The EMBO Journal Vol.2 No.11 pp.1883 – 1891, 1983

Xenopus laevis U2 snRNA genes: tandemly repeated transcription units sharing 5′ and 3′ flanking homology with other RNA polymerase II transcribed genes

Iain W.Mattaj* and Rolf Zeller

Biocenter of the University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland

Communicated by E.De Robertis
Received on 3 June 1983; revised on 19 August 1983

***Xenopus laevis* U2 small nuclear RNA (snRNA) genes were isolated and expressed by microinjection into frog oocytes. The genes are organised in short tandemly repeated units of ~830 bp. Some of the cloned tandem repeats are closely linked to genes coding for U5 snRNA, tRNA and an uncharacterised 7S RNA. No evidence was found for U2 snRNA pseudogenes. Single repeat units are transcriptionally active, showing that all the signals necessary for U2 snRNA transcription are included in an 831-bp segment of DNA. Sequence analysis of a cloned repeat unit showed that *Xenopus* and rat U2 snRNAs are 94% homologous. Flanking regions 5′ and 3′ to the coding sequence were found which shared extensive homology with similarly positioned sequences in human U1 snRNA genes. Part of the 3′ non-coding region homology (consensus TTTNAAAGA$^{A}_{T}$) was found in many other genes transcribed by RNA polymerase II.**

Key words: U2 sn RNA/*Xenopus laevis*/transcription units/RNA polymerase II genes

Introduction

Six U snRNA species (U-rich small nuclear RNAs) are present in the nuclei of most eukaryotic cells in amounts ranging from 10^4 to 10^6 copies/cell, and they have been highly conserved both in terms of size and sequence throughout evolution (Busch *et al.*, 1982). The RNAs are capped at the 5′ end but not polyadenylated and range in size from 107 to 214 bases (Busch *et al.*, 1982). Genes coding for U snRNAs have been isolated from a variety of eukaryotes. In most cases (Manser and Gesteland, 1982; Wise and Weiner, 1980; Roop *et al.*, 1981) they are found in multiple copies dispersed throughout the genome, but recently it has been reported that the genes coding for N1 and N2, two sea urchin snRNAs, are arranged in tandem repeats (Card *et al.*, 1982). In mammals (man being the best studied case) many U snRNA pseudogenes have been found for U1 snRNA, U2 snRNA and U3 snRNA (Denison *et al.*, 1981) and in fact the ratio of pseudogenes to genes has been reported to be 10:1 (Denison *et al.*, 1981; Bernstein *et al.*, 1983). Our studies on *Xenopus* U snRNA genes were facilitated by the previous finding that cloned human U1 snRNA genes are transcribed when microinjected into *Xenopus laevis* oocytes (Murphy *et al.*, 1982).

We report here the isolation of *X. laevis* U2 snRNA genes and studies on their expression after microinjection into oocyte nuclei. From these studies several conclusions can be drawn: (i) *X. laevis* U2 snRNA genes are tandemly repeated; (ii) single repeats are transcriptionally active, and thus, all the sequence information necessary for the production of a U2 transcript must be contained in an 831-bp DNA fragment; (iii) there are few, if any, U2 pseudogenes in *Xenopus*; (iv) some U2 gene tandem arrays are flanked by other genes coding for tRNA, U5 snRNA or an uncharacterised 7S RNA; (v) sequence comparison with human or rat U1 snRNA genes shows that while there is no homology between the 130 bp immediately preceding the cap site, there are three blocks of conserved sequences between 290 and 130 bp upstream of the coding regions of the gene; and (vi) a conserved sequence (consensus TTTNAAAGA$^{A}_{T}$) is found downstream from the U2 gene which is shared by many other snRNA and mRNA-coding genes.

Results

Selection of clones containing transcriptionally active U snRNA genes

As a probe to select clones containing U snRNA genes from a *X. laevis* genomic library in the Charon 4A derivative of bacteriophage λ (Wahli and Dawid, 1980) we used U snRNAs prepared by immunoprecipitation with a Lupus antiserum and ^{32}P-end-labelled with poly(A) polymerase as described in Materials and methods. Each of these clones was rehybridised to individual U snRNAs (U1, U2, U4, U5 and U6). The heterogeneous U5 snRNA-sized bands found in *X. laevis* (Zeller *et al.*, 1983) were pooled. The clones hybridising to individual U snRNAs fell into three classes. Many hybridised only to U2 snRNA, two hybridised only to U5 snRNA and one, λ5, hybridised to U2 and U5 snRNA. None of the clones hybridised to U1, U4 or U6 snRNA. We have since selected a transcriptionally active U1 snRNA gene-containing clone using chicken U1 cDNA (Roop *et al.*, 1981) as probe.

Since we were only interested in transcriptionally active clones, we tested the clones by microinjection into *X. laevis* oocytes and labelling of the synthesised RNAs by [α-^{32}P]GTP. The result of this experiment is shown in Figure 1. Most of the clones which hybridised to U2 snRNA (λ37, λ32, λ30, λ25, λ4, λ3 and λ2) gave rise to U2 snRNA-sized transcripts upon microinjection of DNA. Clone λ13, which hybridised only to U5 snRNA, gave rise to a U5 snRNA-sized transcript (the position of which is shown by an arrow in Figure 1, lane 13). Clone λ5, which hybridised to both U2 and U5 snRNAs, gave rise to both U2 and U5 snRNA-sized transcripts upon microinjection (Figure 1, lane 5), indicating that genes coding for these two snRNAs can be closely linked in the *X. laevis* genome. The different U5 snRNA-sized transcripts encoded by genes on λ5 and λ13 correspond to the sizes of the two major bands in the U5 snRNA population in *Xenopus*. Two additional clones produced more than one transcript. One, λ3, gave rise to both U2 snRNA and a 7S RNA (Figure 1, lane 3). α-Amanitin inhibition experiments have shown that the 7S RNA, in contrast to U2, is not an RNA polymerase II transcript showing that it cannot be a product of readthrough transcription. The second, λ37, not only gives rise to U2 snRNA transcripts but also considerably increases the level of tRNA transcription (Figure 1B, lane 37). Hybridisation to end-labelled tRNA shows that λ37 contains several *Sau*3AI fragments which hybridise with purified tRNA (Figure 1C).

*To whom reprint requests should be sent.

