# Intracellular localization and unique conserved sequences of three small nucleolar RNAs

N. Selvamurugan, Oscar H. Joost, Elizabeth S. Haas<sup>1</sup>, James W. Brown<sup>1</sup>, Nancy J. Galvin and George L. Eliceiri\*

Department of Pathology, St Louis University School of Medicine, St Louis, MO 63104-1028, USA and <sup>1</sup>Department of Microbiology, North Carolina State University, Raleigh, NC 27695-7615, USA

Received November 18, 1996; Revised and Accepted February 26, 1997

DDBJ/EMBL/GenBank accession nos U64695–U64709

## ABSTRACT

Three human small nucleolar RNAs (snoRNAs), E1, E2 and E3, were reported earlier that have unique sequences, interact directly with unique segments of pre-rRNA *in vivo* and are encoded in introns of protein genes. In the present report, human and frog E1, E2 and E3 RNAs injected into the cytoplasm of frog oocytes migrated to the nucleus and specifically to the nucleolus. This indicates that the nucleolar and nuclear localization signals of these snoRNAs reside within their evolutionarily conserved segments. Homologs of these snoRNAs from several vertebrates were sequenced and this information was used to develop RNA secondary structure models. These snoRNAs have unique phylogenetically conserved sequences.

### INTRODUCTION

Processing of rRNA precursors requires several small nucleolar RNA (snoRNA) species (reviewed in 1,2). E1/U17 (3–8), E2 and E3 (3–6) snoRNAs do not belong to the main class of snoRNAs, since they lack the C and D sequence boxes that are present in most snoRNAs, do not show substantial sequence homology with any other snoRNA and do not associate with the nucleolar protein fibrillarin. They are housekeeping RNAs, since they are present in all tissues tested (3). These three snoRNAs may play as yet undetermined roles in ribosome biogenesis, since they interact directly (psoralen crosslink) with unique segments of pre-rRNA *in vivo* (4). E1, E2 and E3 RNAs do not have any of the sequences that are known to be nuclear or nucleolar localization signals for other small nuclear RNAs (9–15). E1 (5,7,8), E2 (16) and E3 (17,5) RNAs, among other snoRNAs (1), are encoded in introns of protein genes.

The intracellular localization and transport of a given RNA species are essential to its function in the cell and are important to the understanding of its various functions and interactions *in vivo* (reviewed in 18–20). Nothing is known about how these three snoRNAs localize in the nucleus and nucleolus. Toward that long-term goal, we injected human and frog versions of these snoRNAs into the cytoplasm of frog oocytes and then monitored their intracellular localization. To study how the various functional domains of an RNA species function, it is important to

identify its evolutionarily conserved nucleotides. We have determined the sequences of these snoRNAs from several vertebrates.

## MATERIALS AND METHODS

#### **General methods**

The following procedures were as described before: cDNA synthesis with avian myeloblastosis virus reverse transcriptase (4); polymerase chain reaction (PCR) amplification of cDNA and genomic DNA (21); thermal cycle DNA sequencing of PCR products (22). Unless indicated otherwise, human E1, E2 and E3 RNA end primers were used for cDNA synthesis and for PCR amplification of cDNA and liver genomic DNA.

#### Microinjection into frog oocytes

PCR amplification was used to synthesize DNA fragments that had a bacteriophage T7 RNA polymerase promoter and fulllength frog E1, E2 or E3 RNA sequences. The frog E1 DNA template was made from a genomic clone (8); the frog E2 and E3 DNA templates were made from cDNA. *Xenopus laevis* has six potential genes for E1 RNA; we used the *f* sequence, since it has been shown to be expressed (8). RNA synthesis *in vitro* was in the presence of  $[\alpha^{-32}P]$ UTP or  $[^{3}H]$ UTP, as indicated, and the cap analog m<sup>7</sup>G(5')ppp(5')G, to cap the 5'-end (23).

 $^{32}$ P-Labeled snoRNAs were injected into the cytoplasm of *X.laevis* oocytes, which were then incubated at 19°C for 20 h.  $^{32}$ P-Labeled frog oocytes were fractionated under oil into nucleus and cytoplasm (24). Their RNA was extracted and then fractionated by 10% polyacrylamide gel electrophoresis.  $^{3}$ H-Labeled snoRNAs were injected into the cytoplasm of frog oocytes. After 20 h incubation, the oocytes were fixed in formaldehyde, embedded in glycol methacrylate and 2 µm sections were used for autoradiography. Exposures were at 4°C, using autoradiography emulsions Kodak NTB-2 and Ilford K.5D.

