The EMBO Journal vol.6 no.10 pp.3071-3078, 1987

Functional characterization of *X. laevis* U5 snRNA genes

Michael Kazmaier, Graham Tebb and Iain W.Mattaj

European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, FRG

Communicated by R.Cortese

***Xenopus laevis* U5 snRNA genes are found in several genomic arrangements, represented by a predominant tandem repeat of 583 bp and other minor repeats. Several copies of the major tandem repeat have been cloned and expressed in *Xenopus* oocytes. The transcripts assemble into U5 snRNPs which are recognized by anti-Sm antibodies. We have identified functional elements in the U5 gene promoter. Although similar in organization to other U snRNA gene promoters, U5 contains significant differences and is more efficiently expressed than the *Xenopus* U2 gene in oocytes. The proximal sequence element (PSE), although homologous to a mammalian consensus for this region (Skuzeski *et al.*, 1984), does not resemble the previously characterized *Xenopus* U1 and U2 PSEs closely in sequence. The ATGCAAAT (octamer) part of the distal sequence element (DSE 1) is found in U5 in the orientation opposite to that in U1 and U2 gene promoters. DNase I protection experiments led to the identification of a third element (DSE 2), situated close to the octamer motif. Analysis of deletion mutants showed that both DSE 1 and 2 are essential parts of the U5 gene enhancer, and provides evidence that U snRNA enhancers are complex structures consisting of more than one site of DNA-factor interaction.**

Key words: enhancer/factor binding/transcription signals/U5 snRNA genes

Introduction

The U5 RNAs belong to the family of abundant U small nuclear (sn)RNAs (U1-U6) that are present in 10^4-10^6 copies in the nuclei of higher eukaryotes. The U1-U5 RNAs are transcribed by RNA polymerase II but are not polyadenylated and have an unusual 2,2,7-trimethylguanosine cap structure (Busch *et al.*, 1982). In contrast, U6 RNAs are transcribed by RNA polymerase III (Kunkel *et al.*, 1986; Reddy *et al.*, 1987; Krol *et al.*, 1987). U1, U2, U4, U5 and U6 RNAs are found *in vivo* in the form of small nuclear ribonucleoproteins (snRNPs) which are recognized by antibodies from patients with certain autoimmune diseases (Lerner and Steitz, 1979).

Interest in the characterization of the family of U snRNAs and their genes in eukaryotes has increased because of the role of snRNPs in a variety of RNA processing steps. Substantial evidence now exists that U1, U2, U4, U5 and U6 snRNPs play a role in mRNA splicing (for reviews see Sharp, 1987; Maniatis and Reed, 1987), and that U7 is required for the processive generation of 3' ends of histone mRNA (Strub and Birnstiel, 1986; Schaufele *et al.*, 1986).

Functional elements have been defined in several U snRNA gene promoters. The proximal sequence element is centred at position -50 to -60 relative to the transcription start site and is required for accurate initiation of RNA transcription (Zeller *et al.*, 1984; Skuzeski *et al.*, 1984; Ciliberto *et al.*, 1985; Ares *et al.*, 1985; Mattaj *et al.*, 1985; Mattaj, 1986). A second enhancer-like sequence (DSE) (Murphy *et al.*, 1982; Skuzeski *et al.*, 1984; Westin *et al.*, 1984; Ares *et al.*, 1985; Ciliberto *et al.*, 1985; Krol *et al.*, 1985; Mattaj *et al.*, 1985; Mangin *et al.*, 1986) is positioned ~230 bp from the cap site and activates transcription 10- to 20-fold independently both of orientation and, to some extent, of position.

Although we show that the U5 promoter is stronger than the U2 promoter in oocytes, it is composed of similar functional units. The U5 DSE can, however, be separated into at least two distinct domains (DSE 1 and DSE 2), both of which are required for enhancer activity and both of which bind to factors *in vitro*. Furthermore we show that removal of the major part of the sequences between the DSE and the PSE does not affect U5 transcription.

Results

Isolation of snRNA genes from Xenopus laevis and their genomic organization

Using dot blot and Southern blot hybridization analysis we have screened a collection of previously isolated recombinant λ clones (Mattaj and Zeller, 1983) with ^{32}P-end-labelled U5 snRNA as probe. Two clones, λ 5 and λ 13, gave strong hybridization signals (data not shown) and were injected into oocytes. Transcripts originating from the injected DNA were of the size expected for U5 RNA. They were assembled into snRNPs since they became precipitable by Sm-antibodies (Figure 3) and their synthesis was sensitive to a low concentration of α-amanitin (data not shown). These clones were therefore likely to contain U5 RNA genes and were further analysed. Digestion of λ 5 DNA with several enzymes (*Dra*I, *Fok*I, *Hin*dIII, *Sau*3AI) resulted in a band of roughly 600 bp which hybridized strongly to a U5 RNA probe. Incomplete digestion of λ 5 DNA with *Hin*dIII (Figure 1A, lanes 2-5) gave a ladder of hybridizing bands regularly spaced at intervals of 600 bp, while complete digestion with *Hin*dIII left a strong band of ~600 bp and a weak band of 1.2 kb (Figure 1A, lane 1). Digestion of *Xenopus laevis* genomic DNA with *Sau*3AI, *Hin*dIII, *Dra*I or *Fok*I (Figure 1B, lanes 1-4) always led to the production of a major hybridizing band of roughly 600 bp, although other weakly hybridizing bands of different mol. wts could be detected. Taken together, these results indicate that *X. laevis* U5 RNA genes, similarly to U1 RNA genes (Zeller *et al.*, 1984; Lund *et al.*, 1984), are found in several genomic arrangements, a major species clustered in tandem repeats of 600 bp and other either minor repeated species or single copy genes. Quantitative Southern analysis using defined amounts of cloned U5 DNA for calibration revealed that U5 RNA genes are present at roughly 100 copies in the haploid genome (data not shown) and therefore are less abundant than U1 or U2 RNA genes which are repeated 500-1000 times per haploid *X. laevis*

