The EMBO Journal vol.3 no.5 pp.1075–1081, 1984

Xenopus laevis U1 snRNA genes: characterisation of transcriptionally active genes reveals major and minor repeated gene families

Rolf Zeller*, Maria-Teresa Carri[1], Iain W.Mattaj and Eddy M.De Robertis

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

*To whom reprint requests should be sent
[1]Present address: Dip. Genetica e Biologia Molecolare, Facolta di Scienze MFN, Universita di Roma 'La Sapienza', 00185 Roma, Italy
Communicated by E.M.De Robertis

***Xenopus laevis* U1 snRNA genes are found in several different genomic arrangements. The major family of genes is organised in tandem repeats of 1.8 kb. The minor U1-family is much less abundant and is present on 1.2-kb *Hinf*I restriction fragments. In addition there are genomic arrangements present in one or very few copies, which could represent the ends of repeating units. There is no evidence for the presence of U1 pseudogenes in *Xenopus*. A cluster of U1 snRNA genes consisting of a member of the minor class of U1 snRNA genes and two of the 'rarely represented' genes was cloned. All three genes were expressed upon microinjection into frog oocytes. A fragment containing 149 bp of 5′ flanking sequence, the RNA coding sequence, and 27 bp of 3′ flanking sequence was shown to be accurately transcribed into U1 snRNA. These oocyte transcripts are assembled into specific U1 snRNPs. Sequence comparison of the regions flanking *Xenopus* U1 and U2 snRNA genes showed the presence of two blocks of homology, which are also conserved in many other U snRNA genes. One of these blocks is found at position −60 to −50 before the coding sequence, and we discuss its possible role in the correct initiation of transcription. The other is 3′ to the coding sequence and may be involved in the accurate production of mature 3′ ends in the RNA.**
Key words: U1 snRNA genes/transcription/major, minor gene families/sequence comparison

Introduction

Six highly conserved U snRNAs (U-rich small nuclear RNAs) are present in the nuclei of higher eukaryotic cells in amounts ranging from 10^4 to 10^6 copies per cell. These RNAs are capped at the 5′ end but not polyadenylated and range in size from 107 to 214 bases (Lerner and Steitz, 1981; Busch *et al.*, 1982). U snRNAs are complexed with 7–8 proteins in U snRNPs and in this form are immunoprecipitable by sera from some patients with autoimmune diseases (Lerner and Steitz, 1979). There is evidence that U1 snRNPs function in hnRNA splicing, and roles for other U snRNPs in RNA processing have been postulated (Lerner *et al.*, 1980; Rogers and Wall, 1980; Ohshima *et al.*, 1981; Yang *et al.*, 1981; Padgett *et al.*, 1983; Galli *et al.*, 1983).

Genes coding for U snRNAs have been isolated from a variety of eukaryotes. In most cases (Manser and Gesteland, 1982; Wise and Weiner, 1980; Marzluff *et al.*, 1983) they were reported to be in multiple copies dispersed throughout the genome. Recently, however, it was found that the genes coding for N1 and N2, two sea urchin snRNAs (Card *et al.*, 1982), as well as *Xenopus laevis* U2 and U5 snRNA genes (Mattaj and Zeller, 1983), and human U2 snRNA genes (Van Arsdell and Weiner, 1984) are tandemly repeated. In addition, at least some copies of the U1 snRNA genes in mouse and rat were found to be clustered (Marzluff *et al.*, 1983; Watanabe-Nagasu *et al.*, 1983) and all the human U1 snRNA genes were shown to be located on a single chromosome (Lund *et al.*, 1983), suggesting that the earlier reports of dispersal throughout the genome may have been due simply to the scale of resolution of the methods employed. When cloned human U1 snRNA genes (Murphy *et al.*, 1982) or *X. laevis* U2 snRNA genes (Mattaj and Zeller, 1983) are microinjected into *X. laevis* oocyte nuclei, they are transcribed at a high level by RNA polymerase II. The 5′ flanking sequences of these genes do not have a TATA box, which is usually present about 30 bp upstream from most eukaryotic genes transcribed by RNA polymerase II (Goldberg, 1979; Corden *et al.*, 1980). Different authors have proposed several elements in the 5′ flanking sequence of U snRNA genes which might substitute for the TATA box on the basis of sequence comparison data or deletion mapping (Murphy *et al.*, 1982; Mattaj and Zeller, 1983; Skuzeski *et al.*, in preparation; Van Arsdell and Weiner, 1984).

We report here the existence of a major tandemly repeated and a minor family of U1 snRNA genes in the *X. laevis* genome. The three U1 genes cloned, including a member of the minor family, were shown to be transcriptionally active after microinjection into oocyte nuclei. Unlike mammals, *Xenopus* does not contain detectable U1 pseudogenes. A 340-bp subclone of one of the three transcriptionally active genes was sequenced. The 149 bp of 5′ flanking sequence show little homology to the *X. laevis* U2 snRNA gene, except for a striking conserved region between positions −50 and −60 which might function in transcription initiation. The 27 bp of 3′ flanking sequence were shown to contain the previously described region of homology (Mattaj and Zeller, 1983) between positions +11 and +23, which might be necessary for the accurate production of mature 3′ ends on U snRNAs.

