The EMBO Journal vol.5 no.11 pp.2931-2937, 1986

Formation of the 3′ end on U snRNAs requires at least three sequence elements

Gennaro Ciliberto[1], Nina Dathan, Rainer Frank, Lennart Philipson and Iain W.Mattaj

European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, FRG and [1]Istituto di Scienze Biochimiche, II Facoltá di Medicina e Chirurgia, Universitá di Napoli, Italy

Communicated by L.Philipson

The structural requirements for 3′ end formation on the *Xenopus laevis* U1B snRNA gene have been studied. Three sequence elements are shown to be required. The first is a conserved sequence element found immediately 3′ of all vertebrate U snRNA genes studied so far. The second is a gene internal sequence potentially capable of forming a stem-loop structure close to the 3′ end of the RNA. The third element lies upstream of these, and may be part of the gene promoter. Experiments designed to investigate the mechanism of 3′ end formation on primary U1B snRNA transcripts failed to find evidence for a processing event.

Key words: snRNA genes/3′ end/formation

Introduction

A variety of different mechanisms have evolved to ensure accurate and efficient 3′ end formation of RNA polymerase II transcripts. This category includes not only mRNAs, but also transcripts belonging to the class of small nuclear RNAs (U snRNAs) (Busch *et al.*, 1982). U snRNAs are found in the nucleus where they play an important role in the processing of precursor RNAs (Padgett *et al.*, 1983; Krämer *et al.*, 1984; Black *et al.*, 1985; Krainer and Maniatis, 1985; Strut and Birnstiel, 1986). The basic mechanisms involved in 3′ end formation of both polyadenylated and non-polyadenylated mRNAs (the latter class being represented mainly by histone mRNAs) have been clarified during the last few years. Two sequences are absolutely required for histone mRNA 3′ end formation: a potential hairpin-forming sequence located within the mature mRNAs at the 3′ end and a short purine-rich sequence located in the 3′ flanking sequence immediately downstream of the gene (reviewed in Birnstiel *et al.*, 1985). This sequence is conserved in all sea urchin histone genes. Full efficiency of 3′ end formation however requires further sequences on the 3′ side of the second element. Histone gene transcription proceeds beyond the purine-rich element. The mature 3′ end is generated by a processing event which requires the action of the U7 snRNP particle (Birchmeier *et al.*, 1984; Krieg and Melton, 1984; Price and Parker, 1984; Galli *et al.*, 1983; Strub and Birnstiel, 1986). The RNA component of this particle is potentially able to form complementary base pairing with portions of the 3′ terminal stem and loop and of the conserved 3′ flanking purine-rich element (Strub *et al.*, 1984).

3′ end formation of polyadenylated mRNAs also requires a processing step dependent on the presence of a polyadenylation signal, AAUAAA, ~10-30 bp upstream of the polyadenylation site. In this case 3′ end formation also appears to require the presence of natural sequences on the 3′ side of the polyadenylation site (reviewed in Birnstiel *et al.*, 1985).

Small nuclear RNA genes, although belonging to class II genes, have a peculiar promoter structure and, like histone genes, are not polyadenylated. The recent cloning of several members of this group has shown that in addition to a potential stem and loop structure at the 3′ end of the U snRNAs there is a conserved sequence ($GTTTN_{0-3}AAA^{G}_{A}NNAGA$) in the immediate 3′ flanking region (Mattaj and Zeller, 1983; Hernandez, 1985; Yuo *et al.*, 1985). 3′ end formation on U snRNAs is thought to proceed in at least two stages. In the first step a precursor RNA is produced which is extended at its 3′ end by several (up to 15) nucleotides. These extra nucleotides are then removed, probably by an exonuclease activity (Zieve and Penman, 1976; Elicieri, 1981; Madore *et al.*, 1984). It is not clear whether the precursor is the primary product of transcription or is formed by processing of a longer RNA. Both transfection studies in HeLa cells and microinjections in *Xenopus laevis* oocytes have indicated that the conserved box in the 3′ flanking region of U1 and U2 genes is the only element absolutely required for precursor 3′ end formation (Hernandez, 1985; Yuo *et al.*, 1985). Alterations of sequences close to the mature 3′ end, of the distance between the mature 3′ end and the conserved flanking sequence, and of the most 3′ gene-internal stem-loop structure have however been shown to affect the efficiency and accuracy of mature 3′ end formation (Hernandez, 1985; Yuo *et al.*, 1985; Mattaj and De Robertis, 1985).

In this paper we study the 3′ end formation of *Xenopus* U1 and U2 RNAs in the homologous oocyte system. This process is found to require gene internal sequences as well as the conserved 3′ flanking sequence. In addition *in vivo* synthesized RNAs which contain internally all the information required for 3′ end formation are not processed at these internal sites. This argues against the existence of a processing event which can be dissociated from transcription in the biogenesis of the U snRNAs. Furthermore, a U1 terminator cloned in a PolII transcription unit driven by a different promoter, namely the SV40 early promoter, does not seem to direct 3′ end formation.

Results

The conserved 3′ box is necessary but not sufficient for U1 3′ end formation in Xenopus oocytes

The role of the conserved downstream element (in this case GTGCAAAAAGTAGA) in termination of transcription of the *X. laevis* embryonic U1B RNA gene (Ciliberto *et al.*, 1985; Krol *et al.*, 1985) was first analysed by precisely deleting this sequence by oligonucleotide-directed mutagenesis. As substrate for mutagenesis we used clone Δ182 (Ciliberto *et al.*, 1985), whose structure is shown in Figure 1A together with the sequence of the Δ20 oligonucleotide used for mutation. Several individual mutants were obtained and sequenced. Clone Δ182Δ20, carrying the desired mutation, was injected into *X. laevis* oocytes. As expected if 3′ end formation is affected, in contrast to Δ182, this mutant fails to induce defined transcripts (Figure 1B, compare