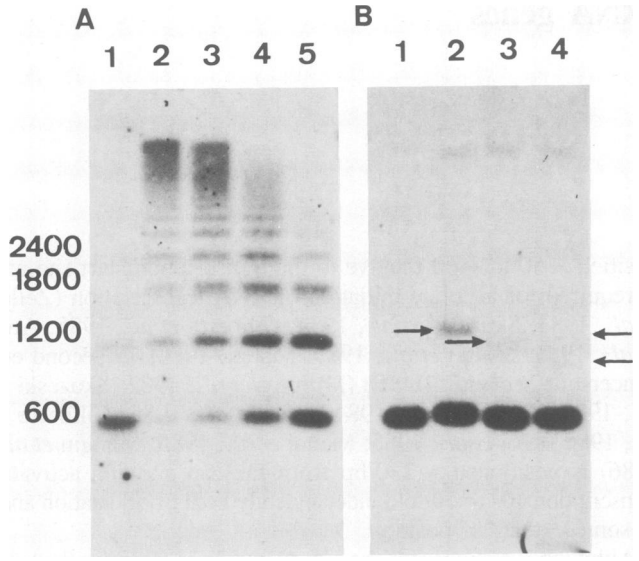


Fig. 1. Analysis of U5 snRNA gene arrangements. *X. laevis* genomic DNA or recombinant λ DNA was digested with different enzymes, separated on an agarose gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled U5 RNA (1A) or nick-translated U5 DNA (1B). The mol. wt of the indicated bands was calculated by comparison with *Hin*dIII digested λ DNA or *Hae*III digested φX174 DNA size markers (not shown). **A.** 2 μg of cloned λ U5 DNA was digested with increasing amounts of *Hin*dIII. **Lane 1:** complete digestion (3 U/μg for 2 h); **lane 2:** 0.1 U/μg; **lane 3:** 0.2 U/μg; **lane 4:** 0.4 U/μg; **lane 5:** 0.8 U/μg. Digests in **lanes 2–5** were carried out for 1 h. **B.** 10 μg of *X. laevis* genomic DNA was digested to completion (3 U/μg for 2 h) with different enzymes: **lane 1:** *Sau*3AI; **lane 2:** *Hin*dIII; **lane 3:** *Dra*I; **lane 4:** *Fok*I. Weakly hybridizing bands that represent minor repeats are indicated by arrows.

X.l.U511H

```
-400
AAGCTTGGCT TATCCAGCAG CCAGGGCATG ACAGAGCCGG AAAGGTGTGC
-350
TTGGCTGCGT GTCTGCTTGC CGGCTGCAAG CAGCAGCAGC AGCAGCAGGA
-300
GGGGCAGGCA AAAGGGCTCT CGCAGCCGGC CCGTGCAAGT CTCGCCTCTA
-250          DSE 1                                  DSE2
GATGCTTAGC TCCATTTGCA TACCCATGCT GCATTAAGAG CGGCTCGGTG
-200
CGCGCCGGTA GCTGTTCTCT TGGACTCCTC TCCTTTCCGC TCCTTTTAAA
-150
CTCGGTCTCG GAGGGAGCTT GTTCAGTTGT CGGCATGATC CTGTCAAAGG
-100                                            PSE
AACTTTGTTT TGGGACATCC AGATAGGCCA CCTTTTGCCT CTCAGCGAGT
-50
GCACGCCACC CTGCTTCGCA GGTCGCCTTT CACCTGGCAG CATTGACGGC
G
ATACTCTGGT TTCTCTTCAA ATTCGAATAA ATCTTTCGCC TTTTACTAAA

GATTTCCGTG GAGAGGAACG ACCATGAGTT TCGTTCAATT TTTTGAAGCC
                                                  +34
TGGTTCACAC CAGGTACTCA TAACAGTTGC AAAGACAGAT GACAAAATGG
                                  +67
CAGCGAGAGA GGTGCTGAAC TAGTTCGGGC AAA
```

Fig. 2. DNA sequence of the RNA-like strand of the 583 bp *Hin*dIII–*Hin*dIII fragment in subclone *X.l.*U511H. The U5 RNA coding sequence deduced from comparison with other U5 RNA sequences (Branlant *et al.*, 1983) and S1 mapping experiments (Figure 3) is underlined. The 5′ end and the major 3′ end are indicated by arrowheads. Brackets indicate short stretches of sequence homology in the non-coding region with other known U RNA gene flanking sequences and their significance is discussed in the text.

genome (Mattaj and Zeller, 1983; Zeller *et al.*, 1984; Lund *et al.*, 1984).

After digestion of λ 5 DNA with *Hin*dIII or *Sau*3AI, 600 bp fragments were inserted into the sequencing vector M13 mp8

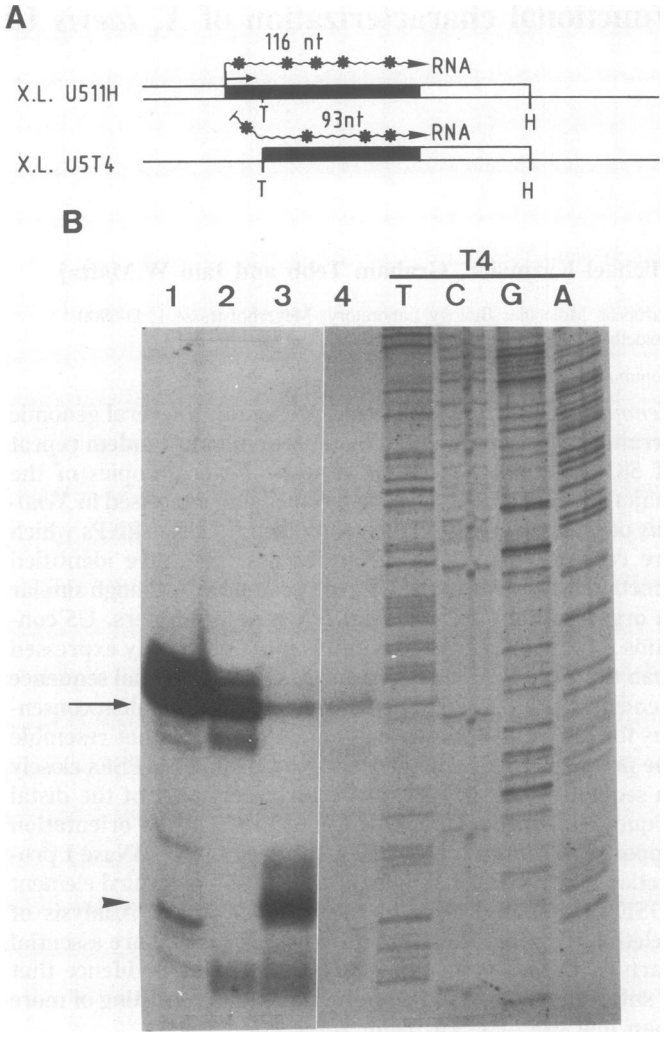


Fig. 3. Determination of the 3′ and 5′ ends of the U5 RNA coding region. **A.** Diagram of the experimental strategy: ssDNA of the clone *X.l.*U511H containing the entire U5 RNA coding region of 116 bp with 5′ and 3′ flanking sequences, or the clone *X.l.*U5T4 containing 93 nucleotides of coding sequence from the gene internal *Taq*I site to the 3′ end and 3′ flanking DNA, were hybridized to *X.l.*U511H transcripts purified by immunoprecipitation (see Materials and methods). The hybridization products were then treated with nuclease S1, denatured, fractionated by gel electrophoresis and autoradiographed. Closed box: coding region. Open boxes: flanking regions. Single line: vector sequence. **B.** Autoradiogram of the protected probes after polyacrylamide gel electrophoresis. 5000 c.p.m. of labelled *X.l.*U511H RNA were used in each assay. In **lane 1** U5 RNA was loaded without any treatment as control for full-length protection. **Lanes 2** and **3** show RNA fragments resistant to nuclease S1 after hybridization with 1 μg ssDNA of the clone *X.l.*U511H **(lane 2)** or *X.l.*U5T4 **(lane 3)**. As a control labelled *X.l.*U511H RNA was treated with nuclease S1 without hybridizing to ssDNA **(lane 4)**. Arrowheads point to the major protected bands, whose length was determined using the T, C, G and A sequencing reactions of *X.l.*U5T4 ssDNA as size markers. The bands corresponding to protected full-length RNA in **lanes 3** and **4** are due to incomplete S1 digestion.