Results

Evidence for the existence of major and minor repeated U1 gene families in Xenopus

We analysed the arrangement of the U1 snRNA genes in the genome of *X. laevis* by whole genome Southern blot analysis using a [^{32}P]nick-translated chicken U1 cDNA clone (Roop *et al.*, 1981) as a probe. Upon complete digestion we found that many enzymes (e.g., *Eco*RI, *Bam*HI, *Bgl*I) gave rise to very large hybridising bands, presumably because they do not cut within the U1 major tandem repeats. Figure 1 shows that partial digestion with *Hinf*I gave rise to a ladder of bands regularly spaced at 1.8-kb intervals, the smallest of which was 1.8 kb. Complete digestion by *Hinf*I leaves only a strongly

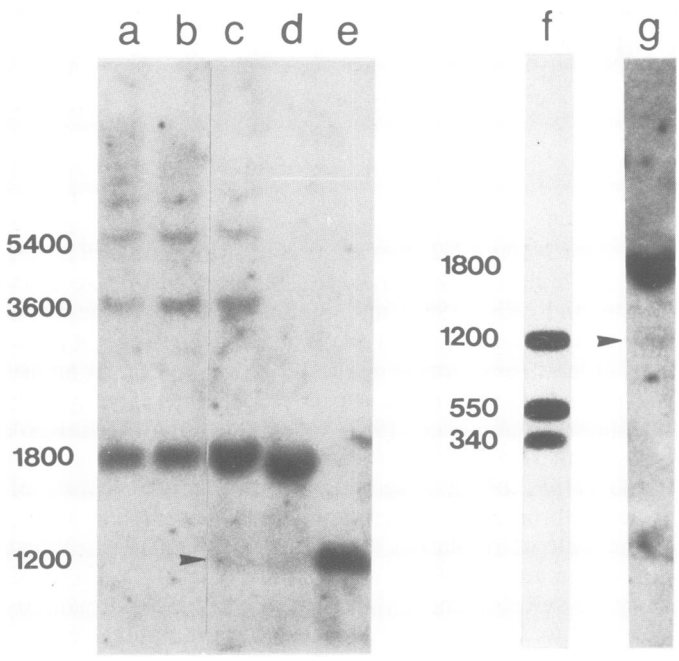

Fig. 1. Analysis of the U1 snRNA gene arrangement in genomic *X. laevis* DNA and in cloned DNA of λU1. The DNAs were digested with different amounts of *Hinf*I, separated on an agarose gel, transferred to nitrocellulose and hybridised to nick-translated cloned chicken U1 cDNA. **Lanes a–d:** genomic *X. laevis* DNA (10 μg/slot) was digested with 10 units of *Hinf*I for increasing amounts of time and analysed on a 0.7% agarose gel: **(a)** 30 min, **(b)** 40 min, **(c)** 60 min, **(d)** complete digest (15 units for 3 h). The arrowhead points to a 1.2-kb band visible in **lanes c** and **d**. **Lane e:** 2 μg of DNA from clone λU1 was digested completely with *Hinf*I (15 units for 3 h) and run on the same gel as samples **a** to **d**. Only the 1.2-kb hybridising fragment of λU1 (see also Figure 2b) is shown because the two small fragments run off the end of this Southern blot (0.7% agarose gel). **Lane f:** clone λU1 completely digested with *Hinf*I run on a 2% agarose gel. All three hybridising bands can be seen. **Lane g:** genomic *X. laevis* DNA completely digested with *Hinf*I run on a 2% agarose gel. Arrowhead points to the weakly hybridising band running at 1.2 kb.

hybridising 1.8-kb band and a more weakly hybridising 1.2-kb band (Figure 1, lanes d and g). This combination of results would be expected if most of the U1 snRNA genes are arranged in a 1.8-kb tandem repeat. Using *Hind*III, similar results suggesting a tandem arrangement were independently found by E.Lund and J.Dahlberg, with the repeat length being measured as 1.9 kb (personal communication).

The minor 1.2-kb band was reproducibly observed in several genome Southern blots, and we believe that the genes on these fragments are also repeated. Although the number of U1 genes was not measured directly in our studies, estimates of others give a number of > 1000 U1 snRNA genes in the *Xenopus* genome (Lund and Dahlberg, 1984). For U2 snRNA genes we estimated between 500 and 1000 copies per haploid genome (Mattaj and Zeller, 1983). Thus, even a band of decreased intensity should represent a repeated gene in our analysis. As will be described in the following section, we cloned a member of this 1.2-kb minor gene family (Figure 1, lane 3). The same bacteriophage λ recombinant also contains two additional U1 snRNA genes (Figure 1, lane f). Despite several attempts we were unable to detect the two latter genes in genomic Southerns (Figure 1, lane g). This suggests that they are present in few or single copies in the *Xenopus* genome and that these 'rare representatives' are less abundant than the minor U1 snRNA gene family.

We conclude that *X. laevis* U1 snRNA genes can be found

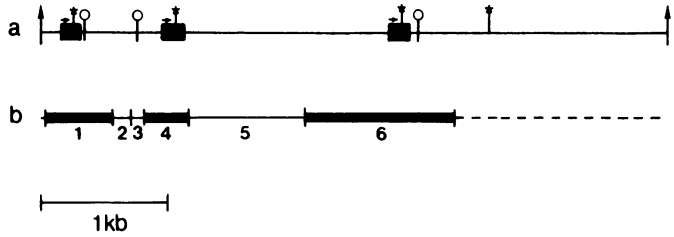

Fig. 2. Restriction map of a 5-kb *Hind*III fragment of clone λU1 which includes three U1 snRNA genes. This restriction map was made by complete and partial digestion of the isolated 5-kb *Hind*III fragment of clone λU1 with the restriction enzymes indicated on the map and Southern analysis of the fragments. The map was confirmed using double digests. **(a)** Map showing all the restriction sites for the following enzymes in the 5-kb *Hind*III fragments: HindIII (cuts outside the U1 snRNA coding sequence); BglII (cuts outside the U1 snRNA coding sequence); HpaII (cuts within the U1 snRNA coding sequence). The black boxes indicate the positions of U1 snRNA homologous regions as determined by hybridisation to ^{32}P-labelled U1 probes; the arrows show their 5′ to 3′ orientation with respect to the U1 snRNA coding sequence (see text). **(b)** Partial restriction map of *Hinf*I sites within the 5-kb *Hind*III fragment. It was not possible to complete this map due to the presence of several clustered *Hinf*I sites towards the right hand side (dashed line). The black bars represent fragments hybridising to ^{32}P-labelled U1 probes. Their sizes are as follows: **1**: 550 bp, **4**: 340 bp, **6**: 1.2 kb (see also Figure 1, lane f). Fragments **2, 3** and **5** do not hybridise to the U1 probe.

in several genomic arrangements: major and minor families of repeated genes and 'rarely represented' genes. The major U1 snRNA gene family is tandemly repeated, with each unit having a length of 1.8 kb.

