The EMBO Journal vol.4 no.1 pp.205-211, 1985

Polyadenylation of the *Xenopus* β1 globin mRNA at a downstream minor site in the absence of the major site and utilization of an AAUACA polyadenylation signal

P.J.Mason, M.B.Jones[1], J.A.Elkington and J.G.Williams

Imperial Cancer Research Fund, Mill Hill Laboratories, Burtonhole Lane, London NW7 1AD, and [1]Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

Communicated by D.Glover

We show that adult reticulocytes of *Xenopus laevis* produce two forms of the β1 globin mRNA that differ in their site of polyadenylation. The minor site of polyadenylation is located 46 nucleotides downstream of the major site and is used in ~1% of mRNA molecules. A fusion gene was constructed containing the promoter from the thymidine kinase gene of herpes simplex virus fused to the protein coding, 3'-non-coding and 3'-flanking sequences of the *X. laevis* β1 globin gene. When injected into the nuclei of *Xenopus* oocytes, transcripts of this fusion gene were accurately and efficiently spliced and polyadenylated. The proportion of fusion gene transcripts terminating at the major and minor polyadenylation sites after injection into oocytes was approximately similar to that found in reticulocytes. When the AATAAA sequence element upstream from the major site was deleted, the minor site was used with a high (>90%) efficiency. Therefore, by comparing the ratio of polyadenylation at the major and minor sites, it is possible to determine the effect of sequence alterations at the major site. In a construct where the AATAAA polyadenylation signal was changed to AATACA a high proportion (35%) of transcripts continued to be polyadenylated at the major site. This suggests a surprisingly high degree of flexibility in the precise polyadenylation signal.

Key words: in vitro mutagenesis/polyadenylation signal/*Xenopus* β1 globin mRNA/*Xenopus* oocyte microinjection

Introduction

Transcription termination and 3' processing in eukaryotic mRNAs are not well characterized (reviewed by Proudfoot, 1984). Transcription of several viral genes (Ford and Hsu, 1978; Nevins *et al.*, 1980) and the mouse β-globin gene (Hofer *et al.*, 1982) proceeds beyond the 3' end of the mature mRNA. If this is a general phenomenon, then 3' end formation must involve two steps; the RNA must be cleaved at the polyadenylation site and a poly(A) tail must then be added. The sequence AAUAAA (Proudfoot and Brownlee, 1976) or the related sequence AUUAAA (Hagenbuchle *et al.*, 1980; Jung *et al.*, 1980) is found between 11 and 30 nucleotides upstream from the polyadenylation site of most eukaryotic mRNAs. Since it is the only stringently conserved feature of 3'-non-coding sequences, this hexanucleotide was postulated to form part of a signal for polyadenylation (Proudfoot and Brownlee, 1976). This was confirmed by Fitzgerald and Shenk (1981) who showed that polyadenylation was blocked in SV40 deletion mutants lacking this hexanucleotide. Single base pair changes within this hexanucleotide have been shown to reduce severely (Montell *et al.*, 1983) or abolish normal polyadenylation (Higgs *et al.*, 1983).

The hexanucleotide AAUAAA cannot, however, be the only recognition signal for the cleavage/polyadenylation process. The sequence occurs unrecognized in some coding regions (Fitzgerald and Shenk, 1981) and, in one case, it is situated in a 3'-non-coding region between two polyadenylation signals which are utilized (Tosi *et al.*, 1981). Moreover, in several systems, tissue-specific or developmentally regulated differences in the site of polyadenylation of a single mRNA have been demonstrated (Rogers *et al.*, 1980; Capetanaki *et al.*, 1983; Rosenfeld *et al.*, 1983).

We have chosen to study mRNA polyadenylation by injecting normal and mutant genes into the nuclei of *Xenopus* oocytes. Here we show that the *Xenopus* major β-globin mRNA is efficiently spliced and polyadenylated in oocytes, that a minor polyadenylation site is used efficiently in the absence of the major site, and that the sequence AAUACA directs polyadenylation with an efficiency only one third that of the unmutated sequence.

Results

Transcripts of a thymidine kinase/β-globin fusion gene are efficiently processed in Xenopus oocytes

The aim of these experiments was to use the well characterized major adult β-globin gene of *X. laevis* (Patient *et al.*, 1983) and microinjection of *Xenopus* oocytes to study sequences required for accurate 3'-terminal processing. The β1 globin gene is, however, not accurately transcribed in oocytes (Bendig and Williams, in preparation) and so we fused the β1 globin gene to a promoter known to be active – that of the thymidine kinase (TK) gene from herpes simplex virus (McKnight and Gavis, 1980). To facilitate subsequent oligonucleotide-directed mutagenesis, the fusion gene (pMTKβS1:2, abbreviated to pS1:2, Figure 1) was constructed using the single-stranded bacteriophage M13mp8 as vector (Messing and Vieira, 1982). In the fusion gene, the bulk of the 5'-non-coding region is from the TK gene but the ATG initiation codon derives from the *Xenopus* β1 globin gene (Figure 1).

The fusion gene, in either single-stranded or double-stranded form, was injected into germinal vesicles of *X. laevis* oocytes. After incubation of the oocytes, total nucleic acid was extracted and nuclease S1 mapping was used to analyse the 5' end of the mRNA (Figure 2A). The results show that the fusion gene is efficiently transcribed and that about half of the mRNA molecules have the correct 5' end – the remainder initiating upstream of the TK cap site. Splicing of the fusion gene transcripts was also analysed by S1 mapping. The gene contains two introns (Patient *et al.*, 1983) and we have examined splicing at three of the four splice junctions. The data presented in Figure 2B show that the splice at the boundary between intron 1 and exon 2 is made correctly and the data presented in Figure 2C show that both splice junctions in intron 2 are processed correctly.

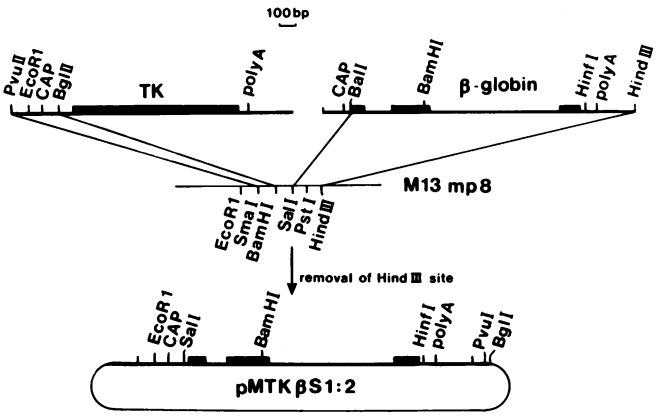

Fig. 1. Construction of a TK/β1 globin fusion gene. The TK promoter was fused to the *X. laevis* β1 globin gene in the vector M13mp8 to generate pMTKβS1. The enzymatic steps involved are described in Materials and methods. pMTKβS1:2 was used as the 'wild-type' gene in oocyte injections and as the starting point for *in vitro* mutagenesis. Restriction enzyme cleavage sites used in plasmid construction and in the preparation of S1 nuclease mapping probes are shown.