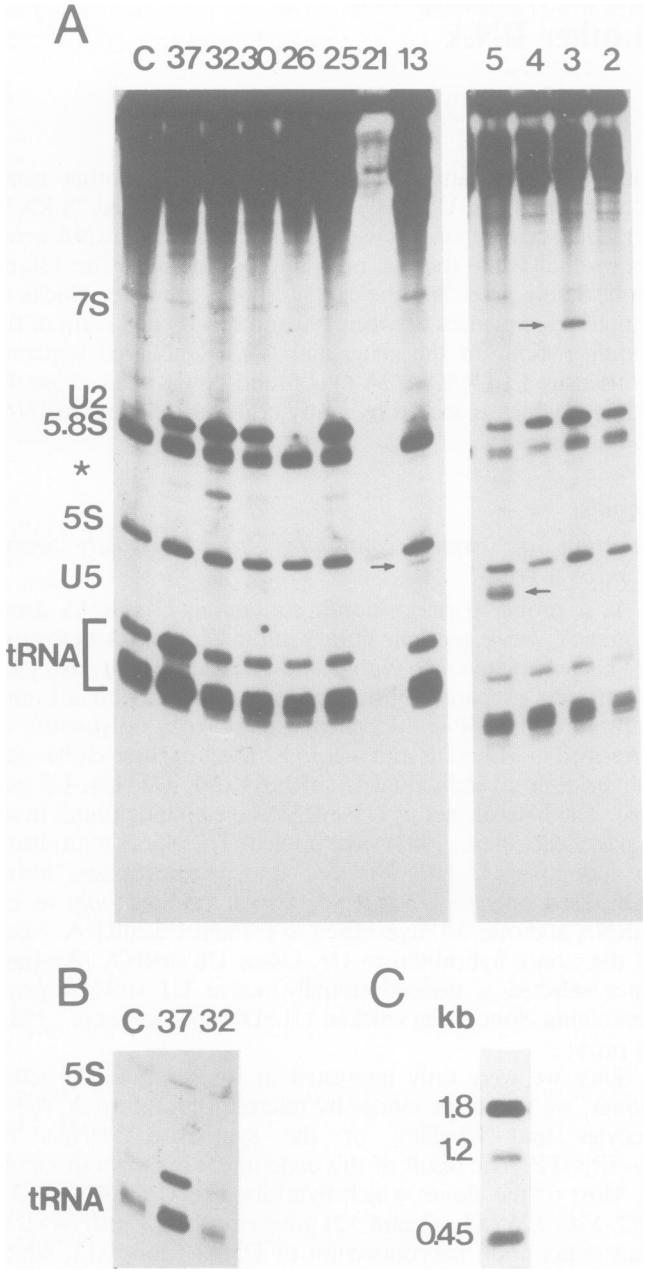

Fig. 1. A: RNAs transcribed in *Xenopus* oocytes after microinjection of DNA from λ clones, which had been selected by hybridisation to U2 or U5 snRNA. 24 h after microinjection of oocytes with cloned λDNA and [α-^{32}P]GTP the RNA was extracted and analysed by polyacrylamide gel electrophoresis (see Materials and methods). Each lane contains an amount of RNA equivalent to that of one oocyte. **Lane c** shows RNA from an oocyte which was injected with [α-^{32}P]GTP only, and shows the endogenous pattern of oocyte RNAs synthesised. The other lanes show the RNA synthesis pattern from oocytes injected with the different λ clones, the numbers correspond to the numbering of the injected λ clones (for detailed description of individual clones see text). The *Xenopus* U snRNAs were identified as described previously (Zeller *et al.*, 1983). The band marked with an asterisk is a U2 snRNA degradation product. **B**: Part of **A** after shorter autoradiographic exposure. The 5S and tRNA-containing region of lanes C, 37 and 32 is shown. At this level of exposure the presence of several tRNA-sized transcripts present only in oocytes injected with λ37 is obvious. **C**: Hybridisation of λ37 DNA to end-labelled tRNA. λ37 DNA was digested with *Sau*3AI, separated on a 1% agarose gel, transferred to nitrocellulose, and hybridised to purified *Xenopus* tRNA (a gift of E.M.De Roberties) which had been end-labelled by *in vitro* polyadenylation (see Materials and methods). The size of the hybridising bands was determined by comparison with λ *Hind*III fragments. Polyadenylated tRNA did not hybridise with *Sau*3AI fragments of λ32 (not shown).

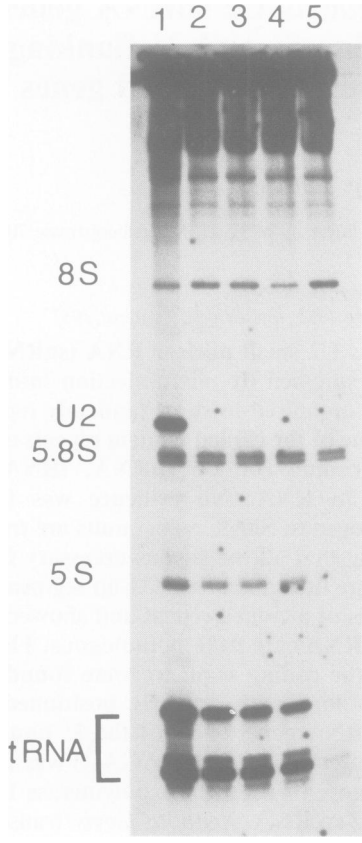

Fig. 2. α-Amanitin inhibition of RNA transcription in oocytes injected with a λ clone that contains U2 snRNA and tRNA genes. Oocytes were co-injected with λ37 DNA (which contains transcribed U2 snRNA and tRNA gense, see text), different concentrations of α-amanitin and [α-^{32}P]GTP. RNA was extracted and analysed on polyacrylamide gels as described in Materials and methods. The concentration of α-amanitin injected was as follows: **Lane 1**: control without α-amanitin, **lane 2**: 1 μg/ml, **lane 3**: 2 μg/ml, **lane 4**: 10 μg/ml, **lane 5**: 200 μg/ml. The final concentrations of α-amanitin in the oocyte lie between 0.1 and 0.05 times these values. The 8S RNA indicated in this figure was shown to hybridise to a cloned *X. laevis* rRNA gene repeat unit (data not shown).

The DNA preparations of clones which did not give rise to either U2 or U5 snRNA transcripts upon microinjection were either contaminated with substances toxic to oocytes, e.g., λ21 (Figure 2, lane 21), or were later shown to have been at too low a DNA concentration for transcription to be detectable (data not shown). These results allow two major conclusions to be drawn. First, since virtually all of the hybridising clones were transcriptionally active, it is unlikely that many U2 or U5 pseudogenes exist in *Xenopus*, and second, that U2 genes in *Xenopus* are closely linked to at least three other gene types coding for small RNA species (U5 snRNA, 7S RNA and tRNA).