Selection of a clone containing U1 snRNA genes

Clones containing U1 snRNA genes were selected from an *X. laevis* genomic library (twice amplified) in the Charon 4A derivative of bacteriophage λ (Wahli and Dawid, 1980) probed with a nick-translated cDNA clone of chicken U1 snRNA (Roop *et al.*, 1980).

Several haploid genome equivalents (~2 x 10^5 plaques) were screened and positive clones were picked and rescreened using [^{32}P]end-labelled U1 snRNA (Mattaj and Zeller, 1983). Only one clone, λU1, hybridised on the rescreen and on further testing it proved to contain transcriptionally active U1 snRNA genes (see below).

Clone λU1 was digested with different restriction enzymes and DNA was transferred to nitrocellulose and hybridised to ^{32}P-labelled U1 probe. The restriction map of part of the insert of λU1 so constructed is shown in Figure 2. Clone λU1 contains three regions which by hybridisation are homologous to U1 snRNA (Figure 1, lane f), all of which are contained within a 5-kb *Hind*III insert fragment.

By the use of different restriction enzymes (some of which are entered in the map in Figure 2) a restriction map of the 5-kb *Hind*III fragment was made: a 340-bp *Hind*III/*Bgl*II fragment contains one U1 snRNA gene, which was subcloned, sequenced and tested for transcriptional activity (see below). A 2.3-kb *Bgl*II/*Bgl*II fragment contains two U1 snRNA genes at its ends (Figure 2A). That there are only two U1 snRNA genes in this fragment was suggested by the fact that *Hpa*II, an enzyme which cuts within the U1 snRNA sequence (as determined from sequence data) only has two restriction sites within the 2.3-kb fragment. This result is supported by the restriction-hybridisation pattern obtained with *Dde*I, another enzyme that cuts within the U1 snRNA sequence and by subcloning the two genes individually. The orientation of the two genes was determined by hybridising single stranded DNA of the *Bgl*II fragment (subcloned into

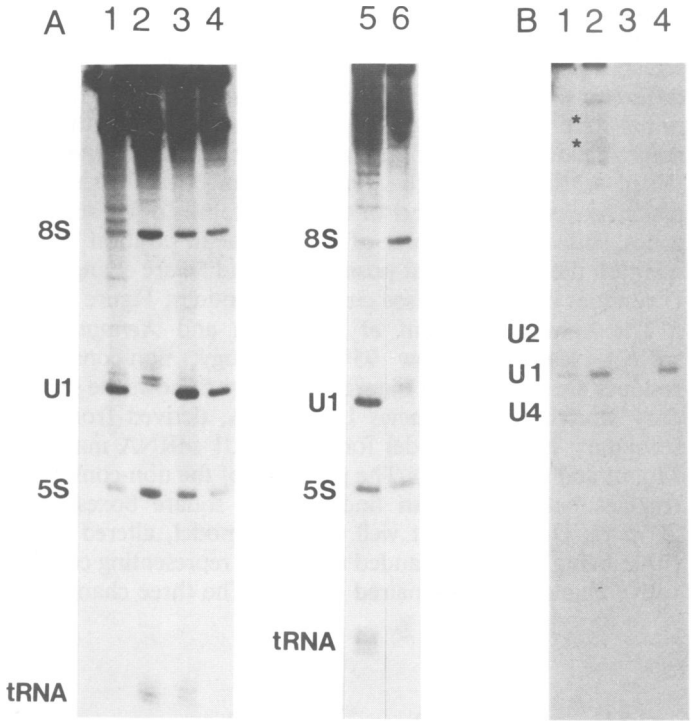

Fig. 3. (A) RNA transcribed in *X. laevis* oocytes after microinjection of DNA from clone λU1 and from various subclones derived from it. 24 h after microinjection of oocyte nuclei with cloned DNA and [α-^{32}P]GTP the RNAs were extracted and analysed by polyacrylamide gel electrophoresis (see Materials and methods). Each lane contains an amount of RNA equivalent to that of one oocyte. **Lane 1:** oocyte injected with DNA from λU1. **Lane 2:** oocyte injected with [α-^{32}P]GTP only, showing the endogenous oocyte RNAs (the two bands running immediately above the U1 snRNA region are 5.8S rRNA. The indicated 8S RNA is an endogenous RNA polymerase I transcript that contains 5.8S RNA sequences). **Lane 3:** oocyte injected with clone *Xl*U1.9, a 340-bp *Hind*III/*Bgl*II fragment (see Figure 1) cloned into M13mp9, containing one U1 snRNA gene. **Lane 4:** oocyte injected with clone *Xl*U1.8, a subclone of the same 340-bp fragment in M13mp8, i.e., in the opposite orientation with respect to the vector. **Lane 5:** oocyte injected with clone *Xl*U1.23, a 2.3-kb *Bgl*II/*Bgl*II fragment (see Figure 1) in M13mp8, which contains two U1 snRNA genes subcloned from λU1. **Lane 6:** oocyte injected with [α- ^{32}P]GTP only. **(B)** Immunoprecipitation of U snRNPs from oocytes injected with and without cloned DNA. 24 h after microinjection of DNA and [α-^{32}P]GTP samples were immunoprecipitated as described in Materials and methods. The equivalent of U snRNAs from one oocyte were loaded and the U snRNAs were identified as described previously (Zeller *et al.*, 1983). **Lanes 1,3:** oocytes injected with [α-^{32}P]GTP only and immunoprecipitated with autoimmune Sm antisera **(lane 1)** or U1-RNP antisera **(lane 3)** to visualise endogenously made U snRNPs. **Lanes 2,4:** oocytes injected with DNA from clone *Xl*U1.9 and [α-^{32}P]GTP and immunoprecipitated with Sm antisera **(lane 2**, for explanation of the asterisks see text) or U1-RNP antisera **(lane 4).**