From experiments in which an S1 mapping probe was used both to analyse RNA from injected oocytes, and a known amount of RNA from adult blood (e.g., Figures 2 and 4) we calculate that each injected oocyte contains the equivalent of 0.5–2.5 ng of β-globin RNA. This is between one tenth and one half of the concentration of TK transcripts found by McKnight *et al.* (1981) after microinjection of a plasmid containing the intact TK gene.

The fusion-gene transcripts are correctly polyadenylated in oocytes

The site of polyadenylation in the β1 globin mRNA has been deduced from the nucleotide sequence adjacent to a segment of the poly(A) tail contained within the cDNA clone pXG8D2 (Kay *et al.*, 1980; Williams *et al.*, 1980). The nucleotide sequence of the 3′-proximal region of the β1 globin gene is known (Patient *et al.*, 1983) and the site of polyadenylation is shown in Figure 3. The probe used in analyzing splicing of intron 2 (Figure 2c) was a fragment from the cDNA clone pXG8D2. This fragment contains a portion of the poly(A) tail and the size of the protected fragment shows that the transcripts must also be correctly terminated. We have used a fragment from the β1 globin gene to confirm this result and this experiment reveals the existence of a minor downstream polyadenylation site which is present both in reticulocyte RNA and in the fusion gene transcripts.

Total cytoplasmic RNA, isolated from erythrocytes of animals which had been rendered anaemic with phenylhydrazine, was used in S1 mapping with a single-stranded probe from the β1 globin gene (Figure 4). The major protected fragments were between 101 and 109 nucleotides in length, which approximates the size expected for an RNA processed at the polyadenylation site determined from the pXG8D2 cDNA clone. (A cluster of S1 nuclease-resistant fragments was observed at the major site in all S1 nuclease mapping experiments. We assume this results from 'breathing' in the AT-rich region adjacent to the polyadenylation site. Elevating the stringency of the digestion conditions increased the proportion of smaller fragments and this is consistent with such an explanation). In heavily exposed autoradiograms there was an additional band derived from a fragment of 151 nucleotides in length. We estimate that mRNA processed at this site (minor site 1) is ~100-fold less abundant than the major species (Table I). This RNA is polyadenylated

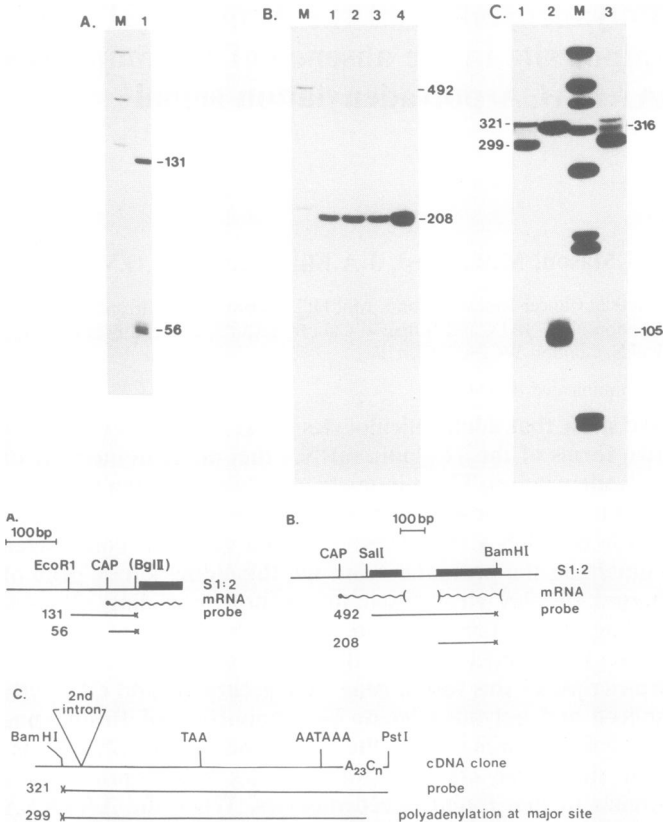

Fig. 2. Transcription and RNA processing in oocytes injected with pS1:2. 4 ng of pS1:2 double-stranded DNA was injected into oocytes and total nucleic acid was extracted after 1 day. Nuclease S1 mapping was carried out as described in Materials and methods and the products of the S1 reactions resolved on 10% **(A)**, 5% **(B)** or 7.5% **(C)** denaturing polyacrylamide gels. **(A) Lane 1.** Two oocyte equivalents of RNA was hybridized with 10 ng of the single-stranded 131 base *Bgl*II/*Eco*RI fragment that spans the cap site of the TK gene. The probe was 5′ end-labelled at the *Bgl*II site. Since, in other experiments (e.g., Figure 2B), 5′ end-labelled probes give no full length protected fragment the labelled fragment of size 131 is due to hybridization with oocyte RNA. The experiment shows that about half of the RNA molecules have the correct cap site whilst the remainder initiate transcription upstream. A *Hinf*I digest of pAT153 was used as a mol. wt. marker. **(B) Lanes 1, 2** and **3.** Two oocyte equivalents of RNA were hybridized with 5, 10 and 20 ng, respectively, of the single-stranded 492 base *Bam*HI/*Sal*I fragment from pS1:2 which spans the first intron/exon junction. The probe was 5′ end-labelled at the *Bam*HI site (10^7 c.p.m./μg). **Lane 4.** 340 ng of total cytoplasmic RNA from adult reticulocytes with 10 ng of the same probe. The result indicates that all the RNA we can detect in oocytes is accurately cleaved at the first intron/exon junction. **(C) Lane 1.** Three oocyte equivalents of RNA were hybridized with an excess of the single-stranded 321 nucleotide *Bam*HI/*Pst*I fragment from cDNA clone 8D2, 3′ end-labelled at the *Bam*HI site. The protected fragment of 299 nucleotides in length derives from molecules in which the 2nd intron is accurately removed from the molecule by splicing and which also are polyadenylated at the same base as the mRNA molecule represented in the cDNA clone. Any RNA with an unspliced second intron is not detected by this probe, therefore we cannot estimate the proportion which is correctly spliced. **Lane 2.** Three oocyte equivalents of RNA hybridized with an excess of the single-stranded 316 base *Hinf*I/*Hind*III fragment from the 3′ end of the β-globin gene (see Figure 1). Analysis of this RNA with a 'read through' probe (Figure 4C) indicates that the band at 316 nucleotides is almost entirely renatured probe. **Lane 3.** 420 ng of total cytoplasmic RNA from adult reticulocytes hybridized with the probe described in C1. This experiment shows that essentially all the RNA detected in injected oocytes is accurately polyadenylated, and that at least a proportion of the RNA molecules have been accurately spliced at the second intron.

