Nucleic Acids Research, Vol. 20, No. 20 5435–5442

Genes for *Xenopus laevis* U3 small nuclear RNA

Rocco Savino[1,2], Youssef Hitti[1] and Susan A.Gerbi[1,*]
[1]Division of Biology and Medicine, Brown University, Providence, RI 02912, USA and [2]Laboratori di Biologia Cellulare e dello Sviluppo, Universita di Pisa, 56010 Ghezzano (La Fontina), Pisa, Italy

Received June 23, 1992; Revised and Accepted August 13, 1992 EMBL accession nos Z12612, Z12613

ABSTRACT

Genomic Southern blots showed there are only 14 to 20 copies of U3 snRNA genes per somatic genome in *Xenopus laevis*, unlike the highly repetitive, tandem arrangement of other snRNA genes in this organism. Sequencing of two U3 snRNA genes from λ clones of a genomic library revealed striking similarity upstream, but much more divergence downstream. Consensus motifs common to other U snRNA genes were also found: a distal sequence element (DSE, octamer motif at −222 to −215), a proximal sequence element (PSE, at −62 to −52) and a 3′ Box (15 or 16 bp downstream of the U3 genes). The DSE of mammals also has an inverted CCAAT motif specific for U3 snRNA genes, and we find this is conserved in the amphibian U3 snRNA genes. The Xenopus inverted CCAAT motif is exactly one helical turn further upstream of the octamer motif than its mammalian counterpart, suggesting interaction of putative transcription factors bound to these motifs. Mutation of the inverted CCAAT motif and part of an adjacent Sp1 site greatly depresses transcription of the mutant U3 snRNA gene in Xenopus oocytes, implying a role in transcriptional efficiency. Electrophoretic mobility shift assays implicate transcription factor binding to this region.

INTRODUCTION

Uridine-rich small nuclear RNAs (snRNAs) found in eukaryotes are utilized for RNA processing. U1, U2, U4/U6 and U5 snRNAs are found in the nuclear sap where, as components of the spliceosome, they mediate removal of introns from pre-mRNAs (1). Until recently, the function of U3 snRNA has remained elusive. The localization of U3 snRNA in the nucleolus (2–4) suggested that it may play a role in rRNA processing which occurs there. Direct proof of this supposition has recently been reported: disruption of U3 snRNA *in vivo* (5–6) and *in vitro* (7) altered the levels of certain rRNA processing intermediates.

U3 snRNA differs from the other highly abundant U snRNAs not only in its cellular location and function, but also in the regulation of its expression. Thioacetamide treatment of rat liver results in preferential synthesis of U3 snRNA compared to the other U snRNAs (8). Moreover, U3 snRNA synthesis is coordinated with rRNA synthesis. When rRNA synthesis stops, U3 snRNA synthesis also decreases much more than U1 and U2 snRNA synthesis (9–12). Similarly, the onset of rRNA synthesis in oogenesis and embryogenesis correlates with an increase in U3 snRNA synthesis (13–14), unlike spliceosomal snRNA synthesis (15–16).

How is differential transcription of U3 snRNA controlled relative to the other U snRNAs? All vertebrate snRNA genes have a proximal sequence element (PSE) at about −55 instead of a TATA Box and a distal sequence element (DSE) at about −220 (reviewed in ref. 17). The PSE determines the site of transcription initiation, while the DSE acts as an enhancer to increase the level of transcription. The DSE of all vertebrate snRNA genes contains an octamer motif, thought to bind the transcription factor Oct-1 (18). All these control elements are the same between mammalian U3 snRNA genes and spliceosomal snRNA genes and so cannot explain their differential regulation. However, the DSE region also can bind other transcription factors that may differ between various snRNA genes. Mammalian U3 snRNA genes always contain an inverted CCAAT motif slightly upstream of the octamer motif (19–23), suggesting that this might be important for differential transcription of U3 snRNA genes.

In contrast to mammals, the U3 snRNA genes of lower eukaryotes lack the PSE and DSE in their 5′ flanking sequences (24–27). Even more striking is the observation that plant U3 snRNA is transcribed by RNA polymerase III, whereas animal U3 genes are transcribed by RNA polymerase II (28).

In order to see if the upstream PSE and DSE are conserved in a non-mammalian U3 snRNA gene, we have cloned and sequenced U3 snRNA genes from the amphibian *Xenopus laevis*. Xenopus U3 snRNA has a 5′ trimethylguanosine cap (29), as expected for a RNA polymerase II product. We report here that U3 snRNA genes of Xenopus retain the PSE and DSE. Like the mammalian U3 snRNA genes, the DSE of Xenopus U3 snRNA genes also has the U3 snRNA gene specific inverted CCAAT motif, which is exactly one helical turn further upstream of the octamer motif than in mammalian U3 snRNAs, thus suggesting an interaction between putative transcription factors bound to these two motifs. The significance of the DSE of U3 snRNA genes has been questioned, as it can be replaced effectively by elements from other genes (30). In the present study we show that substitution of the inverted CCAAT motif and part of an adjacent Sp1 site dramatically decreases the transcription of the U3 snRNA gene when injected into Xenopus oocytes, indicating

* To whom correspondence should be addressed

an important role for this upstream sequence in regulating transcriptional efficiency. Moreover, our electrophoretic mobility shift studies implicate transcription factor binding to this region.

MATERIALS AND METHODS

Isolation and analysis of *Xenopus laevis* U3 genomic clones

A *Xenopus laevis* genomic library, containing partial Sau3AI digested genomic DNA cloned in the λ vector EMBL 4 (kind gift from Doug Melton), was screened (31) with a random primer ^{32}P-labelled RsaI/EcoRI restriction fragment (+37 to +213) from the *X.laevis* cDNA clone pXlU3 (29). Filters were pre-hybridized 3 hours at 65°C in 3×SSC, 5×Denhardt's solution, 0.1% SDS and then hybridized in the same solution plus 50 μg/ml salmon sperm DNA and labelled probe at 270,000 cpm/ml (32). Filters were washed twice for 5 min each in 2×SSC, 0.1% SDS at 25°C, twice for 30 min each in the same solution at 65°C and twice for 1 hour each in 0.1×SSC, 0.1% SDS at 65°C prior to exposure for 36 hours at −80°C with Cronex Lightning Plus intensifying screens (DuPont). Two of the positive clones containing U3 snRNA sequences were confirmed and mapped by Southern blot hybridization (32), using the same DNA probe described above.

Various subclones were generated in the plasmid pBlueScript SK (+) (Stratagene) after digestion with either restriction enzymes or exonuclease III/mung bean nuclease (Stratagene) to produce unidirectional deletions, according to the protocol furnished by Stratagene, except that the exonuclease III/mung bean nuclease-digested DNA was treated with Klenow DNA polymerase in the presence of all four deoxyribonucleotide triphosphates (at 250 μM each) prior to ligation. One of the subclones generated from pXlU3A, which retained only 693 bp upstream of the U3A gene, was subsequently cut with Nde I and Bam HI and religated to the vector after filling in the ends with Klenow DNA polymerase. This construct was called pXlU3A'.