M13mp8, Messing and Vieira, 1982) to both strands of the 340-bp fragment (subcloned into M13mp8 and M13mp9). It was found that only one of the two strands hybridised to the 2.3-kb subclone, demonstrating that both genes have the same orientation. Sequencing the ends of the 2.3-kb subclone (data not shown) confirmed the position and orientation data.

We conclude that λU1 contains a cluster of three regions homologous to U1 snRNA, all of which are in the same orientation (Figure 2A) and represent transcriptionally active U1 snRNA genes (see below). One of these genes corresponds to a member of the minor family of *X. laevis* U1 snRNA genes (on a 1.2-kb *Hinf*I fragment, see Figure 1, lane f and Figure 2B). The other two genes could not be detected in genomic Southern blots (Figure 1, lane g) and must therefore be present in very few copies ('rare representatives'). Perhaps they represent the end of a cluster of U1 snRNA genes.

Clone λU1 contains three transcriptionally active U1 snRNA genes

Having established that the cloned U1 snRNA genes were not members of the major gene family, we tested whether they were pseudogenes by assaying their transcriptional activity upon microinjection into *X. laevis* oocyte nuclei. The newly synthesised RNAs were labelled by [α-^{32}P]GTP. Clone λU1 after microinjection into *Xenopus* oocytes gives rise to U1 snRNA sized transcripts, and therefore contains transcriptionally active U1 snRNA genes (Figure 3A, lane 1). The U1 snRNA gene contained in the 340-bp *Hind*III/*Bgl*II fragment (see Figure 1a) from clone λU1 was subcloned in both orientations using M13mp8 (clone *Xl*U1.8) and M13mp9 (clone *Xl*U1.9) (Messing and Vieira, 1982). Double-stranded DNA of these two clones was shown to be transcriptionally active on microinjection (Figure 3A, lanes 3 and 4). This result indicates that all the sequences necessary for transcription are contained within the 340-bp fragment and that transcription is independent of the orientation of the vector.

The 2.3-kb *Bgl*II/*Bgl*II fragment (see Figure 2a) containing two putative U1 snRNA genes was also subcloned into M13mp8 (clone *Xl*U123) and shown to be transcriptionally active upon microinjection into oocytes (Figure 3A, lane 5) The two genes were further subcloned using an *Eco*RI restriction site close to the middle of the 2.3-kb fragment, and each subclone was injected into oocyte nuclei as single-stranded DNA (Cortese *et al.*, 1980). Both genes were transcribed into RNA of the same size as mature U1 snRNA (data not shown). From this we conclude that all three U1 snRNA homologous regions on clone λU1 represent transcriptionally active genes.

U1 snRNA transcripts from microinjected genes are assembled into U1 snRNPs

To determine whether the U1 snRNA made from the 340-bp subclone *Xl*U1.9 is correctly assembled into U1 snRNPs in oocytes, extracts of injected oocytes were immunoprecipitated with Sm antisera and U1-RNP antisera as described by De Robertis *et al.* (1982). Sm antisera precipitate all U snRNPs, whereas U1-RNP antisera only precipitate U1 snRNPs (Lerner and Steitz, 1979). As shown in Figure 3B the U1 snRNAs made from clone *Xl*U1.9 are immunoprecipitable with both Sm and U1-RNP antisera (Figure 3B, lanes 2 and 4) and migrate to the same position in the gel as endogenously synthesized U1 snRNAs (Figure 3B, lanes 1 and 3). The resolution of the polyacrylamide gels used was determined to be one nucleotide by using sequenced single-stranded DNA as markers, i.e., the U1 snRNA synthesised from clone *Xl*U1.9 is exactly identical in length to endogenous U1 snRNA. Some longer RNA molecules are also precipitated by Sm antisera (Figure 3B, lane 2, indicated by asterisks). These might represent read-through transcripts which assemble with the proteins antigenic for the Sm antisera. These longer RNA molecules are not precipitated with the U1 snRNP-specific U1-RNP antisera (Figure 3B, lane 4). This might indicate that the particles into which these longer RNA molecules are assembled are structurally abnormal or incomplete whereas U1 snRNA molecules of correct length seem to assemble into complete and structurally normal U1 snRNP particles in microinjected oocytes.

Since U snRNA genes are transcribed by RNA polymerase II in *X. laevis* oocytes (Murphy *et al.*, 1982; Mattaj and Zeller, 1983) we can conclude that the 340-bp subclone from λU1 is able to accurately direct the transcription of U1 snRNA molecules by RNA polymerase II and that the transcripts can then assemble into U1 snRNPs.

Sequence analysis of the 340-bp subclone containing a transcriptionally active U1 snRNA gene

The sequence of the 340-bp fragment from clone λU1 was determined using the strategy shown in Figure 4B. Because of difficulties in reading the sequence, particularly between positions −5 and −20 on both strands (most likely due to the formation of secondary structure in the single-stranded DNA), the sequence obtained using the method by Sanger *et al.* (1977) was checked using the method of Maxam and Gilbert (1980) to the extent shown in Figure 4B (bar no. 3).

