© 1994 Oxford University Press

Nucleic Acids Research, 1994, Vol. 22, No. 20 4073–4081

U21, a novel small nucleolar RNA with a 13 nt. complementarity to 28S rRNA, is encoded in an intron of ribosomal protein L5 gene in chicken and mammals

Liang-Hu Qu+, Monique Nicoloso, Bernard Michot, Marie-Claude Azum, Michèle Caizergues-Ferrer, Marie-Hélène Renalier and Jean-Pierre Bachellerie*
Laboratoire de Biologie Moléculaire Eucaryote du CNRS, Université Paul-Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex, France

Received July 26, 1994; Revised and Accepted September 5, 1994 EMBL accession nos Z35310–Z35312 (incl.)

ABSTRACT

Following a search of sequence data bases for intronic sequences exhibiting structural features typical of snoRNAs, we have positively identified by Northern assays and sequence analysis another intron-encoded snoRNA, termed U21. U21 RNA is a 93 nt. long, metabolically stable RNA, present at about 10^4 molecules per HeLa cell. It is encoded in intron 5 of the ribosomal protein L5 gene, both in chicken and in the two mammals studied so far, human and mouse. U21 RNA is devoid of a 5′-trimethyl-cap and is likely to result from processing of intronic RNA. The nucleolar localization of U21 has been established by fluorescence microscopy after *in situ* hybridization with digoxigenin-labeled oligonucleotide probes. Like most other snoRNAs U21 contains the box C and box D motifs and is precipitated by anti-fibrillarin antibodies. By the presence of a typical 5′–3′ terminal stem, U21 appears more particularly related to U14, U15, U16 and U20 intron-encoded snoRNAs. Remarkably, U21 contains a long stretch (13 nt.) of complementarity to a highly conserved sequence in 28S rRNA. Sequence comparisons between chicken and mammals, together with Northern hybridizations with antisense oligonucleotides on cellular RNAs from more distant vertebrates, point to the preferential preservation of this segment of U21 sequence during evolution. Accordingly, this complementarity, which overlaps the complementarity of 28S rRNA to another snoRNA, U18, could reflect an important role of U21 snoRNA in the biogenesis of large ribosomal subunit.

INTRODUCTION

Small nucleolar RNAs constitute a growing class of molecules believed to participate in ribosome biogenesis in eucaryotic cells (1–3). However, even for the best studied snoRNAs, for which this notion is supported by experimental evidence (4–9), the molecular mechanisms underlying their function remain hypothetical at best, in marked contrast to spliceosomal snRNAs. Over recent months snoRNAs have been the subject of intense interest, not only because several new molecular species have been identified in vertebrates but also because most of them have been shown to be encoded in introns of protein coding genes and to result from a novel form of intronic RNA processing (3,10). Most snoRNAs characterized so far in vertebrates share distinctive structural features. Thus, intron-encoded species U14 (11), U15 (12), U16 (13), U18 (14) and U20 (15) RNAs contain the box C and box D sequence motifs, which were originally delineated by the comparison of U3 snoRNA sequences from distant eucaryotes (16) and later also found in U8 and U13 snoRNAs, which are encoded by independent genes (17). Moreover, the secondary structures of four of these intron-encoded snoRNAs U14, U15, U16 and U20 exhibit a common 5′–3′ terminal stem involving the nucleotides flanking immediately the single-stranded box C and box D motifs (11–13, 15). The frequent occurrence of this complex feature in the intronic snoRNAs identified so far prompted us to undergo a systematic search of data bases for intronic sequences exhibiting the potential to form this structure in an attempt to identify other snoRNAs. We have found some intronic sequences meeting this criterion. Since several intronic snoRNAs reported so far show a significant base-pairing potential with mature rRNAs (11,14,15), we have focused our attention on one of these positive sequence, located in intron 5 of the chicken ribosomal protein L5 gene, because of the presence of an extended sequence complementarity to a phylogenetically conserved sequence of 28S rRNA. By direct analyses of cellular RNA with antisense oligonucleotides, we have positively detected a novel small nucleolar RNA, termed U21, which is encoded in this particular sequence. We have also identified its homolog in all vertebrate classes and examined whether this peculiar localization of its coding sequence was phylogenetically conserved, through PCR amplification of mammalian genomic DNAs.

*To whom correspondence should be addressed

+Permanent address: Biotechnology Research Center, Zhong Shan University, Guangzhou, China

MATERIALS AND METHODS

Oligonucleotide probes

Sequences of the U21 RNA antisense oligodeoxynucleotides are given in Figure 1a. The U3 RNA antisense oligonucleotide was complementary to positions 1–24 of mouse (18) or human U3 RNA sequence. The sequence of the 26-mer U18 RNA antisense oligonucleotide was 5′ TCAGAAACACGGACCAWWGAA-GTGGA 3′ (with W=A or T), selected from the sequences of the different *Xenopus laevis* U18 RNA variants (14). They have been used for end-labeling according to standard procedures (19) without further purification.

PCR conditions and sequencing of amplified genomic DNA

Sequences of the forward (20-mer) and reverse (21-mer) primers can be found in Figure 2. They have been selected within portions of rpL5 exon 5 and exon 6 sequences which are particularly conserved, as judged by the comparison of chicken (20), rat (21), *Xenopus laevis* (22) and *S.cerevisiae* (23) rpL5 gene or cDNA sequences.

PCR products were cloned into the *Sma*I site of the pUC 18 plasmid using a surclone ligation kit (PL Biochemicals, Pharmacia). Transformation was carried out as described by Sambrook *et al.* (19) and positive clones were selected after Southern dot-blot hybridization with the 5′-end-labeled oligonucleotide 1 probe.

Single-strand sequencing was performed for both strands of positive plasmid DNA insert after subcloning into M13 mp18 and mp19 vectors. Automated sequencing was carried out on the entire length of both DNA strands.

Synthesis and sequencing of U21 cDNA

Reverse transcription was carried out in 20 μl reaction mixtures containing 15 μg of RNA (purified from chicken liver) and 4 ng of antisense oligonucleotide 1 (Fig. 1a) which had been 5′ (^{32}P) labeled with [γ-^{32}P] ATP. Labeled cDNA was recovered as previously reported (24), following alkaline hydrolysis of RNA, ethanol precipitation and electrophoresis on a 6% acrylamide/7M urea gel. After extraction from the acrylamide gel (25), the 93 nt. long cDNA was tailed by addition of an oligo-(dG) tract, through incubation (37°C, 30 min) with dGTP in the presence of terminal deoxynucleotidyl transferase (19). The 3′-tailed cDNA was then subjected to a PCR amplification (after 4 min at 95°C, 25 cycles of: 1 min at 95°C, 1 min at 50°C and 1.5 min at 72°C, with a 5 min elongation step in the final cycle) using as primers U21 antisense oligonucleotide 1 and the oligonucleotide 5′GGAATTCGGAT(C)$_{16}$3′ (Fig. 1c). The PCR product was purified on a 6% acrylamide/7 M urea sequencing gel and cloned into the pUC18 plasmid vector using the surclone kit (Pharmacia) before automated sequencing of double-strand plasmid DNA.