Plasmid DNA was sequenced using either Sequenase (US Biochemical Corp.) or T7 polymerase (Pharmacia), as directed by the manufacturers, using synthetic oligonucleotides complementary to the T3 and T7 promoters (Stratagene) as primers. Nucleic acid sequences were analyzed using the computer program of Queen and Korn (33). The nucleotide sequence data reported in this paper for XlU3A and XlU3B will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers Z12613 and Z12612, respectively.

Oocyte staging and injection

Xenopus laevis oocytes were prepared and maintained as described elsewhere (5). After collagenase digestion and overnight incubation in Modified Barth's saline (34), the oocytes were segregated as needed into stages III, IV, V, and VI on the basis of the morphological criteria described by Dumont (35).

Plasmid DNA was purified by cesium chloride gradients (32). Plasmid and λ DNA were purified either according to Maniatis *et al.* (32) or using the Qiagen large plasmid preparation kit as directed by the manufacturer. The DNA was extensively dialyzed against TE buffer (10 mM Tris−HCl, pH 8.0, 1 mM EDTA) and in the more recent experiments put into injection buffer (10 mM Tris pH 7.5, 88 mM NaCl). Each oocyte was microinjected with 10−50 nl containing 10 ng of λ DNA or 0.04−2.5 ng of U3 plasmid DNA and in some case 0.25−0.35 μCi of [α-^{32}P] UTP (New England Nuclear). Typically, a batch of 20−30 oocytes were injected per plasmid and incubated for 12−24 hours at 18−25°C before RNA extraction (5).

Analysis of RNA expressed in oocytes

RNA synthesized after oocyte injection was either ^{32}P-labelled, or else it was unlabelled but marked with a reporter tag. In the latter case, the RNA was fractionated (4 oocyte equivalents per lane) on a 6% polyacrylamide-7 M urea gel and electroblotted on a nylon (Nytran) membrane. An antisense oligonucleotide (23-mer) complementary to the reporter ('hinge') tag sequence was 5′ end-labelled with T4 polynucleotide kinase (DuPont) in the presence of [γ-^{32}P]ATP, and was used to probe oocyte RNA for expressed reporter tag-containing RNA. The hybridization conditions were according to the membrane manufacturer's protocols: prehybridization for 2 hours at 42°C in 6×SSPE, 0.1% SDS, 10×Denhardt's, 50 μg/ml *E.coli* tRNA, 50 μg/ml salmon sperm DNA; hybridization for 16 hours at 61°C in 6×SSPE, 1% SDS, 10×Denhardt's; and several washes at room temperature and then at 61°C in 6×SSPE, 0.1% SDS followed by 6×SSPE alone. The reporter tag sequence ensures that only plasmid-encoded, but not endogenous, U3 snRNA is detected.

For hybrid selection two plasmids were used: (1) pXlU1b (kind gift from Jim Dahlberg; 36) which is embryonic *X.laevis* U1 snRNA genomic DNA cloned into pBR322, (2) pBSKU3 which is the Rsa I/Eco RI fragment from the cDNA clone of *X.laevis* U3 snRNA (29) subcloned into pBlueScript KS (+) (the subcloning was performed by Joe Gall). Both plasmids were diluted to a concentration of 1 μg/μl, and heated at 100°C for 10 min; an equal volume of 1 M NaOH was then added for denaturation at room temperature for 10 min. 1/40 volume of phenol red was added as a pH indicator, the DNA solution was neutralized by gradually adding 10×SSC, 0.5 M Tris−HCl pH 7.5, 1 M HCl, and the DNA was immobilized on nitrocellulose filters at 1 μg/mm^2 (25 μg DNA/filter). Filters were rinsed twice in 6×SSC, air dried and baked for 90 min at 80°C under vacuum. The filters were pre-hybridized 3 hours at 50°C in 65% formamide, 20 mM PIPES pH 6.4, 0.4 M NaCl, 0.2% SDS, 100 μg/ml calf liver tRNA, and then hybridized overnight at 42°C in 100 μl of the same solution with 5 oocyte equivalents of total RNA from injected oocytes. Filters were washed 10 times for 10 min each at 65°C in 10 mM Tris−HCl pH 7.5, 0.15 M NaCl, 1 mM EDTA, 0.5% SDS and then washed twice at 65°C in the same solution without SDS. RNA was recovered by boiling for 1 min in 300 μl water containing 0.6 μg calf liver tRNA per square mm of filter surface. After phenol−chloroform extraction and ethanol precipitation, the labelled RNA was fractionated on 5% or 8% polyacrylamide gels containing 7 M urea, 1×TBE (32).

Analysis of genomic DNA

Genomic DNA was extracted from purified nuclei from *Xenopus laevis* livers (kind gift from Claus Jeppesen), digested with restriction enzymes, fractionated on 0.8% agarose gels and transferred to nitrocellulose filters (32). After a 3 hour pre-hybridization, the Southern blots were hybridized 16−36 hours either in 3×SSC, 5×Denhardt's solution, 0.1% SDS, and 50 μg/ml salmon sperm DNA at 65°C, or hybridized in 50% formamide, 5×SSPE, 10×Denhardt's solution, 0.1% SDS and 50 μg/ml salmon sperm DNA at 42°C. In both cases, the probe was the same Rsa I/Eco RI 182 bp fragment of the U3 snRNA

gene described above, labelled with ^{32}P using the random primer labelling kit (Boehringer Mannheim). Filters were washed 4 times for 5 min each at room temperature in 2×SSC, 0.1% SDS, twice for 20 min each in the same solution at 65°C, and twice for 1 hour at 65°C in 0.1×SSC, 0.1% SDS. Filters were exposed either at room temperature or with intensifying screens at −80°C.