Xenopus U2 snRNA genes are transcribed by RNA polymerase II

Murphy *et al.* (1982) showed that microinjected human U1 snRNA genes were transcribed in *Xenopus* oocytes by RNA polymerase II. We wished to determine if this was also the case for the cloned *Xenopus* U2 genes. To do this we were fortunate in having a clone, λ37, from which both tRNA, (transcribed by RNA polymerase III), and U2 snRNA were transcribed. Figure 2 shows the RNA synthesised in oocytes co-injected with λ37 DNA and various concentrations of α-amanitin. U2 snRNA synthesis is abolished by α-amanitin in-

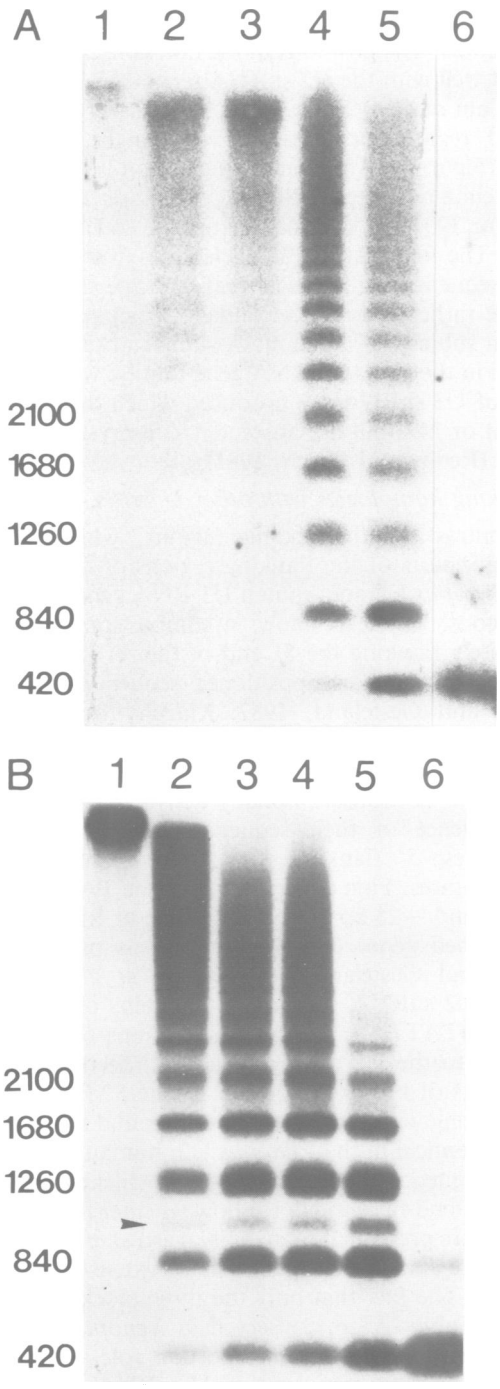

Fig. 3. Analysis of the U2 snRNA gene arrangement in genomic *X. laevis* DNA and cloned DNA. The DNAs were digested with different amounts of *Sau*3AI, separated on a 1% agarose gel, transferred to nitrocellulose and hybridised to radioactively labelled U2 snRNAs (see Materials and Methods). **A.** Genomic *X. laevis* DNA (10 μg/sample) was digested with increasing quantities of *Sau*3AI for 1 h at 37°C (1 unit of *Sau*3AI completely digests 1 μg λDNA in 15 min). **Lane 1**: undigested DNA, **lane 2**: 0.1 units, **lane 3**: 0.2 units, **lane 4**: 0.4 units, **lane 5**: 0.8 units, **lane 6**: 8 units (complete digest). **B.** Cloned DNA of λ32 (2 μg/sample) was digested with 1 unit of *Sau*3AI for increasing times. **Lane 1**: undigested DNA, **lane 2**: 10 min, **lane 3**: 20 min, **lane 4**: 30 min, **lane 5**: 40 min, **lane 6**: 60 min (complete digest). The size of bands was measured by comparison with λDNA *Hind*III fragments and φX174RF *Hae*III fragments. The band marked by an arrowhead is presumed to be composed of the tandemly repeated U2 snRNA gene-containing unit and some adjacent flanking DNA.

jected at a concentration of 1 μg/ml (Figure 2, lanes 1 and 2), consistent with U2 snRNA being an RNA polymerase II transcript (Gurdon and Brown, 1978).

The RNA polymerase III mediated transcription of tRNA and endogenous 5S RNA is completely inhibited by α-amanitin injected at 200 μg/ml (Figure 2, lane 5). 5.8S and 8S RNA, the two RNA polymerase I transcripts visible in Figure 2, are still transcribed after co-injection with 200 μg/ml α-amanitin. From this we conclude that the microinjected U2 genes are transcribed by RNA polymerase II in *Xenopus* oocytes.

The X. laevis U2 snRNA genes are tandemly repeated

We analysed the arrangement of the U2 snRNA genes in the genome of *X. laevis* by restriction enzyme digestion of DNA, Southern transfer to nitrocellulose and hybridisation with labelled U2 snRNA. We found that many enzymes (including *Eco*RI, *Bam*HI and *Hind*III) gave rise to very large hybridising bands. Incomplete digestion with *Sau*3AI (Figure 3A) gave rise to a ladder of bands regularly spaced at 420-bp intervals, the smallest of which was 420 bp. Complete digestion by *Sau*3AI leaves only the 420-bp band (Figure 3A, lane 6). These properties are characteristic of tandemly repeated DNA sequences. In fact, the U2 gene repeat unit is actually ~830 bp long and consists of two *Sau*3AI fragments of similar size (see below). Figure 3B shows the structure of λ32, a clone containing U2 snRNA genes. DNA of clone λ32 was digested with *Sau*3AI for various lengths of time (Figure 3B, lanes 1–6). The hybridisation pattern observed is identical to that with total *X. laevis* DNA (Figure 3A). This indicates that clone λ32 consists of tandemly repeating units identical to the major form found in genomic DNA. We have evidence from R-loop experiments (C.Brack, unpublished results) that the entire 15-kb insert of clone λ32 consists of tandemly repeated U2 snRNA transcription units.

A single 831-bp repeat is a complete U2 snRNA transcription unit

Having established that the *Xenopus* U2 genes are tandemly repeated, we wanted to ask if a single repeat unit would be transcribed into U2 snRNA. We knew from RNA sequence data (Reddy *et al.*, 1981) that rat U2 snRNA coding sequences contain a *Sau*3AI restriction cut site. If this were conserved in *Xenopus*, complete *Sau*3AI digestion products would all be cut within the U2 coding sequence. We therefore subcloned *Sau*3AI partial digestion products of λ37 DNA by extracting DNA from the 840-bp region of a 1% low gelling temperature agarose gel (Maniatis *et al.*, 1982) and ligating them into the *Bam*HI site of the plasmid vector pUC8 (Vieira and Messing, 1982). Figure 4 shows the two different transcription patterns obtained on injecting four different subclones into *Xenopus* oocytes. pX1U2-2 and pX1U2-5 (Figure 4, lanes 1 and 3) give rise to U2 transcription, while pX1U2-1 and pX1U2-6 (Figure 4, lanes 2 and 4) give rise to a smear of transcripts of various lengths. Sequence analysis of pX1U2-5 (Figure 5) provided a simple explanation for these two patterns of transcription which is shown diagramatically in Figure 4. The repeat unit, whose length is 831 bp in pX1U2-5, contains two *Sau*3AI sites (at positions 1 and 384 in the sequence shown in Figure 5). If the *Sau*3AI site internal to the U2 snRNA coding sequence (position 384 in Figure 5) has not been cut during sub-cloning, then an intact gene is present and can be transcribed. This is the case in pX1U2-5 and pX1U2-2 (Figure 4, diagram A). If, however, this *Sau*3AI site