RNA extraction and Northern hybridization analysis

RNA was isolated by the guanidinium– thiocyanate method (26) followed by phenol–chloroform extraction before electrophoresis on 6% or 10% acrylamide/7M urea gels (25). Electrotransfer onto nylon (Hybond N) membranes was followed by a short UV irradiation before hybridization in 5×SSPE (1×SSPE is 150 mM NaCl, 1 mM EDTA, 10 mM NaH_2PO_4, pH 7.4), 1×Denhardt, 1% SDS and 150 μg/ml tRNA carrier. Unless otherwise stated, hybridization was performed at a temperature 15°C below the T_m calculated for a perfectly matched duplex (27). Filters were washed first with 2×SSC, 0.1% SDS (15 min at room temperature), then twice with 0.1×SSC, 0.1% SDS (15 min at room temperature).

Preparation of snRNP extracts. Immunoprecipitations

Preparation of HeLa cells extracts and immunoprecipitation with the monoclonal 72B9 anti-fibrillarin antibody obtained from G.Reimer and E.Tan (28) were performed essentially as described previously (29). Cells were sonicated in NET buffer: 40 mM Tris–HCl (pH 7.5), 150 mM NaCl and 0.05% Nonidet P-40 at the concentration of 4.10^7 cells/ml. The sonication was performed (five 30 sec pulses with 30 sec intervals) using a Branson microtip (setting: 1.5). The sonicated extract was rendered 300 mM in NaCl and immunoprecipitation was carried out on 0.5 ml aliquots in the presence of 10 ml serum. The same buffer was used for binding and washing. Immunoprecipitation was for 90 min at 4°C. The RNA was recovered from Protein A sepharose by phenol–chloroform extraction, and run on 8% acrylamide–urea gel before Northern analysis. Immunoprecipitation with the anti-trimethyl-guanosine (TMG) cap monoclonal antibody, obtained from R.Lührmann (30) was performed on purified nuclear RNA (dissolved in NET buffer) after a heat denaturation step.

In situ hybridization

Mouse NIH-3T3 cells grown on coverslips were fixed in 4% formaldehyde in phosphate-buffered saline (PBS) for 10 min on ice and dehydrated successively in 30, 50, 70 and 95% ethanol. Permeabilization treatments were performed with 0.5% saponin, 1mM EGTA in PBS for 3 min on ice, then with cold (−20°C) methanol for 10 min. After prehybridization in 6×SSC, 5×Denhardt, 10% Dextran in the presence of 10 ng/ml *E.coli* DNA, the oligonucleotide probe, labeled by addition of digoxigenin-11-dUTP (Boehringer–Mannheim) by terminal transferase, was added to a final concentration of 2 ng/ml and incubation performed overnight at 37°C in a humidified chamber. Following hybridization coverslips were washed successively in 6×SSC, 2×SSC, 0.2×SSC, at room temperature. Hybridization sites were immunolabeled using a Rhodamin-anti-digoxigenin conjugate (Boehringer–Mannheim) diluted 1/20 in 0.1M Tris–HCl pH 8.5, 150 mM NaCl, 5% BSA. After washing coverslips were mounted in Mowiol containing DAPI (25 ng/μl) as DNA counterstain.

RESULTS

A snoRNA-related sequence in an intron of the chicken rpL5 gene cross-hybridizes to a small cellular RNA

We have searched the EMBL Data Library for intronic sequences exhibiting the hallmark structure of four of the six intron-encoded snoRNAs detected so far, i.e., the presence of the box C (5′UGAUGA3′) and the box D (5′CUGA3′) sequence motifs immediately flanked by a 4–5 bp stem involving both termini of the RNA molecule (see Materials and Methods). Several positive sequences have been identified (B.Michot, unpublished results), in which the two motifs are separated by 50–140 nt. and flanked by an inverted repetition of at least 4 bp in the same relative arrangement as in U14, U15, U16 or U20. These intronic sequences were further examined for the potential of their box C–box D intervening segment to base-pair with phylogenetically conserved sequences in small or large subunit rRNA. Among the few double positives which have been finally identified we

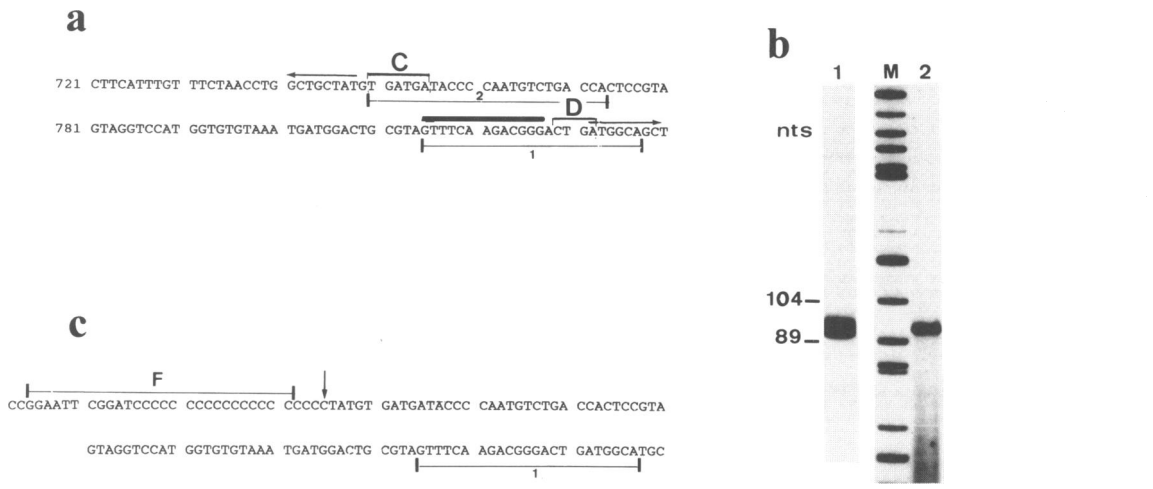

Figure 1. Detection of a small RNA encoded in intron 5 of the rpL5 gene in chicken. (**a**) Region of the chicken intron 5 sequence (20) exhibiting the hallmark structural features already found in snoRNAs U14, U15, U16 and U20 (see text). Box C and box D motifs are boxed and the inverted repetition which flanks immediately the two boxes is denoted by arrows. Nucleotides complementary to a phylogenetically conserved sequence of 28S rRNA are overlined by a thick bar. Location of the two antisense oligodeoxynucleotides used for Northern analysis of chicken RNA (shown in b) is depicted by the numbered bars below the sequence. Sequence numbering is from the 5′ end of intron 5 (20). (**b**) Detection of a cross-hybridizing small RNA in chicken and synthesis of its cDNA. Northern analysis (lane 1): Ten μg of total cellular RNA from chicken liver were separated onto a 6% acrylamide/7M urea gel. After electro-transfer, the nylon membrane was hybridized with the 5′(^{32}P) end labeled antisense oligodeoxynucleotide 1 (depicted in Fig. 1a). cDNA synthesis (lane 2): Reverse transcription was carried out, as detailed in Materials and Methods, using the 5′(^{32}P) end labeled antisense oligodeoxynucleotides 1 as primer. The length of labeled cDNA was determined by reference to the migration of a size marker (lane M). (**c**) Sequence of the amplified and cloned cDNA. The sequence corresponds to the RNA-like strand. The position matching the 3′ terminal nucleotide of the cDNA [before 3′ (dG) tailing] is denoted by a vertical arrow. Position of the forward primer (F) [complementary to the 3′ oligo (dG) tail added to the cDNA] used for the PCR amplification of the cDNA is shown above the sequence (see Materials and Methods for details). The reverse primer (bar under the sequence) for PCR was also used for cDNA synthesis.

have first focused our attention on a remarkable specimen showing a 13 bp complementarity to a conserved segment of 28S rRNA sequence (Fig. 1a). This sequence tract in 28S rRNA maps at positions 1126−1138 in the mouse molecule (31). It has remained invariant in all eucaryotic large subunit rRNAs analyzed so far and corresponds to a single-stranded region of the molecule (32).