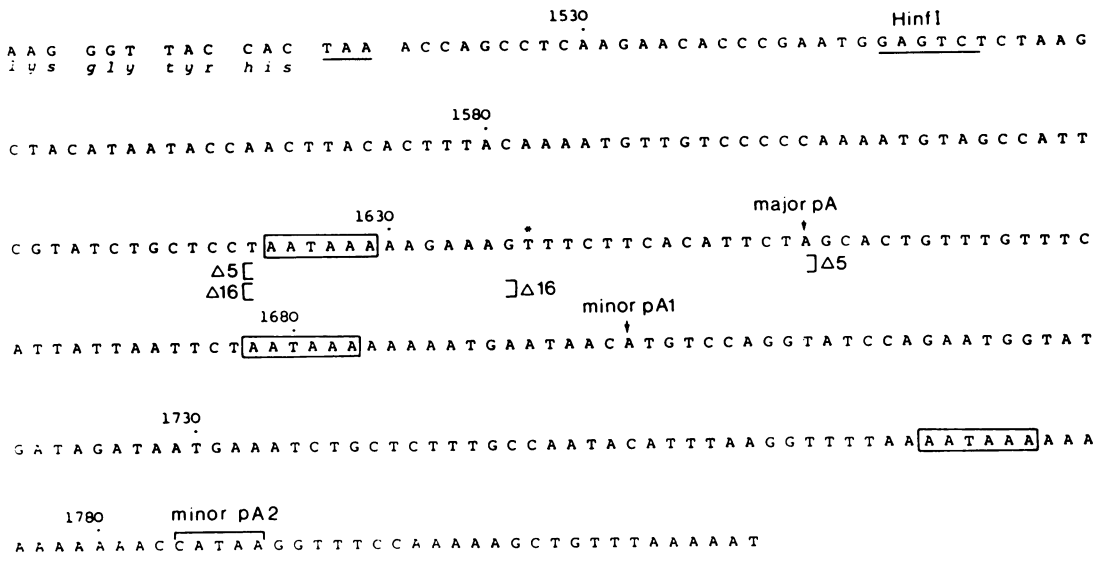

Fig. 3. The nucleotide sequence of the 3′ end of the β1 globin gene of *X. laevis* (Patient *et al.*, 1983). The position of the *Hinf*I site which is labelled in the preparation of many of the S1 mapping probes is indicated. The asterisk indicates the T residue in pS1:2 that was changed to a C residue to generate a *Hind*III site. The three AATAAA sequences are boxed and the sites of polyadenylation indicated (pA). The minor pA1 site was determined to be 14 ± 1 bases downstream of the AATAAA hexanucleotide by size determination of the appropriate fragment after S1 mapping and by comparison of the migration of the fragment with a Maxam/Gilbert sequencing ladder of the probe run on the same gel. The A residue is arrowed since the site of addition of the poly(A) tail is almost always an A in the genomic sequence. The minor site (pA2) has not been determined to the same degree of accuracy. The square brackets show the extent of deletions △5 and △16. Plasmids containing these deletions were recircularized in the presence of *Hind*III linkers (sequence CAAGCTTG) and hence in pS1:2△5 this eight base sequence replaces the sequence within the brackets. In pS1:2△16, which was generated by unidirectional *BAl*31 digestion, the deleted sequence is replaced by a half-linker CAAG. The three A residues at positions 1780–1782 were found to be C residues by Patient *et al.* (1983), but the sequence was derived from the genomic clone λXGαβ103 which is allelic to λXGαβ1. We assume therefore that this difference is a reflection of allelic polymorphism.

as it is selected by oligo(dT)-cellulose chromatography (data not shown). The position within the gene of minor polyadenylation site 1 is indicated in Figure 3. It lies 14 ± 1 nucleotides downstream of an AATAAA hexanucleotide.

Analysis with the same probe of RNA extracted from oocytes injected with pS1:2 DNA showed a pattern very similar to that observed with reticulocyte RNA (Figure 4) with the major and minor RNA species being present in a ratio approximately similar to that observed with reticulocyte RNA. Again oligo(dT)-cellulose chromatography showed that both RNA species were polyadenylated (data not shown).

In the absence of the major polyadenylation site the minor one is used efficiently

The construction of pMTKβS1:2 was such that the final recombinant did not contain a *Hind*III site. We used oligonucleotide-directed mutagenesis (Wallace *et al.*, 1981; Zoller and Smith, 1982) to convert the sequence AAGTTT – which occurs 17 nucleotides upstream of the major polyadenylation site (Figure 3) – to a *Hind*III recognition site (AAGCTT). Using nuclease *Bal*31, a series of small deletions was created around this unique *Hind*III site. Synthetic *Hind*III linkers were then inserted and the DNA recircularized. A series of mutant genes was thus generated in which different blocks of sequence around the polyadenylation signal were replaced by a synthetic linker. The sequence of these mutants was determined by the chain termination method. The sequences of two altered genes (pS1:2△5 and pS1:2△16), in which the AATAAA element is deleted, are shown in Figure 3. These M13 recombinants were injected into oocyte nuclei and the transcription products were analysed by S1 mapping with a 3′-labelled probe (Figure 5A). The major transcription products detected from pS1:2△5 and pS1:2△16 are polyadenylated 14 ± 1 bases downstream of the second AAUAAA sequence, i.e., at minor site 1 (see Figure 3). Hence deletion of the AATAAA sequence abolishes polyadenylation at the major site and leads to a large increase in the amount of polyadenylation at minor site 1. We show that minor site 1 is used with a high efficiency by demonstrating the presence of mutant RNA in amounts comparable with that of the wild-type and the absence of transcripts extending beyond minor site 1.

In the S1 nuclease mapping experiments described above the transcripts from each mutant were analyzed with a different probe. Hence it is difficult to compare the amount of transcript which accumulates to determine whether there is any effect of the mutations on mRNA stability. We therefore assayed the concentration of stable β-globin transcripts in oocytes injected with the various plasmid DNAs using a single probe which was specific for the first intron/exon junction. Results (Figure 5B) of this and other experiments show that, although the concentration of fusion gene transcripts varies between samples, the variation does not correlate with the site of polyadenylation. This variation is probably due to the differences in transcriptional efficiency which we, and others (Asselbergs *et al.*, 1983; Jones *et al.*, 1983), observe between batches of oocytes.