(Messing and Vieira, 1982) and screened with U5 snRNA. All the hybridizing clones that were injected into oocytes were transcriptionally active. We therefore conclude that the 600 bp fragment generated either by *Hin*dIII or *Sau*3AI digestion is a complete U5 RNA transcription unit.

Sequence analysis of U5 RNA genes

The sequence of twelve subclones, each containing a transcriptionally active *Hin*dIII or *Sau*3AI 600 bp fragment, was deter-

mined. The *Taq*I fragment isolated from a *Hin*dIII dimer and cloned into the vector M13 mp8 was used to confirm overlapping sequences. Where necessary, the sequences obtained by the method of Sanger *et al.* (1977) were checked using the method of Maxam and Gilbert (1980). However, we are still unable to tell whether the nucleotide at position −42 is a guanosine or a cytosine (Figure 2), probably due to the formation of a very strong secondary structure in the ssDNA around that position. Figure 2 represents the RNA-like strand of the single *Hin*dIII fragment of the subclone *X.l.*U511H which is 583 bp in length. All of the sequenced subclones were identical except for minor polymorphism in a $(GCA)_n$ motif at position −320 that was repeated six times in most of the subclones, but only three times in one isolate. It is unclear whether or not this represents a cloning artefact. The sequences coding for U5 RNA (underlined in Figure 2) were identified by comparison with the rat U5A RNA (Jacob *et al.*, 1984) and by S1-mapping (Figure 3). For the S1 experiments, we labelled *X.l.*U511H transcripts with [α-^{32}P]GTP by microinjection of *X.l.*U511H DNA into oocytes and immunoprecipitated them with anti-Sm antibodies (Figure 3B, lane 1). This RNA was hybridized to ssDNA containing either the entire coding region or a truncated gene lacking the sequences upstream from the gene-internal *Taq*I site (Figure 3A), prior to digestion by nuclease S1. Using the full-length clone, the protected RNA appears in a prominent band of 116 bp and a series of bands extending up to 120 bp, corresponding exactly to the input RNA (Figure 3B, lanes 1 and 2). A smear of RNA molecules, mostly 93−96 nucleotides in length but extending up to over 100 nucleotides, is protected using the 3′ end probe (Figure 3B, lane 3). The similarity of the size distribution observed when the two different probes are used makes it likely that the multiple U5 transcripts arise from initiation at a defined nucleotide with the production of multiple 3′ ends (see also Figure 4). Comparison with other known sequenced U5 RNAs shows that the sequence of the 5′ stem loop (nucleotides 1−72) is highly conserved in several vertebrates (Branlant *et al.*, 1983), whereas in the 3′ half of the RNA only the AATTTTTTGA Sm-binding site (Mattaj and De Robertis, 1985) is completely conserved. In spite of the lack of conservation of primary sequence the secondary structure predicted for the 3′ half of the molecule is conserved (D.A.M. Konings, unpublished data).

We next analysed the 5′ and 3′ flanking regions for the existence of blocks with sequence homology to other U snRNA genes, some of which have been proven to be functionally important (see Introduction for references). Three short blocks of sequence homology were detected in the 5′ flanking region of the U5 RNA genes. One conserved sequence is located at the distance from the cap site (−50 to −60) expected for the PSE. Strikingly this box in the U5 promoter shows good homology to the consensus sequence for the mammalian PSE (Skuzeski *et al.*, 1984) and only weak homology to the *X. laevis* PSE consensus derived from *Xenopus* U1 and U2 genes (Zeller *et al.*, 1984). Further upstream, at position −240, the octamer motif that is also common to U snRNA genes (Skuzeski *et al.*, 1984; Mattaj *et al.*, 1985; Ciliberto *et al.*, 1985; Ares *et al.*, 1985; Krol *et al.*, 1985; Mangin *et al.*, 1986) is found. Interestingly, this element, which has been shown to modulate promoter activity in an orientation independent manner (Mattaj *et al.*, 1985; Ares *et al.*, 1985; Mangin *et al.*, 1986) is present naturally in the U5 and U4 (Hoffmann *et al.*, 1986) RNA gene promoters in the opposite orientation to that in U1 or U2 RNA. This is illustrated in Figure 5B. At position +11 downstream from the major 3′ end of the U5 RNA is a short block of 14 nucleotides — GTTGCAAAGACAGA — similar to the sequence that has been shown to be involved in the correct 3′ end formation of U1 and U2 snRNAs (Hernandez, 1985; Yuo *et al.*, 1985; Hernandez and Weiner, 1986; Neuman de Vegvar *et al.*, 1986; Ciliberto *et al.*, 1986).