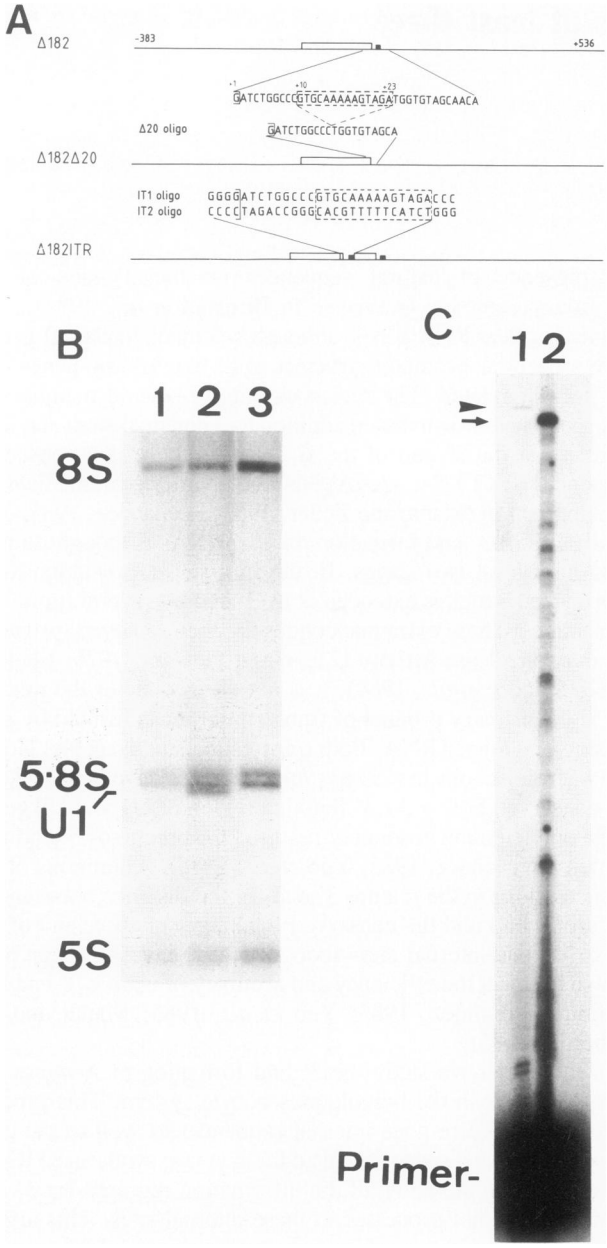

Fig. 1. Oligonucleotide-directed mutations of the *X. laevis* U1B RNA gene and their expression in *X. laevis* oocytes. (**A**) Schematic structure of the wild type U1B gene (Δ182), of the 3′ box deletion mutant (Δ182Δ20) and of the 3′ box duplication mutant (Δ182 ITR). Lines represent flanking regions, open boxes represent U1 coding sequence, black boxes the 3′ consensus sequence, which is also boxed within the sequences. (**B**) Analysis of transcripts from the wild-type U1B gene and from the 3′ box deletion mutant. **Lane 1**, control oocytes; **lane 2**, Δ182; **lane 3**, Δ182Δ20. (**C**) Primer elongation analysis of transcripts from wild type and 3′ deleted U1B gene. Total RNA from two oocytes injected with Δ182 (**lane 1**) or with Δ182Δ20 (**lane 2**) was primed with a 3′ end-labelled *Ava*II–*Sac*I fragment extending from base +57 to +107 of the U1B 3′ flanking sequence (Ciliberto *et al.*, 1985). The primer extended products of Δ182 and Δ182Δ20 are indicated by arrowheads and arrows, respectively.

lanes 2 and 3) as a consequence of the deletion of an element essential for termination of transcription (Hernandez, 1985; Yuo *et al.*, 1985). Mutant Δ182Δ20 is however efficiently transcribed. This was proven with a primer extension assay using total RNA from Δ182Δ20-injected oocytes and as primer a DNA fragment corresponding to the sequence between bases +57 and +107 in the 3′ flanking region of the U1B gene, i.e. from a

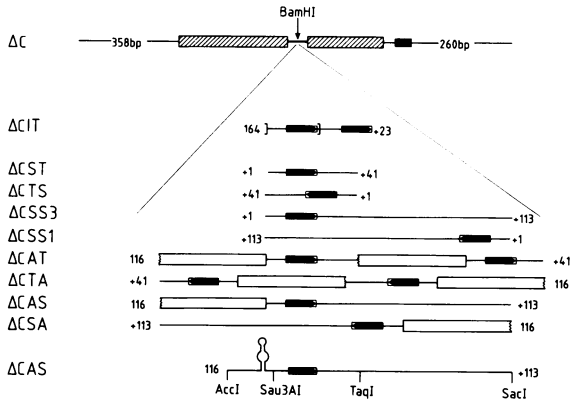

Fig. 2. Schematic diagram of fusion constructs between *X. laevis* ΔC (Mattaj and De Robertis, 1985) and U1B (Cilberto *et al.*, 1985) RNA genes. The structure of the U2 ΔC mutant used as an acceptor of different fragments of the U1B gene is diagrammed in the upper part of the figure. In the lower part the structure of the different U1 inserts is shown and their relative orientation to the ΔC coding sequence is indicated. Hatched boxes: ΔC coding sequence; lines: U1 or ΔC flanking sequences; open boxes: U1 coding sequences; black boxes: 3′ consensus sequence. In the lower part of the figure the U1B AS fragment is drawn with the coding region in its stem–loop configuration. The restriction sites used are also indicated.

portion of the U1B 3′ flanking region downstream of the conserved box. As expected, this gives a 258 base primer elongated product (Figure 1C, lane 2) consistent with an RNA product whose 5′ end corresponds to that of U1B RNA. Primer elongation on RNA from Δ182–injected oocytes also gives a product (Figure 1C, lane 1, indicated by an arrowhead). This is much less abundant than for the mutant and is 12 bases shorter due to the deletion in the mutant.

We then asked if the conserved 3′ box was the only element necessary for efficient 3′ end formation of U1B. Two complementary oligonucleotides reproducing the U1B sequence from base 164 to +23 (IT1 and IT2, Figure 1) were synthesized and inserted in the correct orientation into the U1B gene at a unique *Acc*I site at position 116 of the coding sequence. Injection of this clone (Δ182-ITR, as well as other clones containing monomer and dimer copies of the oligonucleotide in either orientation) yields a single transcript longer than wild type whose 3′ end presumably corresponds to that of wild type U1B (not shown, see below). It appears therefore that the conserved 3′ element is not sufficient *per se* to ensure efficient termination of transcription.

Gene internal sequence are required for Xenopus U1 snRNA 3′ end formation

We decided to further investigate which other sequences contribute to the process of 3′ end formation on U snRNA. To this end fusions between a *X. laevis* U2 gene and the U1B gene were constructed, in which segments of different length from the U1B gene were introduced internally into the U2 gene. The schematic representation of the clones is shown in Figure 2. Details of their construction are given in Materials and methods. We chose to use a U2 mutant (ΔC) that is transcribed and 3′ processed with high efficiency but in which the internal Sm binding site has been substituted by a different sequence containing a unique *Bam*HI site (Mattaj and De Robertis, 1985).