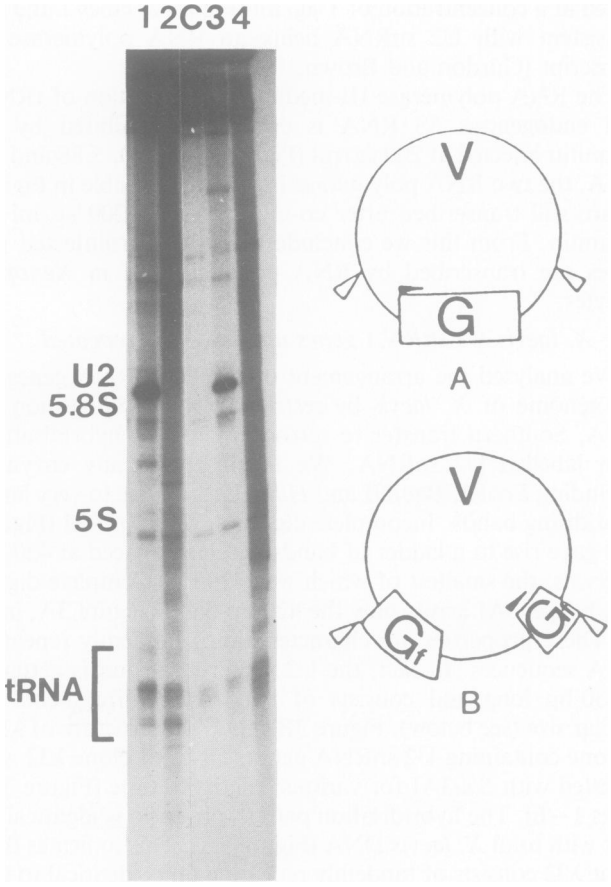

Fig. 4. Transcription of fragments from the U2 gene-containing clone λ37 subcloned into pUC8. Individual subclones were injected with [α-^{32}P]GTP into oocytes, RNA was extracted and analysed on polyacrylamide gels as described in Materials and methods. **Lanes 1–4** show RNA from oocytes injected with different subclones, **lane c** shows the endogenous RNA synthetised in an oocyte injected only with [α-^{32}P]GTP. **Lane 1**: injection of subclone pXLU2-2, **lane 2**: subclone pXLU2-1, **lane 3**: subclone pXLU2-5, **lane 4**: subclone pXLU2-6. Each track contains an RNA amount equivalent to 1 oocyte. **Diagram A** shows the structure of subclones pXLU2-2 and pXLU2-5 (**lanes 1** and **3**), which give rise to U2 snRNA transcription; **diagram B** shows the putative structure of subclones pXLU2-1 and pXLU2-6, which give rise to transcripts of undefined size. G: U2 snRNA coding region (boxed). G_F: U2 snRNA coding region fragments (boxed). V: pUC8 vector sequences. The black arrowhead indicates the start of transcription. Open arrowheads indicate the *Bam*HI sites into which *X. laevis* DNA was inserted.

has been cut and ligated to the pUC8 *Bam*HI restriction site, the two halves of the gene are separated (Figure 4, diagram B). Transcription probably initiates normally and continues through the first 28 residues of the gene into the pUC8 vector sequences where it terminates non- or semi-specifically giving rise to transcripts of various length. These results show that a single 831-bp repeat unit contains all the sequence information necessary for transcription of U2 snRNA.

Sequence analysis of a U2 snRNA transcription unit

Figure 5 shows the sequence of the non-coding strand of the pX1U2-5 insert. The insert is 831 bp long and includes *Sau*3AI sites at positions 1 and 384. The 188 underlined bases correspond to the *X. laevis* U2 snRNA sequence, as determined by comparison with the published rat Novikoff hepatoma U2 snRNA sequence (Reddy *et al.*, 1981) and by S1 mapping of the pX1U2-5 transcript (data not shown). The arrows show the cap site and 3′ end. The rat and *Xenopus* U2 sequences show 94% homology, and non-conserved residues are overlined. The positions of the non-conserved bases in the RNA fit well with the U2 snRNA secondary structure model of Branlant *et al.* (1982), altered bases either being in single-stranded regions or conservative changes in base-paired regions (Figure 6). The non-coding region is punctuated by a 4-bp satellite repeated 19 times which starts 177 bases downstream (or 390 bases upstream) from the end of the coding sequence. The fact that the 72 nucleotides at the 5′ end of the coding sequence are identical may indicate that the primary sequence rather than the secondary structure of this region has been subject to selection. A similar observatin has been reported in the rat U3 snRNA gene family, where two major species of U3 snRNA are produced which differ in 17 positions out of 213, but are absolutely conserved in the first 84 residues (Reddy and Busch, 1981).

5′ Flanking homologies with other U snRNA genes

In contrast to the coding regions, which show little homology, parts of the flanking regions of transcriptionally active *Xenopus* U2 and human U1 RNA genes are strikingly homologous. Table IA shows a comparison of the regions immediately flanking the 5′ end of the *X. laevis* U2 coding sequence with similarly positioned sequences from human (Manser and Gesteland, 1982; Murphy *et al.*, 1982) U1 coding regions, and with rat (Watanabe-Nagasu *et al.*, 1983) and chicken (Roop *et al.*, 1981) sequences from genes presumed to be transcriptionally active on the basis of the correspondence of their sequences with the homologous RNAs. These 5′ flanking sequences contain the following notable features. First of all, they lack the TATA homology, usually found ~25 bp 5′ to the cap site of RNA polymerase II transcribed genes (Goldberg, 1979) as previously noted (Manser and Gesteland, 1982; Roop *et al.*, 1981). [The *Xenopus* U2 snRNA gene repeat contains only one TATA homology (TATTAAA, Figure 5, positions 448–454) which is internal to the U2 coding sequence.] Secondly, there are three blocks of homology present in the 5′ flanking region, whose spacing with respect to each other and to the cap site is virtually identical in the *Xenopus* U2, human U1 and rat U1 flanking sequences. Unfortunately, the chicken U1 sequence does not extend far enough to show whether the homologous sequences are present. Watanabe-Nagasu *et al.* (1983) showed that human and rat U1 genes share extensive 5′ flanking homology. The fact that only the three blocks of homology shown in Table IA are conserved in *Xenopus* U2 genes indicates that they play some important role, perhaps in the recognition and transcription of U snRNA genes by RNA polymerase II.