In a first step, to test the possibility that this particular portion of intron 5 of the chicken rpL5 gene (20) was encoding a stable small cellular RNA, we synthesized two oligodeoxynucleotides (a 23-mer and a 24-mer) complementary to two non-overlapping tracts of the sequence bracketted by the 8 nt. long inverted repetition (shown in Fig. 1a). Northern analysis of chicken cellular RNA with each of these 5′ end-labeled probe did reveal, in stringent conditions of hybridization, a single radioactive signal, migrating at the same position in both cases (shown in Fig. 1b for probe 1). Precise evaluation of the mobility of the cross-hybridizing band pointed to a 93 nt. RNA, which is exactly to the size expected for a small RNA encoded in the sequence depicted in Figure 1a and possessing a 5′−3′ terminal stem similar to U14, U15, U16 or U20. In a next step, we examined the possibility that the signal observed with the two oligonucleotide probes was merely reflecting a close sequence relatedness (rather than a perfect sequence identity) between the 93 nt. long RNA and the rpL5 gene intron 5 segment delineated in Figure 1a. We therefore undertook a sequence analysis of this RNA species, after synthesis of its cDNA. After a reverse transcription primed by oligodeoxynucleotide 1, which maps at the 3′ end of the snoRNA-like sequence in the chicken rpL5 gene intron 5, a single cDNA product was obtained, 93 nts in length (Fig. 1b lane 2). After addition of an oligo (dG) tail at its 3′ end, the cDNA was amplified by PCR, cloned and sequenced (as described in Materials and Methods). The sequence determined for the amplified product between the two PCR primers (Fig. 1c) corresponds perfectly to the rpL5 gene intron 5 sequence over the 70 nts upstream from the tract of complementarity to oligodeoxynucleotide probe 1.

A *Hin*dIII digest of chicken DNA was hybridized with a DNA probe corresponding to the entire intron 5 (plus a few nucleotides of flanking exons) which was obtained by PCR amplification of chicken genomic DNA, (as detailed below). A strong and unique signal was revealed, at about 3.7 kb (not shown), i.e., as precisely expected from the sequenced rpL5 gene *Hin*dIII fragment (20). An identical result was obtained with oligodeoxynucleotide probe 1 confirming that the 93 nt. long RNA, which will be referred to henceforth as U21, does originate from the rpL5 gene intron 5 sequence and not from another copy of this sequence located in another region of chromosomal DNA.

The 93 nt. long RNA-coding sequence is phylogenetically conserved and has a similar intronic location in the rpL5 gene of mammals

Intron-encoded snoRNAs reported so far represent outstanding intronic sequences by their strong conservation during evolution. This aspect could not be examined by analysis of the sequence data bases in the case of the 93 nt. long RNA encoded in the rpL5 gene intron: whereas rpL5 cDNA sequences are available for *Xenopus laevis* (22) and rat (21), the only other species in addition to chicken for which a rpL5 gene sequence is known is yeast *Saccharomyces cerevisiae* (23). However, the yeast homolog, termed YL3, is not informative since this is an intron-less gene. In a search for a potential homolog of the 93 nt. long

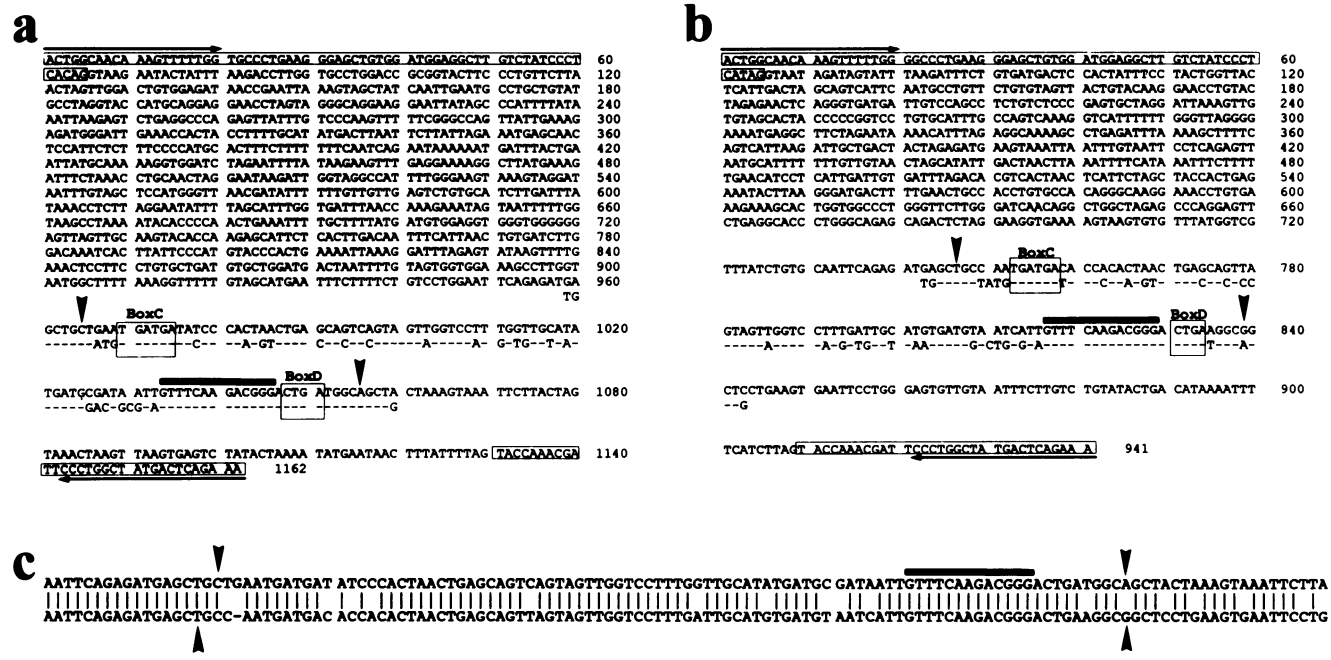

Figure 2. Nucleotide sequence of intron 5 (and part of flanking exons) for the amplified rpL5 gene in humans and mouse. The location of the forward and reverse oligonucleotide primers used for PCR amplification of genomic DNA is shown by arrows above and below the exonic sequences (boxed) respectively. For the portion of the human (**a**) or mouse (**b**) intron over which a significant similarity with the chicken sequence is observed, an aligment of the chicken sequence is shown below each mammalian sequence (with dashes denoting identical nucleotides). (**c**) Aligment of the human (top line) and mouse sequences for the conserved region of intron 5. Termini of the U21 RNA coding sequence (identified as described in Fig. 1) are denoted by arrow-heads. The segment of complementarity to 28S rRNA is overlined by a bar.