Clone ΔCIT (Figure 2) carries an insertion of a dimer of the oligonucleotide containing the conserved 3′ flanking sequence used to construct Δ182-ITR (Figure 1A). Making use of convenient restriction sites within the U1B gene and further downstream we excised various U1B fragments and inserted them into ΔC.

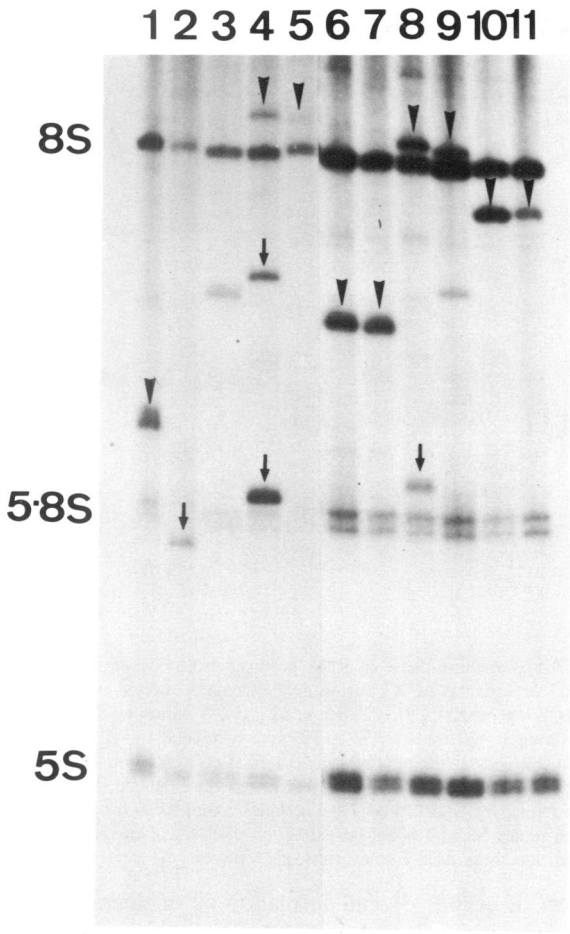

Fig. 3. Analysis of RNA synthesised in *X. laevis* ooxytes injected with U2−U1 fusion genes on 8% polyacrylamide, 7 M urea gels. **Lane 1**, ΔC; **lane 2**, Δ182; **lane 3**, ΔCIT; **lane 4**, ΔCAT; **lane 5**, ΔCTA; **lane 6**, ΔCST; **lane 7**, ΔCTS; **lane 8**, ΔCAS; **lane 9**, ΔCSA; **lane 10**, ΔCSS3; **lane 11**, ΔCSS1. The arrows indicate bands which are discussed in the text. Transcripts with U1B ends are indicated with arrows, those with ΔC 3′ ends with arrowheads. Lanes 1−5 and 6−11 are two different exposures of the same gel. Injections were repeated with two different preparations of each DNA sample, with similar results. In particular the relatively low level of accumulation of transcripts from ΔCIT, ΔCTA, ΔCSA and ΔCSSI was consistently observed, and may be due to instability of these RNAs. The yield of ΔCAT transcripts was not consistently greater than that of ΔC transcripts.

Clones ΔCST (*Sau*3AI−*Taq*I) and ΔCTS and ΔCSS3 (*Sau*3AI−*Sac*I) and ΔCSS1 carry longer segments of the 3′ flanking regions of U1B inserted into ΔC in either orientation as indicated in Figure 2, where the positions of the restriction sites are also shown. Clones ΔCAT (*Acc*I−*Taq*I), ΔCTA, ΔCAS (*Acc*I−*Sac*I) and ΔCSA were made in order to examine the possible contribution of the terminal stem and loop structure of U1 RNA genes shown in diagrammatic form in the *Acc*I−*Sac*I fragment at the bottom of Figure 2. Possible involvement of this structure in 3′ end formation is suggested, as mentioned in the Introduction, by analogy with the case of histone mRNA 3′ end formation. In these constructs the inserted U1 segment extends from the *Acc*I site at position 116 of the coding sequence to the *Taq*I site 19 nucleotides downstream of the 3′ box in constructs ΔCAT and ΔCTA and to the *Sac*I site 90 nucleotides downstream from the box in plasmids ΔCAS and ΔCSA. Furthermore, clones of the ΔCAT class carry two copies in tandem of the U1 insert and were chosen for reasons described in the following section. In all cases both orientations selected were analysed by microinjection into *X. laevis* oocytes (Figure 3).

Lanes 1 and 2 show the transcripts of oocytes injected with the ΔC mutant of U2 RNA and the Δ182 U1B RNA clone whose lengths are 194 and 169 nucleotides respectively. The transcript of ΔCIT5 (Figure 3, lane 3) is of the length expected (258 nucleotides) for an RNA which contains all the ΔC sequences plus an insert of 60 bases. We conclude that this RNA stops at the normal ΔC 3′ end. No transcripts of length around 110 or 140 are seen. These are the lengths expected if either the first or second oligonucleotide had directed 3′ end formation. As mentioned in the previous section this oligonucleotide has been inserted into U1B in Δ182, where it also fails to direct the formation of detectable transcripts shorter than those reading through to the normal Δ182 3′ end. We therefore conclude that the 30-mer oligonucleotide does not efficiently direct 3′ end formation.

We next analysed transcripts of the clones ΔCSS3, ΔCSS1, ΔCST and ΔCTS (Figure 2). The pattern of their transcripts (Figure 3, lanes 6, 7, 10 and 11) is entirely consistent with the idea that the inserted sequences have no effect on 3′ end formation. The lengths of the transcripts (~310 for the SS clones and 240 for the ST and TS clones) and the absence of distinct abundant shorter transcripts indicate that the RNAs terminate at the normal ΔC 3′ end. The absence of shorter transcripts indicates that these inserts cannot efficiently direct 3′ end formation at internal sites.

