its small loop size contrasts sharply with the consensus structure's larger loop and therefore is unlikely to be a structural counterpart. However, the homolog sequence encompassing nucleotides 93–150 (stem–loop III) can be folded into an alternative structure which has limited resemblance to the consensus stem–loop IV in terms of its larger loop size (data not shown).

Analysis of the 5′ end structure of the U2 homolog

A 2,2,7-trimethylguanosine hypermethylated cap structure has been observed in all snRNAs thus far examined, with the exception of U6 RNA (3). The microsporidial U2 RNA homolog was tested for the presence of this cap structure with a monoclonal antibody (K121) which binds specifically to the 2,2,7-trimethylguanosine cap structure (11). As seen in the Northern blot shown in Figure 4, the U2 RNA of *Euglena gracilis* was precipitated by the K121 monoclonal antibody (lane 7, immunosupernatant; lane 8, immunopellet). The *V.necatrix* U2 RNA homolog, however, was not immunoprecipitated by the K121 antibody (Fig. 4, lane 2, immunosupernatant; lane 3, immunopellet). To ensure that this was not due to an artefactual inhibition of precipitation in the case of the microsporidian preparation, immunoprecipitation was attempted with a mixture of the two RNA preparations. In the presence of *V.necatrix* RNA, the *E.gracilis* U2 RNA is precipitated efficiently (Fig. 4, lane 4, immunosupernatant; lane 5, immunopellet), indicating that the lack of precipitation of the *V.necatrix* U2 RNA homolog is not due to inhibition. Immunoprecipitation was also attempted with an antiserum specific for the 7-methylguanosine cap structure. The antibody efficiently precipitated globin mRNA, which contains the 7-methylguanosine cap, but did not immunoprecipitate the U2 RNA homolog (data not shown).