3′ Non-coding homology with other U snRNA and mRNA coding genes

Table IB shows a comparison of the immediate 3′ flanking sequences of the *X. laevis* U2 snRNA gene and the human, chicken and rat U1 snRNA genes (Manser and Gesteland, 1982; Watanabe-Nagasu *et al.*, 1983; Roop *et al.*, 1981). The boxed region indicates a sequence conserved in 13 out of 15 positions between *Xenopus* U2 and human U1, in 10 out of 15 between *Xenopus* U2 and rat U1, and in 11 out of 15 between *Xenopus* U2 and chicken U1. The position of this sequence in relation to the coding region suggests that it may play a role either in transcription termination or in the processing of a longer U2 precursor transcript to the mature size. No other 3′ homology was found, either in the sequences

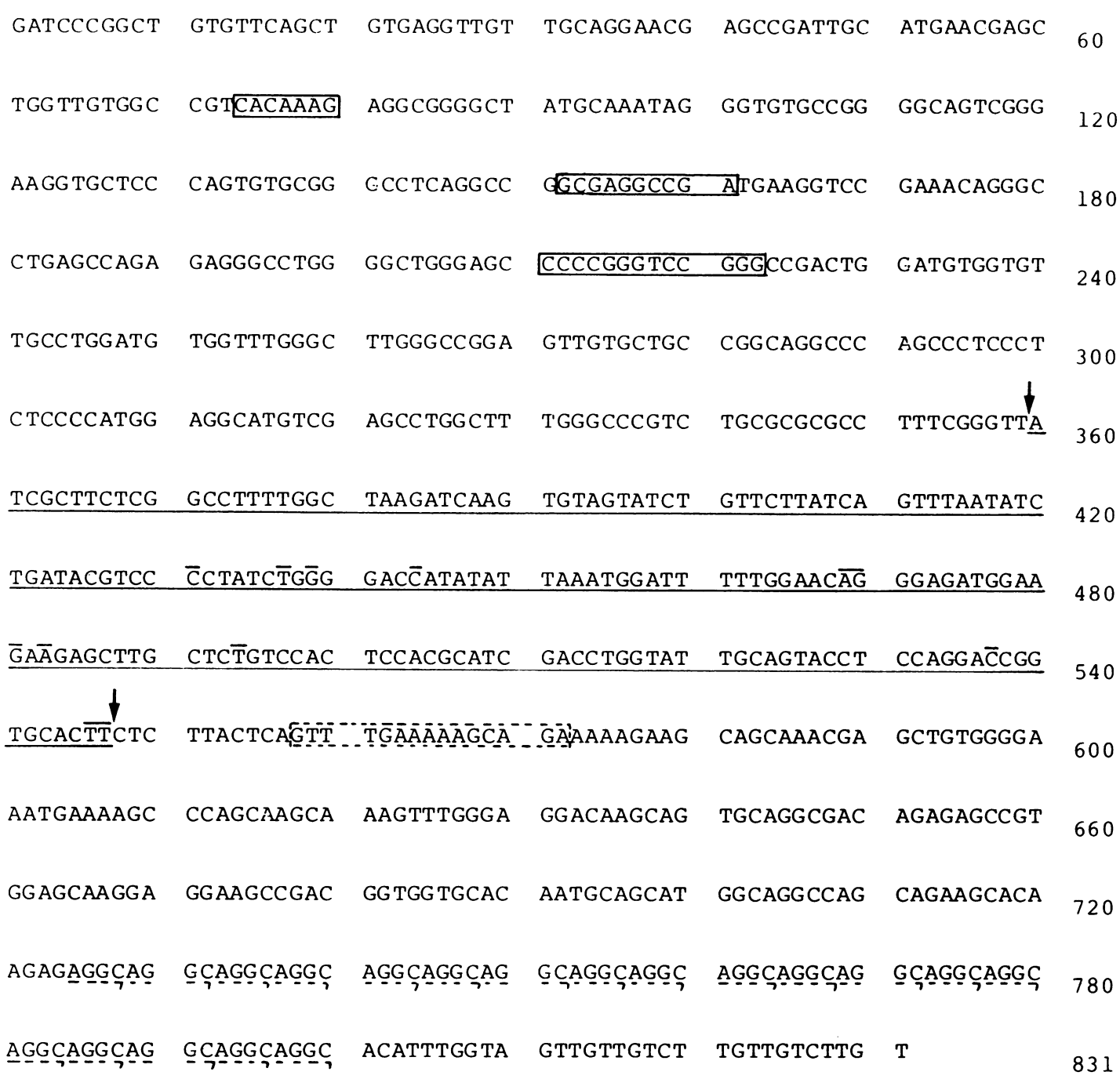

Fig. 5. DNA sequence of the non-coding strand containing the U2 snRNA transcription unit in subclone pXLU2-5. The DNA sequence was determined by the method of Sanger *et al.* (1977). The U2 snRNA coding sequence, determined by comparison with the rat Novikoff hepatoma U2 snRNA sequence (Reddy *et al.*, 1981), is underlined. Non-conserved residues are indicated by dashes above the sequence. The lower case dashed arrows show the position of a 4-bp satellite repeat. Boxed regions are discussed in the text. The arrows indicate the presumptive 5′ and 3′ end of the U2 snRNA coding region.

shown in Table IB or further downstream.

A computer search (carried out by John Shephard) revealed the presence of a sequence homologous to that conserved at the 3′ ends of the four U snRNA genes in the 3′ non-coding region of many eukaryotic genes transcribed by RNA polymerase II. The data are summarised in Table II. The consensus arrived at by comparing the different sequences is TTTNAAAGA$^{A}_{T}$. In several cases the $(A)_nG$ motif is repeated a short distance further downstream as in the *Xenopus* U2 sequence (Figure 5). In three cases shown in Table II, two residues in the N position were allowed in order to improve the homology to the consensus sequence. Apart from the U snRNA genes all the genes shown in Table II code for polyadenylated messages and the base distance given is that from the last A residue of the AATAAA polyadenylation signal (Proudfoot and Brownlee, 1976) to the first T residue of the consensus sequence. We would like to point out, however, that the significance of this conserved sequence in any of these genes will only be definitively determined by direct mutagenesis experiments.

There are two obvious functions which might be fulfilled by a 3′ non-coding conserved sequence. It might either participate in transcription termination or in the processing of precursor transcripts extended in the 3′ direction. The fact

that this sequence occurs near the 3′ ends of genes coding both for polyadenylated and non-polyadenylated RNAs might suggest that it could have a terminator function. There are, however, arguments against the sequence being a terminator. Firstly, in two of the examples given in Table II, the sequence occurs between the protein coding sequence and the first AATAAA. In one of these cases, the *Drosophila* 70-kd heat-shock protein, the sequence occurs twice, once before and once after the AATAAA. The second case, human leukocyte interferon α-1, has an AATAAC sequence 68 residues before the 3′ homology in addition to AATAAA 162 residues after it. In neither of these examples, however, has the 3′ end of the mRNA been experimentally determined. A stronger argument against the 3′ homology having a terminator function is the demonstration by Hofer and Darnell (1981) that transcripts complementary to regions more than 1 kb 3′ to the end of the mature mouse β-major globin gene coding region are found *in vivo*. This gene has a homology to the 3′ consensus sequence reported here at a position 68 bp downstream from the AATAAA sequence (see Table II). If this caused efficient termination, downstream transcripts should not be found.