The constructs which contain U1B gene internal sequences in addition to the 3′ flanking sequences in the normal orientation, ΔCAT and ΔCAS, behave differently. In addition to transcripts of the lengths expected for RNAs with the ΔC 3′ end (~415 and 400 respectively) these constructs also show transcripts of lengths expected if the inserts can direct 3′ end formation within the inserted U1B sequences (Figure 3, lanes 4 and 8, the relevant transcripts are arrowed). ΔCAT has two such shorter transcripts (Figure 3, lane 4), indicating that both of the tandem inserts (Figure 2) can direct 3′ end formation. That these shorter transcripts do indeed have the same 3′ ends as U1B transcripts was shown by S1 mapping (Figure 4, lanes 1−4). The three ΔCAT and two ΔCAS transcripts, labelled *in vivo*, were gel purified. They were then hybridised to single-stranded M13mp8 DNA containing the *Acc*I−*Sac*I fragment of Δ182 (Figure 2), and the hybrids were digested with S1 nuclease. As expected the major S1 protected band of Δ182 is 50 nucleotides long (Figure 4, lane 1). The shortest ΔCAT transcript shows the same pattern (Figure 4, lane 2) indicating that it has the same 3′ end as Δ182. A fragment of the shorter ΔCAS transcript 13 nucleotides longer is protected (Figure 4, lane 4). This again indicates that the 3′ end of this transcript is similar to that of Δ182 since ΔCAS is homologous to 14 nucleotides of the M13mp8 polylinker, as well as to part of the *Acc*I−*Sac*I insert. The 140-nucleotide protected band of the middle transcript of ΔCAT (Figure 4, lane 3) probably arises by protection of this RNA with two molecules of M13mp8 *Acc*I−*Sac*I DNA as discussed in Figure 4, and again indicates that the 3′ end of the transcript is similar to that of Δ182. The same assay was used to show that the longest transcripts of ΔCAT and ΔCAS had the same 3′ end as ΔC (not shown).

The efficiency of the inserted sequences in directing 3′ end formation can be deduced from the intensity of the three ΔCAT transcripts and the two ΔCAS transcripts in relation to their length. The efficiency is similar in both cases, and is ~50−60%. This might indicate that the *Acc*I−*Taq*I and *Acc*I−*Sac*I fragments do not function efficiently out of their natural context, but direct comparison with the wild-type U1B gene (Figure 1C) is difficult

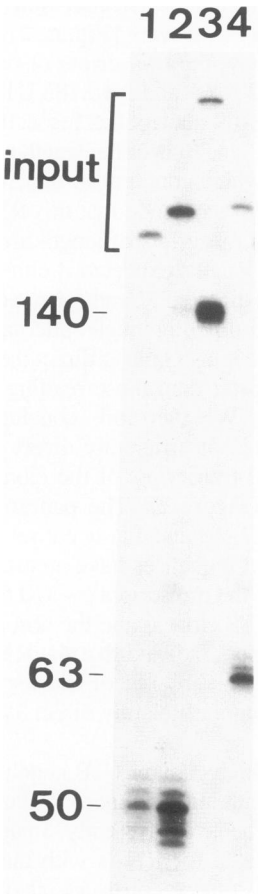

Fig. 4. S1 analysis of the RNA products from U2−U1 gene fusions. The bands indicated with arrows in Figure 3 were purified by excision from preparative scale polyacrylamide urea gels. They were then hybridized with a large excess of single-stranded M13mp8 DNA containing the U1B *Acc*I−*Sac*I insert, and digested with nuclease S1. Protected RNA fragments were analysed on an 8% polyacrylamide, 7 M urea gel. Part of the *in vivo* labelled, gel-purified RNA has remained undigested under the conditions used. These input length RNA bands are bracketed. The lengths of the major S1 protected bands, determined by comparison with an unrelated sequence which was co-electrophoresed, are indicated. **Lane 1**: Δ182; **lane 2**; ΔCAT, lowest band; **lane 3**: ΔCAT, middle band and **lane 4**: ΔCAS, lower band. The pattern of S1 protected bands in lanes 1 and 2 is identical, and the major protected fragment is 50 nucleotides long. The 140 nucleotide RNA seen in lane 3 was not expected. Since ΔCAT contains a dimer *Acc*I−*Taq*I insert, we expected to obtain two S1-protected bands of 91 and 50 nucleotides from the middle-sized ΔCAT transcript. The 91 band from the first inserted *Acc*I−*Taq*I sequence and the 50 band from the start of the second repeat to the second U1B 3′ end. The length of the band obtained, 140 nucleotides, indicates that the entire insert up to the second U1B 3′ end is protected, presumably by the binding of two separate single-stranded M13mp8AS probes to the ΔCAT RNA.

since U1B readthrough transcripts may well be less stable than the longer ΔCAT and ΔCAS transcripts, which have defined, ΔC encoded, 3′ ends. When the U1B fragments, containing the information required to direct 3′ end formation, are inserted into ΔC in the opposite orientation to give the ΔCTA and ΔCSA constructs, no major transcripts which end within 10 bases of the conserved 3′ elements are seen (Figure 3, lanes 5 and 9). The two inverted elements in the ΔCTA transcript are centred at positions 163 and 250, the single inverted element in ΔCSA is centred at position 222. The transcripts shorter than those with the ΔC 3′ end from these constructs therefore do not end in the vicinity of the inverted 3′ consensus sequences, and may either be the

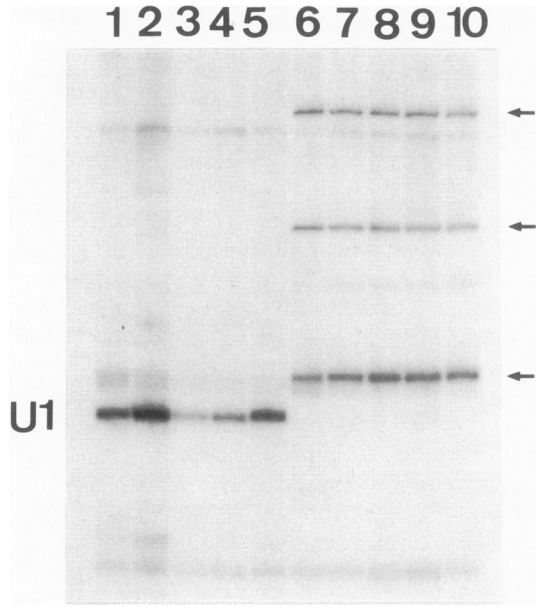

Fig. 5. An α-amanitin chase of RNA products from oocytes injected with X.1 U1.3 or with the ΔCAT fusion gene. Oocytes injected with a minor *X. laevis* U1 gene X.1.U1.3 (Zeller *et al.*, 1984) (**lanes 1−5**) or with ΔCAT (**lanes 6−10**) plus [α-^{32}P]GTP were incubated for 12 h. α-Amanitin was then injected into the cytoplasm to a final concentration of 1 μg/ml and oocytes were either immediately extracted (lanes 1 and 6) or further incubated for 1 h (lanes 2 and 7), 3 h (lanes 3 and 8), 6 h (lanes 4 and 9) and 24 h (lanes 5 and 10) before extraction of RNA. The U1B transcript is indicated, the three ΔCAT transcripts are arrowed.

products of cryptic 3′ end formation or of degradation. This would suggest that 3′ end formation is orientation independent.