### Sequencing of RNA ends

The sequences of the 5'-terminus of *X.laevis* E2 and E3 RNAs were determined by 5'-RACE (rapid amplification of cDNA ends) (25). Complementary DNA was synthesized using a primer corresponding to a conserved internal E2 or E3 RNA sequence. The product was digested with RNase H and purified by gel

<sup>\*</sup>To whom correspondence should be addressed. Tel: +1 314 577 8496; Fax: +1 314 577 8489

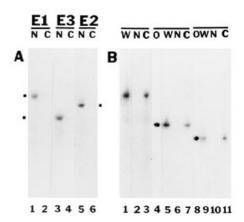
electrophoresis. A tail was added to cDNA with terminal deoxynucleotidyl transferase and dATP. PCR amplification of tailed cDNA was with an anchored dT primer corresponding to another conserved E2 or E3 RNA sequence. The PCR product was purified by gel electrophoresis and its sequence determined by thermal cycle sequencing. The sequences of the 3'-end of frog E2 and E3 RNAs were determined by the following procedure. Cellular RNA was ligated with T4 RNA ligase to an oligodeoxynucleotide that was phosphorylated at the 5'-terminus and was blocked at the 3'-end with cordycepin [SP6Reco, ATAGTGTCACCTAAATGAATTCC(3'-dA)] (26). The product was precipitated with isopropanol in the presence of ammonium acetate. A second 5'-end phosphorylated oligodeoxynucleotide (SP6eco, GGAATTCATTTAGGTGACACTAT), complementary to SP6Reco, was used for cDNA synthesis of the ligation product using reverse transcriptase (26). The product was digested with RNase H and precipitated with isopropanol in the presence of ammonium acetate. The cDNA was amplified by PCR using the SP6eco primer and a primer corresponding to an internal E2 or E3 RNA conserved sequence. The PCR product was gel purified and sequenced by thermal cycle sequencing.

### **RNA** secondary structure

Models for the secondary structures of E1, E2 and E3 RNAs were constructed by a combination of the phylogenetic comparative method (27) and thermodynamic prediction (28). Sequences were aligned manually and searched by computer for conserved potential pairings and co-variation of sequence changes using COVARIA-TION (29). Potential secondary structure elements were predicted thermodynamically using MULFOLD (28) and sorted for consistency between sequences and with the comparative data.

## RESULTS

E1, E2 and E3 RNAs need mechanisms to localize in the nucleolus and nucleus. First, the mature forms of these snoRNAs remain continuously in the nucleolus in interphase cells, instead of being scattered throughout the cell. Second, both the nucleolus and the nuclear membrane break down during mitosis. Then, whether the association of these mature snoRNA molecules with nucleolar structures is interrupted or not in mitosis, there has to be a mechanism to restore or maintain this association. Finally, the pre-mRNA transcription and processing steps that generate these snoRNAs occur at nuclear sites outside the nucleolus. A mechanism is needed to transport these newly made snoRNAs to the nucleolus. We asked first if these snoRNAs, when placed in the cytoplasm, can migrate to the nucleus and, if so, whether the nuclear localization signals of these three snoRNAs are conserved between human and frog. Injections were into the cytoplasm because (i) the mature forms of these snoRNAs are in the cytoplasm during mitosis, (ii) we are interested in both the nuclear and nucleolar localization signals of these snoRNAs and (iii) cytoplasmic injections are less damaging to the oocyte. In vitro transcribed human E1, E2 and E3 RNAs localized in the nucleus after they were injected into the cytoplasm of frog oocytes (Fig. 1A). This localization is RNA sequence specific, since antisense transcripts of frog E1, E2 and E3 RNAs did not migrate to the nucleus (Fig. 1B). To minimize degradation, both the snoRNAs and antisense transcripts were capped at the 5'-end

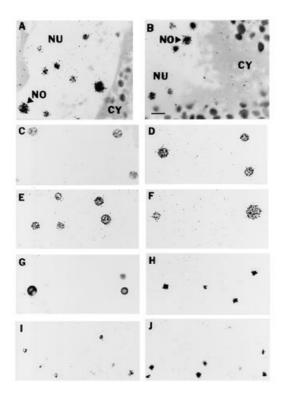


**Figure 1.** E1, E2 and E3 RNAs, but not their antisense transcripts, specifically localize in the nucleus after they are injected into the cytoplasm of frog oocytes. (**A**) <sup>32</sup>P-Labeled human E1, E2 and E3 RNAs were injected into the cytoplasm of frog oocytes, which were then incubated. The oocytes were fractionated into germinal vesicles (nuclei) (N) and cytoplasm (C) and their RNA was extracted and fractionated by gel electrophoresis. Similar cell equivalents were loaded in all lanes. The dots indicate the electrophoretic mobility of the radiolabeled, capped snoRNAs before injection into oocytes. (**B**) Labeled antisense transcripts complementary to frog full-length E1 (lanes 1–3), E2 (lanes 4–7) and E3 (lanes 8–11) RNAs were injected and analyzed as in (A). Lanes 4 and 8 show the original RNA samples before injection (O). Lanes 1, 5 and 9 show RNA samples isolated from whole oocytes (W).

with 7-monomethylguanosine. This cap is not a nuclear localization signal (9,10) and had no effect on this transport, since capped antisense transcripts remained in the cytoplasm (Fig. 1B).

