The EMBO Journal vol.12 no.7 pp.2921 – 2928, 1993

A novel small nucleolar RNA (U16) is encoded inside a ribosomal protein intron and originates by processing of the pre-mRNA

Paola Fragapane, Silvia Prislei[1], Alessandro Michienzi[1], Elisa Caffarelli and Irene Bozzoni[1,2]

Centro Acidi Nucleici of CNR and [1]Istituto Pasteur Fondazione Cenci-Bolognetti, Dipartimento di Genetica e Biologia Molecolare, Università 'La Sapienza', Roma, Italy
[2]Corresponding author

Communicated by F.Amaldi

We report that the third intron of the L1 ribosomal protein gene of *Xenopus laevis* encodes a previously uncharacterized small nucleolar RNA that we called U16. This snRNA is not independently transcribed; instead it originates by processing of the pre-mRNA in which it is contained. Its sequence, localization and biosynthesis are phylogenetically conserved: in the corresponding intron of the human L1 ribosomal protein gene a highly homologous region is found which can be released from the pre-mRNA by a mechanism similar to that described for the amphibian U16 RNA. The presence of a snoRNA inside an intron of the L1 ribosomal protein gene and the phylogenetic conservation of this gene arrangement suggest an important regulatory/functional link between these two components.

Key words: intron/processing/r-proteins/snRNA/splicing

Introduction

Ribosome biosynthesis is a complex process which requires the coordinated synthesis of many different components and their precise temporal and stoichiometrical assembly. Until a few years ago studies were mainly focused on the regulation of expression of rRNAs and r-proteins. Recently there has been a growing interest in small nucleolar RNAs (snoRNAs) which are found associated with the pre-ribosome in the nucleolus. Although the number of known snoRNAs is continuously increasing (Reddy *et al.*, 1979; Trinh-Rohlik and Maxwell, 1988; Tyc and Steitz, 1989) very little is known about their function; U3 is the only one for which a specific role in rRNA processing has been demonstrated (Kass *et al.*, 1990; Savino and Gerbi, 1990). Thus, the study of snoRNAs and of their function with respect to the synthesis of the structural components of the ribosome has acquired a pivotal role for clarifying the molecular steps of ribosome biogenesis. In the present paper we deal with this aspect of gene expression coordination as we found a physical association between an r-protein gene and that of a snoRNA.

We previously reported (Caffarelli *et al.*, 1987; Fragapane *et al.*, 1992) that the third intron of the L1 ribosomal protein gene of *Xenopus laevis* has a peculiar behaviour upon injection into oocyte nuclei: it splices very inefficiently and undergoes site specific endonucleolytic cleavages with the accumulation of truncated molecules. The low splicing efficiency depends on the presence of suboptimal splice sites (Caffarelli *et al.*, 1992; Fragapane *et al.*, 1992) while cleavages depend on an internal, 110 nucleotide (nt) long, intron region. Phylogenetic comparison showed that both the splice site sequences and the internal intron region are conserved in the second copy of the *X.laevis* L1 gene and in the single copy of *X.tropicalis*. In this paper we report an extension of the phylogenetic analysis showing that both the splice site sequences and the 110 nt intron region are conserved in the human species. In addition, we show that the internal intron region, when excised from the pre-mRNA, becomes a small RNA which is associated with fibrillarin and is localized inside the nucleolus. This previously uncharacterized species (U16) originates not from independent transcription but from a processing event.

Results

The third intron of the X.laevis L1 r-protein gene contains an internal region which is highly conserved in the corresponding intron of the L1 human gene

In order to gain more information on the possible role of the conserved sequence of the L1 third intron, we isolated the corresponding sequences from the human genome. Two oligonucleotides, spanning 19 nt in the 3′ portion of the third exon and 20 nt in the 5′ portion of the fourth exon, were used as primers for PCR amplification on total genomic human DNA [the sequence was derived from the human L1 cDNA clone isolated by C.Bagni, F.Annesi and F.Amaldi (personal communication)]. The amplified fragments were cloned under the T7 promoter of the Bluescript plasmid and sequenced. Figure 1a shows the sequence deduced for the entire intron and part of the flanking exons. The nucleotides corresponding to the conserved intron region of the L1a gene copy of *X.laevis* are indicated in bold characters and appear to be highly conserved with the exception of a few substitutions, deletions and insertions which are indicated above. The rest of the intron differs completely in length and sequence from the *Xenopus* intron, with the only exception of the 5′ splice junction which is identical (AC/GTATC). This is quite interesting because this sub-optimal splice site was previously shown (Fragapane *et al.*, 1992) to be responsible for the low splicing efficiency of the *X.laevis* L1 third intron. The evolutionary conservation of this suboptimal splice sequence, which confers low splicing efficiency also in the human species (see below), suggests that it must have a fundamental role for the correct expression of this gene.

We previously reported that the corresponding introns isolated from the second copy of the L1 gene in *X.laevis* and from the single copy of the *X.tropicalis* gene behaved similarly when microinjected in oocytes: namely, very little splicing occurred and specific endonucleolytic cleavages converted the majority of the input RNA into truncated