Readthrough transcripts are not further processed

There is general agreement that the mature 3′ ends of U snRNAs are formed by a processing event in which a few (7−15) overhanging nucleotides are removed by an exonuclease activity (Elicieri, 1981; Madore *et al.*, 1984; Yuo *et al.*, 1985; Hernandez, 1985). There is less information about how these precursors are formed (see Discussion). We have attempted to use defined transcripts containing an internal 3′ end, such as those obtained from ΔCAT and ΔCAS, to determine whether they arise by a processing mechanism like that observed in histone 3′ end formation (Birnstiel *et al.*, 1985). In our first experiments we gel-purified such transcripts and re-injected them into oocytes. No processing products were ever seen, independent of whether injection was nuclear or cytoplasmic (data not shown). Since it is possible to argue that processing depends on the RNA having a particular structure which could be disrupted during purification we devised the experiment shown in Figure 5. A U1 clone (X.I.U1.3 from Zeller *et al.*, 1984) and ΔCAT (Figure 2) were injected into oocytes along with [α-^{32}P]GTP. After overnight incubation oocytes were injected with α-amanitin at 1 μg/ml, a concentration which inhibits snRNA transcription (Mattaj and Zeller, 1983). Immediately or at later times RNA was extracted from these oocytes. The U1 injections showed that over the time period of the experiment U1 RNA is stable (Figure 5, lanes 1−5). The ΔCAT experiment (Figure 5, lanes 6−10) shows that all three ΔCAT transcripts are also stable for at least 24 h. ΔCAT contains two copies of U1B information for 3′ end formation. If 3′ end formation were a consequence of a processing event which could use the longer ΔCAT transcripts as a substrate we would expect an accumulation of the shortest transcript and a reduction

in quantity of the longer transcripts during the course of the experiment. Since no such change occurs this result suggests that U snRNA precursors are formed by either termination or a processing event tightly coupled to transcription, or that once formed, the longer ΔCAT transcripts can no longer be recognised as substrates for such a processing event.

To determine whether limiting factors involved in 3′ end formation bind to 3′ sequences from the U1B gene either at the DNA or the RNA level, we attempted to compete out the ability of oocytes to form 3′ ends on ΔCAT by co-injecting the ΔCAT DNA with an excess of either DNA or *in vitro* synthesized RNA corresponding to the *Acc*I−*Taq*I and *Acc*I−*Sac*I inserts (Figure 2). In no case was competition successful (data not shown). The ΔCAT clone should provide a very sensitive assay for competition since a weak competition for factors involved in 3′ end formation would result in a shift in accumulation from the short towards the long transcripts rather than simply an accumulation of undefined readthrough transcripts. The failure to observe competition does not favour either a termination or a processing mechanism of 3′ end formation, but only indicates that the oocyte has an excess of any factors involved in this process, which is not saturable by microinjection.

Promoter specificity of 3′ end formation

U snRNA gene promoters have a structure which is different to that of other genes transcribed by RNA polymerase II (Murphy *et al.*, 1982; Skuzeski *et al.*, 1984; Westin *et al.*, 1984; Mattaj *et al.*, 1985; Ciliberto *et al.*, 1985; Krol *et al.*, 1985; Ares *et al.*, 1985). Since our results indicated that the *Acc*I−*Taq*I and *Acc*I−*Sac*I fragments of the U1B gene contain enough information to direct 3′ end formation when inserted into the ΔC U2 mutant we decided to investigate whether the *Acc*I−*Sac*I fragment would function in the context of a different *Pol*II transcription unit. For this purpose we chose pSV2 CAT (Gorman *et al.*, 1982), which consists of the bacterial chloramphenicol acetyl transferase gene driven by the SV40 early promoter region.

The *Acc*I−*Sac*I fragment, whose structure is given in Figure 2, was inserted into pSV2 CAT in both orientations as diagrammed in Figure 6A. Its ability to direct 3′ end formation was assayed in two ways: (i) by measuring the level of chloramphenicol acetyl transferase activity or (ii) by S1 analysis with uniformly labelled probes. The latter technique would detect 3′ ends whose formation were directed by the *Acc*I−*Sac*I insert. Measurement of chloramphenicol acetyl transferase enzyme activity (Figure 6B) led to the surprising result that, while the inverted insert had little or no effect on activity (Figure 6B, compare lanes 1 and 3), the *Acc*I−*Sac*I insert in the orientation found in U1B appeared to cause an increase in transferase activity (Figure 6B, compare lanes 1 and 2). Transcripts from the pSV2 CAT and pSV2 AS constructs were next analysed with S1 nuclease to directly determine whether the *Acc*I−*Sac*I insert was able to direct 3′ end formation. Uniformly labelled single-stranded probes extending from the *Eco*RI to the *Bgl*I sites of pSV2 CAT (320 nt) and pSV2 AS (507 nt) were prepared. RNA from oocytes injected with pSV2 CAT protected a major band of ~310 nt from S1 digestion when the pSV2 CAT probe was used (Figure 6C, lane 2). Transcription starting at the early-early start sites of pSV2 CAT would initiate 3−10 nt upstream from the *Bgl*I cut site (see e.g. Takahashi *et al.*, 1986). As well as these expected protected products, an additional smear of shorter bands is visible in lane 2, which may arise from aberrant transcription starts or alternatively be due to degradation of the pSV2 CAT transcripts.