Nucleolar localization domains tend to be more complex than nuclear localization domains (30,31). We asked next if these snoRNAs, when injected into the cytoplasm, migrate to the nucleolus. If that was the case, our other question was whether the nucleolar localization signals of these three snoRNAs are conserved between human and frog. Intranuclear distribution was monitored by cell microscopy autoradiography. Whole oocytes were sectioned to minimize losses of non-organellar nucleoplasmic components. Human and frog E1, E2 and E3 RNAs localized in the nucleolus after they were injected into the cytoplasm (Fig. 2). Figure 2G, without many autoradiography silver grains, shows the typical appearance of the many nucleoli present in X.laevis stage 5 and 6 oocytes (32). The clustering of silver grains is easier to see in the absence of staining (Fig. 2H-J). Staining shows that it co-localizes with nucleoli (Fig. 2A-F). The nucleus (germinal vesicle) of a X.laevis stage 5 or 6 oocyte contains ~1500 large (15-20 µm diameter) nucleoli, located mainly in the outer region of the nucleus. B snurposomes are smaller (1-4 µm diameter) and are scattered all over the oocyte nucleus. Sphere organelles are fewer (50-100 per oocyte nucleus), are located primarily in the center of the oocyte nucleus and have a different appearance (a smaller sphere with two B snurposomes on its surface) (reviewed recently in 19). Our cell microscopy autoradiography experiments show silver grain clusters co-localized with every one of the many large bodies located in the outer region of the frog oocyte nucleus (Fig. 2 and data not shown). It is clear that most, if not all, of the nuclear structures where this autoradiographic signal localizes are nucleoli.

Next we wanted to identify the evolutionarily conserved sequences of these snoRNAs, since the results in Figures 1 and 2 indicate that their nuclear and nucleolar localization signals are conserved between amphibians and primates and this information



**Figure 2.** Human and frog E1, E2 and E3 RNAs localize in the nucleolus after they are injected into the cytoplasm of frog oocytes. <sup>3</sup>H-Labeled frog E2 (**A**), E3 (**B**) and E1 (**C**) RNAs, human E1 (**D** and **H**), E2 (**E** and **I**) and E3 (**F** and **J**) RNAs and the antisense transcript of frog E3 RNA (**G**) were injected into the cytoplasm of frog oocytes. For each transcript, similar amounts of RNA and radioactivity were loaded per oocyte. After incubation, the oocytes were fixed, embedded, sectioned and then exposed to autoradiography emulsion for 5 days (A–G) or 3 months (H–J). Some slides were then stained with hematoxylin and eosin (A and B) or toluidine blue (C–G). Other slides were not stained (H–J). NU, nucleus; NO, nucleoli; CY, cytoplasm. The bar represents 10 µm.

would also be useful to study other functional domains of these snoRNAs. Since the cellular levels of these snoRNAs are low (3), we determined their sequences after PCR amplification of cDNA and genomic DNA. When only genomic sequences are available, they are apparently expressed sequences, since they differ from the RNA sequences of other organisms in similar nucleotide positions (Fig. 3). Since the E3 RNA gene resides in intron 8 of the human and mouse protein synthesis initiation factor 4AII (eIF-4AII) gene (17,5), primers corresponding to the flanking exons were used for PCR amplification of genomic DNA from various vertebrates. Sequencing of the PCR products showed the E3 RNA sequence in intron 8 of the eIF-4AII gene in all vertebrates tested except frog (Fig. 3). The nucleotide sequence of the internal section of the frog PCR product could not be determined directly and it was necessary to clone it first. The sequence of one of these plasmid clones showed substantial homology to exons 8 and 9 of the eIF-4AII gene, but E3 sequence homology cannot be detected in between (Fig. 3). These results indicate that the E3 RNA gene has remained in intron 8 of the eIF-4AII gene in mammals and birds, but not in amphibians. There is substantial sequence conservation from mammalian to fish E1 and from mammalian to amphibian E2 and E3 RNAs. Some segments of the fish and frog E1 RNA sequences are absent in mammals and chicken. These segments have the same length in fish and frog and some of their nucleotides are conserved. They

include a 6 base sequence between human E1 positions 124 and 125 and a 9 base segment between positions 129 and 130 (Fig. 3).