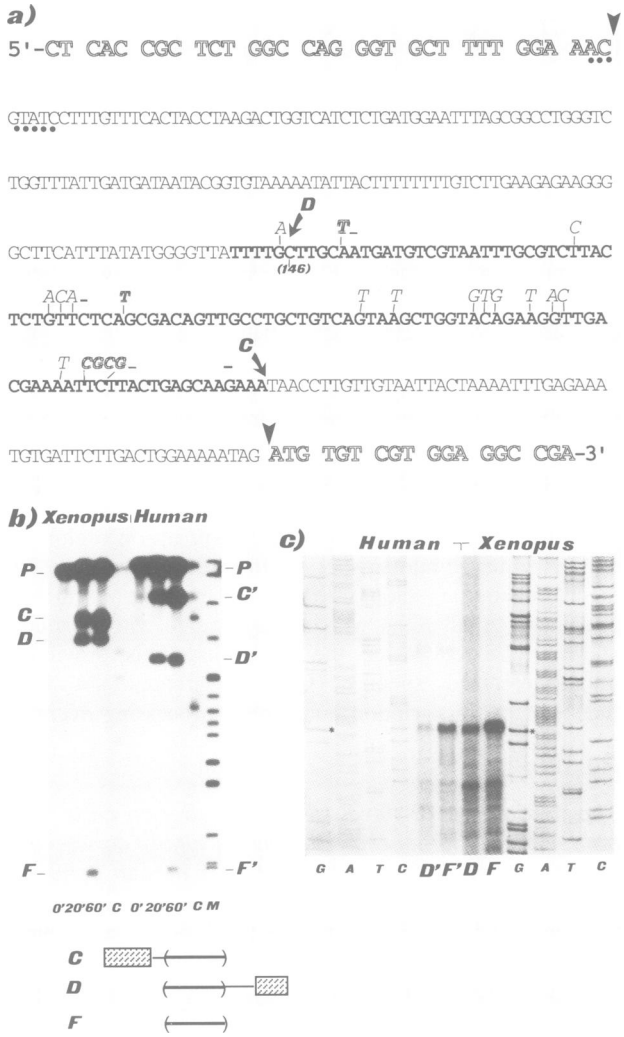

Fig. 1. (**a**) Nucleotide sequence of the third intron and part of the flanking exons (outlined nucleotides) of the L1 human gene. The conserved 5′ splice site sequence is underlined; the sequence homologous to the *X.laevis* intron region is in bold characters. The differences in the *X.laevis* sequence are indicated above: the insertions are represented by shadowed characters and deletions by dashes. Arrows point to the 5′ and 3′ ends of the D- and C-type molecules respectively. (**b**) *In vivo* splicing analysis of the third intron containing precursors of the *X.laevis* and human L1 genes. The ^{32}P-labelled transcripts were injected into the germinal vesicles of *X.laevis* oocytes and incubated for 0, 20 and 60 min; the nuclei were manually dissected and RNA was extracted and loaded on to a 10% denaturing polyacrylamide gel. C, cytoplasmic RNAs of the 60 min incubations; M, *Msp*II-digested pBR322 plasmid. C, D and F cleavage products are indicated at the side and schematically represented below. (**c**) Reverse transcriptase elongations on gel purified D and F molecules of human (D' and F') and *Xenopus* (D and F) using the γ-h and γ primers which are complementary to the 19 terminal nucleotides of the intron's conserved regions. The products were run in parallel with the sequence (G, A, T and C) performed with the same oligos on 003h and 003a plasmid DNAs. The extended products comigrate with the G (indicated by the asterisk) corresponding to the C of the coding strand at positions +146 and +36 of the human and *X.laevis* introns respectively (for the 003a sequence see Prislei *et al.*, 1992).

molecules (Prislei *et al.*, 1992). We performed the same experiment with the human sequence by microinjecting ^{32}P-labelled pre-mRNA containing the third intron (003h) into the nuclei of *X.laevis* oocytes and analysing the nuclear RNA products after 20 and 60 min of incubation. Figure 1b shows that the human third intron behaves exactly like the amphibian sequence: very little splicing occurs and site specific cleavages produce C- and D-type molecules with the final release of a 106 nt molecule (band F) which is stably accumulated. All these different RNAs were eluted from the gel and analysed by RNase H digestion and reverse transcriptase elongation using the γ-h and γ oligonucleotides which are complementary to the terminal 19 nt of the human and *Xenopus* conserved intron regions. The RNase H analysis (not shown) demonstrated that in analogy with the amphibian intron (Fragapane *et al.*, 1992), D-type molecules extend from the beginning of the conserved region to the 3′ end of the precursor, C-type molecules extend from the 5′ end of the transcripts up to the end of the conserved region and the F molecules originate from double C and D cleavage of the pre-mRNA (see schematic representation in Figure 1b). The reverse transcriptase elongation analysis allowed the precise identification of the 5′ end of the D and F molecules. Panel c shows that the human and *Xenopus* D and F molecules give extension products of the same size. The sequences run in parallel allow the 5′ ends to be mapped at the Gs indicated by asterisks. These Gs correspond, on the coding strands, to the C at the beginning of the conserved intron region: position +146 of the human intron and position +36 of the *Xenopus* one (Prislei *et al.*, 1992). The bands of the reverse transcriptase products are not as sharp as those of the sequences run alongside because they originate from gel purified material; independent experiments have confirmed the comigration with the indicated Gs.

Reverse transcriptase analysis of the human D and F molecules shows two minor extension products mapping 6 and 30 nt upstream of the major site. As already described for the *Xenopus* sequence (Fragapane *et al.*, 1992) and as recently observed *in vitro* (manuscript in preparation) it appears that the primary cleavages that produce D-type molecules occur upstream of the mature end. Similarly, C-type molecules also originate by downstream primary cleavage(s). Mature D and C termini will then derive from rapid exo/endo trimming of the precursor molecules.

The highly conserved region hybridizes to a snoRNA that is associated with fibrillarin

The peculiar conservation of both the sequence and the processing of the third intron internal region stimulated the search for corresponding sequences among endogenous oocyte transcripts. Figure 2 shows that when the conserved intron region is utilized as a probe (α-γ DNA, referred to as U16, see Materials and methods) on Northern blots it indeed identifies an snRNA species (lane Nu). The size, calibrated with RNA markers and by comigration with U6 RNA, is in the range 105−110 nt. Hybridization with RNA probes corresponding to each of the two DNA strands was also performed, confirming that the identified snRNA corresponds to the coding strand of L1 (not shown).

Since it is known that many snRNAs are associated with very well characterized antigens, we screened specific snRNP particles for the presence of RNA sequences corresponding to the conserved intron region. Antibodies against Sm, trimethylguanosine cap (m_3G) and fibrillarin antigens were used to immunoprecipitate *X.laevis* nuclear RNPs and subsequent Northern analysis of the recovered RNA. Figure 2 shows that the U16 probe gives specific hybridization with the anti-fibrillarin immunoprecipitation sample (Fibr). In order to control the specificity of the

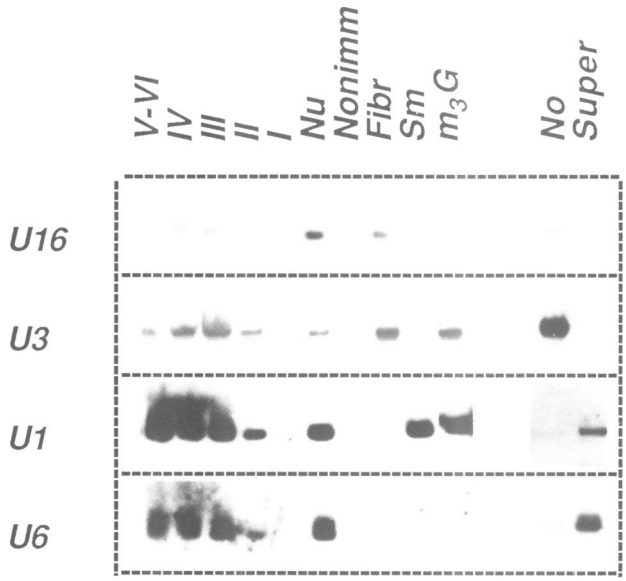

Fig. 2. (**Left**): Northern analysis of RNAs from 10 *X.laevis* oocytes at different stages (I–VI, according to Dumont, 1971), 10 nuclei (Nu), or 30 nuclei immunoprecipitated with non-immune rabbit serum (Nonimm), anti-fibrillarin (Fibr), anti-Sm (Sm) and anti-trimethylguanosine cap (m_3G) antibodies. (**Right**): Northern analysis of RNAs from nucleolar (No) and nucleoplasmic (Super) preparations from 50 nuclei. The same filters were utilized for subsequent hybridizations with the different snRNA probes indicated on the left.

