The EMBO Journal vol.4 no.6 pp.1529–1535, 1985

The two embryonic U1 RNA genes of *Xenopus laevis* have both common and gene-specific transcription signals

Alain Krol[1], Elsebet Lund and James E.Dahlberg

Department of Physiological Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA

[1]Present address: Institut de Biologie Moleculaire et Cellulaire du C.N.R.S., F 67084 Strasbourg, Fracne

Communicated by J.-P.Ebel

We have cloned and sequenced the 1842-bp repeat DNA encoding the two *Xenopus laevis* embryonic U1 RNAs, xU1a and xU1b. Although these two U1 RNAs are almost identical in sequence and are coordinately expressed during early embyrogenesis, the flanking sequences of their genes show very little homology. Both genes contain two short conserved sequences, centered around positions −55 and +19, that probably are essential for 5′ and 3′ end formation of U1 RNAs, respectively. Efficient transcription of either gene in stage VI oocytes requires gene-specific promoter elements, located upstream of position −220. In the xU1b gene, these required 5′-flanking sequences include an 18-bp palindrome that has potential for Z-DNA formation. When injected separately into stage VI oocytes, the xU1a and xU1b genes are equally well transcribed, but co-injection of the two genes, either as the full length repeat or as two separate subclones, results in preferential accumulation of xU1b RNA. This competitive advantage of the xU1b gene in injected oocytes apparently is the result of preferred binding of one or more transcription factors that are limiting in these oocytes.

Key words: U1 small nuclear RNA genes/embryonic U1 RNAs/oocyte injection/transcription enhancer/*X. laevis* U1 RNA synthesis

Introduction

Like all higher eukaryotic cells, *Xenopus laevis* cells contain $10^5 - 10^6$ molecules of U1 small nuclear RNA (U1 RNA) (for a review, see Busch *et al.*, 1982). We recently showed that the U1 RNAs of *X. laevis* fall into two classes, which we call embryonic and somatic U1 RNAs (Forbes *et al.*, 1984). Two embryonic U1 RNAs, xU1a and xU1b, are the major species synthesized during early embryogenesis (Forbes *et al.*, 1984). The stockpile of 5000 – 10 000 somatic cell equivalents of U1 RNAs accumulated in oocytes (Forbes *et al.*, 1983; Fritz *et al.*, 1984) also comprises these two snRNAs (E.Lund and J.E.Dahlberg, in preparation). In contrast, only low levels of the somatic U1 RNAs are expressed in mature oocytes and in early embryos (Forbes *et al.*, 1984); however, during later stages of development the relative level of these RNAs increases gradually so that somatic RNAs are predominant in tissues from large tadpoles and frogs (E.Lund and J.E.Dahlberg, in preparation.)

The two classes of *X. laevis* U1 RNA genes differ significantly in their organization. The two embryonic U1 RNA genes, comprising >90% of the U1 genes of *X. laevis*, are found in 1.9-kb tandem repeat units that are reiterated >500-fold per haploid genome. The 50 – 100 copies of the somatic U1 RNA genes appear to be more dispersed in the genome, but the precise organization of these genes has not yet been determined (Lund *et al.*, 1984; Zeller *et al.*, 1984; see also review by Mattaj *et al.*, 1985b).

To study the mechanism of differential control of *X. laevis* U1 RNA expression we have cloned embryonic xU1a and xU1b genes (Lund *et al.*, 1984) and determined the complete nucleotide sequence of the 1842-bp repeat unit. We find that the flanking sequences of the two genes are very divergent, but they show strong homology with each other (and with other snRNA genes) in regions that appear to be essential for correct transcription initiation (Skuzeski *et al.*, 1984; Zeller *et al.*, 1984; Mattaj *et al.*, 1985a; Early *et al.*, 1984; Ciliberto *et al.*, accompanying paper) or for 3′ end formation (Mattaj and Zeller, 1983; Mattaj, 1984; H.Neuman de Vegvar, E.Lund and J.E.Dahlberg, in preparation). Further upstream regions that are required for efficient transcription are similar in structure to transcription activators of the other U1 and U2 genes (Skuzeski *et al.*, 1984; Early *et al.*, 1984; Westin *et al.*, 1984; Mattaj *et al.*, 1985a; Ciliberto *et al.*, accompanying paper; Ares *et al.*, 1985) and to the SV40 transcription enhancer (Benoist and Chambon, 1981).

We previously noted that when a subclone of the embryonic U1 gene repeat was injected into *X. laevis* oocytes the xU1b gene was trancribed much more efficiently than the xU1a gene (Lund *et al.*, 1984). This has recently been confirmed by Ciliberto *et al.* (accompanying paper). We demonstrate here that there were two reasons for this difference. First, the injected xU1a gene, with only 220 bp of 5′-flanking sequence, lacked a far upstream promoter element required for efficient transcription, whereas the xU1b gene contained all the necessary upstream sequences. Secondly, U1 genes compete for transcription factor(s) that are limiting in the injected oocytes. We propose that the competitive advantage of the xU1b gene may be due, at least in part, to the presence of sequences within and near the xU1b gene transcription activator that have potential for Z-DNA formation (Nordheim and Rich, 1983).

Results

Our previous analysis of *X. laevis* genomic DNA demonstrated that the embryonic U1 RNA genes were located in a 1.9-kb segment of DNA that is tandemly repeated several hundred-fold. We reported the cloning and partial characterization of several isolates of a 1.5-kb *Hind*III fragment that corresponds to the major portion of this repeat (Lund *et al.*, 1984). These *Hind*III fragments contain all of the xU1b gene, but lack sequences >220 bp upstream of the xU1a gene. Similar clones have been independently isolated by Ciliberto *et al.* (accompanying paper). To compare the structures and template activities of these two coordinately expressed genes, we cloned a copy of the entire repeat unit, as described in Materials and methods.