We cannot distinguish between read-through transcription, and reannealed probe in the experiment shown in Figure 5A. Hence it remained possible that a proportion of the transcripts were polyadenylated at other sites in M13-derived RNA. A probe which would allow this distinction to be made was constructed by inserting fragments spanning the 3′ end of pS1:2 and pS1:2△16 into the plasmid vector pAT153. These recombinants lose homology with the original M13 constructs downstream of the β-globin gene. Hence that fraction of the RNA which is not polyadenylated within the *Xenopus* derived sequence gives a discrete band on the auto-

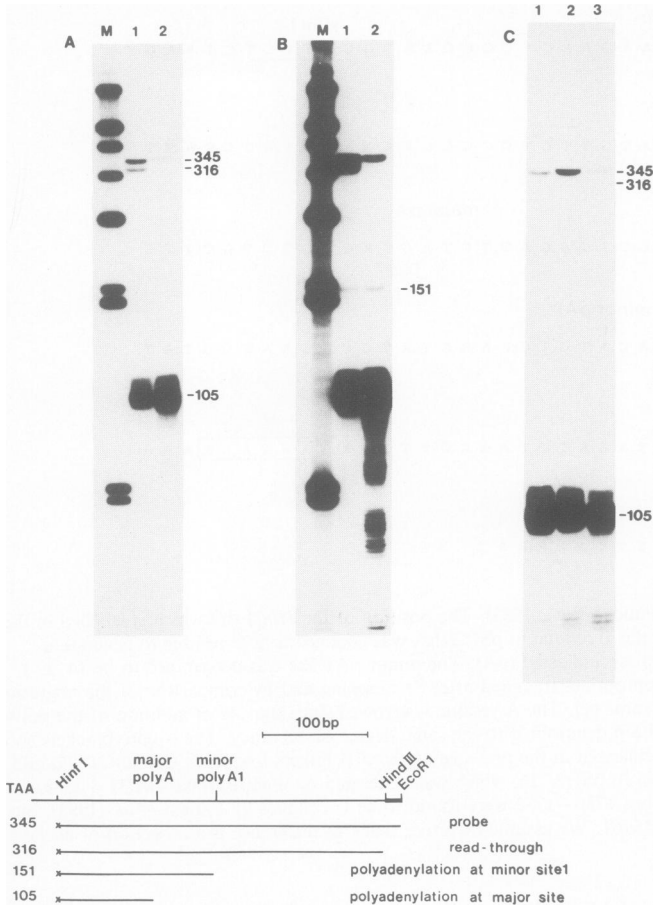

Fig. 4. Nuclease S1 mapping of the 3′ end of β-globin RNA from adult reticulocytes and of RNA from oocytes injected with pS1:2. S1 mapping was carried out as described in Materials and methods and the products of the reaction resolved on 7.5% denaturing polyacrylamide gels. Unless specified two oocyte equivalents of nucleic acid and 5 – 10 ng of single-stranded probe were used per reaction. In all the experiments shown here the probe was the 345 base *Hinf*I/*Eco*RI fragment from plasmid pβBH1. This was 3′ end-labelled at the *Hinf*I site and the coding strand (with respect to globin mRNA) was purified by strand separation. Since this probe is only homologous to the injected DNA for a distance of 316 bases from the labelled end a band of that size represents either 'read-through' RNA or hybridization of the probe to the injected DNA. The RNAs analysed were: **(A** and **B). Lane 1**, total RNA from oocytes injected with 2 ng/oocyte single-stranded pS1:2 DNA; **lane 2**, 400 ng total cytoplasmic RNA from reticulocytes of anaemic frogs. Autoradiography was for 16 h **(A)** or 1 week **(B). (C) Lane 1**, total RNA from oocytes injected with 4 ng/oocyte double-stranded pS1:2 DNA (5 ng probe); **lane 2**, as **1** but with 10 ng probe; **lane 3**, 250 ng total cytoplasmic RNA from frog reticulocytes. This experiment shows the presence of major and minor polyadenylation sites in adult β-globin RNA and in RNA from injected oocytes. The band of 316 nucleotides seen in **A1** and **B1** is mainly due to hybridization of the probe with residual single-stranded DNA as the band is much weaker when double-stranded DNA is injected **(C1, C2)**. M, mol. wt. markers were as in Figure 2.

radiogram (Figures 4 and 5C). The fraction of the various transcripts, including this 'read-through' RNA, was determined by cutting out the gel bands and counting them in a scintillation counter. The results (Table I) show that most of the RNA (>90%) from pS1:2 is polyadenylated at the major site and most of the RNA (>90%) from pS1:2△16 is polyadenylated at the minor site. These results show that the potential efficiency (i.e., the efficiency when the major site is removed) of the minor polyadenylation site is similar to that of the major one.

In the deletion mutant pS1:2△16 a small proportion of transcripts became polyadenylated at a point further downstream than minor polyadenylation site 1 (minor polyadenylation site 2 in Figure 3). This generates an S1 nuclease-resistant fragment of 230 nucleotides in length. This cryptic site also lies downstream of an AATAAA sequence element (Figure 3) but it is observed only in oocyte injection when the major site is deleted and not in erythroid RNA.

Table I. Relative utilization of polyadenylated sites in β-globin mRNAs from wild-type and mutated genes

Source of RNA	Polyadenylation site used (percent)[a]			
	Major pA	Minor pA1	Minor pA2	Read through
Adult reticulocytes	99	1	–[b]	–
pS1:2 injected oocytes	98	1	–	1
pS1:2△16 injected oocytes	–	91	3	6
pS1:2△19 injected oocytes	35	60	2	3

[a]Nuclease S1 mapping was performed on the RNA preparations indicated. Autoradiograms of the resulting gels are shown in Figures 3A, 5C and 6B. Bands were excised from the gels and the radioactivity in each determined by direct scintillation counting. The figures given show the c.p.m. in each band as a percentage of the total counts per minute in all bands.
[b]None detected.

The sequence AAUACA acts as a polyadenylation signal

In one of the mutants (pS1:2△19) the *Bal*31 deletion removed two nucleotides from the 3′-proximal portion of the AATAAA sequence and subsequent insertion of the *Hind*III linker generated the sequence AATACA (Figure 6A). This mutant was injected into oocytes and the RNA was subjected to S1 mapping with the results shown in Figure 6B. Again, a 'read-through' probe was used to allow precise estimation of the proportion of transcripts terminating at the two polyadenylation sites (Table I). Approximately 35% of the transcripts are processed at the major site and 60% are processed at minor site 1. In this experiment RNA was selected by oligo(dT)-cellulose chromatography and transcripts terminated at both sites were found exclusively in the poly(A)$^+$ fraction (Figure 6C).

Discussion

We have shown that transcripts from a thymidine kinase/β-globin fusion gene are accurately spliced and polyadenylated in *Xenopus* oocytes. Previously, correct processing of mRNA sequences in frog oocytes has been inferred from the detection of the protein product (Rungger and Turler, 1978; Wickens *et al.*, 1980) but accurate splicing and polyadenylation by direct RNA analysis had only been shown in the case of oocytes injected with SV40 DNA (Wickens and Gurdon, 1983). The fusion between the TK gene promoter and the β1 globin gene was made in the 5′-non-coding region of both genes. Approximately half of the stable transcripts detected initiate at the TK cap site. These molecules are therefore identical to authentic β1 globin transcripts apart from the 5′-non-coding region. The remaining transcripts initiate upstream of the TK promoter. Since essentially all of the stable transcripts detected are correctly processed the 5′-non-coding sequences do not appear to influence RNA processing.