Oligonucleotides

The letters 'S' or 'AS' indicate whether the oligonucleotide contains identity to the sense (S) or antisense (AS) strand of the pXlU3A insert. Oligonucleotides used were:

(A) *ΔU3 Box Mutation*

XlU3−692S	5′-CTATAGGGCCAATTGGGTACCGGG-3′ (22 nt at the 5′ end are from BlueScript 633 to 654)
XlU3−60AS	5′-CGGACCCTCTTTACTCTCGAGGCG-3′
XlU3−219S	5′-GCGGAATTCTTTCGGGTCAGGTGGGTGTGAC-ACATGC-3′
XlU3−276AS	5′-GCGGAATTCAAATCAGGAGGAAACATCTGT-CGC-3′

(B) *Mutation of the hinge region of U3*

BSK+769AS	5′-GGGAACAAAAGCTGGAGCTCCACCGCG-3′ (in pBSK II + vector)
XlU3+58S	5′-GCGCGAATTCAGACCAAACCACGAGGAAGAGC-GTCAGTG-3′
XlU3−82S	5′-GTGTAACTATGGATTTATGTCGCC-3′
XlU3+71AS	5′-GCGCGAATTCACCTTTCGAGCACATTTCACC-AGG-3′

(C) *Reporter tag oligo used to probe RNA from microinjected oocytes*

Hinge-TAG	5′-GTTTGGTCTGAATTCACCTTTCG-3′

(D) *Oligonucleotide pairs for double-stranded probes or competitors for electrophoretic mobility shift assays*

Wild type U3 Box, Wild type GC Box (WW)

XlU3−261AS	5′-GCACACCTGACCCGCCCACCAATCAATCAGGA-3′
XlU3−261S	5′-TGCTCCTGATTGATTGGTGGGCGGGTCAGGTG-3′

Mutant U3 Box, Mutant GC Box (ΔΔ)

XlU3-ΔU3/GCAS	5′-GCACACCTGACCCGAAAGAATTCAAATCAGGA-3′
XlU3-ΔU3/GCS	5′-TGCTCCTGATTTGAATTCTTTCGGGTCAGGTG-3′

Mutant U3 Box, Wild type GC Box (ΔW)

XlU3-ΔU3AS	5′-GCACACCTGACCCGCCCGAATTCAAATCAGGA-3′
XlU3-ΔU3S	5′-TGCTCCTGATTTGAATTCGGGCGGGTCAGGTG-3′

Wild type U3 Box, Mutant GC Box (WΔ)

XlU3-ΔGCAS	5′-GCACACCTGACCCGAAAACCAATCAATCAGGA-3′
XlU3-ΔGCS	5′-TGCTCCTGATTGATTGGTTTTCGGGTCAGGTG-3′

Unrelated oligonucleotide (from David Jackson)

TGT3-top	5′-GCTCCGAACGTGTTTGCCTTGGCC-3′
TGT3-bottom	5′-AGCGGCCAAGGCAAACACGTTCGG-3′

The names of the annealed oligonucleotide pairs are indicated above each pair.

Oligonucleotides were prepared by Charles Setterlund using a Biosearch 8600 DNA synthesizer (phosphoramidite method), or were purchased from Genosys (The Woodlands, Texas). Full length products of all oligonucleotides were purified by electrophoresis in denaturing 20% polyacrylamide gels.

Site-directed mutagenesis

First, for site-directed mutagenesis (37) of the U3 Box, PCR fragments were generated using 40 pmols each of oligonucleotides XlU3−692S and XlU3−276AS or XlU3−219S and XlU3−60AS, 20 ng of subclone pXlU3A′ as template in 100 μl of 50 mM KCl, 10 mM Tris−HCl pH 8.4, 1.5 mM $MgCl_2$, 20 μg/ml gelatin and 2.5 units Thermas aquaticus (Taq) DNA polymerase (Perkin-Elmer). The reaction went through 35 cycles (1 min at 94°C, 2 min at 50°C, 3 min at 72°C) in a Coy Model 50 Tempcycler. After passage through a Bio-Gel P-6 spin column (Bio-Rad) and gel electrophoresis, equimolar amounts of the two PCR-generated DNA fragments were mixed, digested with Eco RI, phenol−chloroform extracted, ethanol precipitated, and ligated with T4 DNA ligase (HPLC-pure; Pharmacia). Ligation products were heated at 65°C for 20 min and digested with Kpn I and Xho I. The desired recombinant product was purified from an agarose gel by electroelution (32) and ligated to pXlU3A′ that had been previously digested with Kpn I and Xho I, thereby generating the pXlΔU3 Box construct.

Second, the hinge domain within the coding sequence of the U3 gene was similarly mutated. This domain was selected because the sequence is not conserved (29) and appears to have no bearing on the function of U3 snRNA (unpublished observations). PCR fragments were generated (with pXlU3A′ as a template) using oligonucleotides BSK+769AS and XlU3+58S, or XlU3+71AS and XlU3−82S. The two PCR fragments were purified, digested with Eco RI, ligated and re-introduced after restriction into the upstream Xho I site and downstream Bst XI site within pXlU3A′. The resulting construct was named pXlU3A′.T.

Finally, a construct bearing both the reporter tag in the hinge domain and the mutated U3 Box region was made by 'cassette-switching' between the two constructs described above. Fragments generated with Xba I-Xho I double restriction cuts of pXlU3A′.T and pXlΔU3 Box were purified and cross-ligated. The resulting plasmid has both the reporter tag hinge and the mutated U3 Box region and is referred to as pXlΔU3 Box.T.

Electrophoretic mobility shift assay

Nuclear extracts were prepared from *Xenopus laevis* germinal vesicles which were mass isolated exactly as described by Ruberti *et al.* (38), and were stored at −70°C in a buffer consisting of 70 mM NH_4Cl, 7 mM $MgCl_2$, 0.1 mM EDTA, 2.5 mM DTT, 10 mM Hepes, pH 7.4, and 10% glycerol. Nuclear extracts were prepared in the same buffer made to 0.4 M in NaCl (39) and were stored in aliquots at −70°C. Protein concentration in the extracts was determined by the Bradford assay (Bio-Rad) to be 0.3 μg/ml.

For preparation of double-stranded oligonucleotide probe and competitors, pairs of complementary 32-mer oligonucleotides representing either wild type (XlU3−261AS and XlU3−261S), or mutated U3 Box and/or mutated GC Box areas were synthesized containing a 3-nt overhang at their 5′ ends. After annealing each pair, the double-stranded oligonucleotide was gel-purified and 10 ng were labelled by filling in the ends with [α-^{32}P] dATP or [α-^{32}P] dCTP in the presence of Klenow polymerase I. The labelled DNA was then purified on a Bio-Gel P-6 spin column (Bio-Rad). Similarly, double-stranded oligonucleotides used as cold competitors were filled in with Klenow polymerase I and cold nucleoside triphosphates.

The DNA binding reactions and electrophoretic mobility shift assays were performed essentially as described by Liu *et al.* (40). The binding reactions were carried out in 30 μl containing 10 mM Tris (pH 7.5), 50 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, 1% Ficoll, and 100 ng poly dI:poly dC as a non-specific competitor. Ten minutes after the addition of 3 μg protein of oocyte nuclear extract to the solution above, double-

stranded oligonucleotide probe was added (ca. 0.1 ng; 100,000 cpm) either alone or in combination with cold competitor double-stranded oligonucleotide, and the reaction allowed to proceed for another 15–20 minutes. The reactions were then immediately loaded on an 8% polyacrylamide gel in 200 mM glycine, 25 mM Tris, pH 8.3, and 1 mM EDTA. The gel was subsequently dried and exposed to X-ray film with an intensifying screen at −70°C.