Functional characterization of the U5 promoter

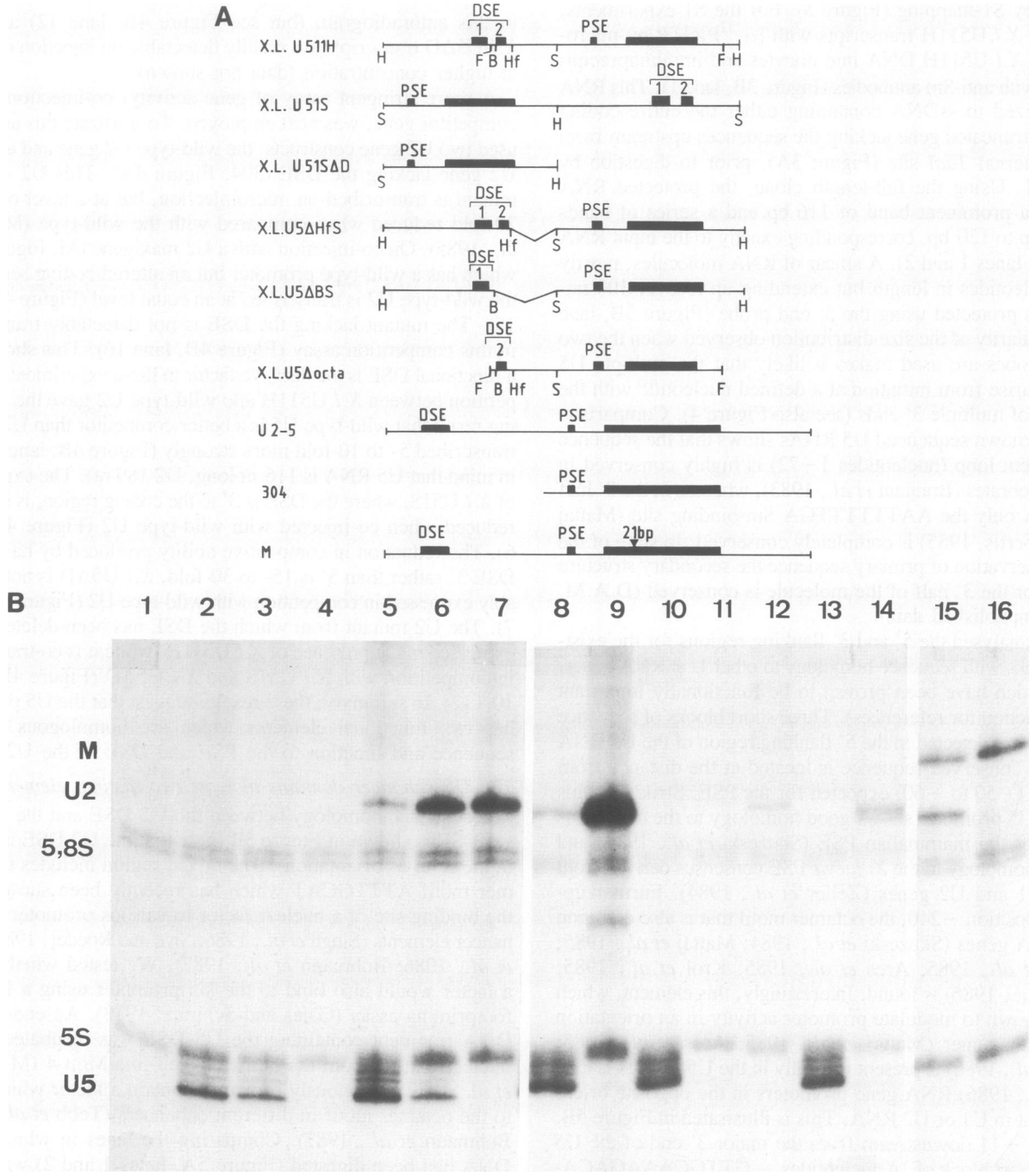
To test whether the DSE and PSE, defined by sequence comparison, were functional domains of the U5 promoter, we performed oocyte injection experiments. Three subclones, two having the DSE either 5′ or 3′ with respect to the coding region (*X.l.*U511H and *X.l.*U51S) and a third in which the DSE had been deleted (*X.l.*U5ΔD), as illustrated in Figure 4A, were injected singly. *X.l.*U511H and *X.l.*U51S are both transcribed, and the effect of moving the DSE from 5′ of the gene to a 3′ position is slight, causing a roughly 2-fold drop in transcription (Figure 4B, lanes 2 and 3). Deletion of the DSE in *X.l.*U5ΔD causes a further drop in transcriptional activity (Figure 4B, lane 4) to a level not visible in this autoradiogram (but see Figure 4B, lane 12) although *X.l.*U5ΔD transcripts are readily detectable on injection of DNA at higher concentration (data not shown).

A more stringent assay of gene activity, co-injection with a competitor gene, was next employed. To illustrate this assay we used two U2 gene constructs, the wild-type U2 gene and a mutant U2 gene lacking the DSE (304, Figure 4A). This U2 deletion mutant is transcribed on microinjection, but at a level of 10- to 20-fold reduced when compared with the wild-type (Mattaj *et al.*, 1985). On co-injection with a U2 maxigene (M, Figure 4A), which has a wild-type promoter but an altered coding sequence, the wild-type U2 is transcribed at an equal level (Figure 4B, lane 15). The mutant lacking the DSE is not detectably transcribed in this competition assay (Figure 4B, lane 16). This shows that a functional DSE is the decisive factor in these experiments. Competition between *X.l.*U511H and wild-type U2 gave the surprising result that wild-type U5 is a better competitor than U2, being transcribed 5- to 10-fold more strongly (Figure 4B, lane 5; bear in mind that U5 RNA is 116 nt long, U2 189 nt). The expression of *X.l.*U51S, where the DSE is 3′ to the coding region, is severely reduced when co-injected with wild-type U2 (Figure 4B, lane 6). The reduction in competitive ability produced by having the DSE 3′ rather than 5′ is 15- to 30-fold. *X.l.*U5ΔD is not detectably expressed in competition with wild-type U2 (Figure 4B, lane 7). The U2 mutant from which the DSE has been deleted is not expressed in the presence of *X.l.*U511H, while it is co-transcribed in competition with *X.l.*U51S and *X.l.*U5ΔD (Figure 4B, lanes 10−12). In summary, these results suggest that the U5 promoter has two functional elements which are homologous both in sequence and function to the PSE and DSE of the U2 genes.

The U5 enhancer contains at least two distinct elements

The sequence homology between the U2 DSE and the putative U5 DSE is shown in Figure 5B (note that the U2 DSE is shown in the 3′ to 5′ orientation). The boxed region includes the octamer motif ATTTGCAT which has recently been shown to be the binding site of a nuclear factor to various promoter and enhancer elements (Singh *et al.*, 1986; Sive and Roeder, 1986; Tebb *et al.*, 1986; Bohmann *et al.*, 1987). We tested whether such a factor would also bind to the U5 promoter using a DNase I footprinting assay (Galas and Schmitz, 1978). An end-labelled DNA fragment containing the U5 DSE was incubated with a nuclear extract from the human T cell-line Molt-4 (Minowada *et al.*, 1972), previously shown to contain a factor which binds to the octamer motif in different enhancers (Tebb *et al.*, 1986; Bohmann *et al.*, 1987). Comparing the lanes in which naked DNA had been digested (Figure 5A, lanes 1 and 2) with those

where DNA had been incubated with protein extract prior to digestion (Figure 5A, lanes 3 and 4), regions of protection are observed on both strands. One region (D1) is seen to be centred around the octamer motif. A schematic drawing of the protected region is given in Figure 5B. At the border of the footprint, there are DNase I hypersensitive sites (arrows in Figure 5A and B). The DNase I protection pattern of the U2 DSE (Bohmann *et al.*, 1987; G.Tebb, D.Bohmann and I.W.Mattaj, in preparation) is shown for comparison (Figure 5B). This, together with the *in vivo* competition data, suggests that a common factor binds to both the U2 and U5 octamer domains. A second short stretch of protected nucleotides (D2, Figure 5A and B) is detected ~25 nucleotides downstream of the octamer motif in the U5 promoter. To test whether these sites of protein–DNA interaction, which we have called DSE 1 and 2, are important for U5 promoter function we constructed three deletion mutants. One, which retains DSE 1 and 2, extends from the *Sau*3AI site at −114 to the *Hin*fI site at −178 (*X.l.*U5ΔHfS; Figure 4A). Another extends 23 bp further from the *Sau*3AI site to the *Bss*h2 site at −200. In this mutant (*X.l.*U5ΔBS; Figure 4A), DSE 2 is deleted while the octamer motif is retained. In the third mutant DSE 1 has been deleted and DSE 2 retained (*X.l.*U5Δocta; Figure 4A). The transcriptional activity of these mutants was tested by microinjection (Figure 4B and C). *X.l.*U5ΔHfS behaves exactly like wild-type U5 when injected alone (Figure 4C, compare lanes 1 and 2), or in competition with either wild-type U2 (Figure 4B, lane 8) or with a U2 mutant retaining only the PSE (Figure 4B, lane 13). This shows that the region between −114 and −178 is unimportant for transcription in the oocyte, and further that alteration of the spacing between the DSE and the PSE has no apparent