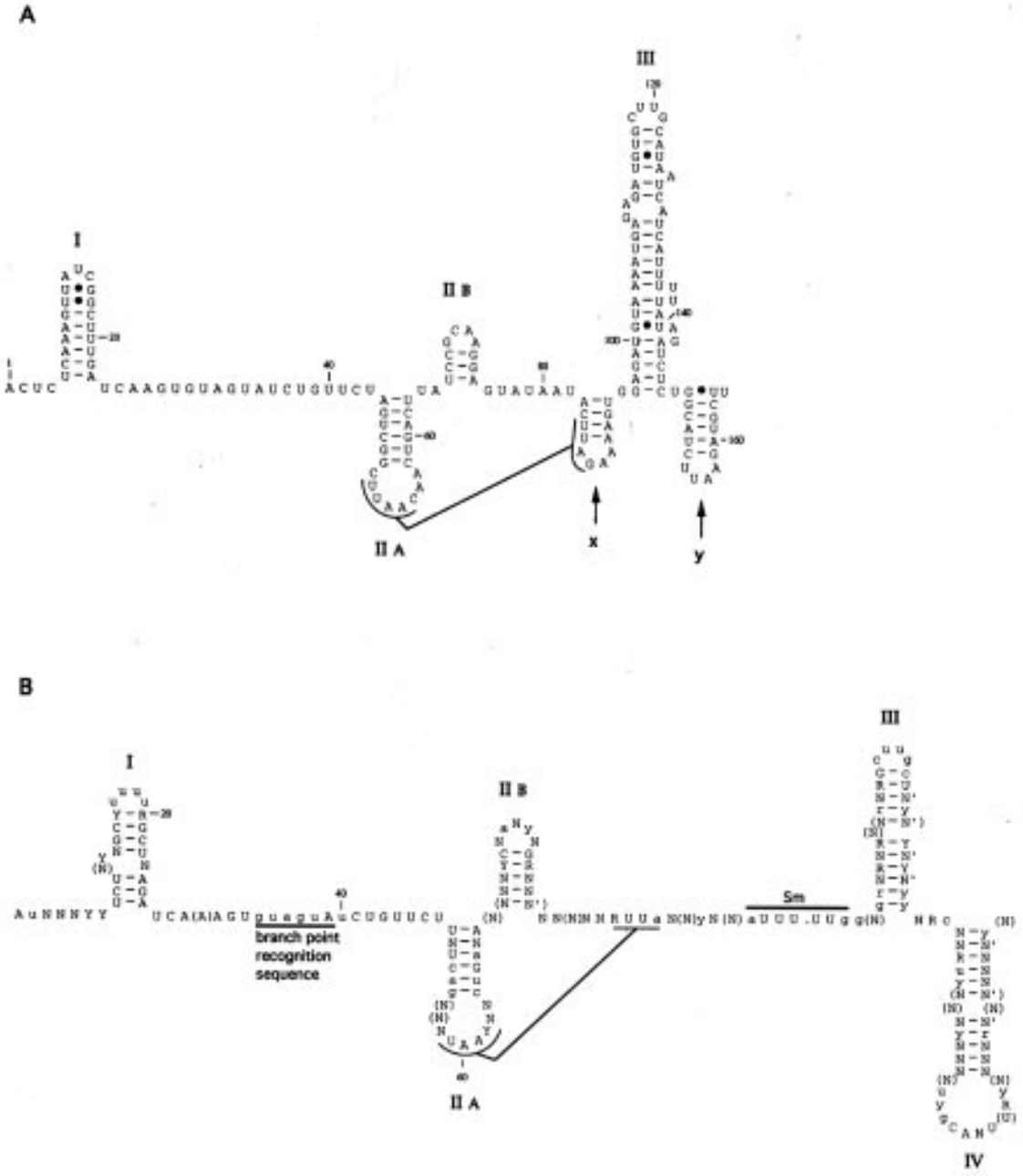

Figure 3. Secondary structure models of the *V.necatrix* U2 RNA homolog and the U2 consensus RNA. **(A)** Secondary structure of the *V.necatrix* U2 RNA homolog. *V.necatrix* U2 secondary structure model was generated by the program Mulfold. Stem–loops which show homology to those of the consensus structure are correspondingly labeled. Other stem–loops not corresponding to those of the consensus structure are labeled x and y. **(B)** U2 consensus secondary structure (numbering is based on the *S.cerevisiae* sequence). The secondary structure is based on the alignment of 12 U2 RNA sequences and is redrawn from Guthrie and Patterson (2). Invariant residues are uppercase. Lowercase letters represent positions which are identical in all of the sequences except one.

While these results appear to rule out the presence of either of these two cap structures in the *V.necatrix* U2 RNA homolog, it is also possible that the putative cap structure is sequestered or otherwise inaccessible to antibodies. To test this possibility, the U2 RNA homolog was exposed to periodate and β-elimination treatment. In RNAs which contain an inverted nucleotide cap structure, this treatment exposes the 5′ triphosphate terminus, which is a substrate for the Vaccinia virus guanylyltransferase capping enzyme (22). Periodate/β-elmininationtreated *V.necatrix* RNA was incubated with guanylyltransferase capping enzyme in the presence of GTP and AdoMet and was subsequently analyzed for immunoprecipitability with anti 7-methylguanosine antibody. The modified *V.necatrix* U2 RNA homolog was immunoprecipitable with this antibody (data not shown). This result suggests that the *V.necatrix* U2 RNA homolog putative cap structure should be accessible to antibodies and that the lack of precipitability of the native *V.necatrix* U2 RNA homolog with either anti-2,2,7-trimethylguanosine or anti-7-methylguanosine antibodies is not due

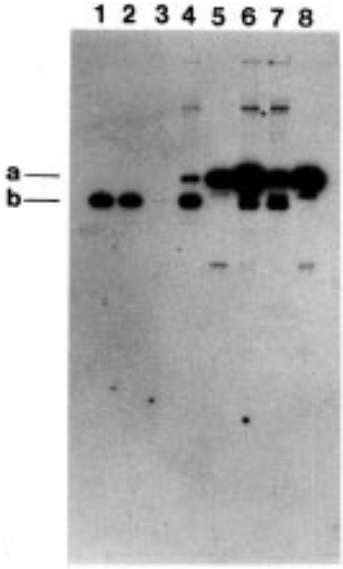

Figure 4. Immunoprecipitability of the *V.necatrix* U2 RNA homolog and of the *E.gracilis* U2 RNA with anti-trimethylguanosine mouse monoclonal antibody (K121). *V.necatrix* and *E.gracilis* total RNA were incubated with K121 antibody as described in the text. RNAs were visualized by Northern blot analysis, using 5′-^{32}P-labeled U2-L15 as a probe. Lane 1, unfractionated *V.necatrix* RNA; lane 2, *V.necatrix* RNA immunosupernatant; lane 3, *V.necatrix* immunopellet. Lane 4, *V.necatrix* + *E. gracilis* RNA immunosupernatant; lane 5, *V. necatrix* + *E. gracilis* immunopellet. Lane 6, unfractionated *E.gracilis* RNA; lane 7, *E.gracilis* RNA immunosupernatant; lane 8, *E.gracilis* immunopellet. a and b indicate the electrophoretic mobility positions of *Euglena* U2 RNA and *V.necatrix* U2 RNA homolog respectively.