The results presented here bear upon the mechanisms of transcription termination and polyadenylation. When pS1:2, pS1:2△16 or pS1:2△19 DNA is transcribed in oocytes a low level of transcripts are detected by our S1 mapping probes that extend to, or beyond, the *Hind*III site that lies 210 bases

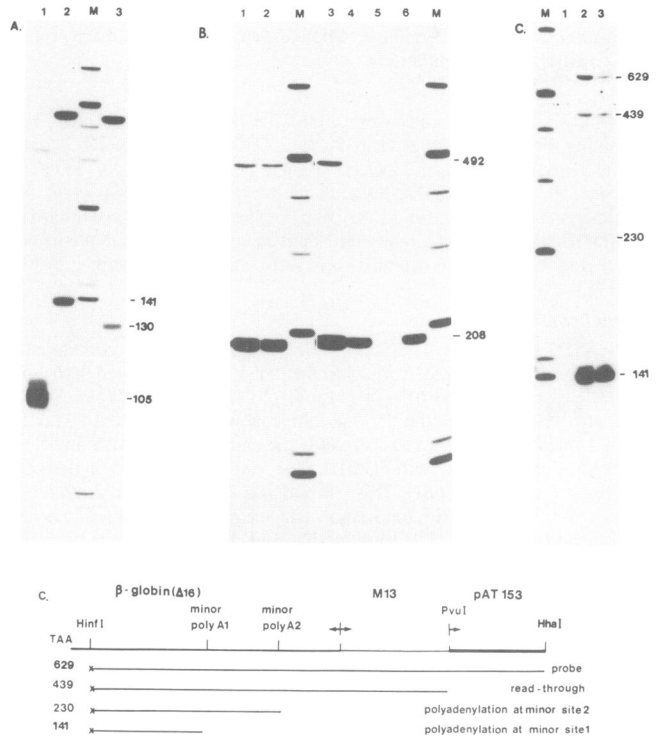

Fig. 5. Transcription and RNA processing in oocytes injected with genes deleted for the β-globin polyadenylation signal. **(A)** S1 nuclease mapping of the 3′ end of RNA from oocytes injected with pS1:2, pS1:2△16 and pS1:2△5 single-stranded DNA. **Lane 1**, two oocyte equivalents of RNA from pS1:2 injected oocytes hybridized with single-stranded *Hinf*I/*Hind*III fragment from plasmid pS1. **Lane 2**, two oocyte equivalent of nucleic acid from pS1:2△16 injected oocytes hybridized with the single-stranded *Hinf*I/*Bgl*I fragment from pS1:2△16 (see Figure 1). **Lane 3**, two oocyte equivalents of nucleic acid from pS1:2△5 injected oocytes hybridized with the single-stranded *Hinf*I/*Bgl*I fragment from pS1:2△5. The bands at 141 bases and 130 bases represent molecules polyadenylated at the minor polyadenylation site 1 for pS1:2△16 and pS1:2△5, respectively. No molecules which are polyadenylated around the major site are detected (expected size 80–90). **(B)** 10 ng of the 492 nucleotide probe specific for the 1st intron (see Figure 4B) (sp. act. 10^7 c.p.m./μg) was used in S1 mapping with nucleic acid (two oocyte equivalents) isolated from oocytes injected with single-stranded DNA from: **1**, pS1:2△16; **2**, pS1:2△19; **3**, pS1:2; **4**, pS1:2 [poly(A)$^+$ RNA]; **5**, uninjected; **6**, pS1:2△5. The fragment of 492 nucleotides in length is due to hybridization with residual single-stranded DNA (cf. Figure 4B). **Lane 4** shows that this material is not bound by oligo(dT)-cellulose. **(C)** RNA (two oocyte equivalents) from oocytes injected with double-stranded pS1:2△16 DNA was hybridized with a 3′ end S1 mapping probe. This was prepared by subcloning the *Pst*I/*Pvu*I fragment of pS1:2△16 into *Pst*I/*Pvu*I cut pAT153 to give plasmid pS1:2△16 RT. The 629-bp *Hinf*I/*Hha* fragment of pS1:2△16 RT spans the 3′ end of the gene and loses homology with pS1:2△16 439 bases from the labelled *Hinf*I site. Hence a band of this size represents RNA that has 'read through' at least as far as this *Pvu*I site. **Lane 1**, minus RNA control; **lane 2**, 10 ng probe; **lane 3**, 5 ng probe. After autoradiography the bands were excised from the gel and the radioactivity in each determined (Table I).