RNA-coding sequence in species relatively distant from chicken, we decided to determine the sequence of intron 5 of the rpL5 gene in two mammals, mouse and humans.

Comparisons of the eucaryotic rpL5 cDNA sequences in sequence data bases allowed us to select two phylogenetically conserved exonic sequences. Two corresponding oligonucleotides (spanning 20 nt. in exon 5 and 20 nt. in the 5′ portion of exon 6) were synthesized and used as primers for a PCR carried out on total genomic DNA from human or mouse cells. PCR primers and conditions for amplification and sequencing of the PCR products were first tested on chicken genomic DNA. The sequence of the entire chicken intron 5 (1032 bp) was determined and found in excellent agreement with the published sequence (20): not a single difference was observed over the segment represented in Figure 1a. The same region of the rpL5 gene was also amplified, cloned and sequenced for humans (Fig. 2a) and for mouse (Fig. 2b). The amplified mouse rpL5 gene intron 5 is 845 nt. long. Sequences at both ends of the amplified mouse DNA fragment (boxed in Fig. 2b) were found almost identical with the corresponding regions of exons 5 and 6 of rat rpL5 gene as inferred from the rat rpL5 cDNA sequence (21). No difference was observed over exon 6; as for the only three nucleotide changes observed over exon 5, they are all conservative in terms of amino acid sequence of L5 protein, thus confirming the specifity of the amplification reaction. For the amplified human DNA fragment too, examination of the nucleotide sequences at both ends unambiguously shows that the PCR product does correspond to the cognate region of the rpL5 gene: while the number of nucleotides differences with the rat exonic sequences is much higher than for mouse (8 over exon 5 and 2 over exon 6), translation to amino acid sequence reveals a perfect identity with the rat L5 protein. The human rpL5 gene intron 5 is 1065 nt. long (Fig. 2a). Within the entire intron the only region of significant sequence similarity between mammals and chicken corresponds precisely to the segment exhibiting structural hallmarks of a snoRNA (shown in Fig. 1a). It is noteworthy that, within this region, the box C and box D motifs and the 13 nt. long complementarity to 28S rRNA are perfectly conserved between chicken and mammals, together with the potential to form a 8−9 bp stem closely bracketting box C and box D.

This region is also the sole portion of the intron to be conserved between mouse and humans. The counterpart of the segment encoding U21 RNA in chicken exhibits a 85% sequence similarity between the two mammals (Fig. 2c). Remarkably the similarity of the mammalian sequences is not exactly restricted to the segment encoding U21 RNA in chicken but extends over a significant number of flanking nucleotides at both ends (17 identical upstream nucleotides and 14 out of 17 downstream nucleotides) as shown in Figure 2c.

Detection of a U21 RNA homolog in more distant vertebrates

Taking advantage of the conservation detected between chicken and mammals for the segment of the intronic sequence encoding chicken U21 RNA, we have looked for a potential homolog of chicken U21 RNA in other vertebrate classes. Northern hybridization of total cellular RNA from diverse species was carried out with synthetic oligodeoxynucleotide 1 which spans the most extended tract of conservation observed between chicken

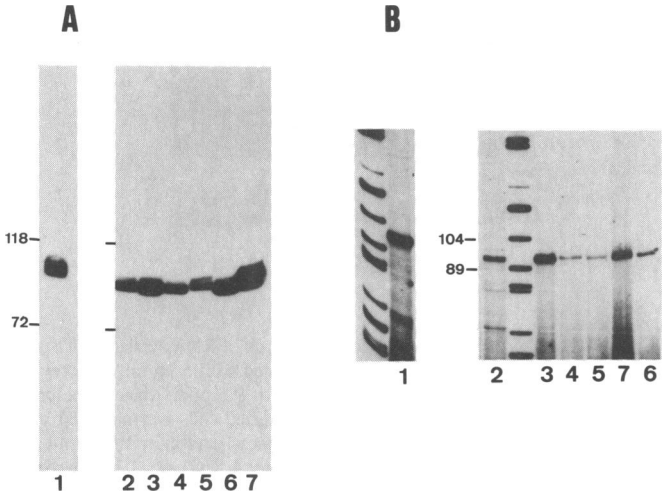

Figure 3. Identification of a U21 RNA homolog in different vertebrates by Northern hybridization. (**a**) For each species, total cellular RNA (10 μg per lane) was separated on a 6% acrylamide/7M urea gel. After electrophoretic transfer, the membrane was hybridized with 5′ (^{32}P) labeled oligonucleotide probe 1 (shown in Fig. 1). Lane 1: rainbow trout *Salmo gairdneri*. Lane 2: snake *Natrix viperina*. Lane 3: chicken *Gallus domesticus*. Lane 4: duck *Anas domesticus*. Lane 5: pigeon *Columba palumbus*. Lane 6: human HeLa cells. Lane 7: mouse NIH 3T3 cells. (**b**) Primer extension: It was carried out with the 5′(^{32}P) end-labeled U21 RNA antisense probe 1 (shown in Fig. 1c) on 10 μg aliquots of total cellular RNA for the same set of species as in (a) and cDNAs were analyzed on a 6% acrylamide/7M urea gel [lanes 1−7, corresponding to the same species as in (a)].

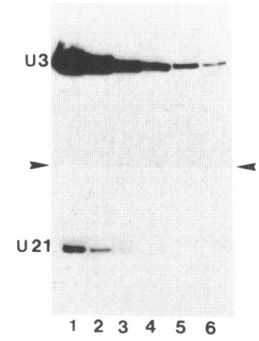

Figure 4. Cellular abundance of U21 RNA in exponentially growing human HeLa cells. Decreasing amounts of total nuclear RNA (10, 5, 2, 1, 0.5 and 0.25 μg in lanes 1−6, respectively) were separated on a 6% acrylamide/7M urea gel. After electrotransfer, the membrane was cut into 2 parts, as denoted by the arrowheads. The upper part was hybridized with a U3 RNA-specific oligonucleotide probe and the lower part with U21 RNA antisense oligonucleotide (shown in Fig. 1). The two probes had been 5′(^{32}P) end-labeled to the same specific radioactivity and hybridizations were performed in identical conditions (same stringency and same concentrations of probes, corresponding to a large excess of probe by reference to membrane-bound U3 and U21 RNAs). Intensities of hybridization signals were quantitated with a Fuji Bas-1000 Imager.