Since these snoRNAs have unique conserved sequences (Fig. 3), the next question was where these segments may be in their possible secondary structures. There were no models of E3 or E2 RNA secondary structure and the only E1 RNA secondary structure models (8,33) were based solely on thermodynamics, an approach that has important limitations. We have used the available sequences to construct working models for the secondary structures of these three snoRNAs (Fig. 4). Thermodynamic predictions were used to guide the comparative analysis of structure. The model of E1 RNA is supported by a number of phylogenetic 'co-variations' (concerted sequence changes that maintain complementarity within a predicted helix). Co-variation at two positions in an uninterrupted helix is generally accepted as 'proof' that a helix exists (34). On this basis, three of the helices in the E1 RNA secondary structure model are proven. The model is also supported by several isolated co-variations and numerous instances of single base changes that retain complementarity through the use of G · U pairings that are consistent with the predicted structure. All of the helical elements in this model are present in structures predicted thermodynamically (e.g. within 10% of the minimum free energy at 37°C) for each sequence individually. Several base paired segments in previous models of E1 RNA secondary structure (8,33) are not compatible with the sequence variation data (Fig. 4). The E3 RNA secondary structure model is supported by two phylogenetic co-variations in separate helices and there are numerous instances of single base changes that retain complementarity. In the model of E2 RNA, one helix is proven by phylogenetic co-variation of 2 base pairs and several others are supported by co-variation of individual base pairs, as well as single base changes that maintain complementarity. Thermodynamic predictions using the available complete and partial sequences are consistent with the structure model; none of the alternatives predicted thermodynamically for any one of the sequences are feasible in the others. All of the helical elements are predicted thermodynamically for each sequence independently. These are the first secondary structure models of E2 and E3 RNAs. The available results suggest that these three snoRNAs have different secondary structures.

In psoralen crosslinking experiments in vivo, there are psoralen adducts at several nucleotide positions of E1, E2 and E3 RNAs that may be crosslinking sites to pre-rRNA (4). Antisense oligodeoxynucleotide-targeted degradation in cell extracts (4) and in frog oocytes (R.Mishra and G.Eliceiri, submitted) revealed accessible snoRNA segments in ribonucleoprotein particles. Some of these accessible sections or possibly crosslinking sites are in evolutionarily conserved sequences located in apparently conserved single-stranded snoRNA regions (Fig. 3). E1, E2 and E3 and other snoRNAs that lack C and D boxes have the sequence ACA near their 3'-ends; the ACA box is required for accumulation and 3'-end formation of a yeast non-intronic snoRNA, snR11 (35). The ACA sequence is conserved and in an apparently conserved single-stranded segment of these three snoRNAs in vertebrates (Fig. 4). The present models of vertebrate E1, E2 and E3 RNA secondary structure (Fig. 4) differ from the model of yeast snR11 snoRNA secondary structure (34).

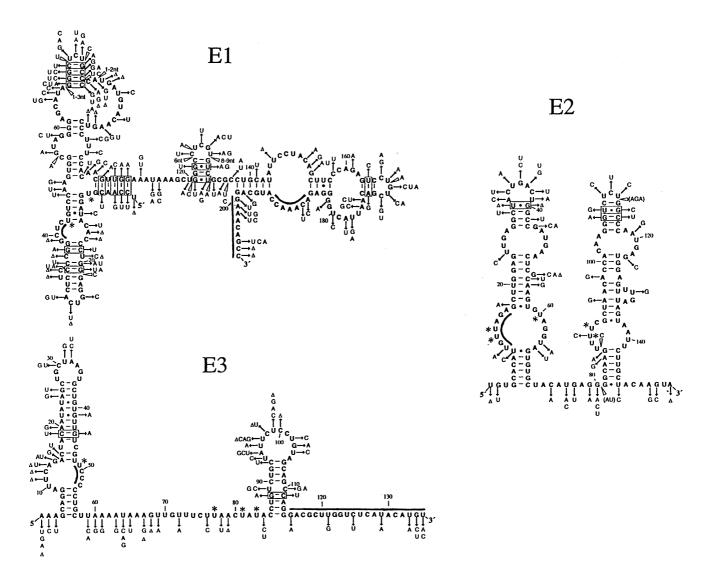
#### DISCUSSION

E1, E2 and E3 RNAs are expected to have novel nucleolar localization signals, since they lack the known nuclear or