The best studied genes from the viewpoint of termination of RNA polymerase II transcription are the histones. Busslinger *et al.* (1979) first reported two regions of 3′ homology between different histone genes. The proximal region was an interrupted inverted repeat, capable of forming a stem-loop structure, after which the 3′ end of the mature mRNA is found (Hentschel and Birnstiel, 1981). The second part of the homology is found a short distance downstream from the 3′ end of the mRNA and is a purine-rich sequence. These homologies were later shown to be common to eukaryotic genes from many species coding for non-polyadenylated histone messages (Hentschel and Birnstiel, 1981). The first homology has been shown to be necessary but not sufficient for the efficient generation of authentic 3′ ends on sea urchin histone H2A mRNA (Birchmeier *et al.*, 1982). The lack of sequences homologous to the inverted repeat in the immediate 3′ flanking region of other genes might indicate that histones have a unique terminator structure. Purine-rich sequences similar to the histone downstream consensus (CAAGAAAGA) have been reported to occur close to the polyadenylation signal in ovalbumin and human interferon γ (Taya *et al.*, 1982). The *Xenopus* U2 repeat (Figure 5) also contains such a sequence, which is directly repeated starting within the TTTNAAAGA$^{A}_{T}$ consensus sequence region. In the future it

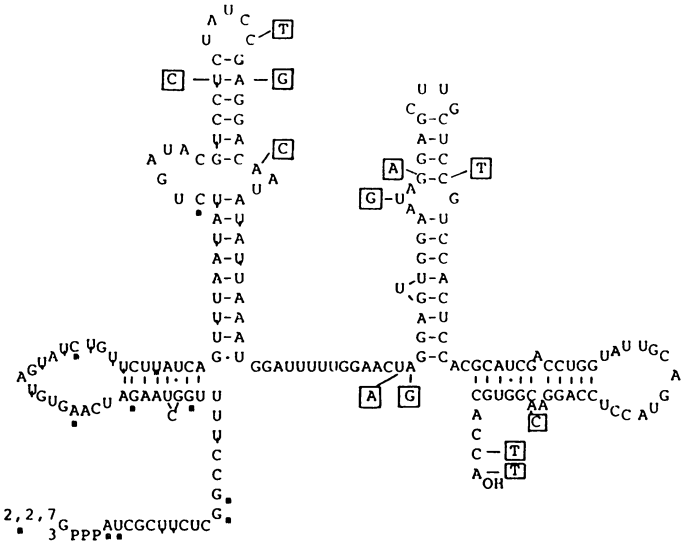

Fig. 6. Secondary structure model of rat U2 snRNA showing the base substitutions in *Xenopus* U2 snRNA. The *Xenopus* U2 base substitutions are boxed. The rat sequence data and the stem-loop structure nearest the 5′ end is from Reddy *et al.* (1981). The rest of the secondary structure model is from Branlant *et al.* (1982). Note that the nucleotide changes are in single-stranded regions or produce conservative changes which preserve the secondary structure.

Table I. Sequence comparison between the 5′ and 3′ flanking sequences of the *X. laevis* U2 snRNA gene, the human U1 snRNA gene and the chicken U1 snRNA gene (references in the text)

A. Comparison of the immediate 5′ flanking sequences

	-280	-198	-137	-41	-31	-21	-11	-1
Xenopus U2	CACAAAG	----- GCGAGGCCGA	-- CCCCGGGTCCGGG	--- GAGGCATGTC	GAGCCTGGCT	TTGGGCCCGT	CTGCGCGCGC	CTTTCGGGTT
	-276	-194	-139	-41	-31	-21	-11	-1
Human U1	CAGAAAG	-- GCGCAGAGGCTGA	-- CCCTGGGAGCGGG	--- GTAAAGAGTG	AGGCGTATGA	GGCTGTGTCG	GGGCAGAGCC	CGAAGATCTC
	-271	-192	-136	-41	-31	-21	-11	-1
Rat U1	GAGAAAG	GCGCAGGGTCTGCCGG	------ GGGAGCGCG	--- TAAGAGTGGA	GTGGCGGCGT	CCGTGAGTCG	GGGCTGTGCG	GTAGAAAAGC
				-41	-31	-21	-11	-1
				--- GGTGCGGGCT	GGTGGTGGGG	CGTGGGGAGC	GGGGCGGGCG	AGAGCAAAGC

B. Comparison of the immediate 3′ flanking sequences

	10	20	30	40	50
Xenopus U2	CTCTTACTCA	GTTTGAAAAA	GCAGAAAAAG	AAGCAGCAAA	CGAGCTGTGG
Human U1	ACTTTCTGGA	GTTTCAAAAA	CAGACCGTAC	GCTAAGGGTC	ATGTCTTTTC
Rat U1	GCATTTCTGG	TATGAGAAAG	TAAGAGTTTC	TAAGCTGTCT	TGCCTGTTGT
Chicken U1	ATTTGCGCGG	TTCAAAGACA	GAACGCTGCT	CTTCACCTGT	ATTCCTCGCT

Sequences common to all the genes and found at roughly equal distance from the coding sequences are boxed and explained in the text.

Table II. A sequence homologous to that conserved at the 3′ ends of the U snRNA genes is found in the 3′ non-coding region of many RNA polymerase II transcribed genes

Gene														bp dist. from 3′ end or AATAAA to 1st T of consensus
Chicken ovalbumin		C	T	T	T	C	T	A	A	G	C	A	T	C	234
Human β-globin		A	T	T	T	A	A	A	A	C	A	T	A	A	81
Human δ-globin		T	T	T	T	AC	A	A	A	G	A	G	T	A	103
Human embryonic ε globin		T	G	T	T	A	A	A	A	G	G	A	A	A	60
Human cl.1 transpl. antigen (HLA)		T	T	T	T	T	A	A	A	G	G	A	A	G	343
Human leu.-interferon (α-1)		A	T	T	T	C	A	A	A	G	A	C	T	C	−162
Human leu.-interferon (λα-2)		C	T	T	T	A	A	A	A	T	G	A	A	A	79
Human fibro.-interferon (β-1)		T	T	T	T	T	A	A	A	A	T	A	T	A	99
Human immune interferon (γ)		C	T	T	T	C	T	A	A	G	A	T	A	C	209
Human Ig κ-light chain (const. region)		T	T	T	T	CA	A	A	A	G	A	A	G	A	351
Human preproinsulin		T	T	T	T	T	A	G	A	G	T	T	A	T	375
Mouse α-globin		G	T	C	T	GC	A	A	A	G	G	T	G	T	39
Mouse β-globin		G	T	T	T	T	C	A	A	G	A	T	A	C	68
Mouse IgG H-chain (const. region)		C	T	T	T	C	C	A	A	G	G	T	A	T	42
Mouse Ig κ-L chain		C	T	T	C	T	A	A	A	G	A	A	G	T	80
Rabbit β-1 globin		A	T	T	T	A	A	A	A	C	A	T	C	A	79
Rat growth hormone		T	T	T	T	T	T	A	A	G	G	C	G	T	39
Maize zein (19 kd protein)		G	T	T	T	T	T	A	A	G	C	T	A	G	252
Drosophila hsp70-1 87C		T	A	T	T	T	A	A	A	G	A	T	A	A	12
		T	T	T	T	A	A	A	A	G	T	G	A	T	−126
Drosophila hsp22		A	T	T	T	G	A	A	A	A	G	A	C	T	2
Drosophila hsp26		T	T	T	T	G	A	A	A	G	A	G	G	C	62
Drosophila hsp27		A	T	T	T	A	A	A	A	G	A	A	G	A	?
Dictyostelium D_2 snRNA		A	A	T	T	A	A	A	T	G	A	A	A	A	10
Chicken U1 snRNA		G	G	T	T	C	A	A	A	G	A	C	A	G	9
Human U1 snRNA		G	T	T	T	C	A	A	A	A	A	C	A	G	11
Xenopus laevis U2 snRNA		G	T	T	T	G	A	A	A	A	A	G	C	A	11
		|	|	|	|	|	|	|	|	|	|	|	|	|	
		27	27	27	27	30	27	27	27	27	27	27	27	27	
	A	6	2	0	0	9	21	26	26	4	15	10	14	11	
	T	10	23	26	26	8	4	0	1	1	3	9	4	7	
	C	5	0	1	1	9	2	0	0	2	2	4	3	5	
	G	6	2	0	0	4	0	1	0	20	7	4	6	4	
Consensus:		-	T	T	T	N	A	A	A	G	A	A/T	(A)	-	