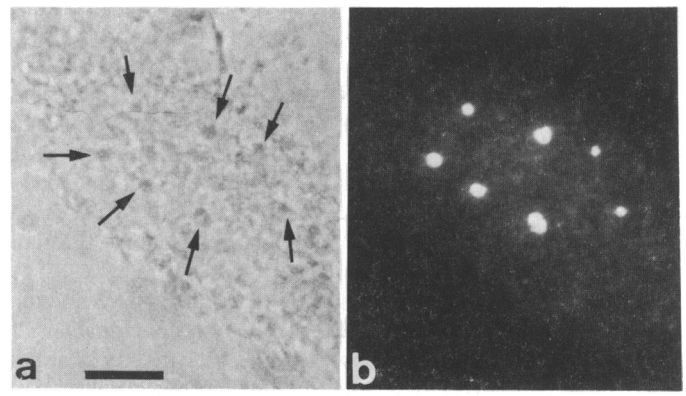

Figure 5. Intranuclear distribution of U21 RNA in rodent cells. Fixed mouse NIH 3T3 cells were permeabilized and observed either by phase-contrast microscopy (**a**), or by fluorescence microscopy (**b**) following *in situ* hybridization with U21 RNA antisense oligodeoxynucleotide probe 1 labeled with digoxigenin. Fluorescence detection was obtained by use of a TRITC anti-digoxigenin conjugate. All the fluorescence signals (b) precisely correspond to the contours of the different nucleoli (darker spots denoted by arrows) seen by phase-contrast microscopy in (a). Bar: 15 mμ.

and the two mammals, i.e., the complementarity to 28S rRNA and the downstream box D motif.

As shown in Figure 3a, a strong and unique signal can be detected in representatives of all vertebrate classes after hybridization with this probe in moderate conditions of stringency (hybridization temperature 25°C below the T_m of a perfectly matched duplex). However for some samples a minor band could also be distinguished, immediately vicinal to the main signal, likely reflecting a 1−2 nt. terminal heterogeneity of U21. In all cases, the major signal corresponds to an RNA of 93−96 nt., as judged by its mobility in denaturing acrylamide gel, with only very minor size differences among the species (96 nt. for trout, 94 nt. for mouse and pigeon, 93 nt. for snake, chicken, duck and human).

To confirm the identity of this RNA species, we again carried out primer extensions with the same oligonucleotide probe (as reported in Fig. 1b for chicken) for the same set of vertebrate species. A major extension product of nearly identical size is detected with all species (Fig. 3b). It is noteworthy that the minor size differences of the cDNA which can be detected among the various species precisely match the size differences observed for the cross-hybridizing RNA bands in the Northern analysis between the same species (thus trout cDNA is 3 nt. longer that chicken cDNA). A more accurate determination of the 5′ terminal nucleotide of the RNA was carried out for chicken, human and mouse through another primer extension with a more proximal oligodeoxynucleotide, i.e., probe 2 (see Fig. 1c) or its mammalian version, selected from the human/mouse sequence aligment (Fig. 4c). Results (not shown) are in full agreement with the previous observations, indicating that the 5′ end of oligodeoxynucleotide probe 1 matches the 3′ end of the small RNA for all these vertebrate species.

U21 RNA is relatively abundant and localizes to the nucleolus

Results of the Northern hybridizations shown in Figure 3a indicate that the relative cellular content in U21 is roughly similar among the different vertebrate species analyzed in that experiment. To obtain a rough estimation in exponentially growing HeLa cells, a Northern hybridization was repeated with oligonucleotide probe 1, using a series of lanes loaded with decreasing amounts of total cellular RNA (Fig. 4). To normalize the data by reference to U3 RNA, known to be present at about 2×10^5 copies per HeLa

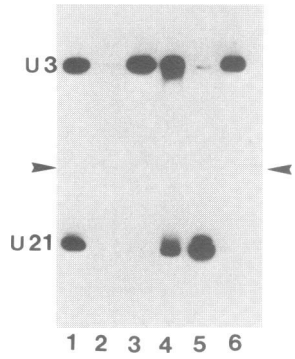

Figure 6. Northern analysis of nuclear RNA immunoprecipitated by anti-trimethylcap and anti-fibrillarin. RNAs were analyzed on a 8% acrylamide/7M urea gel. After transfer, the membrane was hybridized with both U21 RNA antisense probe 1 (lower part) and with a U3 RNA-specific antisense oligonucleotide (upper part). Lane 1: 10 μg of total nuclear RNA. Lane 2: 10 μg of nuclear RNA precipitated by a non-immune serum. Lane 3: 10 μg of nuclear RNA recovered by anti-TMG cap immunoprecipitation. Lane 4: RNA from an aliquot of the HeLa cell extract corresponding to 2.10^6 cells. Lane 5: RNA recovered by immunoprecipitation with anti-fibrillarin of a an aliquot of the HeLa cell extract corresponding to 2.10^7 cells. Lane 6: RNA from the fraction of the HeLa cell nuclear sonicate immunoprecipitated by the anti-TMG cap antibody corresponding to 2.10^7 cells.

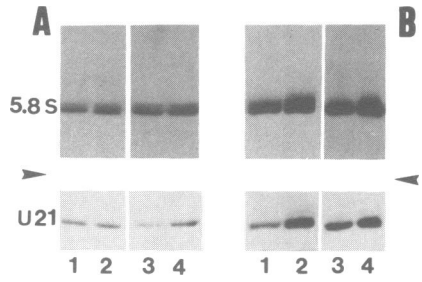

Figure 7. Stability of U21 RNA following a block of RNA synthesis or a protein synthesis inhibition. (**A**) HeLa cells were treated with 5 μg/ml actinomycin D (A) for 0.5 hour (lane 2), 3.5 hours (lane 3) or 9.5 hours (lane 4) before cell recovery and RNA isolation. RNA from untreated cells was analyzed in lane 1. (**B**) HeLa cells were treated by 10 μg/ml (lanes 1 and 2) or 100 μg/ml (lanes 3 and 4) cycloheximide for either 0.5 h (lanes 1) or 3.5 h (lanes 2) before recovery and RNA extraction. RNA aliquots were analyzed on a 6% acrylamide/7M urea gel. After electrotransfer, the membranes were cut in two parts (as denoted by arrow-heads). The lower part was hybridized with U21 anti-sense probe 1 and signals normalized for equal amounts of RNA loaded in each lane using 5.8S rRNA as an internal reference (the upper part of the membrane was hybridized with a 5.8S rRNA-specific labeled oligonucleotide).

cell (33), the corresponding part of the membrane was hybridized with a U3-specific oligonucleotide. A quantitation of hybridization signals obtained with the two probes (labeled at the same specific radioactivity) was performed. For each lane, the U21 RNA signal was about 20-times lower than the U3 RNA signal. Consistent with this observation, the U21 RNA signal in a given lane was roughly identical to the U3 RNA signal detected in a lane loaded with 20 times less cellular RNA. These data point to the presence of about 10^4 molecules of U21 RNA per exponentially growing HeLa cell.