RESULTS

Genomic organization of the U3 RNA coding loci in Xenopus

Total Xenopus liver DNA was digested with several restriction enzymes which, except for Sac I, do not have a restriction site within the U3 cDNA coding sequence (29), and hybridized to a probe derived from Xenopus U3 cDNA. A complex but reproducible pattern of bands appeared (Figure 1). Except for the Sac I lane, each band was at least as intense as the band in a single copy genomic reconstruction, which together with the reproducibility of the banding pattern for each enzyme argues against the complexity of the pattern being simply due to partial restriction digests. By comparison with 1, 2, 4 and 8 copies of genomic reconstructions, there are in the Xenopus soma 14 to 20 U3 hybridizing sequences per haploid genome; the same U3 copy number was obtained in slot blot experiments (data not shown).

Characterization of Xenopus DNA clones containing U3 RNA coding sequences

Out of the 24 positive clones isolated from an unamplified genomic library of *Xenopus laevis* DNA, two (λ 17.3 and λ 20.1) were randomly chosen for DNA preparation and further characterization. Preliminary restriction mapping showed a different restriction pattern for each, suggesting that they came from different genomic sequences. We tested if these λ clones contain transcriptionally active sequences by microinjection into Xenopus oocytes and labelling the synthesized RNAs with [α-^{32}P]UTP. As shown in Figure 2A (lanes 1 and 2), U3 snRNA-sized transcripts appeared in the RNA samples prepared from oocytes injected with the two λ clones, whereas, when the oocytes were injected with [α-^{32}P]UTP alone, only the endogenous 5.8S and 5S rRNA transcription could be detected (Figure 2A, lane 5: control). The transcripts after oocyte injection were hybrid-selected by U3 cDNA sequences. As a negative control, filters containing U3 sequences selected nothing from RNA of oocytes injected with [α-^{32}P]UTP alone (Figure 2B, lane 6). Also, filters containing U1 sequences (XlU1b; 34) instead of U3 selected nothing from oocytes injected with phage λ 20.1 (Figure 2B, lane 1: mock selection). These controls suggest that the transcripts about 220 nt long seen in the other lanes (Figure 2B, lanes 2–5) are indeed U3 snRNA transcribed from the injected recombinant phages.

Since the signal produced by phage λ20.1 is stronger than the signal produced by phage λ17.3 in both Figure 2A (compare lanes 1 and 2) and Figure 2B (compare lanes 2 and 3); we chose phage λ20.1 for further characterization. A detailed restriction map (Figure 3) was determined and two different U3 snRNA coding regions were identified within the 18.7 kb of the insert (Figure 3). Both regions have been subcloned in pBlueScript SK(+), generating pXlU3A and pXlU3B; this nomenclature is arbitrary and it does not relate to the different U3A and U3B snRNA genes found in rat (19, 41) and mouse (21, 23, 41). Upon injection of either plasmid into Xenopus oocytes, transcripts the size of U3 snRNA were synthesized (Figure 2A, lanes 3 and 4), and they were hybrid-selected by a U3 cDNA fragment (Figure 2B, lanes 4 and 5), thus suggesting that each plasmid subclone contains a functional U3 gene.

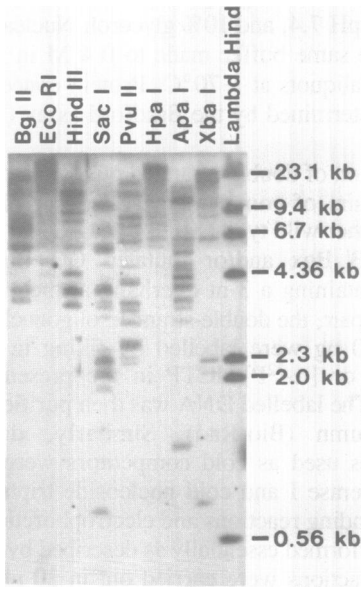

Figure 1. Genomic organization of U3 genes. *X.laevis* liver DNA was restricted with the enzymes shown and the Southern blot hybridized to U3 coding sequences, resulting in 14–20 bands/lane. The marker was Hind III digested λ DNA (sizes indicated).

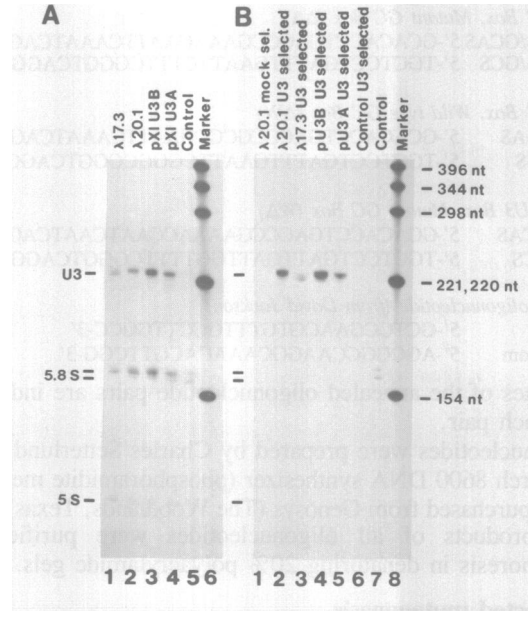

Figure 2. Transcriptional activity of U3 clones. Xenopus oocytes were injected with [α-^{32}P]UTP along with the λ or plasmid DNA indicated or no DNA (lanes marked 'control'). In some cases the isolated RNA was hybrid selected with U3 sequences (panel 2B, lanes 2–6), or U1 sequences (panel 2B, lane 1: 'mock selected'); note that the endogenous 5.8S and 5S RNA (panel 2A) are not selected and are absent (panel 2B). Sizes of the Hinf I digested pBR322 marker are shown.

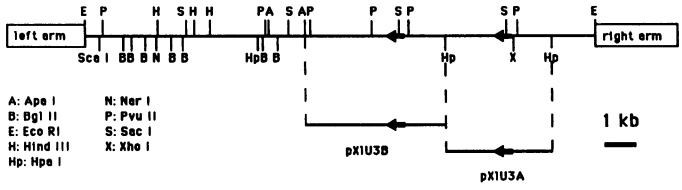

Figure 3. Restriction map of λ 20.1. Sites cleaved in the genomic Xenopus DNA insert are marked for the indicated restriction enzymes. Two transcription units, pXlU3A and pXlU3B, are present (5′ to 3′ direction of transcription shown by arrows) and were subcloned (brackets).

Structure of *Xenopus laevis* U3 genes

Various deletions of pXlU3A and pXlU3B were used for sequence analysis. We sequenced the entire coding regions of both genes, more than 250 nucleotides downstream of the 3′ end and about 700 nucleotides upstream of the 5′ end. The results are summarized in Figure 4.