Structure and sequence of the DNA

A restriction map of the repeat unit is shown in Figure 1A, along with several subclones that were used in the sequencing or tran-

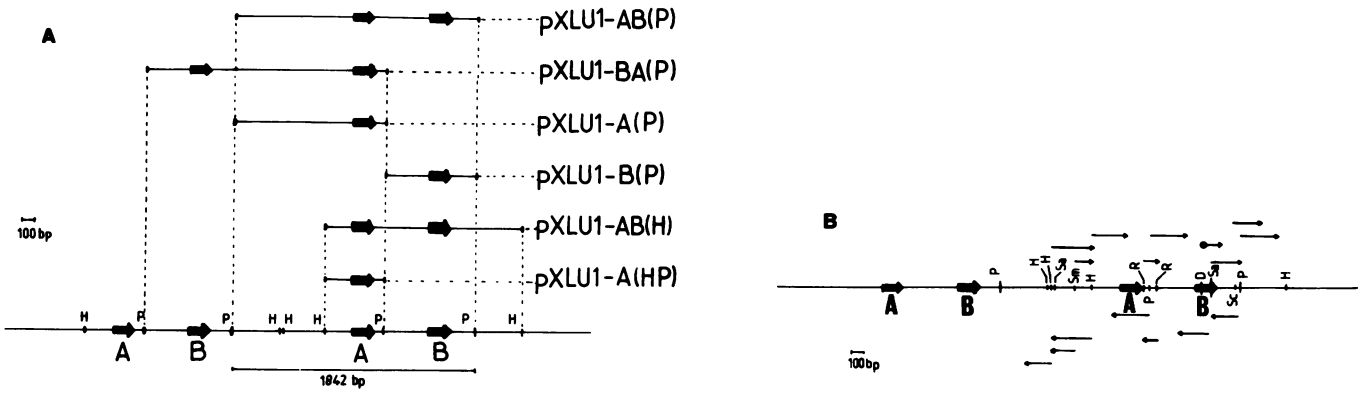

Fig. 1. Structure of the *X. laevis* embryonic U1 repeat DNA. **(A)** The general structures of the pXlU1-A and -B clones discussed in the text. The restriction enzyme cleavage map as well as clones pXlU1-A(HP) and pXlU1-AB(H) (formerly pXlU1A and pXlU1-H) were from Lund *et al.* (1984). The positions and orientations of the two U1 RNA coding regions are indicated by heavy arrows. P = *PstI*; H = *HindIII*. **(B)** Sequencing strategy. Lines above and below the map show the extents of sequences determined on the non-template and the template strands, respectively, using the method of Sanger *et al.* (1977). The filled circle indicates the end-labeled fragment that was sequenced according to the method of Maxam and Gilbert (1980). D = *DdeI*; H = *HindIII*; P = *PstI*; Sa = *SalI*; Sc = *SacI* (*SstI*); Sm = *SmaI*.

scription analyses. These clones include constructs containing both embryonic U1 RNA genes, arranged in different permutations of the repeat, as well as constructs with only one of these genes. The nucleotide sequence of the full-length repeat was determined, using the sequencing strategy shown in Figure 1B. In addition to sequencing one entire repeat unit that was representative of the majority of the cloned isolates, we also analyzed local regions of several variant clones (Lund *et al.*, 1984).

Figure 2 presents the sequence of the 1842 nucleotides which constitute the non-template strand of the *X. laevis* embryonic U1 RNA gene repeat. Apart from the U1 RNA coding regions (thin underline) and two short conserved sequences, upstream (position −55, heavy overline) and downstream (position +19, heavy underline) of both coding regions, the two genes appear to be very different in structure. As predicted from the RNase T1 fingerprints of xU1a and xU1b RNAs (Lund *et al.*, 1984), the xU1a (positions 923−1086) and the xU1b (positions 1513−1676) coding region sequences are almost identical.

Inspection of the flanking region sequences shows that, as with other snRNA genes, no TATA boxes are located ~30 nucleotides upstream of the initiation sites. Instead, both genes contain a sequence located between positions −50 and −60 that matches very closely the consensus sequence found in comparable positions of other *X. laevis* snRNA genes (Zeller *et al.*, 1984; Mattaj *et al.*, 1985b). Recent results obtained by others indicate that this sequence is essential for transcription of *X. laevis* U1 and U2 genes (Ciliberto *et al.*, accompanying paper; Mattaj *et al.*, 1985a); we propose that this conserved region functions like a similar promoter element of mammalian snRNA genes (cf. Table I) by fixing the site of transcription initiation (Skuzeski *et al.*, 1984).

About 15 nucleotides downstream of the positions corresponding to the 3′ end of the mature U1 RNAs, both genes contain a very highly conserved sequence found at the ends of almost all snRNA genes (Mattaj and Zeller, 1983; cf. Table I). For a human U1 RNA gene, we have found that this sequence is essential for correct 3′ end formation (H.Neuman de Vegvar, E.Lund and J.E.Dahlberg, in preparation); similar findings have been reported for a *X. laevis* U2 RNA gene (Mattaj and De Robertis, unpublished results, cited in Mattaj, 1984).

Direct and inverted repeats

The nucleotide sequences of the 1842-bp repeat displays several interesting features, which might function as replication or transcription control signals. A number of direct repeats are indicated in the region between positions 360 and 640 of Figure 2. One of these (wavy lines, positions 421−462) contains a regular, short nucleotide repeat with only pyrimidines on one strand. Thus, it resembles the strong nuclease S1-sensitive sites found near human U1 and U2 genes (Htun *et al.*, 1984; H.Htun, E.Lund, G.Westin, U.Pettersson and J.E.Dahlberg, in preparation). Treatment of pXlU1-AB(P) DNA with nuclease S1 resulted in cleavage at this site, as expected (A.Krol and H.Htun, unpublished data).

A repeat of particular interest, located upstream of the xU1b gene (positions 1264 and 1280), is composed of an 18-bp palindrome; it contains four consecutive copies of the sequence T-G-C-A, resulting in a continuous string of eight pyrimidine-purine doublets. Such a structure of alternating pyrimidine-purine sequences would be expected to form Z-DNA (Nordheim and Rich, 1983); the possible functional significance of this region will be discussed below.

Structures of xU1a and xU1b RNAs

The 1842-bp repeat was sequenced completely for one clone and partially for three other clones that were suspected of having slightly altered structures (clones pXlU1-H-3, -6 and -8 of Lund *et al.*, 1984). Analyses of RNAs synthesized after injection of the cloned DNA into *X. laevis* oocytes, indicated that mutant U1 RNAs were encoded by two of these latter clones (E.Lund, unpublished data). In clone H-6 part of the xU1a RNA coding region was deleted, whereas in clone H-3 the encoded xU1b RNA was normal length but migrated abnormally in non-denaturing polyacrylamide gels. The third variant, clone H-8, contained deletions in the non-coding regions of the repeat. The alterations in sequence are indicated in Figure 2.