immunoprecipitations the same filter was re-hybridized with the U3, U1 and U6 RNA probes. The different panels show that each probe identifies the corresponding sequence in the appropriate sample: the U3 RNA probe hybridizes to fibrillarin and m_3G immunoprecipitates (Tyc and Steitz, 1989) whereas U1 RNA is positive with Sm and m_3G antibodies. U6 RNA is negative for all these antibodies because it does not possess the m_3G cap and is not associated with Sm antigens (for a review see Reddy and Bush, 1988). No indirect reactivity of U6 RNPs with these antibodies, due to the U4 association, is expected since in the oocyte most of the U6 RNA is present in U6 snRNPs rather than in U4/U6 snRNP complexes as in somatic cells (Hamm and Mattaj, 1989).

The analysis of expression during oogenesis shows that the four different RNAs behave very similarly: expression increases from stage I to III (Dumont, 1972) and thereafter remains at a plateau with some decrease in mature oocytes.

The next effort was directed towards identifying the nuclear localization of this snRNA. 50 manually dissected nuclei were treated according to Peculis and Gall (1992) in order to fractionate nucleoli from nucleoplasm. The RNA from the nucleolar pellet was extracted and analysed by Northern blotting (lane No) in parallel with the supernatant fraction (lane Super). The right panel of Figure 2 shows that a specific hybridization with the U16 probe is detected only in the pellet fraction such as the control hybridization with the U3 probe. On the other hand, control hybridizations with non-nucleolar snRNAs, such as U1 and U6, show that these RNAs are found predominantly in the nucleoplasm supernatant.

The conserved third intron sequence is represented among the endogenous oocyte RNAs

In order to verify the sequence correspondence between the third intron conserved region and products endogenous to the oocyte we proceeded to clone the small RNA identified by hybridization. RNA from a 4–8S sucrose gradient fraction was separated on a 6% polyacrylamide gel and the RNA corresponding to a size of 100–120 nt was eluted and utilized as substrate for poly(A) tailing with poly(A) polymerase and reverse transcription in the presence of oligo(dT) primers. The cDNA was ligated to adaptor oligos and PCR amplified with primers complementary to the synthetic tails (see schematic representation of Figure 3). The amplified DNA was cloned into the *Eco*RI and *Xho*I sites of the Bluescript vector and the clones were screened for hybridization to the intron conserved region (α-γ probe). Three positives were isolated from more than 1000 clones analysed: all three have the same 3′ end: it maps at one of the five A residues at the end of the intron conserved region. Due to the poly(A) tailing cloning procedure it is impossible to determine which of the five As is the 3′ terminus. As far as the 5′ end is concerned the isolated clones appear to be products of partial reverse transcriptase elongation, the longest extending up to the G indicated by the asterisk in Figure 3. In order to verify the real 5′ end of the endogenous U16 RNA, reverse transcriptase elongation was performed with the γ-primer which is complementary to the terminal 19 nt of the conserved intron region (see schematic representation and lower panel of Figure 3). The elongation products (lane U16) extend up to the C at the beginning of the conserved intron region (position +36 of the *Xenopus* L1a intron). The extension is the same as that obtained with the same primer on the D and F molecules (see Figure 1, lanes D and F) indicating that the 5′ ends of all these different molecules are the same. Thus our longest cDNA clone is only three nucleotides shorter than the endogenous RNA. These results demonstrate that in the endogenous oocyte RNA population, sequences corresponding to the conserved region of the L1 third intron are present. We propose to call this previously uncharacterized species U16 RNA. The three U16 clones analysed contained sequences corresponding to the L1a intron; in consideration of the limited number of clones analysed it cannot be excluded that the L1b intron also contributes to the U16 RNA pool (U16b). RT-PCR amplification performed with primers internal to the U16 RNA sequence on the same RNA preparation utilized for the cloning, showed the presence also of L1b sequences (not shown). It is then likely that the U16 RNA is composed of two subpopulations: the U16a and U16b RNAs.

The U3, U8 and U13 nucleolar snRNAs contain two highly conserved sequences, a 9 nt box C and a 6 nt box D (Tyc and Steitz, 1989). The U16a and U16b sequences show that boxes C and D are present and that box D is also situated a few nucleotides from the 3′ end (see Table I). The same boxes are also present in the putative U16 sequence identified in the L1 human intron, although we do not have a formal proof that a corresponding RNA exists in the snRNA population.

The only coding sequences for U16 are inside the L1 gene

A Southern blot analysis was performed on *X.laevis* genomic DNA to check if, outside the L1 gene, there was any sequence contributing to the accumulation of U16 RNA. *X.laevis* genomic DNA was digested with *Eco*RI and *Bam*HI and hybridized with the L1a (003a) and L1b (003b) third introns and with an L1 cDNA probe. Figure 4 shows that