Figure 4A shows the sequence of the non-coding strand of the 340-bp fragment. The 164 underlined bases correspond to the U1 snRNA coding sequence. The *X. laevis* U1 snRNA sequence obtained was kindly compared by E.Lund and J.Dahlberg to *X. laevis* somatic cell U1 snRNA fingerprints obtained from tissue culture cells (personal communication). Except for the U residue at position 78 of the U1 snRNA no differences between the DNA sequence and the RNA fingerprints were found. Sequencing (one strand only) of the U1 snRNA coding region of the transcriptionally active gene contained within the 1.2-kb *Hinf*I fragment (see Figure 2b), which represents a member of the minor family of U1 snRNA genes, showed the presence of a C residue at position 78. In addition the C residues at positions 14 and 54 are changed to U residues in this gene (see circled nucleotides, Figure 5).

The human (Branlant *et al.*, 1980) and *Xenopus* U1 snRNA sequences show 95% homology, non-conserved residues are overlined in Figure 4A. Figure 5 shows the secondary structure of *X. laevis* U1 snRNA, derived from the secondary structure model for human U1 snRNA made by Mount and Steitz (1981). The positions of the non-conserved residues between human (indicated in square boxes) and *X. laevis* U1 snRNA fit well with the model, altered bases either being in single-stranded regions or representing conservative changes in base-paired regions. The three changes in

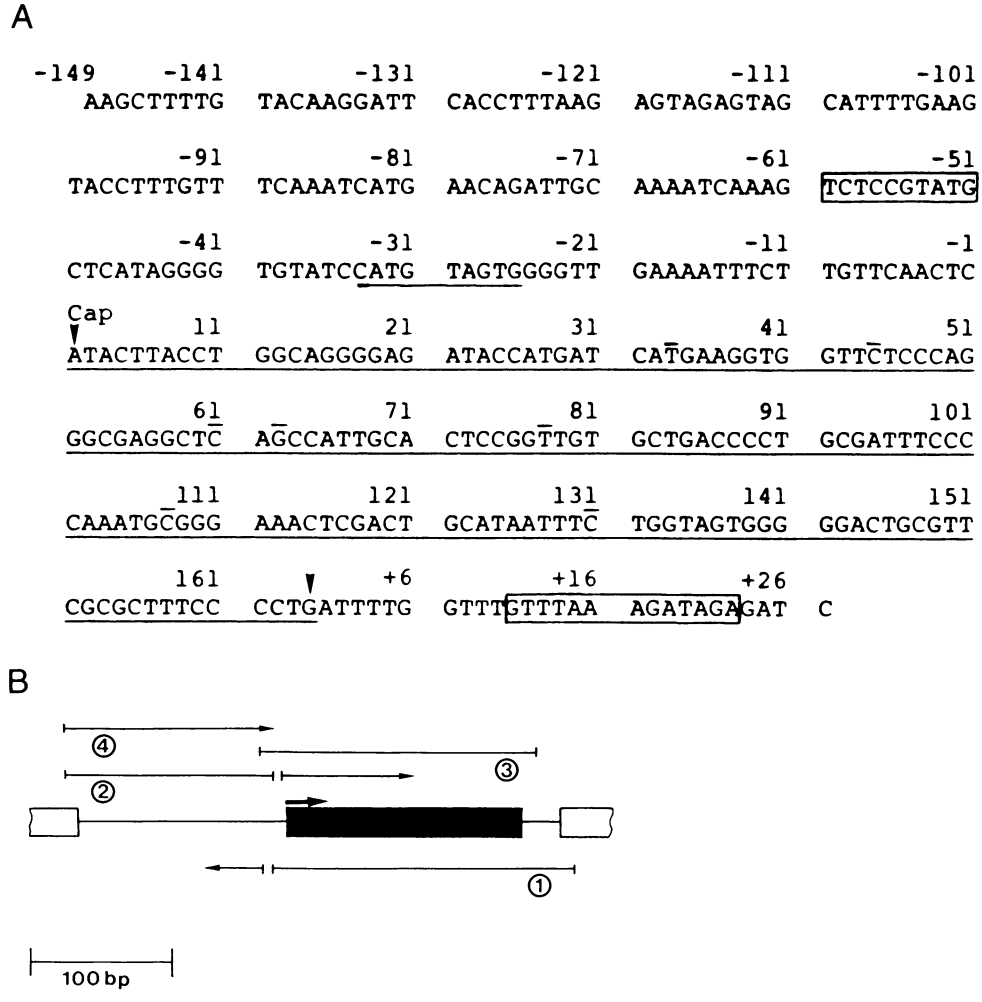

Fig. 4. (A) DNA sequence of the non-coding strand of the 340-bp fragment in subclones *Xl*U1.8 and *Xl*U1.9. The U1 snRNA coding sequence is underlined. Residues not conserved between *Xenopus* and human U1 snRNA are indicated by dashes above the sequence. Boxed and additional underlined sequences are discussed in the text. The arrowheads indicate the presumptive 5′ and 3′ ends of the U1 snRNA coding sequence, deduced from the human U1 snRNA sequence (Branlant *et al.*, 1980). **(B)** Sequencing strategy: DNA sequence data determined by the method of Sanger *et al.* (1977): ① Clone *Xl*U1.9 sequenced reading the coding strand. ② Clone *Xl*U1.8 sequenced reading the non-coding strand. Blanks in ① and ② represent unreadable sequences (see text). ④ Sequence data obtained from an additional subclone made from *Xl*U1.8. DNA sequence determined by the method of Maxam and Gilbert (1980) is indicated by bar ③.

the RNA coding region of the U1 snRNA gene contained in the 1.2-kb *Hinf*I fragment (see circled residues) do not alter the secondary structure.

In the 5′ and 3′ flanking regions sequence homologies (boxed in Figure 4A) to the previously sequenced *X. laevis* U2 snRNA gene-flanking sequences (Mattaj and Zeller, 1983) were found (Table I). These homologies will be discussed below.