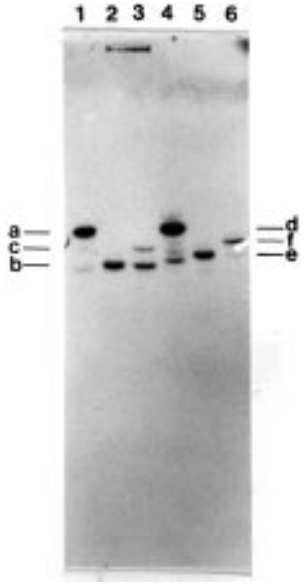

Figure 5. Changes in electrophoretic mobility of the *V.necatrix* U2 RNA homolog and *E.gracilis* U2 RNA after removal of the 5′ cap structures and *in vitro* capping. *Vairimorpha necatrix* and *E.gracilis* total RNA were subjected to periodate and β-elimination treatment. Aliquots of the treated RNAs were then incubated in the presence of guanylytransferase and GTP. Samples were run on an aminophenylboronate gel prepared as described in the Methods. RNAs were visualized by Northern blot analysis using 5′-^{32}P-labeled U2-L15 as a probe. Lane 1, untreated *V.necatrix* RNA; lane 2, periodate/β-elimination treated *V.necatrix* RNA; lane 3, periodate/β-elimination + guanylyltransferase treated *V.necatrix* RNA. Lane 4, untreated *E.gracilis* RNA; lane 5, periodate/β-elimination treated *E.gracilis* RNA; lane 6, periodate/β-elimination + guanylyltransferase treated *E.gracilis* RNA. Band a, *V.necatrix* U2 RNA homolog without treatment; band b, with periodate/β-elimination treatment; band c, with periodate/β-elimination + guanylyltransferase treatment. Band d, *E.gracilis* U2 RNA without treatment; band e, with periodate/β-elimination treatment; band f, with periodate/β-elimination + guanylyltransferase treatment.

to inaccessibility. The fact that after periodate/β-elimination treatment the RNA became a substrate for the capping enzyme suggests that an inverted cap structure may be present in the native *V.necatrix* U2 RNA homolog.

The low level of U2 RNA homolog present in *V.necatrix* and the technical difficulty of *in vivo* labeling with [^{32}P]orthophosphate precluded direct RNA analysis of the cap structure. Therefore, the technique of boronate affinity gel electrophoresis was exploited, allowing detection of cis-diols such as would be found in an inverted nucleoside cap stucture (12). In this technique, aminophenylboronate (APB) derivatives are incorporated into a polyacrylamide gel by copolymerization. The mobility of RNAs through these gels is retarded as a function of the number of cis-diols present in the molecule, which form reversible covalent adducts with the polyacrylamide bound boronates. *Vairimorpha necatrix* RNA and *E.gracilis* RNA were analyzed by this technique. As seen in Figure 5, the U2 RNAs of both preparations increased in mobility to the same degree as a result of periodate/β-elimination pretreatment, which removes nucleosides containing a free cis-diol (e.g. 5′ nucleoside cap and a 3′ unphosphorylated terminal nucleoside; compare lane 1, band a and lane 2, band b for *V.necatrix* U2 RNA homolog and lane 4, band d and lane 5, band e for *E.gracilis* U2 RNA). For both *V.necatrix* and *E.gracilis*, treatment of the oxidized U2 RNAs with guanylyltransferase and GTP resulted in a U2 RNA product with a lower electrophoretic mobility than that of the oxidized U2 RNA (compare lane 2, band b and lane 3, band c for *V.necatrix* U2 RNA homolog and lane 5, band e and lane 6, band f for *E.gracilis*). This mobility shift indicates that a cis-diol containing guanosine was added to the oxidized RNA. Additional experiments showed that the native *V.necatrix* U2 RNA homolog was not a substrate for guanylyltransferase capping enzyme (ruling out a di- or triphosphate 5′ terminus), for the adenylylation reaction of RNA ligase in the presence of ATP (ruling out a monophosphate 5′ terminus), or for polynucleotide kinase (ruling out an unphosphorylated 5′ terminus, data not shown). These results, together with the boronate gel analysis, strongly suggest that a cis-diol containing structure is present at the 5′ end of the native *V.necatrix* U2 RNA homolog. This conclusion rules out the possibility of a γ-monomethyl phosphate cap structure, which has been demonstrated for U6 RNA (23). The fact that the electrophoretic mobility of the RNA products which were oxidized and then treated with guanylyltransferase was greater than that of the original native RNAs also warrants consideration. This result suggests that the original 3′ terminus is unphosphosphorylated and contains a free cis-diol. Periodate/β-elimination treatment would remove this 3′ unphosphorylated nucleoside leaving an RNA containing a 3′ phosphorylated terminus, which would not react with the gel-bound boronates.