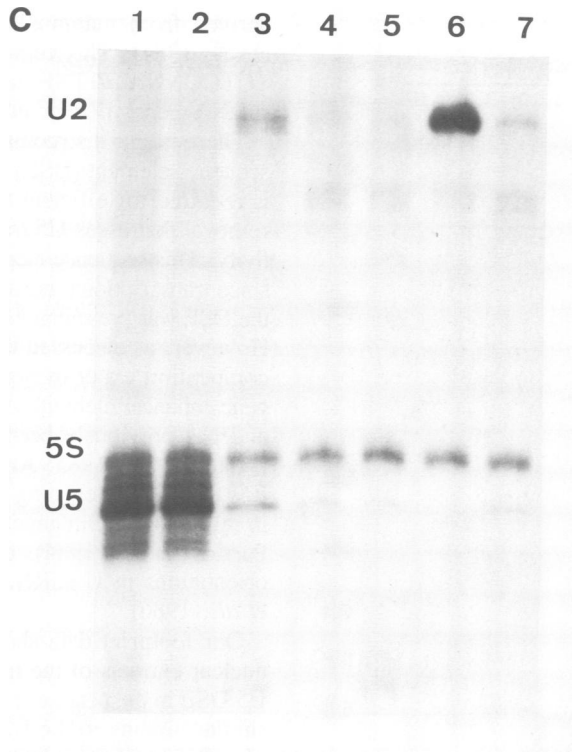


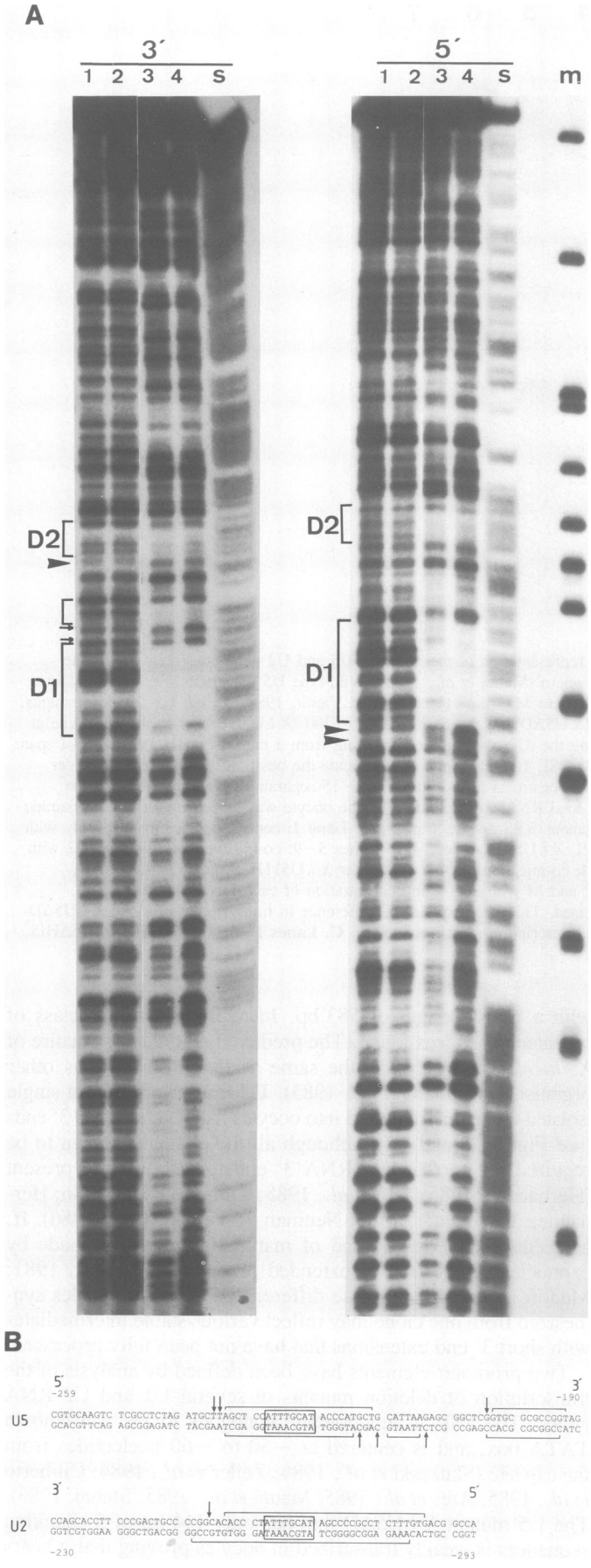
Fig. 4. Transcriptional activity of U5 gene constructs in oocytes. **A.** Diagrammatic representation of the different U5 and U2 gene constructs used. The construction of the wild-type U2 gene and the mutants 304 and M is described elsewhere (Mattaj *et al.*, 1985). Wild-type U5 constructs used were obtained by cloning a single *Hin*dIII repeat (*X.l.*U511H) or a single *Sau*3AI repeat (*X.l.*U51S) into M13 mp8 (Messing and Vieira, 1982). In the U5 deletion mutants, 5′ flanking sequences from the *Sau*3AI site at −114 to the *Hin*dIII site at −400 (*X.l.*U5ΔD), to the *Bss*h2 site at −200 (*X.l.*U5ΔBS) or to the *Hin*fI site at −178 (*X.l.*U5ΔHfS) were removed. The mutant U5Δocta was constructed by cloning the 372 bp fragment (resulting from a partial *Fnu*4H I digest) that spans the U5 transcription unit from position −224 to +35 (see Figure 2) into M13 mp8. DSE 1, DSE 2 and PSE indicate the position of functional promoter elements that are described in the text. **B.** Transcriptional activity of the different U5 constructs in the oocyte. 10−15 separate oocytes injected into the nucleus with each sample were pooled (see Materials and methods). The amount of total RNA corresponding to one oocyte was loaded on an 8% acrylamide−7 M urea gel. U5, U2 and maxigene transcripts are indicated, as well as the endogenous 5.8S and 5S transcripts. **Lane 1:** control oocyte, injected only with [α-^{32}P]GTP. **Lanes 2−4:** oocytes injected with the U5 clones *X.l.*U511H, *X.l.*U51S, *X.l.*U5ΔD respectively. **Lanes 5−9:** co-injection of wild-type U2 with either *X.l.*U511H, *X.l.*U51S, *X.l.*U5ΔD, *X.l.*U5ΔHfS or *X.l.*U5ΔBS. **Lanes 10−14:** co-injection of 304 with either *X.l.*U511H, *X.l.*U51S, *X.l.*U5ΔD, *X.l.*U5ΔHfS or *X.l.*U5ΔBS. **Lanes 15** and **16:** co-injection of M with wild-type U2 and M with 304. The concentration of each DNA was 200 μg/ml, therefore in co-injection experiments twice the amount of DNA per oocyte was injected. This may explain the difference in transcription of clone *X.l.*U5ΔD in **lanes 4** and **12**. Injection of *X.l.*U5ΔD alone at 400 μg/ml resulted in detectable transcription (data not shown). **C. Lanes 1−5:** *X.l.*U511H, *X.l.*U5ΔHfS, *X.l.*U5ΔBS, *X.l.*U5Δocta and *X.l.*U2-5. **Lane 7:** *X.l.*U5Δocta co-injected with 304.