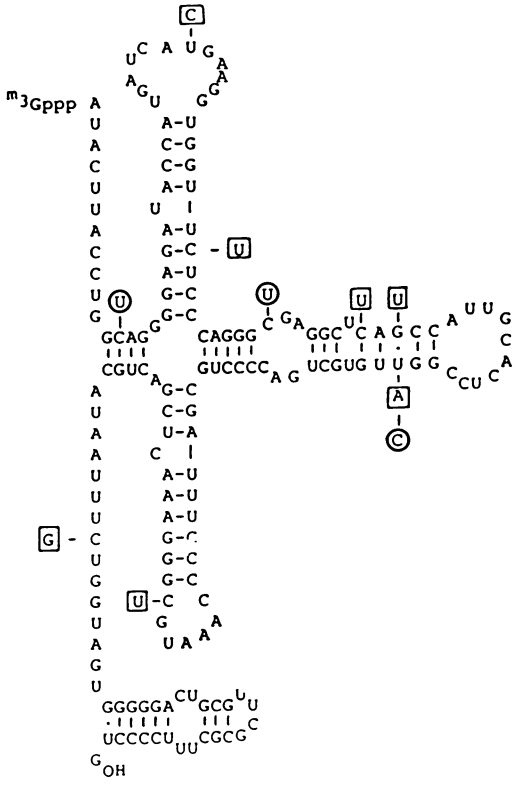

Fig. 5. Secondary structure model of *Xenopus* U1 snRNA constructed using the model of Mount and Steitz (1981) from human U1 snRNA. The sequence indicated is that of the U1 snRNA gene contained in the 340-bp fragment. Added nucleotides in square boxes represent base substitutions in human U1 snRNA. Circled residues correspond to changes found in the coding region of the U1 snRNA gene contained in the 1.2-kb *Hinf*I fragment, which represents a member of the minor family of repeated genes.

Discussion

The existence of alternative genomic arrangements for U1 snRNA genes in *X. laevis* was established by genomic Southern blots. The majority of the *X. laevis* U1 snRNA genes are organised in tandem repeats with a repeat length of ~1.8-kb. *X. laevis* U2 and U5 snRNA genes were previously shown to be tandemly repeated (Mattaj and Zeller, 1983), as are sea urchin *Lytechinus variegatus* N1 and N2 snRNA genes (Card *et al.*, 1982) and human U2 snRNA genes (Van Arsdell and Weiner, 1984). In addition to the major tandemly repeated genes, a minor class of repeated genes was found. A number of 'rarely represented' genes for U1 snRNA was also found, but their degree of repetition in the genome and possible role in development remains to be established. A clone, λU1, contains clustered U1 snRNA genes, one of which is a member of the minor repeated family whereas the others probably represent 'unique' genes in the genome, which were not detectable in the genomic Southern blots. It is conceivable that these 'unique' genes may represent the ends of clusters of genes, which then tend to diverge in their restriction enzyme cleavage pattern. All three U1 snRNA genes contained on clone λU1 accurately transcribe U1 snRNA which is assembled into U1 snRNPs. A subclone from λU1 containing a transcriptionally active U1 snRNA gene, was shown to contain in addition to the coding sequence, only 149 bp of 5′ and 27 bp of 3′ flanking sequence (Figure 4A). When the flanking sequences of this U1 snRNA gene were compared to the flanking sequences of the *X. laevis* U2 snRNA gene cloned previously (Mattaj and Zeller, 1983) and to those of other U snRNA genes, some homologies were observed (see Table I).

Murphy *et al.* (1982) and Skuzeski *et al.* (in preparation) have shown by deletion experiments that the region between position −231 and −203 is essential for transcription of a human U1 snRNA gene both *in vitro* and after microinjection into oocytes. Starting at position −212 the sequence TATGTAGATG was found. It was suggested that this sequence might provide a polymerase entry site or might act as an enhancer-like element. A similar sequence was found star-

Table I. Sequence comparison between the 5′ and 3′ flanking sequences of the *X. laevis* U1 and U2 snRNA genes

A. Homologies found in the 5′ flanking region of *X. laevis* U snRNA genes

	−70	−60	−50	−40	−30
X. laevis U1	AAAATCAAAG	TCTCCGTATG	CTCATAGGGG	TGTATCCATG	TAGTGGGGTT
	−70	−60	−50	−40	−30
X. laevis U2	CAGCCCTCCC	TCTCCCCATG	GAGGCATGTC	TAGCCTGGCT	TTGGGCCCGT
Consensus	-------G	TG/CACCGT/CGTG	NGTT/AAAG/TGTG	-------	

(Skuzeski *et al.*, 1984)

B. A homologous sequence found in the 3′ flanking region of U snRNA genes

	+11													
Xenopus U1	G	T	T	T	-	A	A	A	G	A	T	A	G	A
	+11													
Xenopus U2	G	T	T	T	G	A	A	A	A	A	GC	A	G	A
	+n													
Consensus	-	T	T	T	N	A	A	A	G	A	A/T	-	-	-

(Mattaj and Zeller, 1983)

ting at position −38 and −48 in human U2 snRNA genes (Van Arsdell and Weiner, 1984) and in a rat U2 snRNA gene starting at position −36 (Tani *et al.*, 1983). In the human U1 snRNA gene an additional sequence similar to the one at position −212 is repeated closer to the coding sequence (between positions −37 and −51, see Lund and Dahlberg, 1984), but it is not sufficient for transcription (Murphy *et al.*, 1982; Skuzeski *et al.*, in preparation). The *Xenopus* U1 and U2 snRNA genes both have sequences homologous to TATGTAGATG, but they are located much closer to the RNA coding sequence (starting at position −26 or −38, respectively; see underlined sequences in Table IA).