Although the coding sequences for both genes are in good agreement with the U3 cDNA sequence (29), it is clear that XlU3A is closer than XlU3B to the cDNA sequence: the former differs from the cDNA only in two positions (C instead of T at position 99; G instead of A at position 194 of the cDNA); in

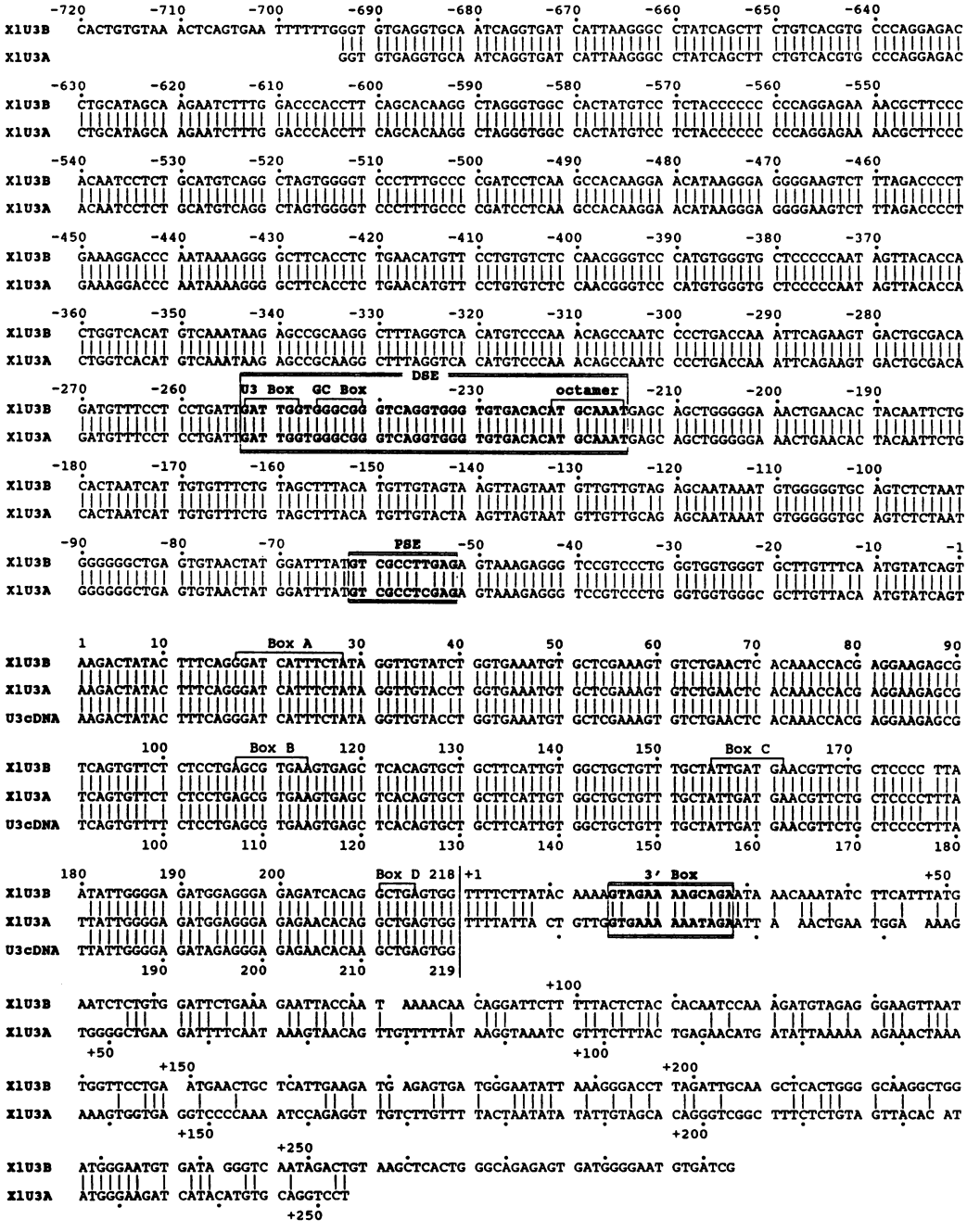

Figure 4. Sequences from pXlU3A and pXlU3B. Evolutionarily conserved regions upstream (DSE = distal sequence element containing the U3 Box which is an inverted CCAAT motif, GC Box which is a putative Sp1 binding site and octamer motif; PSE = proximal sequence element), within the U3 gene (Boxes A–D), and downstream (3′ Box) are indicated. The U3 cDNA sequence is from Jeppesen *et al.* (29).

		DSE	
		U3 Box	octamer
		GATTGG	ATGCAAAT
Xenopus U3B	-253	GATTGG TGGGCGGGTCAGGTGGGTGTGACAC	ATGCAAAT
Xenopus U3A	-253	****** **************************	********
Human U3	-247	****** CT*T*ATTCAGTA*T----------	****T***
Rat U3B.7	-252	****** CTACTC**CG*CCGT----------	****T***
Rat U3D	-230	****** CT*C*ATC*TC*A-T----------	***TT***
Mouse U3A	-265	****** CTATTATCCTCAA-T----------	***TT***
Mouse U3B.1	-253	****** CT*TTC**CG*CCGT----------	****T***
Mouse U3B.2	-253	****** CT*TTC**CG*CCGT----------	****T***
Mouse U3B.3	-253	****** CT*TTC**CG*CCGT----------	****T***

		PSE consensus
		G T G RCCNTRN G C C C
Xenopus U3B	-62	GTCGCCTTGAGAGTAAAGA
Xenopus U3A	-62	*******C**********
Human U3	-63	C*****A*C**TTA****T
Rat U3B.7	-62	******G***CT---****
Rat U3D	-62	***A**A*TT*T--***A*
Mouse U3A	-63	***A**A*T**T-****A*
Mouse U3B.1	-63	******G***CT---****
Mouse U3B.2	-63	******G***CT---****
Mouse U3B.3	-63	******G***CT---****

Figure 5. Comparison of vertebrate U3 sequences for the DSE region (upper) and PSE region (lower). Asterisks indicate identity to the Xenopus U3B nucleotides; dashes indicate lack of nucleotides relative to Xenopus U3B. U3 sequences shown are for Xenopus (this study), human (20, 22), rat (19 with corrections from 30) and mouse (21, 23).