Following the separation of nuclear and cytoplasmic fractions from exponentially growing HeLa cells and after analysis by Northern hybridization, we observed that U21 RNA was almost exclusively recovered in the nuclear fraction (result not shown). To study to intranuclear distribution of U21 RNA, we carried out *in situ* hybridization with synthetic oligonucleotide probe 1. This antisense probe was labeled by digoxigenin and after *in situ* hybridization the location of the hybridized probe was followed by light microscopy using indirect immunofluorescence. As shown in Figure 5, the immunofluorescence signal was unambiguously restricted to the nucleolus, indicating that U21 is actually a snoRNA.

U21 RNA interacts with fibrillarin and is devoid of 5′ TMG cap

Chicken, human and mouse U21 RNA sequences contain the box C sequence motif, required in U3 RNA for the binding of fibrillarin, the protein common to the major family of snoRNPs (33). We have examined if U21 RNA was also involved in a direct or indirect interaction with fibrillarin (Fig. 6). After isolation of an RNP extract from HeLa cell, an immunoprecipitation with fibrillarin antibodies was carried out and the immunoprecipitated fraction was assayed for U21 RNA content by Northern hybridization with 5′(^{32}P) end-labeled antisense probe 1 (delineated in Fig. 1). As shown in Figure 6 (lane 5), U21 RNA was efficiently precipitated by the antifibrillarin antibody. It is noteworthy that at the ionic strength that was selected for the immunoprecipitation (300 mM), the recovery of U21 RNA was markedly more efficient than that of U3 RNA chosen as a positive control in that assay. A very similar result was observed for U20 RNA, another intronic snoRNA, encoded in the nucleolin gene in mammals (15).

The presence of a trimethyl-guanosine cap is a common feature of spliceosomal snRNAs transcribed by RNA polymerase II, also shared by snoRNAs encoded by independent genes such as U3, U8 or U13 (17,33). By contrast the novel intron-encoded small nucleolar RNAs reported so far are devoid of this structural signature (which demarks the 5′ end of the primary transcript for spliceosomal or snoRNAs synthesized from independent transcription units) in line with their resulting from processing of intronic RNA. We have examined whether U21 RNA possesses or not a 5′ trimethyl-cap by utilization of an anti-trimethyl-guanosine cap antibody (30). Purified Hela cell nuclear RNA was put in presence of the anticap antibody and the immunoprecipitated fraction was analyzed by Northern hybridization with U21 antisense probe 1 (Fig. 6). U3 RNA was again used as a positive control of the efficiency of the immunoprecipitation assay. Unlike U3 RNA, which was efficiently immunoprecipitated, U21 RNA was not recovered in that fraction (Fig. 6, lane 3), indicating that U21 RNA is devoid of a 5′ trimethyl cap. We also performed the immunoprecipitation assay with the anti-TMG cap antibody using a HeLa cell RNP extract instead of purified RNA as previously. With a RNP extract isolated at 150 mM NaCl and after an immunoprecipitation performed at 300 mM U3 RNA was again efficiently precipitated, as it was from a purified RNA fraction. As for U21 RNA, in this case again it could not be detected in the immunoprecipitated fraction (Fig. 6, lane 6). This result seems to rule out the existence of a stable interaction between U21 RNA (or a U21-containing snoRNP) and U3 snoRNP in these ionic conditions.

U21 is a stable snoRNA species

To obtain an indication about the half-life of U21 RNA, HeLa cells were treated by actinomycin D (5 μg/ml), in order to block

immediately all RNA synthesis. The treatment was prolonged for various periods of time and the cellular abundance of U21 RNA was monitored by Northern hybridization for different time points. To normalize the results, we selected a stable low molecular weight RNA, 5.8S rRNA, as an internal standard for each lane (Fig. 7) through Northern hybridization with a 5.8S rRNA-specific 5′ end-labeled oligonucleotide. As shown in Figure 7A, the relative level of the U21 RNA signal, as compared to 5.8S rRNA, is not significantly affected by the block in RNA synthesis, even 9.5 hr after addition of actinomycin D, indicating that U21 RNA does not undergo a significant turnover.

We also studied the effects of a prolonged inhibition of protein synthesis on the intracellular pool of U21 RNA. After treatment of HeLa cells by concentrations of cycloheximide efficiently blocking protein synthesis, no significant change was observed in the relative amount of U21 RNA (again taking 5.8S rRNA as a reference) even for the longest period of treatment (3.5 h), as shown in Figure 7B.

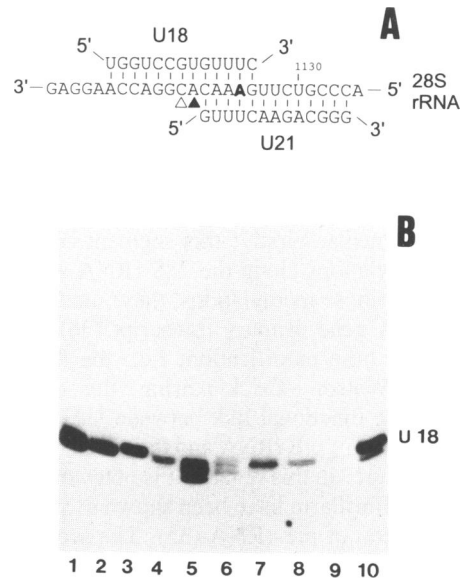

Figure 8. Overlapping sequence complementarities to U21 and U18 RNAs in 28S rRNA . (**A**) Representation of the overlapping complementarities of U21 and U18 snoRNAs to 28S rRNA. Nucleotide numbering in 28S rRNA refers to the mouse sequence (31). The A in bold face is a m^1A conserved among yeast *S.cerevisiae* and vertebrates (36). The nucleotide denoted by a filled triangle is a site of 2′-O ribose methylation conserved between yeast *S.cerevisiae* and vertebrates, while the open triangle denotes a site of 2′-O ribose methylation present in *S.cerevisiae* but absent in vertebrates (36). The 13 nt. of U18 RNA shown here are perfectly conserved in different Xenopus species and in humans (14). (**B**) Phylogenetic conservation of the corresponding sequence in U18 RNA revealed by Northern hybridization of RNAs from distant metazoans with an antisense oligonucleotide spanning the 13 nt. segment of complementarity to 28S rRNA. For each species, 10 μg of total cellular RNA were analyzed on a 6% acrylamide/7M urea gel. After electrophoretic transfer, the membrane was hybridized with a 5′(^{32}P) end-labeled 26-mer oligonucleotide complementary to the central region of *Xenopus laevis* U18 RNA (14), which contains the 13 nt. complementarity to 28S rRNA depicted in A. Lane 1: amphibian *Xenopus laevis*; lane 2: amphibian *Pleurodeles waltii*; lane 3: snake *Natrix viperina*; lane 4: mollusc, mussel *Mytilus edulis*; lane 5: fish, trout *Salmo gairdneri*; lane 6: fish, carp *Cyprinus carpio*; lane 7: insect *Drosophila melanogaster*; lane 8: bird, pigeon *Columba palumbus*; lane 9: bird, duck *Anas domesticus*; lane 10: human HeLa cells.