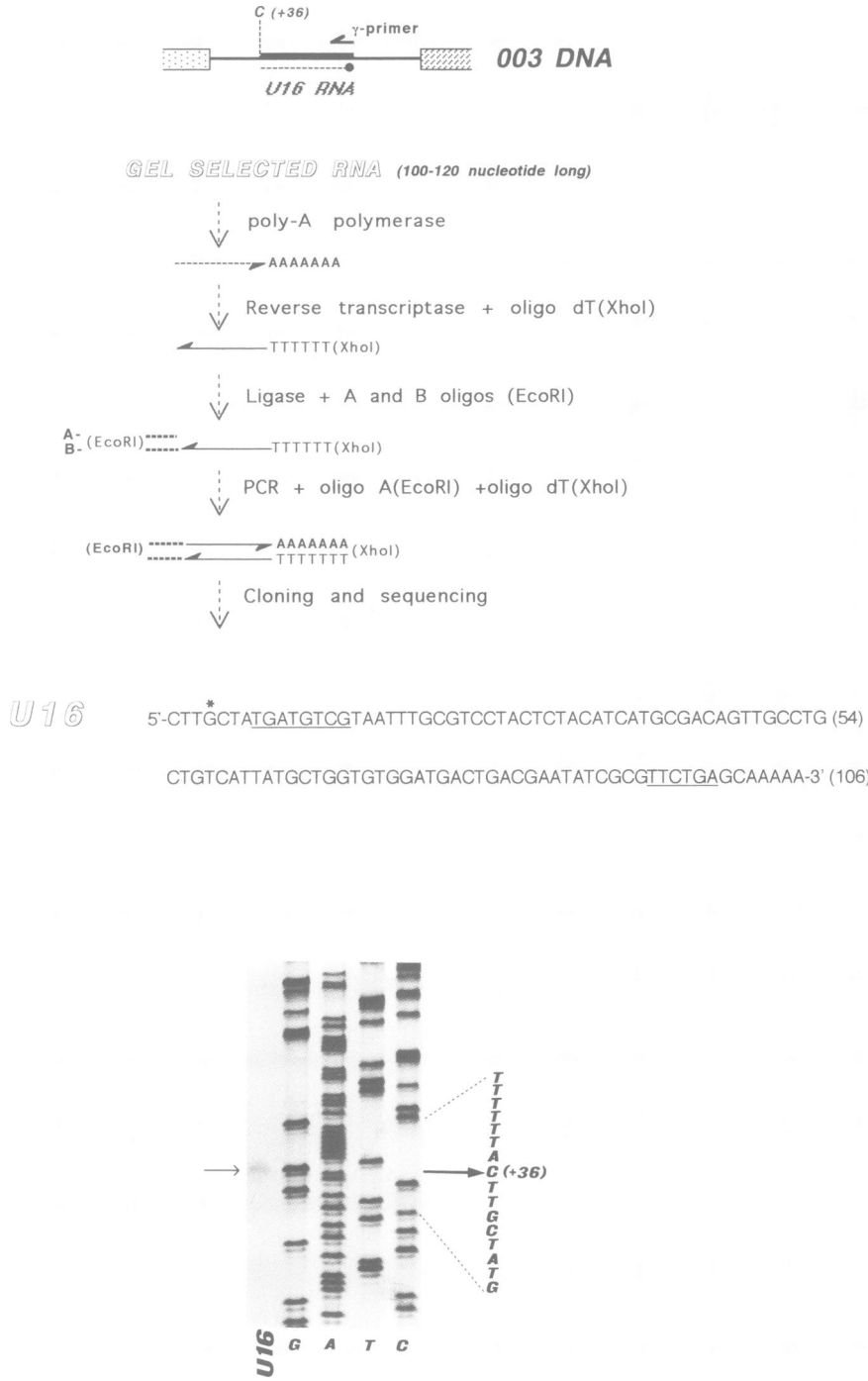

Fig. 3. (**Above**) schematic representation of the procedure used to clone U16 RNA sequences. The underlined regions correspond to the C and D boxes. (**Centre**) the nucleotide sequence of U16 RNA; the asterisk indicates the 5′ end of the longest cDNA clone isolated, the three additional nucleotides are deduced from the RT experiment. (**Below**) identification of the 5′ end of U16 RNA. End-labelled γ primer was annealed with 7 μg of gel selected RNA and elongated with reverse transcriptase. The product (lane U16) was run in parallel with the sequence (G, A, T and C) performed with the same oligo on 003a plasmid DNA; it comigrates (indicated by the arrow) with the G corresponding to the C of the coding strand at position +36 of the *X.laevis* third intron. On the right the deduced complementary nucleotides to the shown sequence are given.

the two intron probes identify two bands of 2 and 1.7 kb. They correspond to the *Eco*RI−*Bam*HI fragments containing the third intron of the two gene copies (these restriction fragments extend from the middle of the first intron to the middle of the fifth one). As a control, the same filter was hybridized to the L1 cDNA probe. The L1a and L1b genes are single copy genes (Bozzoni *et al.*, 1981), so the hybridization with the cDNA probe, which had almost the same specific activity as the other two probes, represents a control for a single copy gene. The cDNA probe recognizes, in addition to these two bands, a 7 kb fragment which contains the remaining part of the gene downstream of the fifth intron (Loreni *et al.*, 1985). The *Eco*RI−*Bam*HI fragment containing the first exon is not visualized because the cDNA probe lacks this sequence. The asterisked band, which is recognized by all probes, is a specific product of partial digestion (the first exon-containing fragment plus the downstream flanking segment). The fact that the cDNA

Table I. Comparison of C and D box sequences of U16 RNAs with known examples of snoRNAs

snRNA	Organism	C box	D box
U3	human	UGAUGAUUG	GUCUGA
U3	*Xenopus*	UGAUGAACG	G-CUGA
U8	human	UGAUGAUCG	AUCUGA
U13	human	UGAUGAUUG	GUCUGA
U16a	*Xenopus*	UGAUG-UCG	UUCUGA
U16b	*Xenopus*	UGAUG-UCG	UUCUGA
U16	human	UGAUG-UCG	UUCUGA A

U3 is the only other snoRNA known in *Xenopus* (Jeppesen *et al.*, 1988).

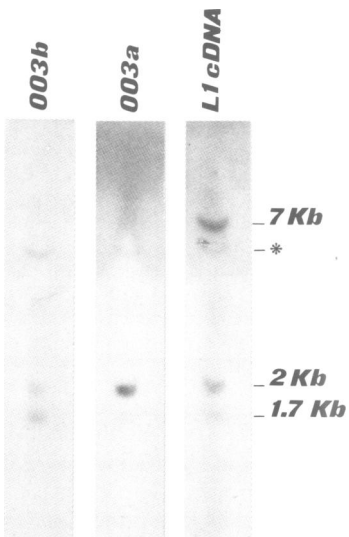

Fig. 4. Southern blot analysis of total genomic DNA digested with *Eco*RI and *Bam*HI. The probes utilized are indicated above: 003a and 003b contain the third intron plus a few nucleotides of the flanking exons of the L1a and L1b genes (see Materials and methods). The L1 cDNA contains all the L1 coding region except the first exon. The size of the hybridization bands are indicated on the right. The asterisked band is a product of partial digestion.

probe recognizes all the bands identified with the intron probes indicates that the only coding sequences for U16 RNA are inside the L1 gene.