The homologous sequences in the 3′ non-coding region of the genes shown in this table were obtained by a computer search using sequencing data of the Los Alamos National Laboratorium Sequence Data Bank (Genbank, 1982) and Southgate *et al.* (1983, for the *Drosophila* heat shock sequences). The consensus sequence shown was derived from all sequences contained in this table. N stands for any nucleotide (for detailed discussion see text). The question mark by *Drosophila* hsp27 indicates that no AATAAA has been found in the 3′ non-coding region so far sequenced.

should be possible to determine the function of these sequences by *in vitro* manipulation.

Discussion

A transcriptionally active *X. laevis* U2 snRNA gene has been cloned. All of the sequence information required for production of the correct transcription product, the coding sequence for U2 snRNA, a promoter site and either a termination or a 3′ end processing signal, are included in an 831-bp repeated unit.

The *Xenopus* U2 snRNA genes, as has been recently reported for the sea urchin *Lytechinus variegatus* N1 and N2 snRNA genes (Card *et al.*, 1982), are arranged in tandem repeats. SnRNA genes in other eukaryotes studied including man (Manser and Gesteland, 1982), chicken (Roop *et al.*, 1981) and *Dictyostelium* (Wise and Weiner, 1980) have been found dispersed throughout the genome. A further feature of the genomic organisation of the *X. laevis* U2 genes is their close linkage to other genes coding for short RNA species. So far clones containing U2 genes linked to U5 genes, tRNA genes and to an as yet uncharacterised 7S size RNA gene have been isolated. Like the U2 genes, the *Xenopus* U5 genes are also tandemly repeated (our unpublished data), as are some *X. laevis* tRNA genes (Clarkson and Kurer, 1976). Whether the juxtaposition of different genes or tandem repeats has any functional significance remains to be seen.

U snRNA pseudogenes have been reported to occur in mammals, the best-studied case being man (Denison *et al.*, 1981; Van Arsdell *et al.*, 1981) where a ratio of 10 pseudogenes to 1 gene has been reported for U1 (Bernstein *et al.*, 1983). There is evidence suggesting that the situation in *Xenopus* is very different. Firstly, no λ clones selected by hybridisation to U2 or U5 were proven to be negative in transcription studies. Secondly, hybridisation of U2 snRNA to *Sau*3AI-digested *X. laevis* genomic DNA gives rise to a single band (Figure 3A, lane 6). This band has been shown to correspond to ~500–1000 U2 gene copies/haploid genome by titration against cloned U2 genes on Southern blots (unpublished data). If ten times this number of pseudogenes were present, as in the human, we would see a smear of hybridisa-

tion to differently sized fragments, although a low number of single copy pseudogenes might not be detected. It is possible that *Xenopus* has no U2 pseudogenes, although our results do not prove this. Hosbach *et al.* (1983) have published evidence that *X. laevis* possesses larval α-globin pseudogenes, and the reason why frogs should have fewer U snRNA pseudogenes than mammals remains an intriguing mystery.

As mentioned in the Introduction, cloned human U1 snRNA genes are transcribed when microinjected into *Xenopus* oocytes and the 5′ flanking sequences necessary for this transcription have been studied (Murphy *et al.*, 1982). *In vitro* deletion of all sequences further than 100 bp upstream from the U1 coding region prevents transcription of these genes in *Xenopus* oocytes. Removal of the 100 nucleotides between 6 and 106 nucleotides 5′ to the cap site, but leaving the further upstream sequences, allows transcription of RNA complementary to a human U1 DNA probe at the same level as is obtained from the intact gene. Furthermore, *in vitro* transcription (Manley *et al.*, 1980) of these genes gives rise to transcripts initiating 183 bp upstream from the cap site (Murphy *et al.*, 1982). These results lend support to the idea that the upstream sequences conserved between *Xenopus* U2 and human U1 genes (Table IA) may well be involved in the initiation of transcription. The homology ending at position −194 in the human U1 snRNA gene is only 11 bp upstream from the site of transcription initiation *in vitro*. Further analysis of *in vitro* constructed deletion mutants of the U2 snRNA gene should enable us to determine the importance of these sequences in transcription.

Previous work (De Robertis *et al.*, 1982; Zeller *et al.*, 1983) has shown that *Xenopus* oocytes, unlike somatic cells, accumulate the RNA and protein components of snRNP (small nuclear ribonucleoprotein) particles non-coordinately. This was shown to be due to a relative lack of accumulation of snRNAs with respect to snRNP proteins. The resulting excess of snRNP proteins are located in the oocyte cytoplasm and microinjection of snRNAs results in the migration of the snRNP proteins into the oocyte nucleus (reviewed by De Robertis, 1983). This process occurs *in vivo* during early embryonic development. Just after the first synthesis of snRNAs (Newport and Kirschner, 1982), the snRNP proteins translocate from the cytoplasm into the nucleus (Zeller *et al.*, 1983). Two possible reasons why the oocytes might become depleted of snRNAs relative to the snRNP proteins are that either snRNAs are not actively synthesised in oocytes, or they are made but are unstable in oocytes. We have shown (Figure 1) that microinjected U2 genes are actively transcribed in oocytes, although we do not yet know how actively the endogenous genes are being transcribed. When U2 snRNA produced by microinjection is reinjected into the cytoplasm of a second oocyte, it migrates to the nucleus, where it appears to be stable (I.Mattaj, R.Zeller and E.De Robertis, unpublished results). Further experiments using the cloned genes described here should help in understanding the mechanism by which snRNA and snRNP proteins are accumulated non-coordinately, and may also help to elucidate the mechanism by which the translocation of the snRNP proteins from the cytoplasm to the nucleus takes place.