DISCUSSION

In this paper we have identified and characterized a homolog of U2 spliceosomal RNA from the microsporidian *V.necatrix*, one of

the most primitive eukaryotic organisms yet identified (5). Indeed, this organism is thought to have diverged from the eukaryotic line prior to the introduction of mitochondria. The presence of a U2 RNA homolog in this organism suggests that the eukaryotic pre-mRNA splicing machinery arose very early in the eukaryotic lineage and may indeed be a universal feature of eukaryotes. The presence of a U2 RNA homolog in this amitochondrial organism suggests that either this organism must have once had mitochondria, or that spliceosomal snRNAs were present prior to the appearance of mitochondria. If the latter is true, it would contradict the hypothesis that spliceosomal snRNAs arose from fragmentation of group II self-splicing introns (24). In contrast, in the amitochondrial organism *Giardia lamblia*, no homologs to spliceosomal snRNAs appear to be present, although several trimethylguanosine capped nucleolar snRNA homologs have been identified (25).

The U2 RNA homolog of *V.necatrix* is relatively small (165 nt) in relation to other U2 RNAs with only those of trypanosomes seen to be smaller (13,14). As would be expected, the RNA possesses primary sequence homology to U2 RNAs primarily in the more highly conserved 5′ half of the molecule. The *V.necatrix* homolog contains the highly conserved GUAGUA branchpoint recognition sequence which is involved in a base pairing interaction with the pre-mRNA and is seen in all U2 RNAs except for those of the trypanosomes (13,14). The functionality of the U2 homolog in splicing will be predicated on ascertaining the presence of introns in specific *V.necatrix* genes and the mechanism of their removal. This is a subject which still awaits investigation. If splicing does occur, the question of whether homologs to the other requisite spliceosomal snRNAs are present is also pertinent. In preliminary experiments we have detected the presence of a number of small RNAs in the 80–200 nucleotide size range which contain 5′ cap structures. Among these is an RNA whose partial sequence shows homology to U6 RNA (preliminary results).

A highly conserved U2 RNA sequence which has been implicated in a base pairing interaction with U6 RNA (helix I) in *Saccharomyces cerevisiae* (26) is present in the U2 homolog. The U2 homolog sequence GCUUUGAUCA (positions 17–26) is identical, with the exception of U21, to the *S.cerevisiae* sequence GCUUAGAUCA (positions 21–30). The underlined portions of the *S.cerevisiae* sequence pair with U6 RNA to form the helix I structure, with U24 and A25 forming a two nucleotide bulge. The *V.necatrix* sequence matches the *S.cerevisiae* sequence exactly in the paired regions. However, the invariant bulged A25 of the yeast corresponds to U21 in the *V.necatrix*. This finding is consistent with the apparent tolerance of *in vitro* or *in vivo* splicing to substitution at the A25 site observed in *S.cerevisiae* (27,28). This nucleotide in yeast is thought to participate in a tertiary interaction with the G52 of U6 RNA (28). Identification and characterization of a *V.necatrix* U6 homolog would be necessary to ascertain whether a pairing homologous to the above U2:U6 pairing occurs in *V.necatrix*. As mentioned above, we have preliminary evidence for the existence of a U6 RNA homolog.