The U16 RNA originates by processing of the L1 pre-mRNA and not by autogenous transcription

To investigate further the mechanism responsible for the synthesis of U16 RNA, and to test its correlation with the F molecule (see Figure 1), we analysed the fibrillarin association of the different RNA products obtained after oocyte microinjection of ^{32}P-labelled 003a RNA. Figure 5 shows that in comparison with samples immunoprecipitated with preimmune serum (lanes 1) and to control nonimmunoprecipitated samples (lanes 3), the samples immunoprecipitated with anti-fibrillarin antibodies (lanes 2) show specific reactivity that follows the time course of production of F molecules: at 30 min the D band is almost totally precipitated together with small amounts of F molecules; the latter become the major products of immunoprecipitation at 60 min when their accumulation reaches a maximum. It is more difficult to interpret the reactivity of the C-band because of its low level of

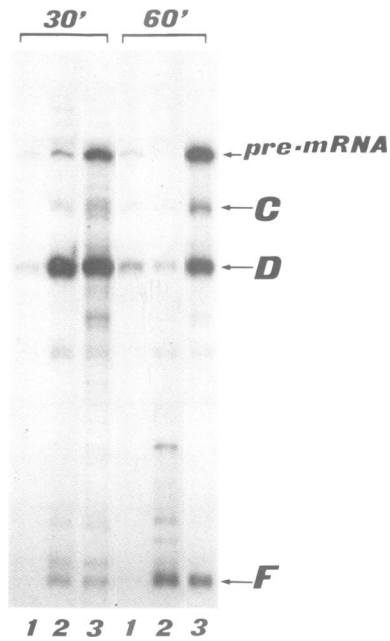

Fig. 5. Analysis of fibrillarin association with the products of 003 RNA injection. ^{32}P-labelled 003 RNA was injected into *X.laevis* nuclei, incubation was allowed to proceed for 30 or 60 min and nuclei were manually purified. RNA was extracted from 15 nuclei immunoprecipiteted with preimmune serum (lanes 1), 15 nuclei immunoprecipitated with anti-fibrillarin antibodies (lanes 2) and 15 untreated nuclei (lanes 3). The different C, D and F products are indicated on the right.

accumulation. We have previously shown that the production of the C molecules varies considerably depending on the batch of oocytes (Fragapane *et al.*, 1992). Finally, the pre-mRNA shows reactivity similar to the background. Thus, fibrillarin appears to be associated with both the mature F RNA (U16) and its precursor D molecule.

In 003a RNA, which is a poor splicing substrate and produces very low levels of the lariat form, F molecules originate predominantly from cleavage of the pre-mRNA. In any case, the lariat form does not seem to be a substrate for cleavage and F molecule production as shown by the analysis of mutant F4 (Fragapane *et al.*, 1992), which produces high levels of the lariat form due to the conversion of the suboptimal 5′ splice site into a canonical site. This substrate shows no significant conversion of the lariat into F molecules; equimolar amounts of lariat and mature RNA can in fact be visualized (and quantified by densitometric scanning) at all times of incubation (Fragapane *et al.*, 1992, Figure 3); if conversion of the lariat form into F molecules was occurring, one would have expected a lower accumulation of lariat with respect to mature RNA. In addition, no reactivity of the lariat has been observed with anti-fibrillarin antibodies (not shown).

These results demonstrate that the F molecules are present in RNP complexes containing fibrillarin, such as the endogenous U16 sequences; it is reasonable to think that the endogenous U16 RNA originates from a processing event analogous to that leading to the production of F molecules. The results presented above also suggest that, since the U16 RNA can be released only from the unspliced intron and not from the excised lariat, each precursor molecule will give rise to either a molecule of U16 RNA or a molecule of L1 mRNA.

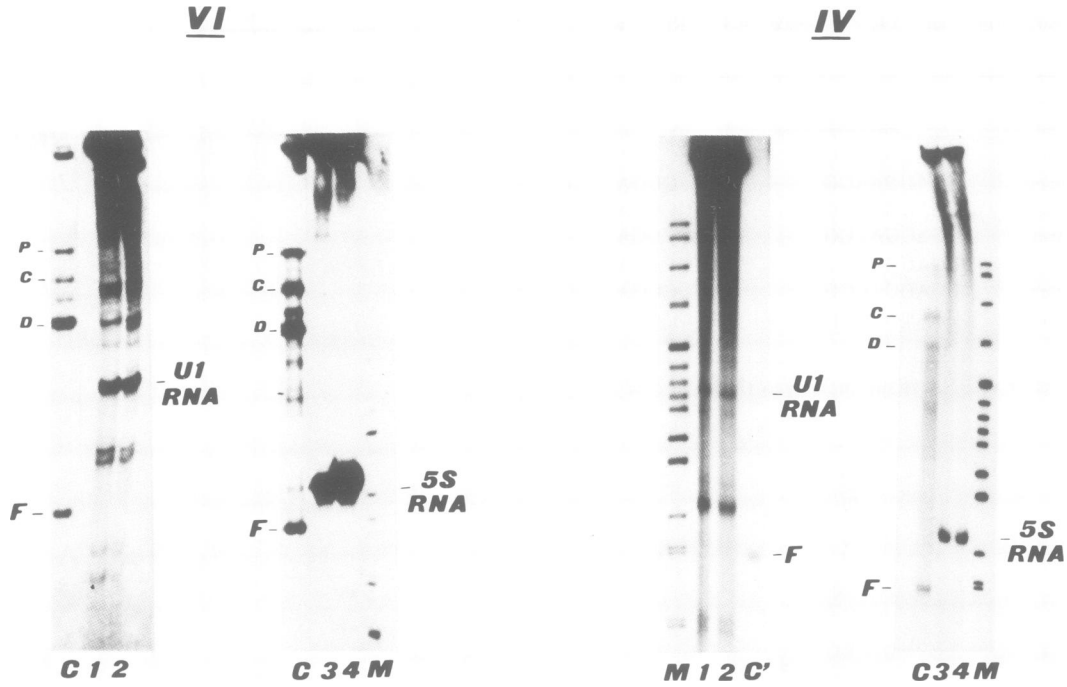

Fig. 6. Analysis of ^{32}P-labelled transcripts produced after microinjection of different DNA plasmid combinations in stage VI and stage IV oocytes. For pol II analysis U1 DNA was injected with 003a DNA (lanes 1) or with Bluescript plasmid (lanes 2). For pol III analysis 5S DNA was injected with 003 DNA (lanes 3) or with Bluescript plasmid (lanes 4). The RNA was extracted after 6 h of incubation and run on 10% denaturing polyacrylamide gels in parallel with the RNA products originating from the injection of ^{32}P-labelled 003a RNA (C) or with the gel purified F molecule (C'). The newly transcribed 5S and U1 RNAs are indicated. M, plasmid pBR322 digested with *Msp*II.