DISCUSSION

U21, another member of an expanding family of snoRNAs

The chicken rpL5 gene intronic sequence selected through computer searches of Sequence Data Bases does encode a novel snoRNA. Searches were carried out for the presence of the typical box C and box D motifs bracketed by an inverted repetition of at least 4 bp. Remarkably, not only this particular segment of the rpL5 gene intron is expressed as a stable RNA, but the termini of U21 RNA along the intronic sequence are precisely located as would have been predicted for allowing the formation of a 5′−3′ terminal structure previously shown to be shared by U14 (11), U15 (12), U16 (13) and U20 (15) snoRNAs. Both ends of all these molecules are held together by 4−5 base-pairings which bring in close vicinity the box C (5′UGAUGA3′) and box D (5′CUGA3′) sequence motifs. This hallmark structure is known to be required for the accumulation of stable U14 RNA in yeast cells (7).

As expected from the presence of the box C−box D terminal stem, U21 RNA is immunoprecipitated from cellular extracts by antifibrillarin antibodies like U14, U15, U16 and U20 RNAs. Fibrillarin is an essential nucleolar protein the expression of which is directly coupled with the rate of ribosome production during development in xenopus (34). In yeast, fibrillarin has been shown to be involved in multiple aspects of ribosome biogenesis (35) and it interacts with most of the snoRNAs identified so far (1−3). By *in situ* hybridization with an antisense oligonucleotide highly specific for U21 RNA (even in moderately stringent conditions of Northern hybridization), we have established that U21 RNA does localize exclusively to the nucleolus in exponentially growing mouse cells. The cellular abundance of metabolically stable U21 RNA (about 10^4 copies per HeLa cell) appears very similar to that measured for most snoRNAs in vertebrates: with the exception of the most abundant U3 RNA (2×10^5 copies) and to a lesser extent of U8 and 7-2 MRP RNAs (about 4×10^4 and 3×10^4 copies respectively), most of these species are present at about 10^4 copies per cell (3,10,15).

Structural conservation and potential function of U21 RNA

An homolog of U21 RNA can be detected in all vertebrate classes, which exhibits an almost identical size in the different species analyzed so far (Fig. 3). Comparisons of U21 RNA sequences determined for the two mammals (mouse and human) and for chicken plainly show that the 13 nt. long complementarity to 28S rRNA corresponds to the most strongly conserved segment of U21 RNA, in addition to the box C and box D motifs and to the nucleotides forming the terminal stem. It is noteworthy that utilization of the same antisense probe in less stringent conditions of Northern hybridization than those used in Figure 3a experiment has recently allowed us to detect and characterize the U21 RNA homolog in an invertebrate metazoan, *Drosophila melanogaster* (Renalier, M.H., Nicoloso, M. and Bachellerie, J.P., unpublished results), further confirming the preferential conservation of this particular segment of U21 RNA sequence.

The presence of a 13 nt. long perfect sequence complementarity between a molecule as short as U21 RNA and either the large or the small subunit rRNA is unlikely to result from random chance occurrence (its probability to occur between a pair of randomized sequence of such lengths is lower than 0.01). Its biological significance is reinforced by the phylogenetic conservation of the two matching sequences. Moreover, the

preferential and outstanding conservation of this particular segment of U21 RNA during vertebrate evolution must reflect the key importance of the pairing for snoRNA function, similar to what has been observed for U14 (7).

The exclusively nucleolar localization of U21 RNA in growing cells indicates that the base-pairing must take place with 28S rRNA-containing precursors and not with mature rRNA. The identity of the U21 RNA-matching sequence in 28S rRNA deserves two comments. Firstly, this segment contains one of the five base-methylations along the 28S rRNA sequence (Fig. 8a). Like the 2′-O ribose methylations, these modifications take place on the rRNA gene primary transcript (36). Considering that this particular base modification, i.e., the formation of a m^1A, suppresses Watson−Crick pairing, this co-localization might suggest some functional link between U21 RNA pairing to (or its unpairing from) pre-rRNA and the covalent modification of pre-rRNA structure. In this regard, it is noteworthy that some mutations of yeast fibrillarin have been shown to strongly inhibit nucleolar methylation of pre-rRNA (35). The second comment deals with the intriguing overlapping (Fig. 8A) observed between the 28S rRNA segment complementary to U21 RNA and the tract complementary to another recently reported intronic snoRNA, U18 (14). Remarkably, the overlapping portion of the two segments (which are both 13 nt. long) contains the above-mentioned m^1A base methylation. Moreover, the two nucleotides in 28S rRNA which are immediately downstream for the segment of complementarity to U21 RNA (and are part of the complementarity to U18 RNA) are sites of 2′-O ribose methylation of rRNA (36). Northern hybridizations of total cellular RNAs from a series of vertebrates and invertebrate metazoans, using a U18 RNA antisense oligonucleotide probe (Fig. 8B), were consistent with a strong conservation of the 13 nt. complementarity to 28S rRNA in U18 RNA too. Such a clustering of two likely snoRNA binding sites raises the possibility that the two snoRNAs act in some concerted fashion during the nucleolar stages of ribosome biogenesis. Identifying the particular rRNA precursors to which U21 RNA and U18 RNA bind should definitely provide information on the temporal order of formation of the two distinct base-pairings. It should also help to settle the temporal relationship between methylation of the rRNA transcript and its base-pairing association with the snoRNAs.

Biosynthesis of U21 RNA and significance of its intronic location

Southern analysis of chicken genomic DNA (not shown) indicates that U21 RNA cannot be encoded elsewhere than in rpL5 gene intron 5. Moreover, the absence of a 5′ trimethyl-guanosine cap strongly suggests that the 5′ terminal nucleotide of U21 RNA does not correspond to the 5′ end of a primary transcript and that U21 RNA must result from RNA processing. The extensive sequence divergence observed between mouse and human for rpL5 gene intron 5 outside the U21 coding region and its immediate flanking nucleotides (Fig. 2) seems also to rule out that U21 RNA originates from an independent transcription unit located within intron 5. This conclusion is reinforced by the absence in these intronic sequences of the classical sequence motifs acting as transcriptional signals for vertebrate snRNA genes (37). All the available evidence point to U21 RNA resulting from the novel form of intronic RNA processing already reported for the four other intron-encoded snoRNAs with which U21 is structurally more particularly related, i.e., U14, U15, U16 and U20 (11−13, 15) as mentioned above. In this regard, the very significant conservation observed between mouse and humans for the nucleotides which flank immediately the U21 RNA coding region (Fig. 2c) seems noteworthy: interestingly in the two mammals and in chicken, the 4 bp long 5′−3′ terminal stem of mature U21 RNA can be extended to 8−9 bp in the intronic RNA precursor. As observed recently (12), a similar situation also occurs in the case of U14 and U15 RNAs and this extended stem structure might represent a common processing signal for this subset of intronic snoRNA.