The *V.necatrix* U2 RNA homolog lacks the canonical Sm binding site sequence (Fig. 3) which occurs in all U2 RNAs except for those of trypanosomes (13,14). Future experiments will determine whether polypeptides with Sm antigenicity are associated with the U2 RNA homolog. Preliminary experiments have shown the U2 RNA homolog is not immunoprecipited from crude *V.necatrix* extracts with anti-Sm antisera (unpublished data). This preliminary result may be due either to the absence or extreme heterologous nature of Sm polypeptide homologs associated with the *V.necatrix* U2 RNA homolog.

The U2 RNA homolog can be folded into a structure which has partial similarity to the consensus structure proposed for U2 RNAs (2). The folded structure model of the U2 homolog, which was generated based on energy minimalization, possesses three of the four principal stem–loops proposed for the U2 RNA consensus structure (Fig. 3). The stem–loop I structure of the U2 RNA homolog resembles its consensus U2 counterpart in size and location. The U2 homolog loop sequence (AUC), however, differs from the uridine rich (uuuu) of the consensus U2 (Fig. 3). This characteristic is consistent with the loop region being functionally tolerant to mutation in yeast (27). Phylogenetic support for the pairing near the base of stem I is provided by the A7:U21 covariant base pair seen in the *V.necatrix* U2 homolog. These nucleotides, which correspond to U10:A25 of the consensus sequence, were previously found to be invariant (2). Structures corresponding to the consensus stem–loops IIa and IIb are also present in the U2 homolog. Mutational analysis of stem–loop IIa has been implicated it to have a role in at the level of U2 snRNP or spliceosome assembly (29). As seen for U2 snRNAs (27), potential pairing of the U2 homolog IIa loop structure with downstream sequence to form a pseudoknot is possible. Based on energy minimalization, the 3′ region of the molecule is folded into a stem–loop III like structure. Alternatively, the 3′ region can be folded into a structure which has some resemblance to a stem–loop IV domain. These secondary structure alternatives are also encountered in the 3′ region of an *S.cerevisiae* U2 RNA functional deletion derivative (30). In *Xenopus* oocytes it has been shown the stem–loop IV domain is necessary for the binding of U2 snRNP specific proteins (31). The 3′ region of the trypanosomal U2 RNAs contain a stem–loop IV like structure but lacks a stem–loop III structure (13,14). However, the trypanosomal U2 RNAs may not be strict functional homologs to other U2 RNAs since apparently only trans-splicing occurs in these organisms.

One of the most unexpected findings of this work is the presence of a cis-diol containing cap structure on the U2 homolog which is neither 2,2,7-trimethylguanosine nor 7-methylguanosine. Trimethylguanosine cap structures are present in all spliceosomal snRNAs except U6, which has a γ-monomethyl phosphate cap (23). For trimethylguanosine caps, the *N*-7 methylation occurs in the nucleus and the additional *N*-2 methyl groups are acquired after transport of the snRNA into the cytoplasm (32). The lack of a trimethylguanosine cap is apparently not restricted to the U2 RNA homolog in *V.necatrix*. Immunoprecipitation experiments conducted with 3′-end-labeled total *V.necatrix* RNA indicate that no other trimethylguanosine capped RNAs are present (unpublished data). The apparent lack of an Sm site in the U2 RNA homolog may be related to the lack of the trimethylguanosine cap. In *Xenopus* oocytes it has been shown that an Sm site and associated Sm proteins constituting the Sm core are a prerequisite for cap hypermethylation (32). *Trypanosoma brucei* U2 RNA, however, which lacks the Sm site, has a trimethylguanosine cap (13). The hypermethylated cap structure has been shown to have a role in intracellular snRNA transport (33–35) However, the identification of the cap structure in the U2 RNA homolog and other snRNA homologs will be a necessary first step in ascertaining any such functional role for the cap. It will also be of interest to examine whether anomolous cap structures are present in microsporidial mRNAs.

ACKNOWLEDGEMENTS

We thank Dr Adrian Krainer for kindly providing the anti-trimethylguanosine antibody (K121 monoclonal), Dr Elizabeta Ullu for kindly providing the anti-7-methylguanosine antisera, and Dr William Mason for providing *N*-acryloyl-3-aminophenylboronic acid. We thank Dr S. Ivey for his comments on the manuscript and especially thank Dr Carl Woese for his advice and input throughout the course of this study. This work was supported by grants to P.D. from the NIH (1R15 GM39980-01) and NASA (NAG-248; JOVE program) and to C.V. from the NIH (AI36153).