To rule out the possibility that F molecules could derive from transcriptional events starting inside or close to the third intron, we injected 003a plasmid DNA along with equimolar amounts of U1 RNA genes (lanes 1) or 5S RNA genes (lanes 3) into the nuclei of stage VI and stage IV *X.laevis* oocytes (Figure 6) in the presence of [^{32}P]UTP. These coinjected DNAs serve as internal standards for calibrating pol II and pol III transcription efficiencies. Control experiments were performed by coinjecting U1 DNA or 5S DNA together with Bluescript plasmid vector (lanes 2 and 4). Oocytes were incubated for 6 h and the RNA was extracted and analysed on 10% polyacrylamide gels. The results show that RNA bands comigrating with the F molecule are not detected in either stage VI or stage IV oocytes. From these data it is possible to conclude that the biogenesis of U16 RNA depends on specific cleavage events which operate on the precursor RNA.

Discussion

Since the first identification of snRNAs, defined as stable RNAs found in protein complexes in the nuclei of all types of eukaryotic cells (reviewed by Busch *et al.*, 1982), many new species have been characterized and much progress has been made in defining their structure, function and biosynthesis. SnRNAs are divided into different classes depending on their function, RNP composition, 5′ and 3′ end modifications and transcriptional dependence on different RNA polymerases.

In mammalian cells, the U3, U8, U13 and U14 RNAs are the major representatives of a specific class of snRNPs, also known as snoRNPs, characterized by their compartmentalization inside the nucleolus (Reddy *et al.*, 1979; Parker and Steitz, 1987; Trinh-Rohlik and Maxwell, 1988; Tyc and Steitz, 1989; Liu and Maxwell, 1990) and by their reactivity with serum, from a scleroderma patient, containing autoantibodies against a 34 kDa protein known as fibrillarin (Ochs *et al.*, 1985). Besides these well characterized RNAs, other anti-fibrillarin precipitable RNAs have been identified, like RNAs X (now called U15 RNA, K.Tyc and J.Steitz, personal communication) and Y (Tyc and Steitz, 1989). Because of their cellular localization, snoRNPs have been thought to participate in rRNA processing; indeed, it has been demonstrated that the most abundant species, the U3 RNP particle, has a function in rRNA processing (Kass *et al.*, 1990; Savino and Gerbi, 1990). In yeast, a larger number of snRNPs seem to be localized in the nucleolus and associated with the fibrillarin equivalent protein Nop1 (Schimmang *et al.*, 1989). For the U3 equivalent several pieces of evidence suggest that it also has a specific role in rRNA processing (Beltrame and Tollervey, 1992).

In this paper we describe the identification of a novel snoRNA which is encoded in the third intron of the *X.laevis* L1 r-protein gene and which originates from site specific processing of the intron within the pre-mRNA. The interest in this sequence initially came from its peculiar splicing phenotype in different amphibian species (*X.laevis* and *X.tropicalis*) and in human: besides being a poor splicing substrate, the third intron undergoes site specific cleavages leading to the accumulation of truncated molecules. In particular, an RNA species corresponding to an internal intron region was released by all different precursors when tested for splicing in the *in vivo* oocyte system (Prislei *et al.*, 1992 and this paper). The high degree of conservation (only a few nucleotide substitutions differentiate the amphibian sequences from the human one) suggested a specific

functional role of this region *per se* and stimulated the search for endogenous equivalents.

Probing of Northern blots with the L1 third intron conserved region proved the existence of homologous sequences in the endogenous RNA population of the oocyte. This RNA species is associated with RNP complexes containing the fibrillarin antigen, does not possess an m_3G cap and colocalizes with U3 RNA in nucleolar fractionations. By synthetic polyadenylation followed by reverse transcriptase it was possible to clone a cDNA version of this RNA corresponding to the conserved sequence present in the third intron of the L1a gene copy. In consideration of the limited number of clones isolated it cannot be excluded that the L1b copy also contributes to the U16 RNA pool. Independent cloning performed with RT-PCR amplification with primers internal to the U16 sequence has in fact shown the presence of transcripts deriving also from the L1b copy. We have called this previously uncharacterized RNA U16 and it very probably comprises the U16a and U16b subspecies. In addition, their 5′ end is identical to one of the D- and F-type of molecules obtained upon injection of the RNA precursor containing the third intron. Immunoprecipitation with fibrillarin antibodies of nuclei microinjected with ^{32}P-labelled 003a RNA showed that the F molecules are associated with this antigen, as the U16 RNA. These results, together with the evidence that no F molecules are produced when a plasmid DNA containing the third intron plus flanking sequences is microinjected into oocyte nuclei, demonstrates that the U16 RNA originates by processing of the L1 pre-mRNA and not by independent transcription. Several pieces of evidence suggest that the F molecule, and as a consequence U16 RNA, derives from the unspliced pre-mRNA and not from the excised lariat: (i) the F4 mutant, which has a high splicing efficiency, shows equimolar accumulation of lariat and ligated exons suggesting that no conversion of lariat into F-type of molecules occurs (Fragapane *et al.*, 1992); (ii) reinjection of gel purified lariat does not produce any molecule similar to F-RNA (unpublished).

The production of U16 RNA represents quite a peculiar example of snRNA biosynthesis; in fact, the snRNAs characterized so far are produced by pol II or pol III transcription (for a review see Dahlberg and Lund, 1988). Two snRNAs originating not by independent transcription but by RNA processing have been recently described: U14 RNAs (Leverette *et al.*, 1992) and U15 RNA (K.Tyc and J.Steitz, personal communication). The common feature of these RNAs is that they are all localized in nucleoli. The interesting feature of U16 is that it is encoded inside a ribosomal protein intron which undergoes specific regulation (Caffarelli *et al.*, 1987; Pierandrei-Amaldi *et al.*, 1987). It will obviously be important to identify the specific function of U16 RNA in order to understand its functional and physical association with the r-protein L1 and to correlate it with rRNA synthesis. However, if the presence of snRNAs inside ribosomal protein genes is found to be a more widespread phenomenon, some additional functions beside rRNA processing can be envisaged. An intriguing hypothesis is that they could be involved in the assembly of rRNA with ribosomal proteins. This is quite a complex process which requires many structural rRNA rearrangements. In analogy with what is emerging from studies on spliceosome function, where productive structural rearrangements are achieved by alternative snRNA interactions, it might be that snoRNAs, by specifically interacting with rRNA, regulate the assembly of specific r-proteins (for instance, those coexpressed on the same pre-mRNA). The molecular characterization of these sequences and the availability of suitable *in vivo*/*in vitro* systems will hopefully allow in the near future a better understanding of the unclarified aspects of ribosome biosynthesis.