When the 5′ flanking sequence of the *X. laevis* U1 and U2 snRNA genes were compared, they were found to be highly divergent. However, one striking homology conserved not only in sequence but also in position, was found between positions −50 and −60 (see boxed sequence in Table IA). A clone in which all sequences upstream from position −55 of the U2 snRNA gene are removed is no longer transcriptionally active, whereas a deletion removing everything upstream from position −82 is still active. This suggests that something within or closely upstream of the 10-bp homology is essential for U2 snRNA transcription in *Xenopus* oocytes (I.Mattaj and E.De Robertis, unpublished results). Skuzeski *et al.* (in preparation), by sequence comparison of several mammalian U snRNA genes, were able to find a second consensus sequence for the region between −40 and −60 (see Table IA). The 10-bp homology in *X. laevis* U1 and U2 snRNA genes is strikingly similar in sequence and position to part of their proposed consensus (Table IA). Taken together, these data lend support to the idea that an element essential for faithful U snRNA transcription (at least in *Xenopus* oocytes) might be contained within this −50 to −60 region.

Table IB shows the sequence between positions +11 and +23 in the 3′ flanking region of the sequenced U1 snRNA gene, and compares it to the sequence of the identical region of a *X. laevis* U2 snRNA gene and to a consensus sequence found in the 3′ flanking region of many RNA polymerase II transcribed genes (Mattaj and Zeller, 1983). Homologous sequences can be found starting within 20 bp of the 3′ end of the coding sequence of all U snRNA genes so far sequenced (data not shown). Deletion of part of this sequence leads to a decrease in the efficiency of production of correct 3′ ends on transcripts of a U1 snRNA gene injected into oocytes (I.Mattaj, R.Zeller and E.De Robertis, in preparation) lending support to the hypothesis that this sequence plays a role in the accurate production of 3′ ends on U snRNA transcripts.

In future it will be very important to compare the structure of this U1 snRNA gene to the structure of a U1 snRNA transcription unit from the major tandemly repeated gene family, which has recently been cloned (E.Lund and J.Dahlberg, personal communication). In humans, a ratio of ten U1 snRNA pseudogenes to one gene has been reported (Monstein *et al.*, 1982; Bernstein *et al.*, 1983). In contrast we have no evidence of U1 snRNA pseudogenes in *X. laevis*. Although the genes we have cloned do not belong to the major family of U1 snRNA genes, all of them are transcriptionally active. These minor genes might have some special features, e.g., they might be expressed in specific tissues or differentially during development, although at present there is no evidence supporting this view. The nucleotide differences found in the RNA coding sequence of these genes might serve as markers to detect the expression of these minor species of U1 snRNAs *in vivo*.

Because U snRNA genes in *X. laevis* are developmentally controlled and synthesis of U snRNAs is first turned on during mid-blastula transition (Newport and Kirschner, 1982; Forbes *et al.*, 1983), these genes might provide tools to study the factors exerting negative (Newport and Kirschner, 1982) or positive control on gene expression during early development.

Xenopus oocytes and eggs should be a good source of transcription factors for the switching on of U snRNA genes during early development. In addition *Xenopus* U1 snRNA genes should be useful for transcription studies, since they provide the shortest RNA polymerase II transcription unit (340 bp) known until now.

Previous work from our laboratory (De Robertis *et al.*, 1982; Zeller *et al.*, 1983; Fritz *et al.*, in preparation) has shown that *X. laevis* oocytes and early embryos contain an excess of free U snRNP-binding proteins in the cytoplasm. These assemble into nuclear-migrating U snRNPs after synthesis of U snRNAs in embryos or after microinjection of U snRNAs into oocytes (reviewed by De Robertis, 1983). The *X. laevis* U snRNA genes cloned here and earlier (Mattaj and Zeller, 1983) are transcribed into U snRNAs which are assembled into U snRNPs immunoprecipitable by Sm and U1-RNP autoimmune antisera. *In vitro* deletion mutagenesis and creation of recombinant U1 and U2 snRNA genes will enable the study of the nucleotide sequences in U snRNAs that interact with the different protein components of U snRNPs. In particular, it will be interesting to establish the relationship between the binding of specific U snRNP proteins and the migration of U snRNAs into the oocyte nucleus after microinjection.

Materials and methods

Hybridisation using DNA and RNA probes

Transfer of DNA to nitrocellulose filters was done by the method of Southern (1975) as described by Maniatis *et al.* (1982). Transfer of plaques to nitrocellulose filters (Schleicher and Schüll) was done using the method of Benton and Davis (1977). Hybridisation to a [^{32}P]nick-translated U1 cDNA clone, a gift from D.R.Roop and B.W.O'Malley (Roop *et al.*, 1981), was performed as described by Maniatis *et al.* (1982). Hybridisation of [^{32}P]U1 snRNA labelled with poly A polymerase (Mattaj and Zeller, 1983) was done by the methods of Humphries *et al.* (1978).

Restriction digestions, 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.*, 1983; M13 DNA: Messing and Vieira, 1982) was 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 DNAs was 200−300 μg/ml for λ clones and 500−1000 μg/ml for M13 DNA. The volume microinjected was 30−50 nl.

Immunoprecipitation of U snRNAs

Extracts from injected oocytes were immunoprecipitated with human Sm and U1-RNP antisera and the precipitated ^{32}P-labelled U snRNAs were extracted and analysed on polyacrylamide gels as described by De Robertis *et al.* (1982). The Sm and U1-RNP antisera used were obtained from human SLE patients (Matter *et al.*, 1982).

DNA sequencing

DNA was cloned into M13mp8 and M13mp9 vectors (Messing and Vieira, 1982) and sequenced using the dideoxynucleotide chain terminator method of Sanger *et al.* (1977) to obtain the sequence of both strands. Because of some difficulties in reading the sequence (most likely caused by secondary structure in the sequenced clones), the clone was also sequenced using the method of Maxam and Gilbert (1980).