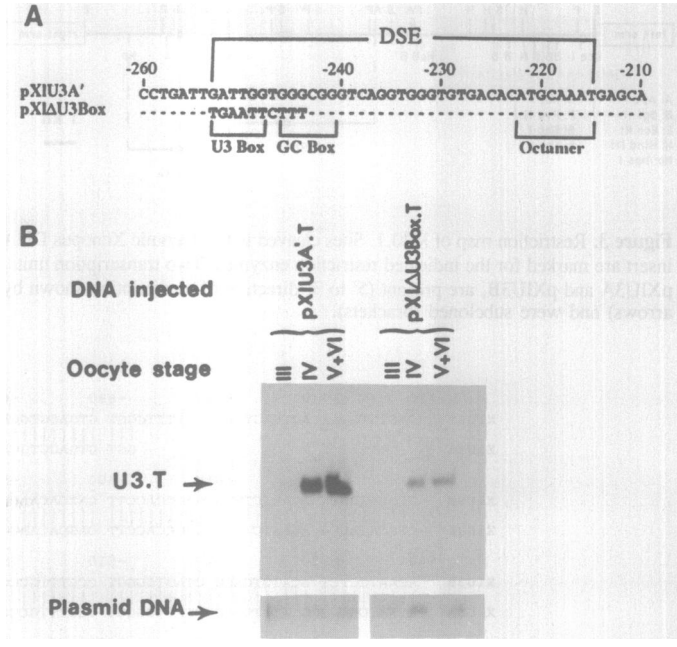

Figure 6. Mutation of the U3 Box. Panel A: wild type sequence (pXlU3A') and mutated sequence (pXlΔU3Box) of the *X.laevis* U3 Box region. Dashes indicate nucleotides that are the same in pXlΔU3Box and pXlU3A'. Panel B: Northern blot of tagged U3 snRNA transcripts (U3.T) after injection of 40 pg plasmid DNA per oocyte of the wild type (pXlU3A'.T) or mutated (pXlΔU3Box.T) U3 gene. To normalize the amount of DNA injected, the same preparations were probed by Southern blotting for plasmid vector sequences; these results are shown at the bottom of the figure. Correction for the amount of DNA injected shows that U3.T RNA first appears at stage IV and increases in amount by stages V and VI in both cases, but the net amount of U3.T RNA is greatly reduced in the mutant plasmid as compared to the wild type.

addition to these two differences, XlU3B differs from the cDNA at three other positions and is one nucleotide shorter. However, none of these differences falls in Boxes A, B, C or D (29, 43), which are areas of U3 sequence conservation among species first noted by Wise and Weiner (24). No difference is seen between the transcript lengths of the two genes in Figure 2 because such gels are not able to discriminate a one nucleotide difference (the two marker fragments, 220 and 221 nucleotides long, migrate as a single band). Every time *X.laevis* U3 snRNA has been analyzed by reverse transcription on a sequencing gel next to a U3 cDNA-generated sequence ladder as a size marker (Figure 2 in ref. 29; Figure 3 in ref. 5), one single band 219 nucleotides in size appeared. The predicted size for RNA transcribed by the XlU3B gene would be 218 nucleotides instead (Figure 4); therefore, XlU3B either codes for a minor U3 snRNA species or the missing nucleotide is a cloning artifact.

While it is not surprising that the coding regions of XlU3A and XlU3B differ from each other only in 4 positions out of 219 (98% identity), it is remarkable that the two upstream sequences are even more conserved than the coding ones; in the 693 nucleotides sequenced for both genes there are only 3 differences (99.5% identity). Computer analysis revealed the presence of two conserved regions important for the proper and efficient transcription of U snRNA genes (17): the proximal sequence element (PSE: consensus G/C T G/C RCCNTRN G/C) located between −62 and −52, and the octamer motif (ATGCAAAT) of the distal sequence element (DSE) located between position −222 and −215 (Figure 4). In addition to these two motifs, shared by all U snRNA genes, the DSE of mammalian U3 snRNA genes contains the sequence GATTGG (complementary to the classical 'CCAAT' Box in reverse orientation) found slightly upstream of the octamer motif. We will refer to this inverted CCAAT Box, unique to the U3 snRNA genes, as the 'U3 Box', in the same sense as used by Yuan and Reddy (22). In contrast, other authors have previously referred to the entire DSE (octamer plus inverted CCAAT motifs) of U3 snRNA genes as the 'U3 Box' (19, 21). As shown in Figure 4, Xenopus U3 snRNA genes also have the U3 Box (= inverted CCAAT motif), but the spacing from the octamer motif is altered from that found in mammals; this difference will be discussed below. An alignment of all the known higher eukaryotic U3 snRNA gene sequences over the regions just mentioned, together with the consensus sequences for the PSE and DSE, is shown in Figure 5. Finally, a GGGCGG motif or GC Box (= Sp1 binding site; 44), is also found in Xenopus U3 genes between positions −246 and −241 (Figure 4).

The remarkable similarity which extends over the coding regions and the 5′ flanking sequences ends abruptly at the 3′ end of the RNA coding region (Figure 4). A partial similarity between the two genes is found 15 bp downstream of the mature 3′ end of the transcript for XlU3A and 16 bp downstream for XlU3B (Figure 4) and matches the consensus called the 3′ Box ($GTYYN_{0-3}AAARRYAGA$), which has been shown to be important for the formation of the 3′ end of the RNA (17).

Functional analysis of the U3 Box region upstream of the octamer motif in the DSE

The XlU3A gene was chosen to study the function of the U3 Box region found upstream of the octamer in the DSE of both genes (Figures 4 and 5), because of the possibility that XlU3B might code for a minor U3 snRNA species. A subclone, pXlU3A', was used since its entire sequence is known: it contains the entire XlU3A gene coding sequence, 693 nucleotides of upstream sequence and 226 nucleotides of downstream sequence. Even though this construct retains only 693 bp of 5′ flanking sequence (compared to 1.5 kb of pXlU3A), it is still transcribed

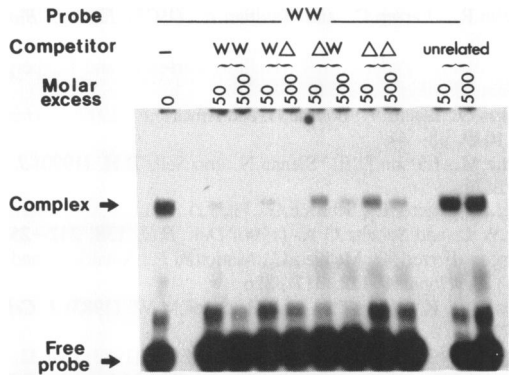

Figure 7. Gel retardation of an oligonucleotide spanning the U3 Box region. The ^{32}P labelled oligonucleotide containing wild type U3 Box and GC Box sequences (WW) was incubated alone (far left lane) or with unlabelled competitor oligonucleotide (all other lanes) in a Xenopus oocyte germinal vesicle extract. The positions of the labelled oligonucleotide when free or bound to protein from the extract is indicated. An unlabelled oligonucleotide of an unrelated sequence does not compete, the mutant oligonucleotides ΔW or ΔΔ hardly compete, but the mutant oligonucleotide WΔ and the wild type oligonucleotide WW do compete for protein binding to the labelled wild type oligonucleotide. Sequences of all oligonucleotides are presented in the Materials and Methods section.