			E3	exon 8 intron 8	
E1	• • • • •		human	CUGAGAAGAUGCAUGCCAGAGACUUCACAGUUUCUGCUCUGFUAAGAGGUGUUCUAA	57
human		59	rabbit	-*************************************	
rat	**************************************		Tac	*G*****A******************************	
mouse	***************************************			-A*****A******************************	
rabbit chicken	**************************************				
frog a	**************************************		-		
frog b	U***U*******U****U_*****U*****AU_**********		_	· · · · · · · · · · · · · · · · · · ·	
frog c	****U*******U***UA_******U***UU_********		human	AAUGUCUG-GAUUUCCACUAAAGCAGGAUUCAGACUAC	94
frog d	*************U*UCAU_*****U****AU_**********UUG*****_***U		rabbit	CUUAACAUU**CA**A*-U**GUA**A**U*G**************	
frog e	************U*UCAU_*****U*****AU_*******UUG*****		rat mouse	GAUUUAGU****CUAAUU**A	
frog f	************U*UCAU-*****U*****AU-A*******UG******UG*****		chicken	CACAGCUUCUGA****G*A*CU*CAUAC*GAAAAG**A*GC**********	
zebrafish	. U		frog RNA	CU******UAGU***	
puffer c	*UG******C*U**-A*C***C*-G******U-**-******UG*****-***U			CCATAGCAACCAC*CC*AGCAACCCCCCG*CACTGGGC***G*CCTG**TTGTG**	
puffer d	**G*******C*U**-A*C***C*-G******U-**-********************				
puffer e puffer f	**G******C*U**-A*C***C*-G*****U-**-*********************				
putter 1	G C 0		human	AAUAUAGCUGCUAAGUGCUGUGUGUGUGCGUUCCCCCUGCUUAAAAUAAAGUUGUUUCUUAA	154
			rabbit	**************************************	
human	GGGAGCAUAGGGC-UCU-GCC-CCAUGAUGUACAAGUCCCUUUCCACAACGUUGG	110	mouse	***************************************	
rat	*******=****-A**_*A**A*********G_**********		chicken	G*********UGC*******A**A****************	
mouse	******U******-A**-**GG*******G-**********		frog RNA	*****U***GC****************************	
rabbit	******C*****_***_*G****C*****G_*****C*****GC******		frog DNA	TCGGCTT**TA**TC**GC*G*GGCAGAGGAAAAG*C*A*GTTCATCCCAGCGCAGCCCC	
chicken				A	
irog a	******G*CGCCCU-GUG-ACGCG**-G***C***U**-****C****************		<b>.</b>		
	**************************************		rabbit	CUAUACCUGUCUGCUAUUCUCCUGUAGCAGCCAGGGACGCUUGGUCUCAUACAUGUU-GA	213
frog d	*****G*CGCCCU-GUG-AGGCG**-G***C***U**-****C****C****C****			**************************************	
frog e			mouse	**************************************	
frog f	******G*CGCCCU-GUG-AGGCA**-G***C***U*****C**************		chicken	*****C****AC******UAGU****CAC*****GU*******G***U**********	
zebrafish	*****G*CAA*CU-**G-AGG-UGUUG***C*****G-****G*****UG******		frog RNA	****U****G***U*C***A*****A****A****A**********	
puffer c	******G**GCUUUUCAGCAGA-G*C*****U-****G******C***CA*		frog DNA	**G*G*T**A*GA**GG*TCATGCAGTGG*GACAT***CG****GT*AT*GAGGAG	
puffer d	*****G**GCUUU*CAGCAGA-G*C*****U-****G*******C***C**				
	*****G**G-CUUUUCAGCAGA-G*C*****U-****G******C***C***C*** *****G**GCUUUUCAGCAGA-G*C*****U-****G*******************		human		
puller i	******G***G***C***C*******************		rabbit	UUAAAAUUAAAUACCGACCUGACUGGUGAUUCUUGAAUUAAGGUUUAU *****C°G*GACCUGACUGGUGAUUCUUGAAUUAAGGUUUAU	261
	• • • •		rat	AAG*******GAAUGCCU******A*A*G****GC*****G**-*G**-**G**	
human	AAAUAAAGCU-GGGCC-UCGUGUCU-GCG-CCUGCAUAUUCCUAC	151	mouse	AAUU**A*GU**G**UGAAUAGUAUG*****A*A*A********	
rat	*G**G*****-*****		chicken	ACUUCUGA*G***UUCUGGUCCUG***UACU*GA****CCAUGG***UUC*CAUGU*	
mouse	*G**G*****_****U-***A****-***-*********		frog DNA	***TCTCAC*CCTTG**CTTC**	•
rabbit	*G*GG*********-***A****-***-			intron 8 exon 9	
chicken	*G**G*****_**************_******		human	UA-AUUUGCAGCAUGGUGACAUGGACCAGAAGGAGAGAGAGUGUUAUCAUGAGGGAAUU	318
frog a frog b	*G*AC****A-U*U*CCCGGC*A*UCGGCUGGUGUG*G*A-U*AC****U******** *G*AC****A-U*U*CUCGGC*A*UAGGCUGUUGUG*G*A-***U************************		rabbit	*UU*********************************	1 218
frog c	*G*AC****A-U*U*CCCAGC*A*UAGGCUGUGUG*G*A-**AU****U********************		rat	GU*********************************	
frog d			mouse	*GC**********************************	
frog e	*G*AC****A-U*U*CCCGGC*A*UCGGUUGGUGUG*G*A-**ACA***U********		chicken	CUU*AC********************************	
frog f	*G*AC****A~U*U*CCCGGC*A*UCGGCUGGUGUG*G*A-U*AC****U****_****		frog DNA	TCCCT*****C**A*********G****************	
zebrafish	*U*GC*****-U*U*CCCGGC*-**UGGUUG-UGUG*A**-**U****U**********				
puffer c	*G*G******CU*U*CCCGGC*-**UGGCUGGCGUG*A*GA***U****U*********		human	CCGGUCAGGGUCAA	
puffer d			rabbit	*********	332
	*G*G*****CU*U*CCCGGC*-**UGGCUGGCGUG*A*GA***U****U************ *G*G*****CU*U*CCCGGC*-**UGGCUGGGGUG*A*GA***U****U***********		rat	****	
purrer r	G G G G G G G G G G G G G G G G G G G		mouse	****A********	
	• • • • •		chicken	UA*A*****A****	
human	AGCUUCCCAGAGUCCU-GUCGACAAUUACUGGGGAGACAAACCAUGCAGGAAACAGCC	208	ITOG DNA	******	
rat	GA****U*********-**G***C**G****A*****				
mouse	GA****U*************G***C**G****A*****	•	むつ	• • • • • •	
rabbit	**********************G*****G*********		L'Z human	UGUGCACAUUGUUAGAGCUUGGAGUUGAGGCUACUGACUG	60
chicken	**************************************		2020220		
frog h	*A********A**C*AA-AAU*G*UC********************************		rat mouse	**************************************	
frog c	*A**********C****G*G****AU*************************		frog	U******CA******A******C******-CU*UU*A-***A**********	
frog d	*A*********C***CG*G****G****A**********				
frog e	*A*********C*G*CG*G**G*G**C***********		-	• • • • •	
frog f	*A*********C***CG*G*G*G*C*************		human	AGGUAGUGUGCUACAUGAGGGGCAAG-UUUUCGCUAACACCACAAGGGUCUCUGG	114
zebrafish	*U********C***-U*G*G*C*****		rabbit	*****G********************************	
puffer c	*U*****************AC*G****CG**********CU**********		rat	**************************************	
putter d	*U**************AC*G**CG**********CU*********UGC***U *U**********		frog	***AU************U*ACAU****GAC*********G**********	
	*U*******************AC*G****CG**********		1109	V AGA OF GALL GALL GALLAR AGA OF AGA C	
puret t				• • • •	
			human	CCCAAUGAGUGGAGUUUGAUAGUAAUUCUUGCUACAAGUA	154
			rabbit	***************G***	
			Irog	U***G***C******************************	