Figure 3 shows the nucleotide sequences of the embryonic U1 RNAs, drawn in the standard secondary structure model proposed by Branlant *et al.* (1981) and Mount and Steitz (1981). The two RNAs differ from each other by only two nucleotides; as indicated, these nucleotide changes are not compensatory, so the apparent stability of the stem II structure is lower for xU1b than for xU1a RNA. That may explain the differential electrophoretic mobilities of the two RNAs observed both in non-denaturing and partial denaturing gels (Lund *et al.*, 1984; cf. Figures 4 and 6, below). The variant xU1a gene had a 23-bp deletion of the region encoding most of stem-loop II. The xU1b variant RNA had a U rather than a C in stem-loop III; in this respect, it resembles the major human U1 RNA, that also has

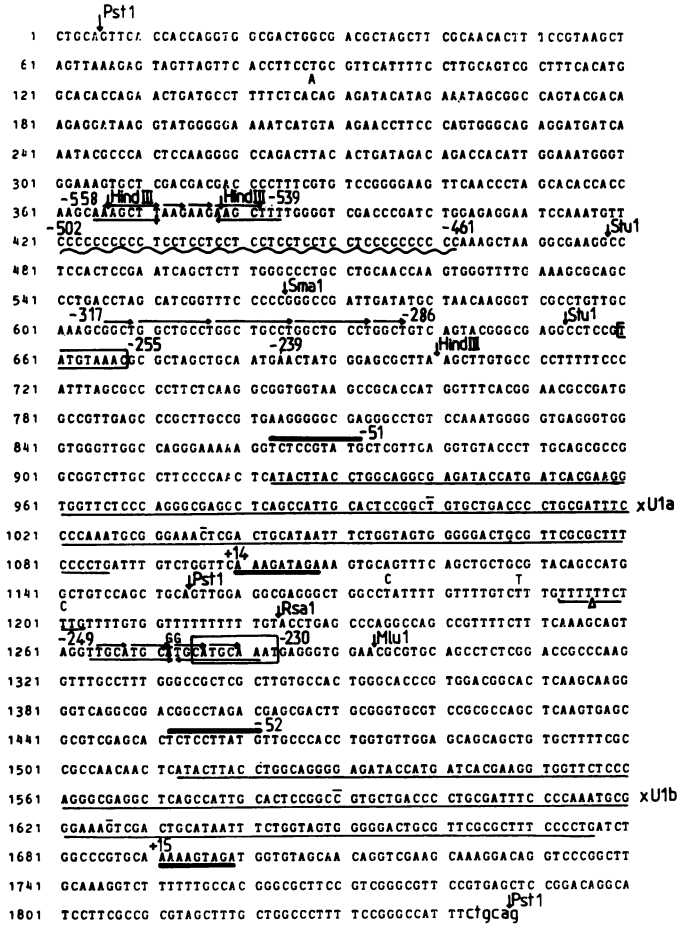

Fig. 2. DNA sequence of the non-template strand of the 1842-bp repeat encoding the two *X. laevis* embryonic xU1a and xU1b RNAs. The U1 RNA coding sequences are denoted by thin underlines; residues not conserved between the two coding regions are indicated by dashes above the sequence. Micro-heterogeneities and a deletion (△) found in other isolates of the repeat are shown under the sequence. Sequences conserved on the flanking regions of snRNA genes (cf. Table I) are indicated by the boxes and heavy underlines and overlines. Other regions, denoted by the wavy line and arrows, are discussed in the text.

a U in this position (Branlant *et al.*, 1980).

Template activities of the full-length xU1a and xU1b RNA genes

The template activities of the cloned embryonic genes were assayed by injection of the DNAs into *X. laevis* stage VI oocytes, as shown in Figure 4. Injection of templates containing the xU1b gene plus 363 bp of 5′-flanking sequences resulted in the synthesis of high levels of xU1b RNA (lanes 2 and 3). Likewise, injection of the xU1a gene template with 922 bp of 5′-flanking region led to highly efficient synthesis of xU1a RNA (lane 5). As reported earlier (Lund *et al.*, 1984), the xU1a gene with only 222 bp of upstream sequences was a poor template (lanes 2 and 4), being only ~5% active as the longer gene. Thus, in spite of their divergent 5′-flanking sequences, the xU1a and xU1b genes [as *Pst*I-subclones pXlUI-A(P) and -B(P), cf. Figure 1A] appear to be equally competent as templates (compare lanes 3 and 5). We note that the template activities of the frog U1 genes were comparable with that of our standard human U1 genes (cf. Figure 6 of Lund *et al.*, 1984; data not shown).

Competition between xU1a and xU1b RNA genes

To test whether the two embryonic genes compete equally well for the snRNA-specific transcription factor(s) that appear to be

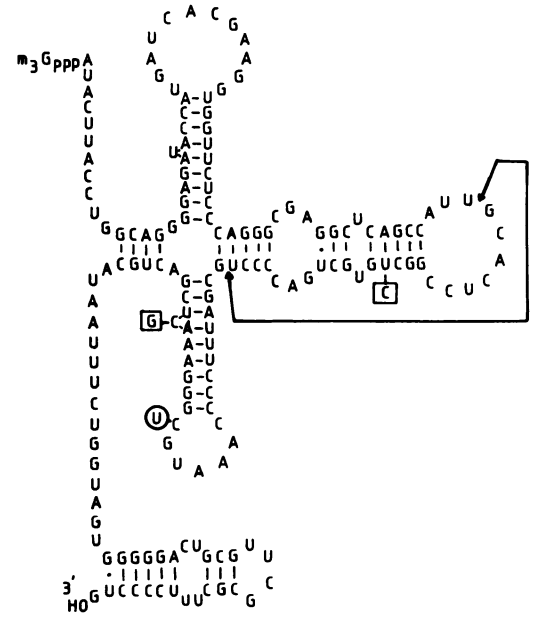

Fig. 3. Secondary structure of the *X. laevis* embryonic xU1a and xU1b RNAs. Boxed residues in the xU1a RNA sequence represent base substitutions found in xU1b RNA. The circled residue indicates a nucleotide change found in a variant of xU1b RNA. Arrows delimit the 32-bp deletion found in the coding region of a variant xU1a gene,