Materials and methods

Polyadenylation of U snRNAs

Total U snRNAs were prepared from immature *X. laevis* ovaries by immunoprecipitation of the U snRNA particles with Sm antisera (Lerner and Steitz, 1979; Zeller *et al.*, 1983) and extraction of the RNAs (De Robertis *et al.*, 1982). The extracted U snRNAs were end-labelled by polyadenylation with poly(A) polymerase (BRL) in the presence of [α-^{32}P]ATP (Amersham, 400 Ci/mmol) as described by Gilvart *et al.* (1975). The specific activity of the probes was ~10^8 c.p.m./μg. When individual U snRNAs were used as probes, the immunoprecipitated and extracted U snRNAs were separated by polyacrylamide gel electrophoresis (De Robertis *et al.*, 1982), individual bands corresponding to the different U snRNAs were extracted according to Maxam and Gilbert (1977) and end-labelled as described above.

Hybridisation using RNA probes

Southern transfer of DNA to nitrocellulose filters was done by the method of Maniatis *et al.* (1982). Transfer of plaques or bacterial colonies to nitrocellulose filters (Schleicher and Schüll) was done using the methods of Benton and Davis (1977). Hybridisation of ^{32}P end-labelled RNA to the nitrocellulose bound DNA and subsequent washing were carried out by the methods of Humphries *et al.* (1978).

Restriction digests, ligations

Restriction enzymes and T4 DNA ligase were obtained from New England Biolabs and used following the procedures of Maniatis *et al.* (1982).

Microinjection of cloned DNA

Purified DNA (λDNA: Garber *et al.*, in preparation, plasmid DNA: Maniatis *et al.*, 1982) was extensively dialysed to reduce toxicity and microinjected into *X. laevis* oocytes together with [α-^{32}P]GTP (Nishikura *et al.*, 1982). 24 h later RNAs were extracted from the oocytes and analysed on polyacrylamide gels (De Robertis *et al.*, 1982). The concentration of microinjected DNA was 200−300 μg/ml for λ clones, and 300−500 μg/ml for plasmid clones. The volume microinjected was 30−50 nl.

DNA sequencing

DNA cloned into M13 mp8 and M13 mp9 vectors (Messing and Vieira, 1982) was sequenced using the dideoxynucleotide chain terminator method of Sanger *et al.* (1977).

Acknowledgements

We would like to thank S.Lienhard for excellent technical assistance and help in preparing the figures. We are indebted to J.Shephard for carrying out the computer search and W.Wahli for generously providing the genomic library of *X. laevis* in bacteriophage λ. We thank R.Garber, W.McGinnis, E.Frei and the members of our laboratory for reading the manuscript and suggesting improvements, E.Weber for help in preparing the manuscript and especially Professor E.M. De Robertis for his enthusiastic support and encouragement throughout the entire course of this work. This work was supported by a grant from the Swiss National Science Fund to E.M.De Robertis.

References

Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
Bernstein,L.B., Mount,S.M. and Weiner,A.M. (1983) *Cell*, **32**, 461-472.
Birchmeier,C., Grosschedl,R. and Birnstiel,M.L. (1982) *Cell*, **28**, 739-745.
Branlant,C., Krol,A., Ebel,J.P., Lazar,E., Bernard,H. and Jacob,M. (1982) *EMBO J.*, **1**, 1259-1265.
Busch,H., Reddy,R., Rothblum,L. and Choi,C.Y. (1982) *Annu. Rev. Biochem.*, **51**, 617-654.
Busslinger,M., Portmann,R. and Birnstiel,M.L. (1979) *Nucleic Acids Res.*, **6**, 2997-3008.
Card,Ch.O., Morris,G.F., Brown,D.T. and Marzluff,W.T. (1982) *Nucleic Acids Res.*, **10**, 7677-7688.
Clarkson,S.G. and Kurer,V. (1976) *Cell*, **8**, 183-195.
Denison,R.A., Van Arsdell,S.W., Bernstein,L.B. and Weiner,A.M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 810-814.
De Robertis,E.M. (1983) *Cell*, **32**, 1021-1025.
De Robertis,E.M., Lienhard,S. and Parisot,R.F. (1982) *Nature*, **295**, 572-577.
Gilvart,C., Bollum,F.J. and Weissmann,C. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 428-432.
Goldberg,M.L. (1979) Ph.D. Thesis, Stanford University, Stanford, CA.
Gurdon,J.B. and Brown,D.D. (1978) *Dev. Biol.*, **67**, 346-356.
Hentschel,C.C. and Birnstiel,M.L. (1981) *Cell*, **25**, 301-313.
Hofer,E. and Darnell,J.E. (1981) *Cell*, **23**, 585-593.
Hosbach,H.A., Wyler,T. and Weber,R. (1983) *Cell*, **32**, 45-53.
Humphries,P., Old,R., Coggins,L.W., McShane,T., Watson,C. and Paul,J. (1978) *Nucleic Acids Res.*, **5**, 905-924.
Lerner,M.R. and Steitz,J.A. (1979) *Proc. Natl. Acad. Sci. USA*, **78**, 2737-2741.
Los Alamos National Laboratorium Sequence Data Bank (Genbank, 1982),

USA.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning,* published by Cold Spring Harbor Laboratory Press, NY.
Manley,J.L., Fire,A., Cano,A., Shamp,P.A. and Gefter,M.L. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 3855-3859.
Manser,T. and Gesteland,R.F. (1982) *Cell,* **29**, 257-264.
Maxam,A. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 560-564.
Messing,J. and Vieira,J. (1982) *Gene,* **19**, 269-276.
Murphy,J.T., Burgess,R.R., Dahlberg,J.E. and Lund,E. (1982) *Cell,* **29**, 265-274.
Newport,J. and Kirschner,M. (1982) *Cell,* **30**, 675-686.
Nishikura,K., Kurjan,J., Hall,B.D. and De Robertis,E.M. (1982) *EMBO J.,* **1**, 263-268.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature,* **263**, 211-214.
Reddy,R., Henning,D., Epstein,P. and Busch,H. (1981) *Nucleic Acids Res.,* **9**, 5645-5657.
Reddy,R. and Busch,H. (1981) *Cell Nucleus,* **8**, 261-306.
Roop,D.R., Kristo,P., Stumph,W.E., Tsai,M.J. and O'Malley,B.W. (1981) *Cell,* **23**, 671-680.
Sanger,F., Nicklen,S. and Coulsen,A.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5463-5467.
Southgate,K., Ayme,A. and Voellmy,R. (1983) *J. Mol. Biol.,* **165**, 35-57.
Taya,Y., Devos,R., Tavernier,J., Cheroutre,H., Engler,G. and Fiers,W. (1982) *EMBO J.,* **1**, 953-958.
Van Arsdell,S.W., Denison,R.A., Bernstein,L.B., Weiner,A.M., Manser,T. and Gesteland,R.F. (1981) *Cell,* **26**, 11-17.
Vieira,J. and Messing,J. (1982) *Gene,* **19**, 259-268.
Wahli,W. and Dawid,I.B. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 1437-1441.
Watanabe-Nagasu,N., Itoh,Y., Tani,T., Okano,K., Koga,N., Okada,N. and Ohshima,Y. (1983) *Nucleic Acids Res.,* **11**, 1791-1801.
Wise,J.A. and Weiner,A.M. (1980) *Cell,* **22**, 109-118.
Zeller,R., Nyffenegger,Th. and De Robertis,E.M. (1983) *Cell,* **32**, 425-434.