When the pSV2 CAT transcripts are analysed with the 507 nt pSV2 AS probe (Figure 6C, lane 4), three major bands are seen, which correspond to undigested probe and protected species of ~250 and 65 nt. The distance from the *Eco*RI site to the *Hin*dIII site, at which the probe and the pSV2 CAT transcripts

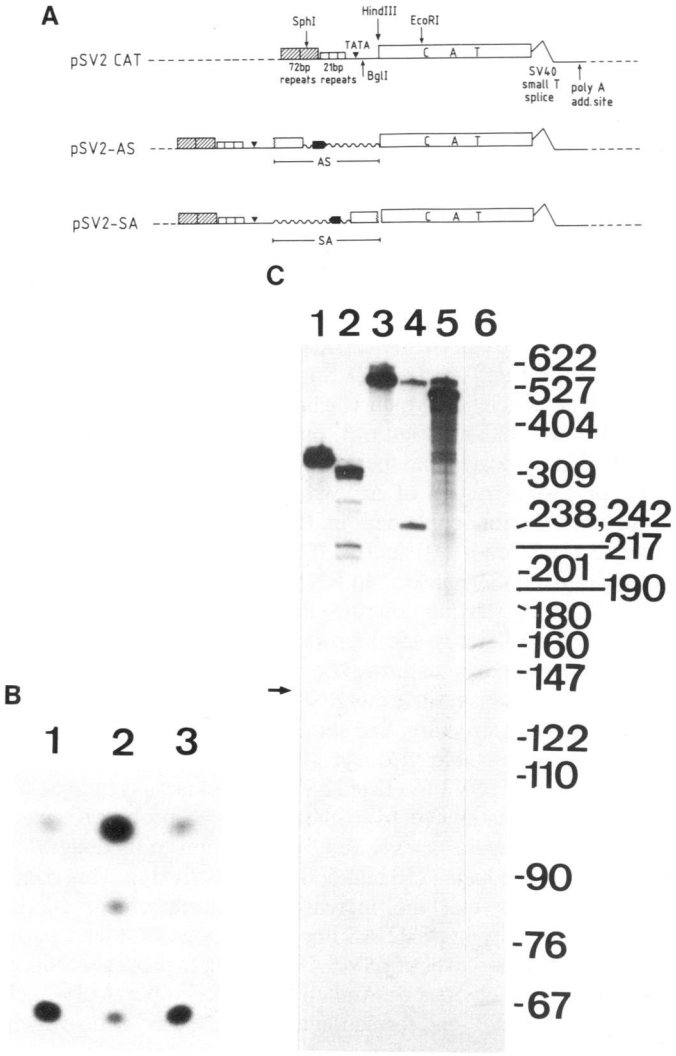

Fig. 6. Structure of U1 terminator CAT gene fusions and their expression in *Xenopus* oocytes. (**A**) Schematic representation of U1B-chloramphenicol acetyl transferase gene fusions. In the upper part the acceptor plasmid pSV2 CAT is represented (Gorman *et al.*, 1982). Hatched boxes: SV40 enhancer 72-bp repeat sequences; small boxes: 21-bp repeats; hatched lines: pBR322 vector sequences. In the lower part are the structures of U1B pSV2 CAT fusion derivatives pSV2 AS and pSV2 SA. Open boxes: U1 coding sequence; wavy lines: U1 3′ flanking sequences; black boxes, 3′ consensus sequence. (**B**) Expression of U1B-chloramphenicol acetyl transferase gene fusions in *X. laevis* oocytes. **Lane 1**: pSV2 CAT; **lane 2**: pSV2 AS, and **lane 3**: pSV2 SA, were injected into oocytes. Extracts were prepared from batches of 10 oocytes 20 h later, and chloramphenicol acetyl transferase activity was assayed. (**C**) S1 analysis of U1B-chloramphenicol acetyl transferase fusion transcription products. Uniformly labelled single-stranded probes corresponding to the *Eco*RI−*Bgl*I fragments of pSV2 CAT and pSV2 AS were used as probe. Protected bands were analysed on an 8% polyacrylamide, 7 M urea gel. **Lane 1**: pSV2 CAT probe, undigested; **lane 2**: pSV2 CAT probe, RNA from pSV2 CAT-injected oocytes; **lane 3**: pSV2 AS probe, undigested; **lane 4**: pSV2 AS probe, RNA from pSV2 CAT-injected oocytes; **lane 5**: pSV2 AS probe, RNA from pSV2 AS injected oocytes, and **lane 6**: *Hpa*II fragments of pBR322. Their lengths are indicated. The arrow is at the position (140 nucleotides) expected of RNA which would terminate at the U1B 3′ end included in pSV2 AS. RNA was prepared from batches of 20 oocytes. Each S1 digestion reaction contained RNA from 2 oocyte equivalents.

diverge (Figure 6A), is 254 nt. The smear of bands of ~65 nt would correspond to transcripts initiating at the SV40 early-early start sites and protecting the probe from this point up to the *Hin*dIII site. RNA from oocytes injected with pSV2 AS gives rise to a major protected species slightly shorter than the undigested *Eco*RI–*Bgl*I pSV2 AS probe (Figure 6C, lane 5) again consistent with transcripts initiating at the early-early start sites and reading through to the *Eco*RI site. As was the case for pSV2 CAT transcription (Figure 6C, lane 2) a smear of shorter protected species is also seen. If the *Acc*I–*Sac*I insert had directed 3′ end formation, we would expect to see defined species of ~140 nt with the early-early cap sites and the U1B 3′ end. Since the *Acc*I–*Taq*I and *Acc*I–*Sac*I fragments direct 3′ end formation with an efficiency which is easily detectable in the context of the U2 ΔC mutant (Figure 3, lanes 4 and 8) the lack of the 140 nt RNA indicates that there are sequences either in the promoter or coding region of ΔC which are required in addition to the information contained in the *Acc*I–*Sac*I fragment to generate U1 3′ ends efficiently. If U1 3′ ends were generated by a processing mechanism, an RNA protecting ~360 nt of the pSV2 AS probe starting downstream of the U1 3′ end and ending at the *Eco*RI site would be produced. No such band is visible. This result cannot however be used to argue against a processing mechanism since this RNA, if it were produced, would not have a 5′ cap structure, and some uncapped RNAs have been shown not to be stable in oocytes (Green *et al.*, 1983).

The amount of pSV2 CAT and pSV2 AS transcripts in injected oocytes can be estimated from the amount of probe protected (Figure 6C, compare lanes 4 and 5) in relation to the length of the protected products (250 and 500 nt respectively). This comparison suggests that the increased chloramphenicol acetyl transferase activity in pSV2 AS injected oocytes is at least partly due to the higher level of pSV2 AS transcripts produced. Since this result has no obvious bearing on the ability of the *Acc*I–*Sac*I fragment to direct 3′ end formation additional experiments which would be required to confirm and explain this effect have not yet been undertaken.