Figure 3. Sequences of vertebrate E1, E2 and E3 RNAs and E3 RNA genes. The chicken and zebrafish E1 sequences and the rat and mouse E2 sequences are from genomic DNA. The other new sequences are from cDNA. Human sequences (3,4) and frog and pufferfish E1 sequences (8,33) were reported earlier. The rabbit E1 and E3 cDNA sequences, nt 21-134 of the frog E2 sequence and nt 30-110 of the frog E3 cDNA sequence were confirmed by those of genomic DNA. To sequence E3, genomic DNA samples from the species indicated were amplified by PCR using primers corresponding to exons 8 and 9 of the host gene, eIF-4AII (5). The PCR product of X. laevis DNA was cloned in plasmid and the sequence of one of the clones is shown (frog DNA). An asterisk means a residue in the sequence of the species indicated that is identical to that in human; a dash indicates a missing residue. The nucleotide positions in the human E1 and E2 sequences and the entire human E3 coding and flanking sequences are numbered on the right side (there is a dot over every tenth residue). The human E3 coding region is underlined and its nucleotide positions are numbered in italics. Exons 8 and 9 of the eIF-4AII gene are boxed. The bold line between the rat and mouse E3 flanking sequences indicates a 26 base segment that is located immediately upstream of the E3 coding region and is identical in rat and mouse.