limiting in injected oocytes (Westin *et al.*, 1984; Ares *et al.* 1985; Mattaj *et al.*, 1985a; J.M.Skuzeski, J.T.Murphy, E.Lund and J.E.Dahlberg, in preparation) we co-injected the xU1a and xU1b genes, either as part of the same plasmid DNA or as two separate DNAs. As shown in the right hand panel of Figure 4, the xU1b gene was 5- to 10-fold more efficient than the xU1a gene when the two templates were injected in a 1:1 ratio (lane 7), whereas approximately equal amounts of xU1a and xU1b RNAs were produced when the ratio of xU1a to xU1b templates was 10:1 (lane 8). This competitive advantage of the xU1b gene is not simply a function of the overall chromatin structure of pXlU1-B(P) DNA, since injection of plasmid DNAs containing both genes in tandem (as in the naturally occurring repeats) also resulted in preferential synthesis of xU1b RNA, independent of the relative orientation of the two U1 genes (lanes 9 and 10). Controls, confirming that the individual xU1a and xU1b templates were both efficiently transcribed in the absence of competitor, are shown in lanes 11 and 12. RNase T1 fingerprint analyses of the xU1b RNAs from lanes 7, 9 and 10 established that no xU1a RNA (or xU1a degradation products) co-migrated with the xU1b RNA; furthermore, analyses of hybrid-selected U1 RNAs from the same samples confirmed that no major xU1b degradation products were present (data not shown). We conclude that the relative abundance of the two embryonic U1 RNAs resulted from preferential transcription of the xU1b gene rather than from specific breakdown of xU1a RNA under these assay conditions. The competitive advantage of the xU1b gene, therefore, is probably the result of preferred binding of one or more limiting transcription factors to a xU1b gene-specific promoter element.

5′-flanking sequences required for efficient transcription

To determine the extent of 5′-flanking sequences that were necessary for template activity, a series of subclones of the xU1a and xU1b genes were constructed as described in Materials and methods. The structures of the deleted templates are shown in

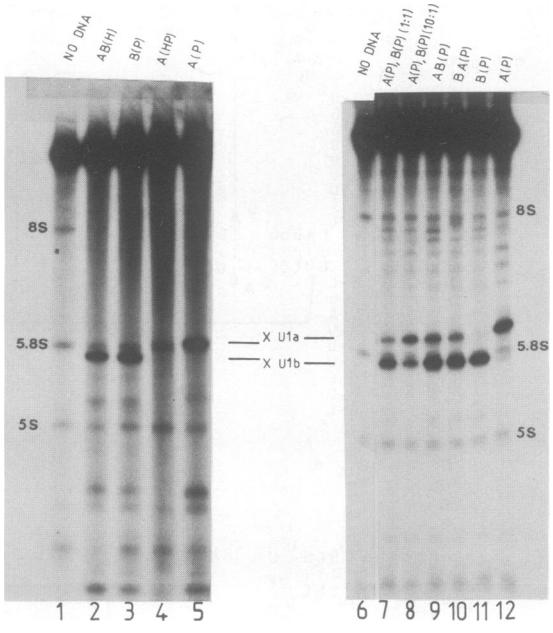

Fig. 4. Transcription of the embryonic xU1a and xU1b RNA genes after injection of cloned DNA templates into *X. laevis* oocytes. Cloned xU1a and xU1b genes were co-injected with [α-^{32}P]GTP into *X. laevis* stage VI oocytes 20 h prior to RNA extraction. Ten to 20 oocytes were injected with 8 ng per oocyte of each of the following DNA templates. pXlU1-AB(H) **(lane 2)**, pXlU1-B(P) **(lane 3)**, pXlU1-A(HP) **(lane 4)** or pXlU1-A(P) **(lane 5)**. Competition experiments between co-injected xU1a and xU1b genes were performed in two ways. First, by co-injection of xU1a and xU1b DNA templates in a 1:1 ratio, 4 ng each of the pXlU1-A(P) and pXlU1-B(P) templates per oocyte **(lane 7)** or in a 10:1 ratio of xU1a over xU1b templates, a total of 8 ng of a 10:1 mixture of the same two templates per oocyte **(lane 8)**. Second, by injection of the full-length repeat DNAs, pXlU1-AB(P) **(lane 9)** and pXlU1-BA(P) **(lane 10)**, each at 8 ng per oocyte. Separate injections of pXlU1-A(P) **(lane 11)** and pXlU1-B(P) **(lane 6)** DNAs, at 8 ng per oocyte, were included as controls. Total RNAs from pooled oocytes were analyzed in a 12% (30:0.8) polyacrylamide gel containing 7 M urea; electrophoresis was at 10–12 V/cm for 18–20 h at room temperature. Each lane contains one oocyte-equivalent of total RNAs isolated from pooled oocytes. Analyses of total RNAs isolated from five individual oocytes gave the same results (data not shown). Autoradiograms are shown. The prefix pXlU1- has been omitted for clarity of the figure.

Figure 5; as indicated, subclones were named by the restriction enzymes used to generate the ends of the cloned fragments (cf. Figure 1A). Template activities were assayed as above and the results of these analyses are shown in Figure 6.

Highly efficient transcription of the xU1a gene was obtained when the template contained 747 (PΔSS, lane 3), 358 (SmP, lane 4) or 269 (SP, lane 5) bp of 5′-flanking sequences. In all cases the template activities were comparable with that of pXlUI-A(P), containing 922 bp of upstream sequences (P, lane 2). In contrast, transcription was reduced ~20-fold when the template had only 222 bp of flanking sequences (HP, lane 6; cf. lane 4 of Figure 4). We conclude that at least part of the sequences between positions −222 and −269 are required as an enhancer for highly efficient transcription of the xU1a gene after injection into stage VI oocytes, but that this transcription activator is not absolutely essential for template activity.

In the case of the xU1b gene we find that the xU1b template containing 289 bp of upstream sequences (RP, lane 8) is as efficient as one containing the entire 5′-flanking region back to position −363 (P, lane 7). However, a subclone with only 219 bp of 5′-flanking sequences (MP, lane 9) is a very poor template, reducing the level of xU1b transcription ~20-fold. Similar observations have been made by Ciliberto *et al.* (accompanying paper). Thus, the xU1b gene contains a transcription activator which resembles that of the xU1a gene with respect to both its enhancing activity and its relative location. A more detailed comparison of the far upstream sequences of the two genes is presented in Table I.