Discussion

Structural requirements for U1B 3′ end formation

We have demonstrated a requirement for three separate elements which are all necessary for faithful 3′ end formation on *Xenopus* U1 snRNA. The first, a conserved 3′ sequence, had previously been shown to be essential for 3′ end formation on human U1 and U2 snRNAs (Hernandez, 1985; Yuo *et al.*, 1985). There is in addition a requirement for gene internal sequences which have the potential to form a stem–loop structure. Additional U snRNA gene sequences upstream of the stem–loop are also required. A role for the 3′ stem–loop structure in the production of mature U snRNA 3′ ends had previously been suggested (Yuo *et al.*, 1985; Hernandez, 1985; Mattaj and De Robertis, 1985), since it was found that mutant U snRNA transcripts which had lost this structure were unable to form 3′ ends efficiently and at the normal position. The results presented here on the differences in 3′ end formation between the pairs of clones ΔCAT and ΔCST and ΔCAS and ΔCSS3 (Figure 3) suggest however that precursor 3′ end formation also requires gene-internal sequences. This conflicts with the results of Yuo *et al.* (1985) and Hernandez (1985) who found that the 3′ stem–loop structure was not required for the formation of primary transcripts of human U2 snRNA in *Xenopus* oocytes or of human U1 snRNA in HeLa cells. A possible explanation for this disagreement would be that the inserted U1B fragments in ΔCSS3 and ΔCST do direct 3′ end formation but the RNAs thus formed are unstable. Although this may explain our observations, many mutant U2 and U1 snRNA transcripts, including some unable to form 3′ stem–loop structures, have been analysed in oocytes and HeLa cells, and found to be stable enough to be detectable (Mattaj and De Robertis, 1985; Yuo *et al.*, 1985; Hernandez, 1985). Since our experiments all rely on the use of gene fusions, another possible explanation is that the information for 3′ end formation in any given gene is redundant. Further experiments will be required to clarify this issue.

The fact that the *Acc*I–*Sac*I fragment of U1B could not direct 3′ end formation when inserted into the chloramphenicol acetyl transferase coding region, although it functioned efficiently when inserted into the ΔC mutant of U2, indicates a third requirement for 3′ end formation. From the data presented here this third element could potentially be either 5′ or 3′ of the insertion point in ΔC, however 36 bp of 3′ flanking sequence of the ΔC clone are enough to direct efficient 3′ end formation (Mattaj *et al.*, 1985) and a human U1 gene with only 35 bp of flanking sequence (Murphy *et al.*, 1982) is also transcribed correctly in oocytes (Skuzeski *et al.*, 1984). Since the *Acc*I–*Sac*I fragment includes a much longer 3′ flanking region this third element is likely to be upstream of the insert point. The only sequences conserved between U2 and U1 in the upstream coding region are the Sm binding sites, which cannot be required for 3′ end formation since ΔC, lacking this region, terminates normally. Additionally, several mutations spanning most of the U2 coding sequences 5′ to the insertion site chosen (Mattaj and De Robertis, 1985; I.W.Mattaj, unpublished results) do not affect 3′ end formation. This raises the interesting possibility that the third element is part of the ΔC promoter. Two elements have so far been identified in the promoters of U snRNA genes. An orientation-independent distal sequence element (DSE) and a proximal sequence element (PSE) centered at ~ −55 relative to the cap site (Murphy *et al.*, 1982; Skuzeski *et al.*, 1984; Westin *et al.*, 1984; Mattaj *et al.*, 1985; Ciliberto *et al.*, 1985; Krol *et al.*, 1985; Ares *et al.*, 1985). Deletions which remove the DSE from U2 or U1 genes of *X. laevis* do not appear to impair 3′ end formation (Mattaj *et al.*, 1985; Ciliberto *et al.*, 1985; Krol *et al.*, 1985). If the factor required for 3′ end formation recognises one of the previously identified promoter elements it is therefore more likely to be either the proximal sequence element, or the combination of proximal plus distal sequence elements. Both of these structures appear to be unique to U snRNA genes. Since the proximal sequence element is essential for U snRNA transcription (Ciliberto *et al.*, 1985; Mattaj *et al.*, 1985; Ares *et al.*, 1985) it will be difficult to test its involvement in 3′ end formation directly.

Termination versus processing

U snRNAs acquire their 3′ ends in two stages. In the first a precursor with a short 3′ extension is formed. This is later processed, probably by an exonuclease activity, to the mature length (Elicieri, 1981; Madore *et al.*, 1984; Hernandez, 1985; Yuo *et al.*, 1985). Our results suggest, but do not prove, an involvement of all three identified sequence elements in the first step of this process, i.e. precursor formation. Whether the precursor is formed by termination or processing is still unknown. Kunkel and Pederson (1985) recently showed that nuclear transcripts of human U1 genes decreased markedly in abundance close to (maximally 60 base pairs) the 3′ end of these genes, arguing in favour of a termination event near this region. Our results of re-injection of RNAs containing all the signals required for 3′ end forma-

tion, and most convincingly, the α-amanitin chase experiment (Figure 5) with ΔCAT show that readthrough transcripts cannot subsequently be processed, although they are *in vivo* transcripts which presumably contain any modifications and structures necessary for 3′ end formation. Our results therefore argue for a termination event or a 3′ processing event which is coupled to transcription. At present we can see no way to distinguish between these two mechanisms.

Materials and methods

General methods

Bacterial enzymes, T4 DNA ligase, polynucleotide kinase, DNA polymerase I, and Klenow fragment were purchased from BRL, Anglian Biological Laboratories or New England Biolabs and used following the procedures of Maniatis *et al.* (1982). AMV reverse transcriptase was purchased from Boehringer. Transformations and preparations of ds and ss DNA were as described (Messing, 1983). DNA sequence analyses were carried out using the dideoxy chain termination method (Sanger *et al.*, 1977).

Microinjection

Oocytes were injected aiming at the nucleus with DNA dissolved in distilled water at a concentration 200−300 $\mu g/ml^{-1}$. DNA concentration was determined spectrophotometrically and checked by examination following agarose gel electrophoresis. The injection volume was 30−50 nl. Oocytes were then fractionated and extracted, and the extracted RNAs were analysed as described (Mattaj and De Robertis, 1983). α-Amanitin experiments were carried out as described (Mattaj and Zeller, 1983).