nucleolar localization elements of other nuclear RNAs (9-15). The intracellular distribution of human E1, E2 and E3 RNAs that were injected into the cytoplasm of frog cells indicates that the nuclear and nucleolar localization signals of these snoRNAs reside within their evolutionarily conserved segments. An AGA triplet is found just downstream of the 5'-terminus folded domain of many snoRNAs that have the ACA box (35), but is absent from the conserved sequences of E1, E2 and E3 RNAs. The ACA sequence is the only conserved sequence shared by these three snoRNAs. Much longer sequence elements are needed for intracellular localization of the snoRNAs whose localization signals are known. For example, the nucleolar localization of MRP snoRNA requires a 40 base snoRNA sequence element (15) and the nuclear localization of U3 snoRNA requires both a 13 base snoRNA sequence element and a 5 bp stem (12). These observations suggest that E1, E2 and E3 RNAs each may have different nucleolar localization signals. Extensive conserved nucleotide sequences, which are absent in E1, E2 and E3 RNAs, are required by other snoRNA species to function in pre-rRNA processing or for snoRNA processing from pre-mRNA introns



**Figure 4.** Models for the secondary structures of E1, E2 and E3 RNAs. These models were constructed by a combination of phylogenetic comparative analysis and thermodynamic prediction (see text). The human RNAs are shown; sequence differences present in the remaining organisms are indicated by the arrows. Nucleotides absent in one or more sequences are indicated by  $\Delta$ . Watson–Crick (A=U or G=C) pairs are indicated by dashes, G-U pairs are indicated with dots. The sequence is numbered 5' $\rightarrow$ 3'. Sequence co-variations are boxed; all other sequence changes in base paired regions are consistent with the use of G-U base pairs in RNA. Asterisks indicate sites of psoralen adducts in psoralen crosslinking experiments *in vivo* which may crosslink to pre-rRNA (4). Lines show single-stranded RNA segments that are in accessible sites, by antisense oligodeoxynucleotide-targeted degradation in cell extracts (4) or in frog oocytes (R.Mishra and G.Eliceiri, submitted).

(36,37). For example, several U8 snoRNA conserved sequences are needed for its function in pre-rRNA processing, including five sequences consisting of 4–8 nt each (36). These observations suggest that the E1, E2 and E3 RNA *cis*-acting elements required for these functions may be novel and possibly also different among these three snoRNAs. It is anticipated that proteins that interact with these elements participate in the mechanism of intracellular localization of these snoRNAs.

A 26 base intron sequence is identical in rat and mouse and in both lies immediately upstream of the E3 RNA gene (Fig. 3). As expected for species whose ancestors split  $\sim$ 15 million years ago, there are many mismatches in the rest of the intron, except near the splice sites. This sequence is not part of the 5'-splice site or the branchpoint site, because it is not sufficiently near them. It is unlikely that a non-functional 26 base sequence would be fully conserved after so many million years. This sequence is not conserved in other vertebrates, but this is true for functional domains that have

co-evolved with their functional partners (38,39). The genes for two intronic snoRNAs, human E2 (5) and frog U16 (40), both show two identical sequences in the same positions: CTACCTA, 123 nt upstream of the snoRNA coding region, and GAGAAATG, 27 bases downstream of the snoRNA coding sequence. One or more of these three flanking sequences might have a role.

# ACKNOWLEDGEMENTS

We thank Noel Daly for PCR amplification of zebrafish E1 DNA, Thomas E.Dahms and Patricia L.Farrar for fresh tissues, Francesco Amaldi for the *X.laevis* E1 RNA genomic DNA clone, Elsebet Lund and Philip L.Paine for advice on the isolation of frog oocyte germinal vesicles and cytoplasm, Charles A.O'Brien for advice on 5'- and 3'-RACE and Paola Pierandrei-Amaldi for instructions on fixation and sectioning of frog oocytes. We also thank Gloriosa Go and Lynne Mann for technical assistance, Chiyen Miller for assistance in microautoradiography, Andrew Grainger for computer searching and Clifford Pollack and Jesse Urhahn for photography. This work was supported by a grant from NIH.