REFERENCES

1 Green, M.R. (1991) *Annu. Rev. Cell Biol.*, **7**, 559–599.
2 Guthrie, C. and Patterson, B. (1988) *Annu. Rev. Genet.*, **23**, 387–419.
3 Reddy, R. and Busch, H. (1988) In Birnstiel, M. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer Verlag, Berlin, pp. 1–37.
4 Madhani, H.D. and Guthrie, C. (1994) *Annu. Rev. Genet.*, **28**, 1–26.
5 Vossbrinck, C.R., Maddox, J.V., Friedman, S., Debrunner-Vossbrinck, B.A. and Woese, C.R. (1987) *Nature*, **326**, 411–414.
6 Vossbrinck, C.R. and Woese, C.R. (1986) *Nature*, **320**, 287–288.
7 Vossbrinck, C.R., Baker, M.D., Didier, E.S., Debrunner-Vossbrinck, B.A. and Shadduck, J.A. (1993) *J. Euk. Microbiol.*, **40**, 354–362.
8 Ares, M. (1986) *Cell*, **47**, 49–59.
9 Sambrook,J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (2nd ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10 Ares, M., Mangin, M. and Weiner, A. (l985) *Mol. Cell. Biol.*, **5**, 1150–1160.
11 Krainer, A.R. (l988) *Nucleic Acids Res.*, **16**, 9415–9428.
12 Igloi, G. and Kossel, H. (1985) *Nucleic Acids Res.*, **13**, 6881–6899.
13 Tschudi, C., Richards, F. and Ullu, E. (l986) *Nucleic Acids Res.*, **14**, 8893–8903.
14 Hartshorne,T. and Agabian,N. (1990) *Genes Dev.* **4**, 2121–2131.
15 Ciliberto, G., Raugei, G., Costanzo, F., Dente, L. and Cortese, R. (1983) *Cell*, **32**, 725–733.
16 Sharp, S., DeFranco, D., Dingermann, D., Farrell, P. and Soll, D. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 6657–6661.
17 Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Haendler, B. and Jacob, M. (1982) *EMBO J.*, **1**, 1259–1265.
18 Zuker, M. (1989) *Science*, **244**, 48–52.
19 Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 7706–7710.
20 Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) *Methods Enzymol.*, **183**, 281–306.
21 Ares, M. and Igel, A.H. (1989) in *UCLA Symp. Mol. Cell Biol.*, **94**, 13–23.
22 Monroy, G., Spencer, E. and Hurwitz, J. (1978) *J. Biol. Chem.*, **253**, 4490–4498.
23 Singh, R. and Reddy, R. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 8280–8283.
24 Cavalier-Smith, T. (1993) *Microbiol. Rev.*, **57**, 953–994.
25 Nie, X.H., Hartshorne, T., He, X.Y. and Agabian, N. (1994) *Mol. Biochem. Parasitol.*, **66**, 49–57.
26 Madhani, H.D. and Guthrie, C. (1992) *Cell*, **71**, 803–817.
27 McPheeters, D.S. and Abelson, J. (1992) *Cell*, **71**, 819–831.
28 Madhani, H.D. and Guthrie, C. (1994) *Genes Dev.*, **8**, 1071–1086.
29 Zavanelli, M.I. and Ares, M. (1991) *Genes Dev.*, **5**, 2521–2533.
30 Igel, A.H. and Ares, M. (1988) *Nature*, **334**, 450–453.
31 Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J. and Mattaj, I.W. (1990) *Nature*, **345**, 502–506.
32 Mattaj, I.W. (1986) *Cell*, **46**, 905–911.
33 Fischer, U. and Luhrmann, R. (1990) *Science*, **249**, 786–790.
34 Fischer, U. Darzynkiewicz, E. Tahara, S.M., Dathan, N.A., Luhrmann, R. and Mattaj, I.W. (1991) *J. Cell Biol.*, **113**, 705–714.
35 Marshallsay, C. and Luhrmann, R. (1994) *EMBO J.*, **13**, 222–231.