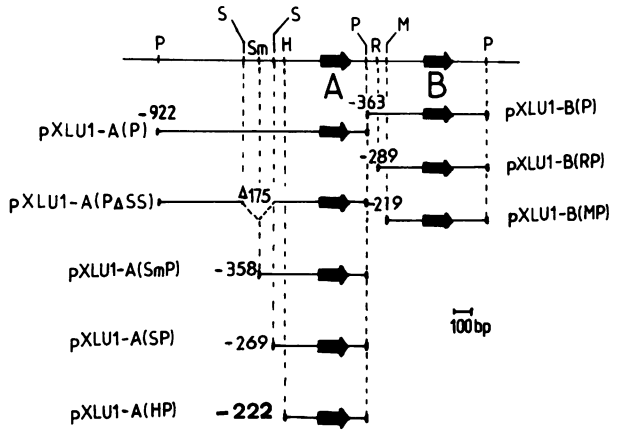

Fig. 5. Structure of xU1a and xU1b genes lacking 5′-flanking sequences. The structures of *X. laevis* DNA fragments of wild-type and deletion mutant templates used in Figure 6 are shown. The end points of the 5′-flanking sequences in the deletion mutants are indicated on the left. P = *Pst*I; H = *Hind*III; S = *Stu*I; Sm = *Sma*I; R = *Rsa*I; M = *Mlu*I.

Discussion

We have cloned and sequenced a full length repeat of *X. laevis* DNA containing the embryonic U1 RNA genes. The 1842-bp repeat unit contains one copy each of the genes for the two embryonic U1 RNAs, xU1a and xU1b, that are synthesized at the onset of transcription at the mid-blastula stage of *X. laevis* embryogenesis (Forbes *et al.*, 1984). Although these RNAs are coordinately expressed and accumulate at comparable levels *in vivo*, their flanking sequences show very little homology with each other.

The two embryonic U1 RNAs differ in only two of their encoded 164 nucleotides and from their predicted secondary structures (Figure 3) one would expect that xU1a would have a more compact structure than XU1b. Nevertheless, xU1a RNA has a lower electrophoretic mobility than xU1b RNA, both in nondenaturing (Lund *et al.*, 1984) and partially denaturing polyacrylamide gels (Figures 4 and 6) indicating that these U1 RNAs have a considerable amount of tertiary structure, that is not immediately evident from the folded structure in Figure 3.

Sequence heterogeneities among various cloned isolates of the U1 repeat

Previously, we showed that several independent clones of the 1.5-kb *Hind*III fragment of the U1 repeat DNA were similar, but not identical to each other (Lund *et al.*, 1984); nucleotide sequence analyses of three of these clones confirmed the presence of sequence difference, located both inside and outside the U1 coding regions. Additional minor sequence heterogeneities were detected when the sequences of Figure 2 were compared with that of the 1.5-kb *Hind*III fragment characterized by Ciliberto *et al.* (accompanying paper).

One variant xU1a gene of particular interest had a 23-bp deletion in the coding region sequences, resulting in the loss of most of the stem-loop II region in the encoded RNA (Figure 3). Although injection of this mutant gene into oocytes leads to

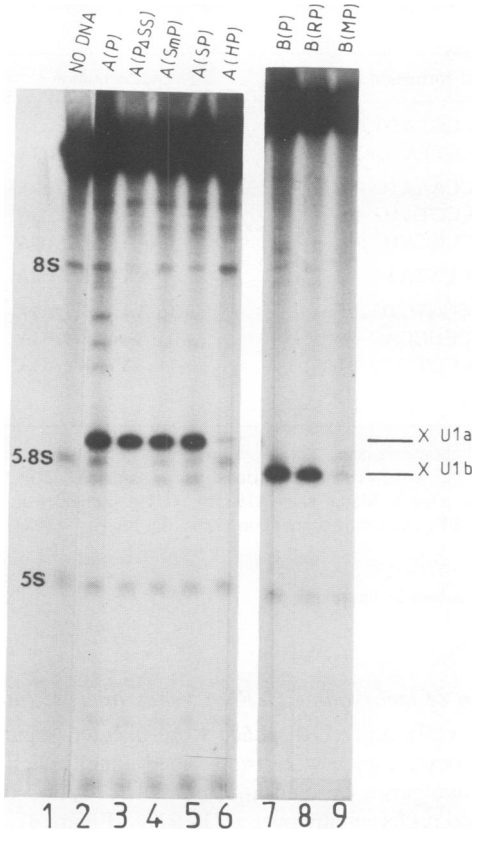

Fig. 6. Deletion mapping of far-upstream transcription activators in the 5′-flanking regions of the xU1a and xU1b genes. The template activities of the deletion mutants shown in Figure 5 were assayed by injection of 8 ng of template DNA per oocyte, as above. Total RNAs from pooled **(lanes 1–6)** or single **(lanes 7–9)** oocytes were isolated and analyzed as in Figure 4. The xU1a gene templates were pXlU1-A(P) **(lane 2)**, -A(PΔSS) **(lane 3)**, -A(SmP) **(lane 4)**, -A(SP) **(lane 5)** and -A(HP) **(lane 6)**. The xU1b genes tested were pXlU1-B(P) **(lane 7)**, -B(RP) **(lane 8)** and -B(MP) **(lane 9)**.

accumulation of the correspondingly shorter xU1a RNA, we feel it is unlikely that this is a true gene. We are currently investigating whether the RNA product can associate with snRNP proteins.

Most of the sequence variants that we detected in the flanking regions were located downstream of the xU1a gene, between positions 1115 and 1214 (Figure 2). This region may not be under strong selective pressure to remain homogeneous, since it is not required for transcription of either gene (cf. Figure 6).

Unusual features of the repeat sequence

The major 1842-bp unit contains several direct and inverted repeats whose functions are unknown. A repetitive sequence that is sensitive to nuclease S1 digestion resembles S1-sensitive sites found near human U1 and U2 snRNA genes (Htun *et al.*, 1984; H.Htun, E.Lund, G.Westin, U.Pettersson and J.E.Dahlberg, in preparation). Since the S1 sites of snRNA genes are not essential for transcription in oocytes (Murphy *et al.*, 1982; Westin *et al.*, 1984; Figure 6, this paper) or in transfected mammalian cells (Schenborn *et al.*, 1985), it is unclear whether any of these cleavage sites correspond to nuclease S1-sensitive sites detected in active chromatin of other genes (Larsen and Weintraub, 1982; Glikin *et al.*, 1983; Schon *et al.*, 1983).

A search of the repeat unit sequence for regions that could possibly encode proteins revealed two long open reading frames (ORFs). ORF 1 starts with an ATG codon at position 138 and extends 166 codons to a TAA at position 1638; the coding region strand is the same as that coding for the U1 RNAs. ORF 2 starts with an ATG at position 1273 and extends 191 codons to TAA at position 700; the coding strand of this ORF is the non-coding strand of the U1 RNAs. We have no evidence that these sequences are transcribed or translated. We note that neither sequence contains a polyadenylation signal (Proudfoot and Brownlee, 1974); in addition, the putative protein genes overlap regions that exhibit substantial variability (deletion between positions 990 and 1012, and point mutations and small deletions between 1115 and 1213).