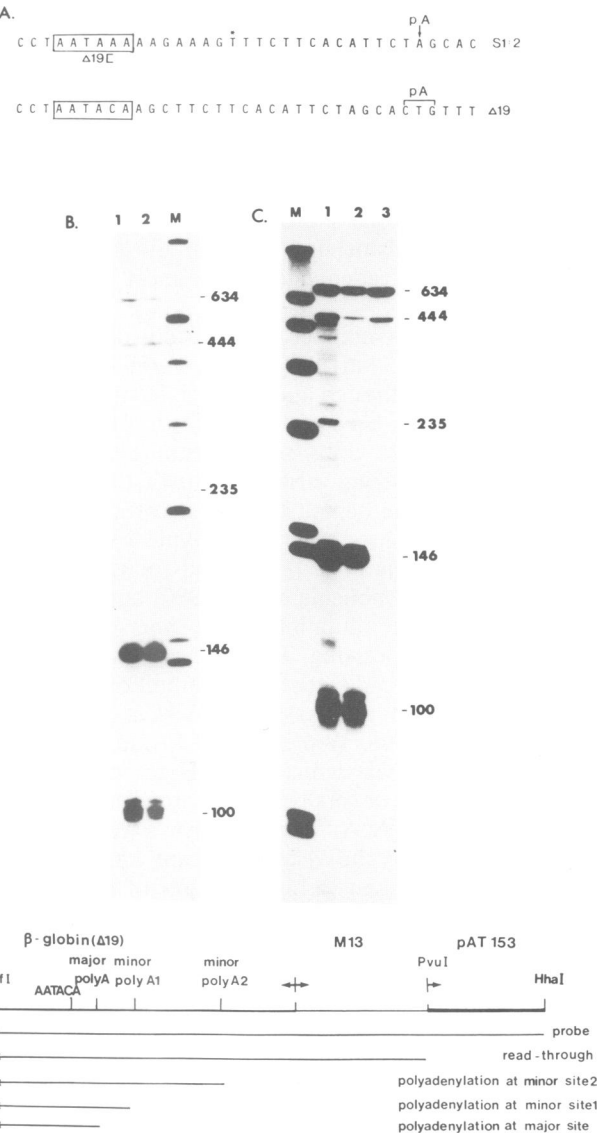

Fig. 6. **(A)** The sequence of pS1:2 and pS1:2△19 around the major polyadenylation site. The asterisk indicates the T residue in pS1:2 that was changed to a C residue to generate a *Hind*III site. The square bracket below the sequence of pS1:2 represents the extent of digestion by nuclease *Bal*31; in this case *Bal*31 digested in only this direction. Addition of *Hind*III linkers and ligation generated pS1:2△19 whose sequence is shown. The polyadenylation signal AATAAA and the related sequence AATACA are boxed. The site of polyadenylation of pS1:2△19 derived RNA was assigned from S1 mapping experiments such as those shown in **B** and **C**. **(B** and **C)** S1 mapping of RNA from oocytes injected with double-stranded pS1:2△19 DNA. The probe in each case was a 'read-through' probe specific for pS1:2△19 and constructed from pS1:2△19 RT as described for the pS1:2△16 probe illustrated in Figure 5. **(B) Lane 1**, 10 ng probe; **lane 2**, 5 ng probe. After autoradiography the bands were excised from the gel and the radioactivity in each determined (Table I). **(C) 1**, total RNA from pS1:2△19 injected oocytes + 10 ng probe; **2**, poly(A)$^+$ RNA from pS1:2△19 injected oocytes + 10 ng probe; **3**, poly(A)$^-$ RNA from pS1:2△19 injected oocytes + 10 ng probe. The experiment shows that polyadenylation at the major site is not abolished in pS1:2△19 RNA and that molecules terminating at both major and minor sites are found exclusively in the poly(A)$^+$ fraction.

downstream of the major polyadenylation site. This implies that there is no transcription terminator in that region and that transcription proceeds beyond this *Hind*III site. Transcription extending 1 kb or more beyond the mature 3′ end of mRNAs has been shown for several viral genes (Ford and Hsu, 1978; Nevins *et al.*, 1980) and for the mouse β-globin gene (Hofer *et al.*, 1982).

Although both major and minor polyadenylation sites can be used efficiently, the more upstream site is used in >95% of molecules in either blood cells or oocytes. The more upstream polyadenylation site is also used preferentially in other systems where there is no apparent differential regulation of polyadenylation (Tosi *et al.*, 1981; Gerlinger *et al.*, 1982). In the case of the immunoglobulin μ gene, two mRNAs, differing in their polyadenylation sites, are produced during different stages in B-cell maturation (Rogers *et al.*, 1980; Early *et al.*, 1980). When, however, a recombinant containing the 3′ region of this gene is transcribed in Cos cells

(Nishikura and Vuocolo, 1984) the upstream polyadenylation site is always preferred and this is the case even when the two sites are exchanged. The cleavage/polyadenylation machinery may, therefore, scan in a 5′ – 3′ direction as suggested by Nishikura and Vuocolo (1984). There is however, an equally plausible alternative mechanism. If cleavage/polyadenylation occurs efficiently at several sites – and providing polyadenylation at one site does not inhibit the subsequent use of more upstream sites – then most of the mature mRNAs will be polyadenylated at the most upstream of the available sites.

There are now a number of naturally occurring, or artificially generated, variants of the AATAAA consensus hexanucleotide. Although the number is small and precise quantitation has not been performed for most of the variants, it is still possible to draw some meaningful comparisons. Thus naturally occurring variants in the first two nucleotides, e.g., uAUAAA (the hepatitis B virus surface antigen gene, Simonsen and Levinson, 1983) and AuUAAA (the chicken lysozyme gene, Jung *et al.*, 1980 and mouse pancreatic α-amylase gene, Hagenbuchle *et al.*, 1980) appear to be fully functional. However mutations in all other positions reduce the efficiency of processing. Thus alteration to the sequences AAgAAA (an artificially generated mutant of the adenovirus E1a gene, Montell *et al.*, 1983), AAUgAA (the human pseudo α-globin gene, Whitelaw and Proudfoot, 1983) and AAUAAg (an α-thalassemia gene, Higgs *et al.*, 1983) all dramatically reduce or totally abolish processing. The variant we describe (AAUAcA) also displays a reduced level of polyadenylation but the very significant amount (35%) of residual activity suggests a higher degree of flexibility in the signal than might have been predicted.

A number of lines of evidence suggest that the consensus hexanucleotide cannot be the sole requirement for polyadenylation. One indirect but compelling argument comes from the existence of AATAAA sequence elements which lie in coding or 3′-non-coding regions but which are not utilized (Tosi *et al.*, 1981). Also, we have introduced an AATAAA sequence into the third exon of the β1 globin gene and find no detectable level of polyadenylation downstream of the mutation (P.Mason *et al.*, in preparation). Direct evidence comes from recent studies in which the deletion of sequences lying downstream of the site of polyadenylation reduced the accuracy and efficiency of polyadenylation (Simonsen and Levinson, 1983; McDevitt *et al.*, 1984; Woychik *et al.*, 1984). Berget (1984) has shown that the consensus sequence CAPyTG is found close to the site of polyadenylation in a large number of genes. In the *Xenopus* β1 globin gene there is a variant of this sequence (CATTc) just upstream of the major site of polyadenylation and a perfect fit to the consensus (CACTG) just downstream of the site. There is a much weaker fit to the consensus (CATcg) at minor polyadenylation site 1 and another equally weak homologue (CATaa) at minor polyadenylation site 2 – a site which is only utilized when the major polyadenylation site is deleted (Figures 5 and 6). In order to assess the importance of sequences such as these, or of other features such as secondary structure which may be involved, additional mutants must be analysed. Such a study is now in progress.

Materials and methods

Enzymes

Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase, the Klenow fragment of DNA polymerase I and *Bal*31 nuclease were purchased from either New England Biolabs Inc. (Beverly, MA) or from Bethesda Research Laboratories, Inc. (Gaithersburg, MD) and were used according to their manufacturers recommendations.

DNAs

The DNAs used in this study were: pHSV106 (McKnight and Gavis, 1980), containing the herpes simplex virus TK (HSVTK) gene and pXG1C3 containing the β1 globin gene from λXGαβ1 (Patient *et al.*, 1980) in pAT153 (Twigg and Sherratt, 1980), and pXG8D2a, β1 globin cDNA clone (Kay *et al.*, 1980; Williams *et al.*, 1980). Plasmid DNA was prepared according to standard procedures. Preparations of phage M13-derived recombinant DNAs in either single-stranded or double-stranded form were prepared as described by Messing (1983).

Plasmid constructions

pMTKβS1:2 (Figure 1). The 250-bp *Pvu*II/*Bgl*II fragment containing the HSVTK promoter derived from plasmid pHSV106 was ligated into *Sma*I/*Bam*HI-cleaved M13mp8 to give plasmid pMTK4. The *Bal*I/*Hind*III fragment of pXG1C3 containing the β-globin gene was purified and ligated into pMTK4 that had been cleaved with *Sal*I, the ends 'filled in' with the Klenow enzyme and then cleaved with *Hind*III. The resulting plasmid was designated pMTKβS1 (abbreviated pS1). This plasmid was cleaved at its unique *Hind*III site, the ends 'filled in' with the Klenow enzyme and then religated in order to remove the *Hind*III site. The resulting plasmid, pMTKβS1:2 was used as the 'wild-type' β-globin gene and as the template for constructing specific mutant genes.