The detection of U21 RNA in an intron of a ribosomal protein gene fits well into the pattern observed so far for intron-encoded snoRNA, i.e., all their host genes encode proteins which are ribosomal, nucleolar or ribosome associated (3,10,38). Moreover, it is noteworthy that the two intron-encoded snoRNAs, U21 and U18, which seem likely to be involved in the biogenesis of the large-subunit of the ribosome (taking into account their long sequence complementarity to 28S rRNA), are located within a gene coding for a ribosomal protein of the large subunit. A similar correlation has been recently observed for U20 (15), which exhibits a 21 nt. long complementarity to 18S rRNA and which is encoded in the gene of a protein involved in the biogenesis of the small ribosomal subunit, nucleolin. However, although this particular form of gene organization and expression might provide the basis for important regulations in the coordinate production of components of each ribosomal subunit, the significance of the above-mentioned correlation remains to be assessed. In this regard, it is important that the presence of the U21 RNA coding sequence in rpL5 gene intron 5 does not extend to all vertebrate classes: a different genomic location has been observed in an amphibian and a fish (Qu, L.H. and Bachellerie, J.P., unpublished results). Moreover, in Drosophila the host-gene of intron-encoded U21 RNA does not encode a ribosomal protein (Renalier, M.H. and Bachellerie, J.P., unpublished results). Identification of the genomic organization of U21 coding sequence in a wider spectrum of eucaryotic species together with further information on the detailed role of U21 RNA are now definitely required to better evaluate the potential links between the function and/or expression of an intronic snoRNA and its host-gene protein.

ACKNOWLEDGEMENTS

We are grateful to N.Joseph for her skillful technical assistance in all the sequence analyses. We thank R.Lührmann for his generous gift of an anticap antibody and G.Reimer and E.Tan for the antifibrillarin antibody. This work received financial aid from the Centre National de la Recherche Scientifique and Université Paul-Sabatier, Toulouse, France and was supported by grants from the Ministère de l'Education Nationale (action Evolution) and from Région Midi-Pyrénées to J.P.Bachellerie.

Note

Sequences of the rpL5 gene intron 5 (and flanking exons) in humans and mouse are available in the EMBL Data Library under accession numbers Z35312 and Z35311 respectively. Sequence of chicken U21 cDNA appears under accession number Z35310.

REFERENCES

1. Tollervey, D. and Hurt, E.C. (1990) Mol. Biol. Rep., **14**, 103−106.
2. Fournier, M.J. and Maxwell, E.S. (1993) Trends Biochem. Sci., **18**, 131−135.

3. Filipowicz, W., and Kiss, T. (1993) Mol. Biol. Rep., **18,** 149–156.
4. Kass, S., Tyc, K., Steitz, A. and Sollner-Webb, B., (1990) Cell, **60,** 897–908.
5. Savino, R. and Gerbi, S.A. (1990) EMBO J., **9**, 2299–2308.
6. Hughes, J.M.X. and Ares, M. Jr. (1991) EMBO J., **10,** 4231–4239.
7. Jarmolowski, A., Zagorski, J., Li, H.V.and Fournier M.J. (1990) EMBO J., **9**, 4503–4509.
8. Peculis, B.A., and Steitz, J.A. (1993) Cell., **73**, 1233–1245.
9. Morissey, J., and Tollervey, D. (1993) Mol. Cell. Biol. **13,** 2469–2477.
10. Sollner-Webb, B. (1993) Cell, **75**, 403–405.
11. Leverette, R.D., Andrews, M.T. and Maxwell, E.S. (1992) Cell, **71**, 1215–1221.
12. Tycowski, K.T., Shu, M.-D. and Steitz, J.A. (1993) Genes Dev., **7**, 1176–1190.
13. Fragapane, P., Prislei, S., Michienzi, A., Caffarelli, E. and Bozzoni, I. (1993) EMBO J., **12**, 2921–2928.
14. Prislei, S., Michienzi, A., Presutti, C., Fragapane, P. and Bozzoni, I. (1993) Nucleic Acids Res., **21**, 5824–5830.
15. Nicoloso, M., Caizergues-Ferrer, M., Michot, B., Azum, M.C. and Bachellerie, J.P. (1994) Mol. Cell. Biol. **14**, 5766–5776.
16. Hughes, J.M.X., Konings, D.A.M. and Cesareni, G. (1987) EMBO J., **6**, 2145–2155.
17. Tyc, K and Steitz, J.A. (1989) EMBO J., **8**, 3113–3119.
18. Mazan, S., Gulli, M.P., Joseph, N. and Bachellerie, J.P. (1992) Eur. J. Biochem., **205**, 1033–1041.
19. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. Kenmochi, N., Maeda, N. and Tanaka, T. (1992) Gene, **119**, 215–219.
21. Chan, Y.-L., Lin, A., McNally, J. and Wool, I.G. (1987) J. Biol. Chem., **262**, 12879–12886.
22. Wormington, W.M. (1989) Mol. Cell. Biol., **9**, 5281–5288.
23. Tang, B., and Nazar, R.N. (1991) J. Biol. Chem. , **266**, 6120–6123.
24. Bachellerie, J.P. and Qu, L.H. (1993) Methods Enzymol., **224**, 349–356.
25. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol., **65**, 499–560.
26. Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem., **162**, 156–159.
27. Meinkoth, J. and Wahl, G. (1984) Anal. Biochem., **138**, 267–284.
28. Reimer, G., Pollard, K.M., Penning, C.A., Ochs, R.L., Lischwe, M.A., Busch, H. and Tan, E.M. (1987) Arthritis Rheum., **30**, 793–800.
29. Lerner, M.R., Boyle, J.A., Hardin, J.A. and Steitz, J.A. (1981) Science, **211**, 400–402.
30. Lührmann, R., Appel, B., Bringmann, P., Rinke, J., Reute, R., Rothe, S. and Bald, R. (1982) Nucleic Acids Res., **10**, 7103–7113.
31. Hassouna, N., Michot, B. and Bachellerie, J.P. (1984) Nucleic Acids Res., **12**, 3563–3583.
32. Gutell, R.R., and G.E. Fox. (1988) Nucleic Acids Res., **16**, r175–r269.
33. Reddy, R., and H. Bush. (1988) Structure and function of major and minor small nuclear ribonucleoprotein particles (Birnstiel, M.L. ed.) Springer Verlag, Berlin, pp. 1–37.
34. Caizergues-Ferrer, M., Mathieu, C., Mariottini, P., Amalric, F. and Amaldi, F. (1991) Development, **112**, 317–326.
35. Tollervey, D., Lehtonen, H., Jansen, R., Kern, H., and Hurt, E.C. (1993) Cell, **72**, 443–458.
36. Maden, B.E.H. (1990) Prog. Nucleic Acid Res. and Mol. Biol., **39**, 241–301.
37. Dahlberg, J.E., and Lund, E. (1988) Structure and function of major small nuclear ribonucleoprotein particles. (Birnstiel, M.L. ed.), Springer-Verlag (Berlin) pp. 38–70.
38. Cecconi, F., Mariottini, P., Loreni, F., Pierandrei-Amaldi, P., Campioni, N. and Amaldi, F. 1994. Nucleic Acids Res., **22**, 732–741.
39. Balakin, A.G., Schneider, G.S., Corbett, M.S., Ni, J., and Fournier, M.J. (1993) Nucleic Acids Res., **21**, 5391–5397.