effect on promoter strength. In contrast to this, *X.l.*U5ΔBS, from which DSE 2 has been deleted, is transcribed much less efficiently when injected alone (Figure 4C, lane 3). In competition with wild-type U2 (Figure 4B, lane 9) this construct is not detectably transcribed but is equally transcribed in competition with 304, the U2 mutant which retains only the PSE (Figure 4B, lane 14). This result is reminiscent of the results of Ares *et al.* (1985) and Ciliberto *et al.* (1987) who demonstrated that short deletions in the octamer region completely inactivated the human U2 and *Xenopus* U1 enhancers respectively. Deletion of the octamer-containing DSE 1 completely abolishes the enhancing effect of the U5 DSE whether tested alone (Figure 4C, compare lanes 4 and 5) or in competition with wild-type U2 (Figure 4C, lane 6) or the DSE-less U2 mutant (Figure 4C, lane 7). These results show that the U5 enhancer consists of at least two non-redundant elements, both of which are required for activity. While deletion of the octamer motif completely inactivates the DSE, removal of DSE 2 may leave a small amount of activity (Figure 4C, compare lanes 3 and 4).

Discussion

We describe here the isolation and characterization of U5 RNA genes. Most of the roughly 100 U5 RNA genes present in the haploid *X. laevis* genome are arranged as tandemly repeated units with a repeat length of 583 bp. In addition a minor class of repeated genes was found. The predicted secondary structure of *X. laevis* U5 snRNA is the same as that from various other organisms (Branlant *et al.*, 1983). U5 transcripts from a single isolated clone microinjected into oocytes had heterologous 3′ ends (see Figures 3 and 4B), although all the elements shown to be required for correct U snRNA 3′ end formation were present (Hernandez, 1985; Yuo *et al.*, 1985; Ciliberto *et al.*, 1986; Hernandez and Weiner, 1986; Neuman de Vegvar *et al.*, 1986). If, as seems likely, the 3′ end of mature U snRNA is made by exonuclease digestion of extended precursors (Elicieri, 1981; Madore *et al.*, 1984), these different U5 RNA molecules synthesized from one clone may reflect various stable intermediates with short 3′ end extensions that have not been fully processed.

Two promoter elements have been defined by analysis of the transcription of deletion mutants of several U1 and U2 RNA genes. The PSE is functionally analogous to the well-known TATA box, and is centered at −50 to −60 nucleotides from the cap site (Skuzeski *et al.*, 1984; Zeller *et al.*, 1984; Ciliberto *et al.*, 1985; Ares *et al.*, 1985; Mattaj *et al.*, 1985; Mattaj, 1986). The U5 mutant *X.l.*U5ΔD that has only 114 bp of 5′ flanking sequences is weakly transcribed in oocytes proving that it bears sufficient information for basic transcription. There is a block of eleven nucleotides in this region very similar to the consensus



derived from mammalian U1 and U2 RNA genes (Skuzeski *et al.*, 1984). This region contains only a poor match to the TCTCCNNATG PSE motif previously found in all *Xenopus* U snRNA genes (Zeller *et al.*, 1984).

There is also a structural and functional analogue to the distal sequence element (DSE) in the *X. laevis* U5 promoter. The DSE is required for efficient transcription of U5 RNA genes and its removal depresses U5 transcription to 5−10% of wild-type activity. This sequence element is somewhat position independent, since transcriptional enhancement was only mildly affected when the DSE was present at its *in vivo* position 3′ of the coding region. However, as discussed later, the enhancer was less effective in competition experiments when 3′ of the gene. The U snRNA gene enhancer element contains the ATGCAAAT octamer motif, and has previously been shown to be orientation independent (Mattaj *et al.*, 1985; Ares *et al.*, 1985; Mangin *et al.*, 1986). It is therefore interesting that, as is the case for the octamer motif in immunoglobulin gene promoters (Falkner and Zachau, 1984; Parslow *et al.*, 1984), this element is found naturally in both orientations in U snRNA gene promoters (see also Hoffmann *et al.*, 1986).

Our footprint data show the existence of a factor present in nuclear extracts of the human Molt-4 cell line that binds to the U5 DSE in the octamer region. We have previously demonstrated similar binding to the U2 DSE (Tebb *et al.*, 1986; Bohmann *et al.*, 1987) and the U1 DSE (Ciliberto *et al.*, 1987). The pattern of protection around the octamer homology is quite similar although the surrounding sequences are divergent. This, along with the transcription competition experiments, suggests that the octamer consensus is the target site for the same transcription factor in the U5, U2 and U1 DSEs. In the U2 DSE protection extends over the sequence GGGCGG, a putative binding site for the transcription factor Sp1 (Gidoni *et al.*, 1984; Dynan and Tjian, 1983; see Mangin *et al.*, 1986). Binding of Sp1 to a number of different promoters increases their transcription *in vitro* (Kadonaga *et al.*, 1986). The *X. laevis* U5 promoter does not contain a consensus Sp1 binding site, but functions at least as well as the U2 promoter in transcriptional enhancement. The U1 and U2 footprints extend over the Sp1 sites, found 3′ or 5′ of the octamer respectively in these genes (Bohmann *et al.*, 1987; Ciliberto *et al.*, 1987). It may therefore be that the U1 and U2 DSEs are also bipartite, being composed of binding sites for both Sp1 and the octamer binding protein(s).