Homologous sequences in the U1 RNA gene flanking regions

When the flanking region sequences of the two embryonic genes were compared, only two short blocks were detected that were homologous both in sequence and position; these sequences of the embryonic genes, centered around position −55 and +19 (Figure 2), fit very well the consensus sequences noted earlier for *X. laevis* snRNA genes (Mattaj and Zeller, 1983; Zeller *et al.*, 1984). A third region of homology, found upstream of positions −220, was identified on the basis of template activities in injected oocytes (see below); the position of this consensus sequence varies between snRNA genes (cf. Table I).

We have previously demonstrated that a function of the consensus sequence located around position −55 of mammalian snRNA genes (cf. Table I) is to fix the site of transcription initiation (Skuzeski *et al.*, 1984). By analogy, we propose that the *X. laevis* consensus sequence T-C-T-C-C-N-T-A-T-G (positions −51 to −60) is a 'TATA-like' promoter element required for correct 5′ end formation. Recent studies have indicated that this sequence is essential for transcription of both the xU1b gene (Ciliberto *et al.*, accompanying paper) and the xU2-5 gene (Mattaj *et al.*, 1985a).

The downstream consensus sequence, A-A-A-Pu-Pu-Py-A-G-A, which appears to be more conserved in U1 and U2 genes than the −55 consensus sequence (cf. Table I) functions in the formation of the 3′ ends of snRNAs. Deletions of two or more nucleotides from the 3′ end of this sequence in a human U1 RNA gene results in accumulation of elongated U1 RNA transcripts (H.Neuman de Vegvar, E.Lund and J.E.Dahlberg, in preparation). Comparable results have been obtained with the *X. laevis* U2 RNA gene (Mattaj and De Robertis, unpublished results, cited in Mattaj, 1984) and other U1 and U2 RNA genes of humans (Hernandez, 1985; Yuo *et al.*, 1985).

Comparison of sequences required for efficient transcription of the xU1a and xU1b genes

In addition to the two consensus sequences discussed above, both embryonic U1 genes require sequences located >220 bp upstream for efficient transcription (Figure 6). Similar transcription activators have also been identified in human and *X. laevis* U1 and U2 genes (Skuzeski *et al.*, 1984; Westin *et al.*, 1984; Ares *et al.*, 1985; Ciliberto *et al.*, accompanying paper; Mattaj *et al.*, 1985a). As the transcription experiments presented here did not establish the sequences responsible for activation, we compared the required 5′-flanking sequences of several U1 and U2 genes (Table I). We propose that a common block of nine nucleotides, Py-A-T-G-Py-A-A-A-T, is the core of the snRNA gene transcription activator (Table I). Similar conclusions have been reached by Early *et al.* (1984), Ares *et al.* (1985), Mattaj *et al.* (1985a) and Ciliberto *et al.* (accompanying paper). As illustrated in Table I and Figure 2, sequences matching this consensus are found around positions −260 and −230 of the xU1a and xU1b genes, respectively. In the human U1 gene, this con-

Table I. Consensus sequences in the 5′- and 3′-flanking regions of U1 and U2 snRNA genes

	Transcription enhancer	5′ end formation	3′ end formation
xU1a	TATGTAAAC -255	TCTCCGTATG -51	+14 AAAGATAGA
xU1b	.tgcatgCATGCAAAT -230	TCTCCTTATG -52	+15 AAAAGTAGA
xU1.3	CATGTAAAT -225	TCTCCAAATG -51	+12 AATGACAGA
xU1.8	NA	TCTCCGTATG -51	+15 AAAGATAGA
xU2-5	TATGCAAAT -262[b]	TCTCCCCATG-51	+16 AAAAGCAGA
X. laevis consensus	**YATGYAAAT**	**TCTCCNTATG**	**AAARRYAGA**
hU1-1	TATGTAGAT -212	TGACCGTGTG -52	+16 AAAAGTAGA
hU2.6[a]	CATGCAAAT -214[b]	TCACCGCGAC -50	+25 AAAAGTAGA
Chicken/mammal	YATGCARAT -219	TCGCCGTGCG -51	+14 AAYRRYAGA[c]
SV40 enhancer	TATGCAAAGcatgcat		

Three blocks of homologous sequences are shown, located upstream of position −200 (transcription enhancer) and around positions −55 and +19 (5′ and 3′ end formation, respectively). Functional studies have been performed on many, but not all, of the sequences of the various genes, as discussed in the text. The sequences shown were obtained from the following references: xU1a and xU1b, this work; xU1.3, Mattaj *et al.* (1985b); xU1.8, Zeller *et al.* (1984); xU2-5, Mattaj and Zeller (1983); hU1-1, Lund and Dahlberg (1984); hU2.6, Westin *et al.* (1984); chicken/mammal consensus, Early *et al.* (1984); SV40 enhancer, Benoist and Chambon (1981).

[a]This gene is identical in sequence to the hU2.24 gene studied by Ares *et al.* (1985).

[b]These sequences were shown by others to be part of functional transcription activators, as discussed in the text.

[c]This is the chicken U1 consensus derived by us from the data of Early *et al.* (1984).

NA = not available. Y, R and N indicate pyrimidines, purines and any nucleotide, respectively.

sensus sequence is centered around position −215 and is flanked by a 10−12 bp direct repeat, as discussed earlier (Skuzeski *et al.*, 1984). In the case of both the human and the *X. laevis* U2 genes, there is strong evidence that at least part of this consensus sequence is required for transcription enhancer function (Ares *et al.*, 1985; Mattaj *et al.*, 1985a).

The flanking sequences of human U1 and U2 genes show good homology with each other and with the *X. laevis* genes in the three consensus regions (Table I); so far, these three regions are the only ones that have been shown to be essential for snRNA transcription. Thus, it is not surprising that the two human genes are efficiently transcribed in *X. laevis* oocytes. We are unable to explain the disrepancy between our results and those of Ciliberto *et al.* (accompanying paper) who report a significant difference between the template activities of the HU1-1 and the xU1b genes. Very subtle factors, such as the age or physiological state of the ooctyes might influence the transcription efficiencies of heterologous genes [as noticed by Skuzeski *et al.* (1984) and Westin *et al.* (1984)].