Acknowledgements

We are indebted to J.Dahlberg and E.Lund for comparing our sequence data with their unpublished U1 snRNA fingerprints data, and for providing us with other unpublished material. D.R.Roop and B.W.O'Malley generously provided us with their chicken U1 c-DNA clone. Furthermore we would like to thank S.Lienhard for excellent technical assistance, T.Bürglin and A.Fritz for help in the analysis of the sequence data, W.McGinnis, C.Sengstag and the members of our laboratory for reading the manuscript and E.Weber for help in preparing the manuscript and figures. The major part of this work was carried out by R.Zeller in fulfilment of the requirements of the PhD studies in Molecular Biology. This work was supported by a grant from the Swiss National Science Foundation. M.T.Carri's stay in our laboratory was supported by an EMBO short-term fellowship.

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.
Branlant,C., Krol,A., Ebel,J.P., Luzar,E., Gallinaro,H., Jakob,M., Sri Widada,J. and Jeanteur,P. (1980) *Nucleic Acids Res.,* **8**, 4143-4154.
Busch,H., Reddy,R., Rothblum,L. and Choi,C.Y. (1982) *Annu. Rev. Biochem.,* **51**, 617-654.
Card,Ch.O., Morris,G.F., Brown,D.T. and Marzluff,W.T. (1982) *Nucleic Acids Res.,* **10**, 7677-7688.
Corden,J., Wasylyk,B., Buchwalder,A., Sassone-Corsi,P., Kedinger,C. and Chambon,P. (1980) *Science (Wash.),* **209**, 1406-1414.
Cortese,R., Harland,R. and Melton,D. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 4147-4151.
De Robertis,E.M. (1983) *Cell,* **32**, 1021-1025.
De Robertis,E.M., Lienhard,S. and Parisot,R.F. (1982) *Nature,* **295**, 572-577.
Forbes,D.J., Kornberg,T.B. and Kirschner,M.W. (1983) *J. Cell Biol.,* **97**, 62-72.
Galli,G., Hofstetter,H., Stunnenberg,H.G. and Birnstiel,M.L. (1983) *Cell,* **34**, 823-828.
Garber,R.L., Kuroiwa,A. and Gehring,W.J. (1983) *EMBO J.,* **2**, 2027-2036.
Goldberg,M.L. (1979) *Ph. D. Thesis,* Stanford University, Stanford, USA.
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.
Lerner,M.R. and Steitz,J.A. (1981) *Cell,* **25**, 298-300.
Lerner,M.R., Boyle,J.A., Mount,S.M., Wolin,S.L. and Steitz,J.A. (1980) *Nature,* **283**, 220-224.
Lund,E., Bostock,C., Robertson,M., Christie,S., Mitchen,J.L. and Dahlberg,J.E. (1983) *Mol. Cell Biol.,* **3**, 2211-2220.
Lund,E. and Dahlberg,J.E. (1984) *J. Biol. Chem.,* **259**, 2013-2021.
Maniatis,T., Fritsch,E.F. and Sambrock,J. (1982) *Molecular Cloning,* published by Cold Spring Harbor Laboratory Press, NY.
Manser,T. and Gesteland,R.F. (1982) *Cell,* **29**, 257-264.
Marzluff,W.F., Brown,D.T., Lobo,S. and Wang,S.S. (1983) *Nucleic Acids Res.,* **11**, 6255-6269.
Mattaj,I.W. and Zeller,R. (1983) *EMBO J.,* **2**, 1883-1891.
Matter,L., Schopfer,K., Wilhelm,J.A., Nyffenegger,T., Parisot,R.F. and De Robertis,E.M. (1982) *Arthritis Rheum.,* **25**, 1278-1283.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.,* **65**, 499-560.
Messing,J. and Vieira,J. (1982) *Gene,* **19**, 269-276.
Monstein,H.J., Westin,G., Philipson,L. and Peterson,U. (1982) *EMBO J.,* **1**, 133-137.
Mount,S.M. and Steitz,J.A. (1981) *Nucleic Acids Res.,* **9**, 6351-6368.
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**, 687-696.
Nishikura,K., Kurjan,J., Hall,B.D. and De Robertis,E.M. (1982) *EMBO J.,* **1**, 263-268.
Oshima,Y., Itoh,M., Odaka,N. and Miyata,T. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 4471-4474.
Padgett,R.A., Mount,S.M., Steitz,J.A. and Sharp,P.A. (1983) *Cell,* **35**, 101-107.
Rogers,J. and Wall,R. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 1877-1879.
Roop,D.K., Kristo,P., Stumph,W.E., Tsai,M.J. and O'Malley,B.W. (1981) *Cell,* **23**, 671-680.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5463-5467.
Southern,E. (1975) *J. Mol. Biol.,* **98**, 503-517.
Tani,T., Watanabe-Nagasu,N., Okada,N. and Ohshima,Y. (1983) *J. Mol. Biol.,* **168**, 579-594.
Van Arsdell,S. and Weiner,A.M. (1984) *Mol. Cell. Biol.,* in press.
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.
Yang,V.W., Lerner,M.R., Steitz,J.A. and Flint,S.J. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 1371-1375.
Zeller,R., Nyffenegger T. and De Robertis,E.M. (1983) *Cell,* **32**, 425-434.

Received on 3 February 1984; revised on 22 February 1984

Note added in proof

The proper citation for the work referred to as E.Lund and J.Dahlberg (personal communication) is now Lund,E., Forbes,D.J. and Dahlberg,J.E., submitted for publication. The work cited as Skuzeski *et al.* (in preparation) was done by Skuzeski,J.M., Lund,E., Murphy,J.T., Steinberg,T.H., Burgess,R.R. and Dahlberg,J.E. (1984), submitted for publication.