To our surprise U5 genes can compete 5- to 10-fold better than

Fig. 5. DNase I protection of the U5 DSE by unfractionated Molt-4 nuclear extract. **A.** DNase I digestion products of naked DNA **(lanes 1** and **2)** or of DNA incubated with nuclear extract **(lanes 3** and **4)** were run along with an A+G specific sequencing **lane S** and an end-labelled *Hpa*II digest of pBR322 **(lane m)**. The 423 nt *Hin*dIII−*Taq*I fragment (Figure 2A) that contains the 5′ flanking region of the U5 gene was end-labelled at the *Hin*dIII site on both strands (see Materials and methods). Regions of protection discussed in the text are indicated by brackets; arrowheads are employed to show increased sensitivity to DNase I in the presence of extract **(lanes 3** and **4)** compared with the control digestions **(lanes 1** and **2)**. **B.** Comparison of the DNase I protection patterns around the octamer motif homologies in the U5 and the U2 DSE. A schematic representation of the footprinting experiment performed on the U5 DSE is compared with the footprint on the U2 DSE (Bohmann *et al.*, 1987). The U2 DNA fragment is aligned 3′ to 5′, in order to have the same orientation with respect to the octamer sequence element as the U5 DSE element. The octamer sequences in both DSEs are boxed and positioned above each other. The brackets indicate protected areas on both coding and non-coding strands. The arrowheads point to DNase I hypersensitive sites.

U2 genes for a limiting transcription factor in *X. laevis* oocytes, whose binding requires the presence of the DSE (Mattaj *et al.*, 1985). This effect is reversed when the DSE is 3′ of the U5 RNA gene. This result can be explained by arguing that the linear arrangement DSE−PSE over a short distance in the promoter favours interaction between bound transcription factors. Disruption of this alignment by moving the DSE 3′ of the gene would not necessarily reduce the affinity of factors for the DNA binding sites, but would weaken the proposed protein−protein interactions. This, however, does not explain the transcriptional dominance of U5 genes over U2 genes. Our attention was therefore directed to a second protected region in the U5 promoter centred around position −200. A mutant was constructed in which this region was deleted without removal of the octamer sequence. This mutant had no transcriptional enhancement activity, demonstrating that the region around −200 is an essential part of the U5 enhancer. We have named this region DSE 2 since it is found close to the octamer motif (DSE 1), and, although it is an essential part of the U5 enhancer, it is not yet clear if a similar region is present in the DSEs of U2 and U1. It is somewhat surprising that removal of DSE 2 almost completely destroys enhancer activity since in this mutant the octamer motif remains intact. Short, octamer containing nucleotides have been shown to exhibit partial DSE activity when attached to U2 mutants from which the DSE has been removed (Mattaj *et al.*, 1985). In addition, footprinting experiments show that these oligonucleotide constructs bind nuclear factors in a manner similar to the wild-type U2 DSE (G.Tebb, D.Bohmann and I.W.Mattaj, in preparation). These results may reflect differences either in the structures of the U5 and U2 DSEs or in the interactions between factors binding to the DSEs and PSEs in the two cases.

There is homology between DSE 2 and sequences found roughly 140 bp upstream of the site of transcription initiation in several U snRNA genes (Mattaj and Zeller, 1983; Murphy *et al.*, 1987; G.Tebb, D.Bohmann and I.W.Mattaj, in preparation) including U5 (Figure 2, compare positions −129 to −137 with −196 to −205). This homology is, however, at least in the case of the *Xenopus* U2 and U5 genes, without apparent functional significance, since site-directed mutagenesis of the −140 region of the U2 promoter has no effect on U2 transcription in the oocyte (G.Tebb, D.Bohmann and I.W.Mattaj, in preparation) and deletion of all the sequences between −114 and −178 has no effect on U5 transcription (Figure 4B). It is interesting to speculate on what causes the observed transcriptional dominance of U5 over U2. Two possibilities suggest themselves: either the combination of factors binding to the DSE 2 and the octamer motif in U5 results in a more active enhancer than the combination of octamer and (proposed) Sp1 factor binding in U2, or the fact that the elements are more widely spaced in U5 results in a greater activity. We are currently testing these possibilities.

Materials and methods

Hybridization using DNA and RNA probes

Transfer of DNA to nitrocellulose filters and hybridization with nick-translated probes was as described by Maniatis *et al.* (1982). RNA probes were prepared by polyadenylating gel purified U5 snRNA with *Escherichia coli* poly A polymerase in the presence of [α-^{32}P]ATP, and used as described previously (Mattaj and Zeller, 1983).

Restriction digestions, ligations

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

DNA sequencing

DNA was cloned into M13 mp8 vectors in both orientations and sequenced either using the chain termination method of Sanger *et al.* (1977) or using the method of Maxam and Gilbert (1980) to obtain the sequence of both strands.

Microinjection of cloned DNA

Purified DNA was microinjected into *X. laevis* oocytes together with [α-^{32}P]GTP (Nishikura *et al.*, 1982). Twenty-four hours later RNAs were extracted from the oocytes and analysed on polyacrylamide gels (De Robertis *et al.*, 1982). The concentration of microinjected DNA was 200−350 μg/ml, the volume microinjected was 30−50 nl.

Immunoprecipitation of U snRNAs

Extracts from injected oocytes were immunoprecipitated with human Sm antiserum (Mattaj and De Robertis, 1985) and the precipitated ^{32}P-labelled snRNAs were extracted and analysed on polyacrylamide gels as described by De Robertis *et al.* (1982). The Sm antiserum used has been described previously (Fritz *et al.*, 1984).

S1 mapping

*X.l.*U511H transcripts were labelled by injecting *X.l.*U511H DNA together with [α-^{32}P]GTP into oocytes and then purified by immunoprecipitation as described above. 5000 c.p.m. of labelled RNA were hybridized with 1 μg of single-stranded DNA in the presence of 10 μg of yeast carrier RNA at 49°C for 3 h in 30 μl of hybridization buffer (Berk and Sharp, 1977). Subsequently the hybridization products were digested with 2000 U of S1 nuclease (Anglian Biological Laboratories) in 300 μl of S1 digestion buffer (Berk and Sharp, 1977). The protected fragments were phenol extracted, ethanol precipitated, resuspended in 95% formamide, denatured and analysed on an 8% denaturing acrylamide gel.

Footprinting

Essentially the procedure of Bohmann *et al.* (1987) was followed. Nuclear extracts used in the experiment were prepared according to Dignam *et al.* (1983) with the modifications introduced by Wildeman *et al.* (1984). 4 μl of nuclear extract (= 33 μg protein) were pre-incubated with 25 ng of linearized pUC plasmid DNA on ice for 15 min in incubation buffer as described by Bohmann *et al.* (1987). 20 000 c.p.m. of DNA end-labelled by means of Klenow DNA polymerase or using T4 polynucleotide kinase were added in 2 μl water and incubated for 10 min at 20°C. 2 μl of freshly diluted DNase I (50 or 100 μg/ml, Worthington DPFF) was added when naked DNA was digested, concentrations of 400 or 800 μg/ml were used where DNA was previously incubated with protein extract. DNase digestion was allowed to proceed for 90 s at 20°C. The reaction mixture was extracted with phenol/chloroform and with chloroform/isoamylalcohol and dialysed on Millipore filters against water for 90 min. The samples were lyophilized, taken up in 2.5 μl of formamide loading buffer and loaded onto a 6% polyacrylamide−7 M urea sequencing gel together with a G + A sequencing reaction (Maxam and Gilbert, 1980) to provide accurate size standards.