In the xU1b gene the consensus sequence partially overlaps an 18-bp inverted repeat (Table I) which is composed of the 4-bp sequence, T-G-C-A, repeated four times. In the 72-bp enhancer region of SV40 DNA, two repeats of a similar sequence (G-C-A-T) abuts the sequence that matches the snRNA gene core consensus (cf. Table I); this 8-bp repeat of alternating purine-pyrimidine nucleotides is able to form Z-DNA (Nordheim and Rich, 1983). It seems very likely that the 16-bp alternating repeat on the upstream side of the xU1b consensus sequence, as well as sequences 10-bp downstream of the core sequence (cf. Figure 2), can also form Z-DNA. Likewise, as discussed by Early *et al.* (1984), several chicken U1 genes have extended stretches of alternating purine-pyrimidine doublets located downstream of their core consensus sequence. Although regions capable of forming Z-DNA may well augment the transcription enhancing activity of the consensus sequence, they do not appear to be essential for its activity since several of the snRNA genes listed in Table I lack nearby sequences with the potential for Z-DNA formation. It is, however, notable that the consensus transcription activator itself consists of several purine-pyrimidine repeats.

Expression of embryonic U1 RNA genes injected into oocytes

When the xU1a and xU1b genes were injected separately into stage VI oocytes, they were transcribed equally well, whereas when the two genes were co-injected in a 1:1 ratio, xU1b RNA accumulated preferentially over xU1a RNA (Figure 4). This competitive advantage of the xU1b gene over the xU1a gene is in marked contrast to the situation *in vivo*, during early embryogenesis, when the level of newly synthesized xU1a RNA is several fold higher than that of xU1b RNA (Forbes *et al.*, 1984). This preferential expression of the xU1b gene in injected oocytes most likely results from competition for one or more limiting snRNA-specific transcription factors (Westin *et al.*, 1984; Ares *et al.*, 1985; Mattaj *et al.*, 1985a; J.M.Skuzeski, J.T.Murphy, E.Lund and J.E.Dahlberg, in preparation). It is possible that the ability of the xU1b gene to compete successfully is due to the presence of sequences that could form Z-DNA, e.g., the 18-bp repeat discussed above. The xU1a gene does not have such long alternating pyrimidine-purine stretches, although it does have one copy of the 4-bp repeat located near a sequence that shares weak homology with the consensus sequence (positions −249 to −239, cf. Figure 2).

The reason that competition between the xU1a and xU1b genes (or between U1 and other snRNA genes) is observed at all may simply be that a vast excess of snRNA genes is injected into oocyte nuclei that have limited amounts of snRNA-specific transcription factor(s). Such U1 gene competition would not be evident under normal *in vivo* conditions, if the required transcription factor(s) are in excess over the endogenous levels of snRNA genes. Under these latter conditions, the efficiency of transcription initiation, rather than the affinity of factor binding, could determine the levels of embryonic xU1a and xU1b RNA synthesis.

It still remains unknown how the expression of the embryonic U1 genes is controlled during the various stages of development. The coordinate accumulation of the two encoded RNAs indicates that control operates at the level of activation (or inactivation) of the repeat units of entire tandem array(s), rather than at the level of the individual U1 genes. Presumably, some of the non-coding sequences in the repeat unit function in that process; the

identification of such sequences would be facilitated by the development of efficient transcription systems from cells that normally express these genes.

Materials and methods

Cloning of the full-length X. laevis U1 DNA repeat encoding the embryonic U1 RNAs

X. laevis genomic DNA enriched in the major (embryonic) U1 RNA genes was obtained as previously described (Lund *et al.*, 1984). Briefly, the total *X. laevis* erythrocyte DNA (a generous gift of D.D.Brown) was digested to completion with a mixture of *Bam*HI, *Bgl*II and *Eco*RI and size-fractionated on a NaCl gradient. Fractions containing large (25 kb) U1 DNA fragments were pooled and these DNA fragments were re-digested with *Pst*I to generate the 1.2-kb *Pst*I fragments containing the xU1a genes (cf. Figure 1). These *Pst*I fragments were purified by preparative agarose gel electrophoresis and ligated to *Pst*I-cut pBR322 DNA. Isolates of the xU1a gene were obtained by transformation of *Escherichia coli* strain HB101 and screening of tetracycline-resistant colonies by hybridization with a human U1 DNA probe (Grunstein and Hogness, 1975; Lund and Dahlberg, 1984).

The 0.69-kb *Pst*I fragment containing the xU1b gene was obtained by ligation of *Pst*I-cut pBR322 DNA to a *Pst*I digest of clone pXlU1-AB(H) DNA [formerly called pXlU1-H in Lund *et al.* (1984; cf. Figure 1)]. Isolates of the xU1b gene [pXlU1-B(P)] were identified by screening of tetracycline-resistant colonies as above.

Reconstruction of the full-length repeat (cf. Figure 1A) was carried out as follows: pXlU1-A(P) and pXlU1-B(P) DNAs were digested to completion with *Bam*HI (cutting only once, in the pBR322 DNA) and then partially cleaved with *Pst*I. From each digest, the two kinds of *Pst*I-*Bam*HI fragments containing the xU1 genes were purified by preparative agarose gel electrophoresis. The smaller *Pst*I-*Bam*HI fragment containing the xU1a (or xU1b) gene was then religated to the larger *Pst*I-*Bam*HI fragment carrying the other U1 gene, i.e., xU1b (or xU1a), so that the resulting recombinant plasmids contained both genes, either xU1a upstream of xU1b [pXlU1-AB(P)] or the reverse [pXlU1-BA(P)] (cf. Figure 1A).

Selection of the reconstructed plasmids was achieved by transformation of HB101 and screening for tetracycline-resistant cells.

Construction of DNA templates for transcription assays

To generate the subclones pXlU1-A(SmP) and -A(SP) (cf. Figure 5), pXlU1-A(P) DNA was digested to completion with *Sma*I or partially with *Stu*I and *Eco*RI linkers were inserted into the *Sma*I (position −358) or *Stu*I (position −269) cleavage sites (position +1 corresponds to the 5′ end of the xU1a coding region, cf. Figure 2). After ligation the DNAs were re-cut with *Eco*RI and re-circularized at low DNA concentration to favor the formation of intramolecular linkages. This DNA was used to transform HB101 cells which were then screened for tetracycline resistance. pXlU1-A(PΔSS), which lacks the sequences between the two *Stu*I sites (positions −269 and −444), was obtained by recircularization of *Stu*I-cut pXlU1-A(P) DNA at low DNA concentration (as above).