at high efficiency when injected into Xenopus oocytes (data not shown). In order to discriminate the transcription product of an injected U3 snRNA gene from U3 snRNA endogenous to the Xenopus oocyte, we replaced nucleotides 60–71 (UGUCUG-AACUCA) with a tag sequence (T) which was GUGAAUUC-AGAC. The tagged U3 snRNA gene is still transcribed efficiently after injection into oocytes (Figure 6B: pX1U3A′.T). As previously reported and also seen by us, U3 snRNA transcription increases considerably after Stage III in Xenopus oocytes (14). To study the function of the U3 Box region found upstream of the DSE octamer, the replacement mutant pXlΔU3Box.T was constructed from the parental pX1U3A′.T clone. As shown in Figure 6A, the U3 Box sequence GATTGG (conserved in all higher eukaryotic U3 genes; see Figure 5) has been completely altered in the mutated plasmid pX1ΔU3Box.T; furthermore, half of the adjacent GGGCGG motif has been mutated too. Sequencing showed that these were the only changes introduced into the mutated plasmid. Figure 6B shows the result of injection of the replacement mutant plasmid into Xenopus oocytes: when a small amount (40 pg) of plasmid DNA is injected per oocyte, a dramatic decrease in transcript of the mutant plasmid pX1ΔU3Box.T to only 15% the level compared to transcript from the parental plasmid pX1U3A′.T is evident (Figure 6B). Injection of the same plasmids at a higher concentration (2 ng DNA per oocyte) also showed deficient expression of U3.T RNA when the U3 Box was mutated as compared to the wild type, though the difference was less pronounced. Therefore, the mutation altering the U3 Box region seems to have a strong negative impact on transcriptional efficiency.

Transcription factor binding to portions of the DSE

Since the U3 Box is an inverted CCAAT motif, it has been surmised that it may bind a CCAAT specific transcription factor, but this presumption has never been tested. To address this, we carried out electrophoretic mobility shift studies. As shown in Figure 7, a 32 bp oligonucleotide containing the wild type U3 Box region is retarded in the gel after incubation with oocyte germinal vesicle extract; this retardation is unaffected by competition with poly dI:poly dC or an unrelated oligonucleotide. However, it can be competed out by an unlabelled oligonucleotide of the same wild type U3 sequence (WW). There is a GC Box adjacent to the U3 Box, and part of the former was also substituted in the mutant plasmid whose U3 snRNA transcription was impaired (Figure 6). In order to dissect the functionally important elements within the U3 Box/GC Box region, we used competitor oligonucleotides for each of these elements in electrophoretic mobility shift studies to ascertain their relative importance in binding transcription factors. As can be seen in Figure 7, an oligonucleotide with the wild type U3 Box sequence but a mutated GC Box (WΔ) competed almost as well as the completely wild type oligonucleotide (WW) for transcription factor binding, suggesting that the GC Box plays little or no role in this factor binding. In contrast, an oligonucleotide with a mutant U3 Box and a wild type GC Box (ΔW) or mutation in both U3 and GC Boxes (ΔΔ) competed very poorly, supporting the notion that the U3 Box is important for factor binding.

DISCUSSION

After 20 years of supposition, the involvement of U3 snRNA in pre-rRNA processing has finally been experimentally demonstrated (5–6). As a first step to understand how the Xenopus oocyte synthesizes the U3 snRNA needed to sustain the massive production of ribosomal RNA that occurs during oocyte maturation, we have isolated and characterized two functional U3 snRNA genes from this amphibian. We have found that there are only 14–20 copies of U3 sequences per haploid genome of the Xenopus soma, and they are not tightly clustered, unlike up to 1000 tandem copies found for other U snRNA genes in the same species (17). It remains to be determined whether the half-life of U3 snRNA is sufficiently long to provide enough molecules in the oocytes for its catalytic role in rRNA processing, or whether there may be U3 snRNA gene amplification during oogenesis.

We investigated the upstream transcriptional signals in Xenopus U3 snRNA genes to begin to understand their regulation. The upstream flanking sequences of the two U3 snRNA genes that we studied show an amazing similarity (only 0.5% divergence versus 2% divergence of the coding regions) as far as the sequence extends (about 700 bp). High levels of upstream sequence identity among U3 genes within the same species have already been reported for mouse (21, 23, 42) and human (20, 22). Nonetheless, our deletion from roughly −1.5 kb to −693 bp for the Xenopus XlU3A gene did not seem to decrease the transcriptional efficiency after injection into stage VI oocytes.

Just as for other vertebrate snRNA genes (17), instead of a TATA Box there is a conserved proximal sequence element (PSE: G/C T G/C RCCNTRN G/C) found at −62 to −52 in Xenopus U3 snRNA genes (Figure 4). In addition, if variable spacing is allowed after the PSE consensus, most vertebrate U3 snRNA genes share the tetranucleotide AAGA (Figure 5). Another region of sequence conservation, which is found upstream of the PSE of mammalian U3 genes, is the distal sequence element (DSE), whose consensus motifs become shorter after inclusion of the amphibian U3 gene reported here. As can be seen in Figure 5, the inverted CCAAT motif (= U3 Box) shrinks from 8 bp to the conserved hexanucleotide GATTGG. Similarly, the octamer motif ATGCAAAT found in the DSE of all vertebrate U snRNA genes (i.e., not just in U3 genes) is reduced from the previous 9 bp consensus. Finally, the spacing between these two conserved

blocks is now 15 bp in mammalian U3 snRNA genes instead of the previously proposed 11–12 bp. The importance of the spacing between the U3 Box and octamer motif was unclear from results of Ach and Weiner (30), perhaps being obscured by their use of a heterologous system. Interestingly, the spacing between the U3 Box and octamer motif in Xenopus U3 snRNA genes is 25 bp, one helical turn more than what is found in mammals. Our observation that evolutionary selection has preserved the DNA helical phase between the two elements among vertebrate species indicates that the spacing between the two elements is important and supports the hypothesis that a transcription factor binds to the U3 Box and interacts with a factor at the octamer motif to regulate the level of U3 snRNA transcription.