Acknowledgements

We should like to thank Nina Dathan for expert technical assistance, Petra Riedinger for help with the figures and Riccardo Cortese, Henk Stunnenberg, Jörg Hamm and Huw Parry for critical reading of the manuscript. M.K. was the recipient of Forschungsstipendium Ka 711/1-1 from the Deutsche Forschungsgemeinschaft for part of this work.

References

Ares,M., Jr, Mangin,M. and Weiner,A.M. (1985) *Mol. Cell. Biol.*, **5**, 1560−1570.

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721−732.

Bohmann,D., Keller,W., Dale,T., Schöler,H.R., Tebb,G. and Mattaj,I.W. (1987) *Nature*, **325**, 268−272.

Branlant,C., Krol,A., Lazar,E., Haendler,B., Jacob,M., Galego-Dias,L. and Pousada,C. (1983) *Nucleic Acids Res.*, **11**, 8359−8367.

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

Ciliberto,G., Buckland,R., Cortese,R. and Philipson,L. (1985) *EMBO J.*, **4**, 1537−1543.

Ciliberto,G., Dathan,N., Frank,R., Philipson,L. and Mattaj,I.W. (1986) *EMBO J.*, **5**, 2931−2937.

Ciliberto,G., Palla,F., Tebb,G., Mattaj,I.W. and Philipson,L. (1987) *Nucleic Acids Res.*, **15**, 2403−2416.

De Robertis,E.M., Lienhard,S. and Parisot,R.F. (1982) *Nature*, **295**, 572−577.

Dignam,J.D., Lebowitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475−1489.

Elicieri,G.L. (1981) In Busch,H. (ed.), *The Cell Nucleus*. Academic Press, NY, Vol. 8, pp. 307−330.

Falkner,F.G. and Zachau,H.G. (1984) *Nature*, **310**, 71–74.
Fritz,A., Parisot,R.F., Newmeyer,D. and De Robertis,E.M. (1984) *J. Mol. Biol.*, **178**, 273–285.
Galas,D.J. and Schmitz,A. (1978) *Nucleic Acids Res.*, **5**, 3157–3170.
Hernandez,N. (1985) *EMBO J.*, **4**, 1827–1837.
Hernandez,N. and Weiner,A.M. (1986) *Cell*, **47**, 249–258.
Hoffmann,M., Korf,G., McNamara,K.J. and Stumph,W.E. (1986) *Mol. Cell. Biol.*, **6**, 3910–3919.
Jacob,M., Lazar,E., Haendler,B., Gallinaro,H., Krol,A. and Branlant,C. (1984) *Biol. Cell.*, **51**, 1–9.
Kadonaga,J.T., Jones,K.A. and Tjian,R. (1986) *Trends Biochem. Sci.*, **11**, 20–23.
Krol,A., Lund,E. and Dahlberg,J.E. (1985) *EMBO J.*, **5**, 1529–1535.
Krol,A., Carbon,P., Ebel,J.-P. and Appel,B. (1987) *Nucleic Acids Res.*, **15**, 2463–2478.
Kunkel,G.R., Maser,R.L., Calvet,J.P. and Pederson,T. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8575–8579.
Lerner,M.R. and Steitz,J.A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5495–5499.
Lund,E., Dahlberg,J.E. and Forbes,D.J. (1984) *Mol. Cell. Biol.*, **4**, 2580–2586.
Madore,S.L., Wieben,D.E. and Pederson,T. (1984) *J. Cell. Biol.*, **98**, 188–192.
Mangin,M., Ares,M. Jr and Weiner,A.M. (1986) *EMBO J.*, **5**, 987–995.
Maniatis,T. and Reed,R. (1987) *Nature*, **325**, 673–678.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
Mattaj,I.W. (1986) *Cell*, **46**, 905–911.
Mattaj,I.W. and De Robertis,E.M. (1985) *Cell*, **40**, 111–118.
Mattaj,I.W. and Zeller,R. (1983) *EMBO J.*, **2**, 1883–1891.
Mattaj,I.W., Lienhard,S., Jiricny,J. and De Robertis,E.M. (1985) *Nature*, **316**, 163–167.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.
Minowada,J., Ohnura,T. and Moore,G.E. (1972) *J. Natl. Cancer Inst.*, **49**, 891–895.
Murphy,J.T., Burgess,R.R., Dahlberg,J.E. and Lund,E. (1982) *Cell*, **29**, 265–274.
Murphy,J.T., Skuzesi,J.T., Lund,E., Steinberg,T.H., Burgess,R.R. and Dahlberg, J.E. (1987) *J. Biol. Chem.*, **262**, 1795–1803.
Neuman de Vegvar,H.E., Lund,E. and Dahlberg,J.E. (1986) *Cell*, **47**, 259–266.
Nishikura,K., Kurjan,J., Hall,B.D. and De Robertis,E.M. (1982) *EMBO J.*, **1**, 263–268.
Parslow,T.G., Blair,D.M., Murphy,W.J. and Granner,D.K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2650–2654.
Reddy,R., Henning,D., Das,G., Harless,M. and Wright,D. (1987) *J. Biol. Chem.*, **262**, 75–81.
Sanger,F., Nicklen,S. and Coulsen,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schaufele,F., Gilmartin,G.M., Bannwarth,W. and Birnstiel,M.L. (1986) *Nature*, **323**, 777–781.
Sharp,P.A. (1987) *Science*, **235**, 766–771.
Singh,H., Sen,R., Baltimore,D. and Sharp,P.A. (1986) *Nature*, **319**, 154–158.
Sive,H.L. and Roeder,R.G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6382–6386.
Skuzeski,J.M., Lund,E., Murphy,J.T., Steinberg,T.H., Burgess,R.R. and Dahlberg,J.E. (1984) *J. Biol. Chem.*, **259**, 8345–8353.
Strub,K. and Birnstiel,M.L. (1986) *EMBO J.*, **5**, 1675–1682.
Tebb,G., Mattaj,I.W., Keller,W. and Bohmann,D. (1986) In Reznikoff,W.S. *et al.* (eds), *RNA Polymerase and the Control of Transcription*. Steenbock Symposium, Vol. 16, 421–426.
Westin,G., Lund,E., Murphy,J.T., Pettersson,U. and Dahlberg,J.E. (1984) *EMBO J.*, **3**, 3925–3301.
Wildeman,A.G., Sassone-Corsi,P., Grundström,T., Zenke,M. and Chambon,P. (1984) *EMBO J.*, **3**, 3129–3133.
Yuo,C., Ares,M. Jr and Weiner,A.M. (1985) *Cell*, **42**, 193–202.
Zeller,R., Carri,M.-T., Mattaj,I.W. and De Robertis,E.M. (1984) *EMBO J.*, **3**, 1075–1081.

Received on April 6, 1987; revised on June 3, 1987

Note added in proof

The manuscript referred to as 'G.Tebb, D.Bohmann and I.W.Mattaj, in preparation' is now in press in *Nucleic Acids Research*.