Construction of deletions. A *Hind*III site was introduced into pMTKβS1:2 by oligonucleotide-directed mutagenesis to give plasmid pS1:2.56 (Figure 3). 1 μg of *Hind*III-cleaved pM56 was treated with nuclease *Bal*31 in 100 μl of a solution containing 12 mM $CaCl_2$, 12 mM $MgCl_2$, 0.2 M NaCl, 20 mM Tris-HCl pH 8.1, 1 mM EDTA, and 1.6 μg of *Hinf*1-cleaved pAT153. The DNA was treated with 0.1 U of *Bal*31 for 1.5 min at 20°C or 0.03 U of *Bal*31 for 2.5 min at 20°C. The reaction was terminated by adding 50 μl of phenol. After phenol and chloroform extraction and ethanol precipitation half of the DNA from each *Bal*31 reaction was treated for 16 h at 10°C in 20 μl of a solution containing 10 mM $MgCl_2$, 50 mM Tris-HCl pH 7.4, 10 mM dithiothreitol, 0.5 mM ATP, 150 ng of ^{32}P-labelled and phosphorylated *Hind*III linkers and 1 U of T4 DNA ligase. After digestion with 100 U of *Hind*III for 2 h at 37°C, the excess linkers were removed on a Sepharose 4B column. The plasmid DNA was recircularized by treatment with T4 DNA ligase as above at a concentration of 1 μg/ml pS1:2.56 DNA. The extent of deletions obtained was determined by restriction enzyme analysis of individual plasmids and those selected for further study were sequenced by the dideoxy chain termination method of Sanger *et al.* (1980) using a synthetic oligonucleotide that primes DNA synthesis downstream of the β-globin polyadenylation site. The sequence of the deletions used in this study is illustrated in Figures 3 and 6.

Plasmids used for the preparation of read-through probes. In order to prepare nuclease S1 mapping probes that enable measurement of the proportion of read-through RNA, fragments of the M13 clones spanning the 3′ end of the β-globin gene were subcloned into pAT153. The *Bam*HI/*Hind*III fragment from pMTKβS1 was subcloned into *Bam*HI/*Hind*III-cleaved pAT153 to give plasmid pβBH1. To prepare a read-through probe from pS1:2△16 and pS1:2△19, the fragments from the *Pst*I site in the TK promoter to the *Pvu*I site in M13 DNA were subcloned into *Pst*I/*Pvu*I-cut pAT153 to give plasmids pS1:2△16 RT and pS1:2△19 RT.

Oligonucleotides

The oligonucleotides used in this study were I 3′TTCTTTCGAAGAAG-TGTAAG5′ (co-ordinates 1630 – 1649), and II 3′-GGTCCATAGATCTT-ACC5′ (co-ordinates 1701 – 1718). The co-ordinates shown refer to the complementary DNA sequence shown in Figure 3. Oligonucleotide I was designed to induce a T→C change at position 1637; and oligonucleotide II was used as a primer for DNA sequence determination by the chain termination procedure. Oligonucleotides were synthesized by the phosphotriester method essentially as described by Gait *et al.* (1982).

Oligonucleotide-directed mutagenesis

Oligonucleotide-directed mutagenesis was performed on single-stranded M13 recombinants as described by Wallace *et al.* (1981) and Zoller and Smith (1982). In later experiments we used the two primer method of generating mutations (Norris *et al.*, 1983) but screened the resulting M13 DNAs directly as described by Zoller and Smith (1982).

To select the M13 recombinants bearing the appropriate mutations, phage were plated in JM103 cells at a density of ~100 per 9 cm plate and several replicas of the plate were made on nitrocellulose by sequentially laying nitrocellulose filters onto the plate and leaving them until wet. The filters were then baked at 80°C for 2 h without further treatment. The filters were then pre-hybridized in 6 x SSC, 0.2% polyvinyl pyrrolidone, 0.2% bovine serum

albumin, 0.2% Ficoll, 50 μg/ml denatured salmon sperm DNA (Denhardt, 1966) for 1 h at room temperature. 10^6 c.p.m./ml of ^{32}P-labelled oligonucleotide was then added to the solution and hybridization continued for 1 h at room temperature. The filters were then washed twice for 15 min at room temperature in 6 x SSC and then for 2 min in the same solution at a range of temperatures around the calculated *T*d value of the perfectly matched hybrid. The *T*d value was calculated using the formula *T*d = 2°C x number of AT base pairs + 4°C x number of GC base pairs (Suggs *et al.*, 1981). Signals from recombinants harboring the appropriate mutations persisted at higher temperatures. The structure of all mutations obtained was confirmed by DNA sequencing through the site of the mutations and by restriction enzyme analysis with several enzymes which cut frequently to verify that no gross rearrangements had taken place.

Microinjection of oocytes and RNA preparation

The methods that we used for the injection of DNA and maintenance of oocytes were those developed by Gurdon (1977) and Kressman *et al.* (1978). Either 2 ng of single-stranded DNA or 4 ng of double-stranded DNA was injected into each of ~50 oocyte nuclei per experiment. The oocytes were incubated at 18°C for 1 day and then total nucleic acids extracted as described by Kressman *et al.* (1978). Cytoplasmic RNA from anaemic *Xenopus* adult blood was prepared as previously described (Hentschel *et al.*, 1979).

Nuclease S1 mapping

Nuclease S1 mapping was performed by the method of Berk and Sharp (1977) and Weaver and Weissmann (1979). Descriptions of the various probes used in particular experiments are given in the figure legends. In all cases the restriction fragment to be used was purified by electrophoresis through a 5% polyacrylamide gel. The fragment was then end labelled either at the 3′ ends by 'filling in' with the Klenow enzyme and α-^{32}P-labelled deoxynucleotide triphosphates, or at the 5′ ends with T4 polynucleotide kinase and γ-^{32}P-labelled ATP. Following labelling the fragments were denatured and electrophoresed through 5% or 7.5% polyacrylamide gels to separate the two strands (Maxam and Gilbert, 1980). Trial experiments were performed to determine which of the separated strands was complementary to the mRNA. An amount of total nucleic acid equivalent to the yield from 1–3 oocytes was mixed with 5–25 ng of single-stranded probe (sp. act. 5 x 10^7 c.p.m./μg) in 10 μl of solution containing 0.4 M NaCl, 10 mM Pipes pH 6.4 and incubated in sealed capillaries at 65°C for 12 h. Samples were digested with 10^3 U of S1 nuclease in 0.2 M NaCl, 0.03 M sodium acetate pH 6.0, 2 mM $ZnSO_4$ in final volume of 200 μl, for 30 min at 37°C. After ethanol precipitation the samples were denatured by heating at 90°C for 2 min in a loading buffer containing 80% formamide and resolved by electrophoresis on 5% or 7.5% denaturing polyacrylamide gels.