Materials and methods

Cloning of third intron sequences from the human L1 gene

The third intron of the human L1 gene was isolated by PCR amplification of genomic human DNA using primers derived from the sequence of the corresponding cDNA (C.Bagni, F.Annesi and F.Amaldi, personal communication).

The forward oligo (5′-GGGAATTCCCAGAGTTCGAGG-3′) includes 19 nt of the 3′ portion of the third exon (underlined) and an *Eco*RI site (bold characters), while the backward primer (5′-CCCTCGAGATGGC-GTATCGTTTTTGG-3′) includes 20 nt of the 5′ portion of the fourth exon (underlined) and an *Xho*I site (bold characters).

PCR reactions were carried out in 50 μl of 50 mM KCl, 10 mM Tris−HCl pH 8.3, 1.5 mM $MgCl_2$, 250 μM dNTP, 0.1 mg/ml BSA, 1 μM of each primer, 4.5 units of *Taq* polymerase and 1 mg of total genomic DNA. Each amplification cycle consisted of 1 min denaturation at 94°C, 2 min annealing at 58°C and 3 min extension at 72°C. Thirty cycles were performed.

The *Eco*RI and *Xho*I sites of the primers allow the amplified DNA to be cloned in the corresponding sites of the Bluescript vector (Stratagene). The recombinant plasmid (003h) was sequenced according to Sanger (1977).

In vitro RNA synthesis and oocyte microinjections

003a RNA was transcribed from plasmid 003a *Sac*I, digested with *Hin*dIII (see Prislei *et al.*, 1992); it contains 89 nt of the third exon, the entire third intron (241 bases), 59 nt of the fourth exon and 82 nt of vector polylinker.

The template for 003h transcription was obtained by *Xho*I digestion of the corresponding plasmid. The transcript includes the 307 nt intron of the human L1 gene. The plasmid also contains 59 nt of exon 3 and 91 nt of exon 4 plus 67 nt derived from the vector polylinker.

The ^{32}P-labelled T7 transcripts were synthesized *in vitro* and gel purified; transcripts were injected into the nuclei of stage VI oocytes and total RNA was extracted from manually purified nuclei as described by Caffarelli *et al.* (1987). Equal numbers of injected counts were loaded on 10% denaturing polyacrylamide gels.

DNA microinjections were performed as follows. 100 ng of 5S (Razvi *et al.*, 1983) and U1 (Zeller *et al.*, 1984) plasmids were mixed with 100 ng of 003a or Bluescript plasmids, precipitated and resuspended in 2 μl of injection buffer in the presence of 20 μCi of [^{32}P]UTP. The nuclei of oocytes were injected with 20 nl of the DNA solution and the oocytes were then incubated for 6 h. Nuclear RNA from stage VI and total RNA from stage IV oocytes was extracted and run on 10% denaturing polyacrylamide gels.

Northern blot analysis

RNAs were electrophoresed on 6% acrylamide−7 M urea gels, electrotransferred to Amersham's Hybond-N paper and UV crosslinked. All hybridizations were carried out in standard conditions but the one with the U3 probe which is a 22mer synthetic oligonucleotide complementary to nt 101−122 of *X.laevis* U3 snRNA sequences (Caizergues-Ferrer *et al.*, 1991). The oligo was 5′ ^{32}P-end-labelled and hybridized according to Caizergues-Ferrer *et al.* (1991). The U16 analysis was performed with the α-γ probe obtained by PCR amplification of the intron conserved region.

Southern blot analysis

Total genomic *X.laevis* DNA was digested with *Eco*RI and *Bam*HI and 15 μg were loaded on single slots of a 0.8% agarose gel. The DNA was transferred to nitrocellulose filters and hybridized with the different DNA probes: 003a and 003b contain the third intron of the L1a and L1b genes respectively (Prislei *et al.*, 1992); The L1 cDNA contains all the L1 coding region except for the first exon (plasmid p103, Loreni *et al.*, 1985).

Cloning of oocyte endogenous U16 RNA and 5′ end mapping

Total RNA extracted from ovaries was fractionated on 10−30% sucrose gradient. 45 μg of the 4−8S fractions were run on a 6% acrylamide−urea gel and the RNA corresponding to the region of U16 was eluted. The RNA

was 3′ elongated with poly(A) polymerase and reverse transcribed in the presence of an oligo(dT)-tailed primer (5′-GCCTCGAGTTTTTTTTTTTTTTTTTTTT-3′) containing an *Xho*I site. The cDNA was ligated to adaptor oligos (A, 5′-GGAATTCGAAGGGGG-3′; B, 5′-CCCCCTTCGAATTCCCCC-3′) containing an *Eco*RI site. The mixture was then PCR amplified utilizing the dT oligo as backward primer and the A adaptor oligo as forward primer. Each amplification cycle consisted of 1 min denaturation at 94°C, 2 min annealing at 50°C and 3 min extension at 72°C. Thirty cycles were performed. The *Eco*RI and *Xho*I sites of the primers allow the cloning of the amplified DNA in the corresponding sites of the Bluescript vector (Stratagene). The clones positive for colony hybridization with the α-γ probe (see above) were sequenced according to Sanger (1977).

To characterize the 5′ end of the U16 RNA, reverse transcriptase elongation was performed on 7 μg of gel selected RNA (see above) utilizing as primer the γ oligonucleotide (5′-CCAAGCTTTTTGCTCAGAACGCGAT-3′) which includes 19 nt of the 3′ portion of the conserved intron region (underlined nucleotides). The same oligo was utilized as primer for the reverse transcription of gel purified D and F RNAs originating from oocyte injection of 003a RNA. The reverse transcriptase elongation of D and F RNAs, obtained after the injection of the human third intron, was performed with the γ-h oligonucleotide (5′-CAAGCTTTCTTGCTCAGTAAGAAT-3′) which includes 19 nt of the 3′ portion of the human conserved intron region (underlined nucleotides). The RNAs and oligo mixtures were heated at 95°C for 5 min, annealed at 55°C for 15 min and incubated on ice for 10 min. 30 units of AMV reverse transcriptase (Biolabs), 20 units of RNasin (Boehringer) and dNTPs to a final concentration of 1 mM each were added to the reactions and incubations were continued at 42°C for 60 min. The products of the reactions were run on 6% denaturing polyacrylamide gels in parallel with a Sanger sequence performed with the same primers on the corresponding plasmid DNAs.