Oligonucleotide synthesis and mutagenesis

Oligonucleotides were synthesized with the phosphoramidite method (Winnacker and Dorper, 1982). Oligonucleotide-directed mutagenesis on ssDNA template was performed according to Zoller and Smith (1983).

Construction of U1 mutants, U1−U2 fusions and U1−CAT fusions

Δ182 ITR: Δ182 (Ciliberto *et al.*, 1985) was linearized with *Acc*I (position 115 of the U1B coding sequence), the ends filled in with Klenow enzyme and then recircularized in the presence of the IT1 and IT2 oligonucleotides. Mutants carrying the inserted oligonucleotides were first isolated by selective hybridization with ^{32}P-kinased IT1 and then sequenced by the dideoxy chain termination method (Sanger *et al.*, 1977).

U2−U1 fusions: ΔC is a U2 mutant in which bases 100−111 of the coding sequence have been deleted and substituted by a 17 base adaptor fragment carrying a unique *Bam*HI site (Mattaj and De Robertis, 1985). ΔC was linearized with *Bam*HI, the ends filled in with Klenow polymerase and recircularized in the presence of various U1B gene segments.

ΔCIT was constructed making use of the IT1 and IT2 oligonucleotides. ST and SS segments were isolated from Δ182 making use of the restriction sites *Sau*IIIA (position +1), *Taq*I (position +41) and *Sac*I (position +113), the ends were filled in by Klenow polymerase and ligated with linear ΔC.

AT (*Acc*I position 116−*Taq*I position +41) and AS (*Acc*I position 116−*Sac*I position +113) segments were first cloned into pSP64 and pSP65 vectors (Melton *et al.*, 1984) respectively and then re-cloned into linear ΔC making use of the *Eco*RI and *Hin*dIII sites in the vector polylinkers.

U1 terminator−CAT gene fusions: pSV2 CAT was linearized with *Hin*dIII, ends repaired with Klenow enzyme and recircularized in the presence of the AS fragment isolated as for the construction of plasmids ΔCAS and ΔCSA. The constructs were all checked by DNA sequencing (Sanger *et al.*, 1977).

S1 analysis and primer elongation

Conditions for formation of DNA−RNA hybrids and for digestion with S1 were as described (Berk and Sharp, 1977). Primer elongation analysis was performed according to Luse *et al.* (1981).

Acknowledgements

We wish to thank Heide Seifert for expert typing of this manuscript. G.C.'s work at the EMBL was supported in part by a grant from the Ministerio delle Pubblica Istituzione to Prof. Franco Salvatore. The work in Italy was supported by a grant from Progetto Finalizzato Ingegneria Genetica e Basi Moleculari delle Malattie Ereditarie, CNR, Rome, Italy.

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.
Birchmeier,C., Schümperli,D., Sconzo,G. and Birnstiel,M.L. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1057−1061.
Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349−359.
Black,D.L., Chabot,B. and Steitz,J.A. (1985) *Cell*, **42**, 737−750.
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.
Elicieri,G.L. (1981) In Busch,H. (ed.), *The Cell Nucleus*. Academic Press, NY, Vol. **8**, pp. 307−330.
Galli,G., Hofstetter,H., Stunnenberg,H.G. and Birnsteil,M.L. (1983) *Cell*, **34**, 823−828.
Gorman,C.M., Moffat,L.F. and Howard,B.H. (1982) *Mol. Cell. Biol.*, **2**, 1044−1051.
Green,M.R., Maniatis,T. and Melton,D.A. (1983) *Cell*, **32**, 681−694.
Hernandez,N. (1985) *EMBO J.*, **4**, 1827−1837.
Krainer,A.R. and Maniatis,T. (1985) *Cell*, **42**, 725−736.
Krämer,A., Keller,W., Appel,B. and Lührmann,R. (1984) *Cell*, **38**, 299−307.
Krieg,P.A. and Melton,D.A. (1984) *Nature*, **308**, 203−206.
Krol,A., Lund,E. and Dahlberg,J.E. (1985) *EMBO J.*, **4**, 1529−1535.
Kunkel,G.R. and Pederson,T. (1985) *Mol. Cell. Biol.*, **5**. 2332−2340.
Luse,D.S., Haynes,J.R., van Leeuwen,D., Schon,E.A., Cleary,M.L., Shapiro,S.G., Lingrel,J.B. and Roeder,R.G. (1981) *Nucleic Acids Res.*, **9**, 4339−4354.
Madore,S.L., Wieben,D.E. and Pederson,T. (1984) *J. Cell. Biol.*, **98**, 188−192.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory.
Mattaj,I.W. and De Robertis,E. (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.
Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green, M.R. (1984) *Nucleic Acids Res.*, **12**, 7035−7056.
Messing,J. (1983) *Methods Enzymol.*, **101**, 28−78.
Murphy,J.T., Burgess,R.R., Dahlberg,J.E. and Lund,E. (1982) *Cell*, **29**, 265−274.
Padgett,R.A., Mount,S.A., Steitz,J.A. and Sharp,P.A. (1983) *Cell*, **35**, 101−107.
Price,D.H. and Parker,C.S. (1984) *Cell*, **38**, 423−429.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.
Skuzeski,J.M., Lund,E., Murphy,J.T., Steinberg,T.H., Burgess,R.R. and Dahlberg,J.E. (1984) *J. Biol. Chem.*, **259**, 8345−8352.
Strub,K. and Birnsteil,M.L. (1986) *EMBO J.*, **5**, 1675−1682.
Strub,K., Galli,G., Busslinger,M. and Birnstiel,M.L. (1984) *EMBO J.*, **3**, 2801−2807.
Takahashi,K., Vigneron,M., Matthes,H., Wildeman,A., Zenke,M. and Chambon,P. (1986) *Nature*, **319**, 121−126.
Westin,G., Lund,E., Murphy,J.T., Pettersson,U. and Dahlberg,J.E. (1984) *EMBO J.*, **3**, 3295−3301.
Winnacker,E.L. and Dorper,T. (1982) In Gassen,H.G. and Lang,A. (eds), *Chemical and Enzymatic Synthesis of Gene Fragments*. Verlag Chemie, Weinheim, pp. 97−102.
Yuo,C.-Y., 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.
Zieve,G. and Penman,S. (1976) *Cell*, **8**, 19−31.
Zoller,M.J. and Smith,M. (1983) *Methods Enzymol.*, **100**, 468−500.

Received on 16 June 1986