# REFERENCES

- 1 Maxwell, E.S. and Fournier, M.J. (1995) Annu. Rev. Biochem., 35, 897-934.
- Lafontaine, D. and Tollervey, D. (1995) *Biochem. Cell Biol.*, 73, 803–812.
   Ruff, E.A., Rimoldi, O.J., Raghu, B. and Eliceiri, G.L. (1993) *Proc. Natl.*
- Acad. Sci. USA, 90, 635–638.
  4 Rimoldi,O.J., Raghu,B., Nag,M.K. and Eliceiri,G.L. (1993) *Mol. Cell. Biol.*, 13, 4382–4390.
- 5 Nag,M.K., Thai,T.T., Ruff,E.A., Selvamurugan,N., Kunnimalaiyaan,M. and Eliceiri,G.L. (1993) Proc. Natl. Acad. Sci. USA, 90, 9001–9005.
- 6 Selvamurugan, N., Nag, M.K. and Eliceiri, G.L. (1995) *Biochim. Biophys.* Acta, **1260**, 230–234.
- 7 Kiss, T. and Filipowicz, W. (1993) EMBO J., 12, 2913-2920.
- 8 Cecconi, F., Mariottini, P., Loreni, F., Pierandrei-Amaldi, P., Campioni, N. and Amaldi, F. (1994) Nucleic Acids Res., 22, 732–741.
- 9 Fischer, U. and Lührmann, R. (1990) Science, 249, 786-789.
- 10 Hamm, J., Darzynkiewicz, E., Tahara, M. and Mattaj, I.W. (1990) Cell, 62, 569–577.
- 11 Fischer, U., Darzynkiewicz, E., Tahara, S.M., Dathan, N.A., Lührmann, R. and Mattaj, I.W. (1991) J. Cell Biol., 113, 705–714.
- 12 Baserga, S.J., Gilmore-Hebert, M. and Yang, X.W. (1992) Genes Dev., 6, 1120-1130.
- 13 Terns, M.P. and Dahlberg, J.E. (1994) Science, 264, 959-961.
- 14 Terns, M.P., Grimm, C., Lund, E. and Dahlberg, J.E. (1995) *EMBO J.*, 14, 4860–4871.
- 15 Jacobson, M.R., Cao, L.-G., Wang, Y.-L. and Pederson, T. (1995) J. Cell Biol., 131, 1649–1658.
- 16 Selvamurugan, N. and Eliceiri, G.L. (1995) Genomics, 30, 400-401.
- 17 Séraphin, B. (1993) Trends Biochem. Sci., 18, 330–331.

- 18 Spector, D.L. (1993) Annu. Rev. Cell Biol., 9, 265-315.
- 19 Gall,J.G., Tsvetkov,A., Wu,Z. and Murphy,C. (1995) Dev. Genet., 16, 25–35.
- 20 Izaurralde, E. and Mattaj, I.W. (1992) Semin. Cell Biol., 3, 279-288.
- 21 Mullis, K.B. and Faloona, F.A. (1987) Methods Enzymol., 155, 335-350.
- 22 Murray, V. (1989) Nucleic Acids Res., 17, 8889.
- 23 Krieg, P.A. and Melton, D.A. (1987) Methods Enzymol., 155, 397-415.
- 24 Lund, E. and Paine, P.L. (1990) Methods Enzymol., 181, 36-43.
- 25 Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Natl. Acad. Sci. USA, 85, 8998–9002.
- 26 O'Brien, C.A. and Wolin, S.L. (1994) Genes Dev., 8, 2891-2903.
- 27 Gutell, R.R. (1993) In Nierhaus, K.H., Franceschi, F., Subramanian, A.R., Erdmann, V.A. and Wittmann-Liebold, B. (eds), *The Translational Apparatus*. Plenum Press, New York, NY, pp. 477–488.
- 28 Zuker, M. (1989) Science, 244, 48-52.
- 29 Brown, J.W. (1991) CABIOS, 7, 391-393.
- 30 Shaw, P.J. and Jordan, E.G. (1995) Annu. Rev. Cell Dev. Biol., 11, 93-121.
- 31 Schmidt, C., Lipsius, E. and Kruppa, J. (1995) Mol. Biol. Cell, 6, 1875–1886.
- 32 Dumont, J.N. (1972) J. Morphol., 136, 153-180.
- 33 Cecconi, F., Crosio, C., Mariottini, P., Cesareni, G., Giorgi, M., Brenner, S. and Amaldi, F. (1996) Nucleic Acids Res., 24, 3167–3172.
- 4 Woese, C.R., Gutell, R.R., Gupta, R. and Noller, H.F. (1983) *Microbiol. Rev.*, 47, 621–669.
- 35 Balakin, A.G., Smith, L. and Fournier, M.J. (1996) Cell, 86, 823-834.
- 36 Peculis, B.A. and Steitz, J.A. (1994) Genes Dev., 8, 2241–2255.
- 37 Watkins, N.J., Leverette, R.D., Xia, L., Andrews, M.T. and Maxwell, E.S. (1996) *RNA*, 2, 118–133.
- 38 Schnapp,A., Rosebauer,H. and Grummt,I. (1991) Mol. Cell. Biochem., 104, 137–147.
- 39 Ishikawa,Y., Safrany,G., Hisatake,K., Tanaka,N., Maeda,Y., Kato,H., Kominami,R. and Muramatsu,M. (1991) *J. Mol. Biol.*, **218**, 55–67.
- 40 Fragapane, P., Prislei, S., Michienzi, A., Caffarelli, E. and Bozzoni, I. (1993) EMBO J., 12, 2921–2928.