Antibodies

Clinical anti-Sm sera were obtained from P.Janteur (Université des Sciences et Techniques du Languedoc, Montpellier), the anti-m_3G cap antibodies were obtained from R.Luhrmann (Philipps University, Marburg) and the anti-fibrillarin sera were provided by M.Caizergues-Ferrer (CNRS, Toulouse).

Immunoprecipitations

The antisera utilized (10 μl of anti-fibrillarin, 10 μl of anti-Sm, 2 μl of anti-m_3G cap and 10 μl of rabbit nonimmune serum) were preadsorbed to 20 μl of protein A–Sepharose CL-4B (Pharmacia, 0.1 g/ml preswollen in IPP_{500}) in 400 μl of IPP_{500} overnight at room temperature and then washed three times with 1 ml of IPP_{500} (Hamm *et al.* 1989). Pools of 30 nuclei were homogenized in 500 μl of 10 mM Tris pH 8.0, 50 mM NaCl and 80 units of RNasin (Boehringer) and added to each antibody suspension. 10 μl of Nonidet P-40 was added and the samples were rolled for 2 h at room temperature. The beads were washed three times for 3–5 min with 1 ml of IPP_{500}, and RNA was released from the beads by digestion with 400 μl of Homomedium (Hamm *et al.*, 1989) The samples were phenol extracted and the RNA was precipitated.

Preparation of nucleoli

Germinal vesicles were manually dissected from *X.laevis* oocytes and treated according to Peculis and Gall (1992) with the only modification of increasing the NaCl concentration to 100 mM. After sonication and centrifugation a pellet fraction containing membranes, chromosomes and nucleoli was separated from a supernatant fraction containing soluble nucleoplasm.

Acknowledgements

We thank Dr Michelle Caizergues-Ferrer for kindly providing the reagents for fibrillarin and U3 analysis; we also thank Mr Arceci, Mr Di Francesco and Mr Ricci for technical assistance and Dr Fabio Riccobono for oligo facilities. This work was partially supported by grants from the Progetti Finalizzati 'Biotecnologie e Biostrumentazioni' and 'Ingegneria Genetica' of CNR and from the Istituto Pasteur Fondazione Cenci-Bolognetti.

References

Beltrame,M. and Tollervey,D. (1992) *EMBO J.*, **11**, 1531–1542.

Bozzoni,I., Beccari,E., Luo,Z.-X., Amaldi,F., Pierandrei-Amaldi,P. and Campioni,N. (1981) *Nucleic Acids Res.*, **9**, 1069–1086.

Busch,H., Reddy,R., Rothblum,L. and Choi,Y.C. (1982) *Annu. Rev. Biochem.*, **51**, 617–654.

Caffarelli,E., Fragapane,P., Gehring,C. and Bozzoni,I.(1987) *EMBO J.*, **6**, 3493–3498.

Caffarelli,E., Fragapane,P. and Bozzoni,I. (1992) *Biochem. Biophys. Res. Commun.*, **183**, 680–687.

Caizergues-Ferrer,M., Mathieu,C., Mariottini,P., Amalric,F. and Amaldi,F. (1991) *Development*, **112**, 317–326.

Dahlberg,J.E. and Lund.,E. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleotide Particles*. Springer-Verlag, Berlin, pp. 38–70.

Dumont,J.N. (1972) *J. Morphol.*, **136**, 153–179.

Fragapane,P., Caffarelli, E., Lener,M., Prislei,S., Santoro,B. and Bozzoni,I. (1992) *Mol. Cell. Biol.*, **12**, 1117–1125.

Hamm,J. and Mattaj,I.W. (1989) *EMBO J.*, **8**, 4179–4187.

Hamm,J., Dathan,N.A. and Mattaj,I.W. (1989) *Cell*, **59**, 159–169.

Kass,S., Tyc,K., Steitz,J.A. and Sollner-Webb,B. (1990) *Cell*, **60**, 897–908.

Jeppesen,C., Stebbins-Boaz,B. and Gerbi,S.A. (1988) *Nucleic Acids Res.*, **16**, 2127–2148.

Leverette,R.D., Andrews,M.T. and Maxwell, S.E. (1992) *Cell*, **71**, 1215–1221.

Liu,J. and Maxwell,E.S. (1990) *Nucleic Acids Res.*, **18**, 6565–6571.

Loreni,F., Ruberti,I., Bozzoni,I., Pierandrei-Amaldi,P. and Amaldi,F. (1985) *EMBO J.*, **4**, 3483–3488.

Ochs,R.L., Lischwe,M.A., Spohn,W.H. and Bush,H. (1985) *Biol. Cell*, **54**, 123–134.

Parker,K.A. and Steitz,J.A. (1987) *Mol. Cell. Biol.*, **7**, 2899–2913.

Peculis,B.A. and Gall,J.G. (1992) *J. Cell Biol.*, **116**, 1–14.

Pierandrei-Amaldi,P., Amaldi,F., Bozzoni,I. and Fragapane,P. (1987) In Firtel,R.A. and Davidson,E.H. (eds), *Molecular Approaches to Developmental Biology*. Alan R.Liss, Inc., New York, pp. 267–278.

Prislei,S., Sperandio,S., Fragapane,P., Caffarelli,E., Presutti,C. and Bozzoni,I. (1992) *Nucleic Acids Res.*, **16**, 4473–4479.

Razvi,F., Gargiulo,G. and Worcel,A. (1983) *Gene*, **23**, 175–183.

Reddy,R. and Busch,H. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleotide Particles*. Springer-Verlag, Berlin, pp. 1–37.

Reddy,R., Henning,D. and Busch,H. (1979) *J. Biol. Chem.*, **254**, 11097–11105.

Sanger,F., Nicklen,S., Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Savino,R. and Gerbi,S. A. (1990) *EMBO J.*, **9**, 2299–2308.

Schimmang,T., Tollervey,D., Kern,H., Frank,R. and Hurt,E.C. (1989) *EMBO J.*, **8**, 4015–4024.

Trinh-Rohlik,Q. and Maxwell,E.S. (1988) *Nucleic Acids Res.*, **16**, 6041–6055.

Tyc,K. and Steitz,J.A. (1989) *EMBO J.*, **8**, 3113–3119.

Zeller,R., Carrì,M.T., Mattaj,I.W. and De Robertis,E.M. (1984) *EMBO J.*, **3**, 1075–1081.

Received on December 15, 1992; revised on April 16, 1993

Note added in proof

The sequence data for L1h have been deposited in the EMBL/GenBank databases under the accession number X72205.