Subclone pXlU1-B(RP) of pXlU1-B(P) was obtained by partial *Rsa*I digestion of pXlU1-B(P) DNA followed by re-ligation at low DNA concentration. After re-cutting with *Pvu*I (to destroy remaining full-length pXlU1-B(P) circular DNAs) the mixture was used for transformation of HB101; selection was for tetracycline-resistant cells. To generate subclone pXlU1-B(MP), pXlU1-B(P) DNA was cleaved with *Mlu*I and the 5′-overhanging ends were filled in using the Klenow fragment of DNA polymerase. *Eco*RI linkers were then inserted at the *Mlu*I site (position −219) and the same procedure as above was used for screening.

DNA sequencing

Restriction fragments from pXlU1-AB(H) and pXlU1-A(P) DNAs were subcloned into M13-mp8, -mp9 or -mp19 vector DNAs (Messing and Vieira, 1982) and sequenced using the dideoxy-chain termination method (Sanger *et al.*, 1977) with [α-^{35}S]dATP as the label. A few fragments were also sequenced using the chemical cleavage technique of Maxam and Gilbert (1980). The sequencing strategy is described in the legend to Figure 1B.

Transcription activity analysis

Injection of recombinant plasmid DNAs into stage VI *X. laevis* oocytes was performed as previously described (Murphy *et al.*, 1982; Skuzeski *et al.*, 1984). [α-^{32}P]GTP was used as label and supercoiled DNAs (at 4−8 ng per oocyte) containing both *X. laevis* and vector DNA sequences were used as templates. Total nucleic acids, extracted from individual (Westin *et al.*, 1984) or pooled oocytes after 20 h of incubation, were analyzed by electrophoresis in 12% (30:0.8) polyacrylamide gels containing 7 M urea (Murphy *et al.*, 1982; E.Lund, unpublished).

Nuclease S1 treatment

Nuclease S1 digestion of supercoiled pXlU1-AB(P) DNA was carried out as described by Htun *et al.* (1984). The cleavage site was mapped by redigestion of the nuclease S1 digestion products with *Hind*III, followed by polyacrylamide gel electrophoresis.

Acknowledgements

We are grateful to James T.Murphy and Henry Neuman de Vegvar for their valuable assistance in the microinjection experiments, to Han Htun for many stimulating discussions and help in the nuclease S1 experiments and to Jacque L.Mitchen for skilled technical assistance. We also thank Philippe Carbon for helpful comments and Genaro Ciliberto, Lennart Philipson, Iain Mattaj, Eddy De Robertis, Emanuel Ares and Alan Weiner for communicating their results prior to publication. This work was supported by NIH grant GM 30220 and NSF grant PCM 83-09618 to J.E.D. and E.L. A.K. was supported by grants from a NIH-C.N.R.S. program for Scientific Collaboration.

References

Ares,M., Mangin,M. and Weiner,A.W. (1985) *Mol. Cell. Biol.*, in press.
Benoist,C. and Chambon,P. (1981) *Nature*, **295**, 622-624.
Branlant,C., Krol,A., Ebel,J.-P., Lazar,E., Gallinaro,H., Jacob,M., Sri-Wadada, J. and Jeanteur,P. (1980) *Nucleic Acids Res.*, **8**, 4143-4154.
Branlant,C., Krol,A., Ebel,J.-P., Gallinaro,H., Lazar,E. and Jacob,M. (1981) *Nucleic Acids Res.*, **9**, 841-858.
Busch,H., Reddy,R., Rothblum,L. and Choi,Y.C. (1982) *Annu. Rev. Biochem.*, **51**, 617-654.
Early,J.M., III, Roebuck,K.A. and Stumph,W.E. (1984) *Nucleic Acids Res.*, **12**., 7411-7421.
Forbes,D.J., Kornberg,T.B. and Kirschner,M.W. (1983) *J. Cell. Biol.*, **97**, 62-72.
Forbes,D.J., Kirschner,M.W., Caput,D., Dahlberg,J.E. and Lund,E. (1984) *Cell*, **38**, 681-689.
Fritz,A., Parisot,R.F., Newmeyer,D. and De Robertis,E.M. (1984) *J. Mol. Biol.*, **178**, 273-285.
Glikin,G.C., Gargiulo,G., Rena Descalzi,L. and Worcel,A. (1983) *Nature*, **303**, 770-774.
Grunstein,M. and Hogness,D.S. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 3961-3965.
Hernandez,N. (1985) *EMBO J.*, in press.
Htun,H., Lund,E. and Dahlberg,J.E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7288-7292.
Larsen,A. and Weintraub,H. (1982) *Cell*, **29**, 609-622.
Lund,E. and Dahlberg,J.E. (1984) *J. Biol. Chem.*, **259**, 2013-2021.
Lund,E., Dahlberg,J.E. and Forbes,D.J. (1984) *Mol. Cell. Biol.*, **4**, 2580-2586.
Mattaj,I.W. (1984) *Trends Biochem. Sci.*, **9**, 435-437.
Mattaj,I.W. and Zeller,R. (1983) *EMBO J.*, **2**, 1883-1891.
Mattaj,I.W., Lienhard,S., Jiricny,J. and De Robertis,E.M. (1985a) *Nature*, in press.
Mattaj,I.W., Zeller,R., Carrasco,A.E., Jamrich,M., Lienhard,S. and De Robertis, E.M. (1985b) *Oxford Survey on Eukaryotic Genes*, Oxford University Press, Vol. 2, in press.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.
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.
Nordheim,A. and Rich,A. (1983) *Nature*, **303**, 674-679.
Proudfoot,N.J. and Brownlee,G.G. (1974) *Nature*, **252**, 359-362.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Schenborn,E., Lund,E. and Dahlberg,J.E. (1985) *Mol. Cell. Biol.*, in press.
Schon,E., Evans,T., Welsh,J. and Efstratiadis,A. (1983) *Cell*, **35**, 837-848.
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.
Westin,G., Lund,E., Murphy,J.T., Pettersson,U. and Dahlberg,J.E. (1984) *EMBO J.*, **3**, 3295-3301.
Yuo,C.-Y., Ares,M. and Weiner,A.W. (1985) *Cell*, in press.
Zeller,R., Carri,M.-T., Mattaj,I.W. and De Robertis,E.M. (1984) *EMBO J.*, **3**, 1075-1081.

Received on 29 March 1985