Unlike the PSE and the octamer motif which are present in all other snRNA genes (17), the U3 Box is unique to vertebrate U3 snRNA genes and therefore is a potential target for their specific regulation. For example, it could be used to boost the synthesis of U3 snRNA during oocyte maturation to accommodate the massive synthesis of ribosomes. Previous attempts to define the role of the U3 Box in mammalian U3 snRNA genes have given mixed results. There was only a minor decrease of 63–77% in transcriptional efficiency after deletion of the U3 Box, but a 35–43% decrease after substitution (30). Surprisingly the entire DSE could be replaced by elements from totally different genes with no ill effects (30). Therefore, it was unclear from that study if the U3 Box is important for transcriptional regulation; perhaps some of the contradictions arose because a mammalian U3 snRNA gene was injected into Xenopus oocytes (the mammalian DSE has some sequence differences from the Xenopus DSE; see Figure 5). In contrast, in the present study using a homologous rather than heterologous system, we found a dramatic decline in transcriptional efficiency when the U3 Box region in the Xenopus U3 snRNA gene was substituted with a different sequence. Moreover, our gel retardation studies implicate protein factors from the germinal vesicle extract binding to this region.

ACKNOWLEDGMENTS

We thank Don Brown, Jim Dahlberg, David Jackson, Joe Gall, Claus Jeppesen and Doug Melton for gifts of experimental material, Josephine Spitzer for technical assistance, Fyodor Urnov for computer graphics and discussions, David Jackson and Ken Zaret for advice on gel shifts, Donna Meinville, Joan Boyles and Leah Pratt for typing and Catherine Denning for her support. The work described here was supported by research grant PHS GM20261 to S.A.G. and a Scuola Normale Superiore fellowship to R.S.

REFERENCES

1. Steitz,J.A., Black,D.L., Gerke,V., Parker,K.A., Kramer,A., Frendewey,D. and Keller,W. (1988) In Birnstiel,M.L (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin. pp. 115–154.
2. Weinberg,R.A. and Penman,S. (1968) *J. Mol. Biol.* **38**, 289–304.
3. Prestayko,A.W., Tonato,M. and Busch,H. (1970) *J. Mol. Biol.* **47**, 505–515.
4. Zieve,G. and Penman,S. (1976) *Cell* **8**, 19–31.
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. Kass,S., Tyc,K., Steitz,J.A. and Sollner-Webb,B. (1990) *Cell* **60**, 897–908.
8. Busch,H., Rothblum,L. and Choi,Y. (1982) *Annu. Rev. Biochem.* **51**, 617–654.
9. Hamelin,R., Larsen,C. and Tavitian,A. (1973) *Eur. J. Biochem.* **35**, 350–356.
10. Frederiksen,S., Pedersen,I.R., Hellung-Larsen,P. and Engberg,J. (1974) *Biochim. Biophys. Acta* **340**, 64–76.
11. Meyuhas,O., Baldin,V., Bouche,G. and Amalric,F. (1990) *Biochim. Biophys. Acta* **1049**, 38–44.
12. Glibetic,M., Larson,D.E., Sienna,N. and Sells,B.H. (1990) *J. Cell Biol.* **111**, 247a.
12. Prather,R., Simerly,C., Schatten,G., Pilch,D.R., Lobo,S.M., Marzluff,W.F., Dean,W.L. and Schultz,G.A. (1990) *Dev. Biol.* **138**, 247–255.
14. Caizergues-Ferrer,M., Mathieu,C., Mariottini,P., Amalric,F. and Amaldi,F. (1991) *Development* **112**, 317–326.
15. Forbes,D.J., Kornberg,T.B. and Kirschner,M.W. (1983) *J. Cell Biol.* **97**, 62–72.
16. Zeller,R., Nyffenegger,T. and DeRobertis,E.M. (1983) *Cell* **32**, 425–434.
17. Dahlberg, J.E. and Lund,E. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, pp. 38–70.
18. Tanaka, M., Grossniklaus,U., Herr,W. and Hernandez,N. (1988) *Genes Dev.* **2**, 1764–1778.
19. Stroke,I.L. and Weiner,A.M. (1985) *J. Mol. Biol.* **184**, 183–193.
20. Suh,D., Busch,H. and Reddy,R. (1986) *Biochem. Biophys. Res. Comm.* **137**, 1133–1140.
21. Mazan,S. and Bachellerie,J.P. (1988) J. Biol. Chem. 263, 19461–19467.
22. Yuan,Y. and Reddy,R. (1989) *Biochim. Biophys. Acta* **1008**, 14–22.
23. Mazan,S., Gulli,M.-P., Joseph,N. and Bachellerie,J.-P. (1992) *Eur. J. Biochem.* (in press).
24. Wise,J.A. and Weiner,A.M. (1980) *Cell* **22**, 109–118.
25. Hughes,J.M.X., Konings,D.A.M. and Cesarini,G. (1987) *EMBO J.* **6**, 2145–2155.
26. Porter,G.L., Brennwald,P.J., Holm,K.A. and Wise,J.A. (1988) *Nucleic Acids Res.* **16**, 10131–10152.
27. Myslinski,E., Segault,V. and Branlant,C. (1990) *Science* **247**, 1213–1216.
28. Kiss,T., Marshalisay,C. and Filipowicz,W. (1991) *Cell* **65**, 517–526.
29. Jeppesen,C., Stebbins-Boaz,B. and Gerbi,S.A. (1988) *Nucleic Acids Res.* **16**, 2127–2148.
30. Ach,R.A. and Weiner,A.M. (1991) *Nucleic Acids Res.* **19**, 4209–4218.
31. Benton,W.D. and Davis,R.W. (1977) *Science* **196**, 180–182.
32. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
33. Queen,C. and Korn,L.J. (1984) *Nucleic Acids Res.* **12**, 581–599.
34. Colman,A. (1984) In Hames,B.D. and Higgins,S.J. (eds.), *Transcription and Translation*, IRL Press, Oxford, pp. 271–302.
35. Dumont,J.N. (1972) *J. Morphol.* **136**, 153–180.
36. Lund,E., Bostock,C.J. and Dahlberg,J.E. (1987) *Genes Dev.* **1**, 47–56.
37. Zaret,K.S., Liu,J.K. and DiPersio,C.M. (1990) *Proc. Nat. Acad. Sci. USA* **87**, 5469–5474.
38. Ruberti,I., Beccari,E., Bianchi,E. and Carnevali,F. (1989) *Anal. Biochem.* **180**, 177–180.
39. Birkenmeier,E.H., Brown,D.D. and Jordan,E. (1978) *Cell* **15**, 1077–1086.
40. Liu,J.-K., Bergman,Y. and Zaret,K.S. (1988) *Genes Dev.* **2**, 528–541.
41. Reddy,R., Henning,D. and Busch,H. (1980) *J. Biol. Chem.* **255**, 7029–7033.
42. Mazan,S., Qu,L.H., Sri-Widada,J., Nicoloso,M. and Bachellerie,J.P. (1990) *FEBS Lett.* **267**, 121–125.
43. Tyc,K. and Steitz,J.A. (1989) *EMBO J.* **8**, 3113–3119.
44. Dynan,W.S. and Tjian,R. (1983) *Cell* **35**, 79–87.