To estimate the amount of label in each band the bands were located by autoradiography, excised with a razor blade and Cerenkov radiation was measured in a liquid scintillation counter.

Acknowledgements

We thank Rita Harris for expert technical assistance. We are grateful to Rita Harris and Brenda Marriott for preparing the manuscript and to D.Banville for his advice during the course of the work and for his helpful criticism of the manuscript.

References

Asselbergs,F.A.M., Smart,J.E. and Mathews,M.B. (1983) *J. Mol. Biol.*, **163**, 209-238.
Berget,S.M. (1984) *Nature*, **209**, 179-182.
Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Capetanaki,Y.G., Ngai,J., Flytzanis,C.N. and Lazarides,E. (1983) *Cell*, **35**, 411-420.
Denhardt,D.T. (1966) *Biochem. Biophys. Res. Commun.*, **23**, 641-652.
Early,P., Rogers,J., Davis,M., Calame,K., Bond,M., Wall,R. and Hood,L. (1980) *Cell*, **20**, 313-319.
Fitzgerald,M. and Shenk,T. (1981) *Cell*, **24**, 251-260.
Ford,J.P. and Hsu,M.-T. (1978) *J. Virol.*, **28**, 795-801.
Gait,M.J., Matthes,H.W.D., Singh,M., Sproat,B.S. and Titmas,R.C. (1982) in Gassen,H.G. and Lang,A. (eds.), *Chemicals and Enzymatic Synthesis of Gene Fragments*, Verlag Chemie, Weinheim (FRG), Deerfield Beach (Florida), Basel, pp. 1-42.
Gerlinger,P., Krust,A., LeMeur,M., Perrin,F., Cochet,M., Gannon,F., Dupret,D. and Chambon,P. (1982) *J. Mol. Biol.*, **162**, 345-364.
Gurdon,J.B. (1977) in Stein,G., Stein,J. and Kleinsmith,L.J. (eds.), *Methods in Cell Biology*, Academic Press, Vol. **16**, pp. 125-140.
Hagenbuchle,O., Borey,R. and Young,R.A. (1980) *Cell*, **21**, 179-187.
Hentschel,C.C., Kay,R.M. and Williams,J.G. (1979) *Dev. Biol.*, **72**, 350-363.
Higgs,D.R., Goodbourn,S.E.Y., Lamb,J., Clegg,J.B., Weatherall,D.J. and Proudfoot,N.J. (1983) *Nature*, **306**, 398-400.
Hofer,E., Hofer-Warkinek,R. and Darnell,J.E.,Jr. (1982) *Cell*, **29**, 887-893.
Jones,N.C., Richter,J.D., Weeks,D.L. and Smith,L.D. (1983) *Mol. Cell. Biol.*, **3**, 2131-2142.
Jung,A., Sippel,A.E., Grez,M. and Schutz,G. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5759-5763.
Kay,R.M., Harris,R., Patient,R.K. and Williams,J.G. (1980) *Nucleic Acids Res.*, **8**, 2691-2708.
Kressmann,A., Clarkson,S.G., Pirrotta,V. and Birnstiel,M.L. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1176-1180.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
McDevitt,M.A., Imperiale,M.J., Ali,H. and Nevins,J.R. (1984) *Cell*, **37**, 993-999.
McKnight,S.L. and Gavis,E.R. (1980) *Nucleic Acids Res.*, **24**, 5931-5948.
McKnight,S.L., Gavis,E.R., Kinsbury,R. and Axel,R. (1981) *Cell*, **25**, 385-398.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20-78.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.
Montell,C., Fisher,E.F., Caruthers,M.H. and Berk,A.J. (1983) *Nature*, **305**, 600-605.
Nevins,J.R., Blanchard,J.-M. and Darnell,J.E.,Jr. (1980) *J. Mol. Biol.*, **144**, 377-386.
Nishikura,K. and Vuocolo,G.A. (1984) *EMBO J.*, **3**, 689-699.
Norris,K., Norris,F., Christiansen,L. and Fiil,N. (1983) *Nucleic Acids Res.*, **11**, 5103-5112.
Patient,R.K., Elkington,J.A., Kay,R.M. and Williams,J.G. (1980) *Cell*, **21**, 565-573.
Patient,R.K., Harris,R., Walmsley,M.E. and Williams,J.G. (1983) *J. Biol. Chem.*, **258**, 8521-8523.
Proudfoot,N. (1984) *Nature*, **307**, 412-413.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-214.
Rogers,J., Early,P., Carter,C., Calame,K., Bond,M., Hood,L. and Wall,R. (1980) *Cell*, **20**, 303-312.
Rosenfeld,M.G., Mermod,J.-J., Amara,S.G., Swanson,L.W., Sawchenko, P.E., Rivier,J., Vale,W.W. and Evans,R.M. (1983) *Nature*, **304**, 129-135.
Rungger,D. and Turler,H. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 6073-6077.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.
Simonsen,C.C. and Levinson,A.D. (1983) *Mol. Cell. Biol.*, **3**, 2250-2258.
Suggs,S.V., Hirose,T., Miyake,T., Kawashima,E.H., Johnson,M.J., Itakura,K. and Wallace,R.B. (1981) in Brown,D.D. and Fox,C.F. (eds.), *Developmental Biology Using Purified Genes*, Academic Press, NY, pp. 683-693.
Tosi,M., Young,R.A., Hagenbuchle,O. and Schibler,U. (1981) *Nucleic Acids Res.*, **9**, 2313-2323.
Twigg,A.J. and Sherratt,D. (1980) *Nature*, **283**, 216-218.
Wallace,R.B., Schold,M., Johnson,M.J., Dembek,P. and Itakura,K. (1981) *Nucleic Acids Res.*, **9**, 3647-3656.
Weaver,R.F. and Weissman,C. (1979) *Nucleic Acids Res.*, **7**, 1175-1193.
Whitelaw,E. and Proudfoot,N.J. (1983) *Nucleic Acids Res.*, **11**, 7717-7733.
Wickens,M.P. and Gurdon,J.B. (1983) *J. Mol. Biol.*, **163**, 1-26.
Wickens,M.P., Woo,S., O'Malley,B.W. and Gurdon,J.B. (1980) *Nature*, **285**, 628-634.
Williams,J.G., Kay,R.M. and Patient,R.K. (1980) *Nucleic Acids Res.*, **8**, 4247-4257.
Woychik,R.P., Lyons,R.H., Post,L. and Rottman,F.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3944-3948.
Zoller,M.J. and Smith,M. (1982) *Nucleic Acids Res.*, **10**, 6487-6500.

Received on 27 September 1984