The EMBO Journal vol.3 no.7 pp.1533 – 1540, 1984

Transcriptional analysis of human zeta globin genes

N.J. Proudfoot[1], T.R. Rutherford[2] and G.A. Partington[3]

[1]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, [2]Nuffield Department of Clinical Medicine, and [3]MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

Communicated by N.J.Proudfoot

The human embryonic α-like globin gene (ζ) and a closely linked pseudogene (ψζ) are located on chromosome 16. The ψζ gene has a nonsense mutation in exon 1 but has identical promoter sequence and RNA processing sites to the ζ gene, raising the possibility that both ψζ and ζ are transcriptionally active. We have studied transcription of the human ζ and ψζ genes in a number of systems to examine their cell type specificity and enhancer requirement. (i) Cloned ζ and ψζ genes transfected into human HeLa or monkey Cos7 tissue culture cells show no transcriptional activity. The presence of an SV40 enhancer does not activate the ζ promoter except at low levels when in very close proximity (<50 bp from the CCAAT box). (ii) In contrast to other tissue-specific genes tested to date, both ζ and ψζ gene promoters initiate transcription efficiently when micro-injected into *Xenopus* oocyte nuclei. We suggest that embryonic-specific factors in the oocyte may permit efficient ζ gene transcription. Furthermore, the ζ promoter sequence from −111 to +38 bp is sufficient for transcription in this system.

Key words: ζ globin genes/promoter analysis/HeLa Cos7 transfection/oocyte micro-injection

Introduction

The human α-like globin gene cluster has been isolated by molecular cloning and revealed to have a relatively simple arrangement (Lauer *et al.*, 1980). Within a 30 kb section of chromosome 16 (Deisseroth *et al.*, 1977), three functional genes and two pseudogenes have been identified and sequenced: ζ-ψζ-ψα-α2-α1 (Michelson and Orkin, 1980; Liehaber *et al.*, 1980; Proudfoot and Maniatis, 1980; Proudfoot *et al.*, 1982). The ζ globin gene is expressed in the yolk sac of early embryos (Huehns *et al.*, 1961). ζ Globin polypeptides combine with ε, the embryonic β-like globin gene product (and γ globin) to form embryonic haemoglobin. During foetal life, ζ globin polypeptides are no longer detected but are replaced by α globin, the product of both the α1 and α2 globin genes. α Globin, together with γ globin, the β-like foetal globin gene product, form foetal haemoglobin in foetal liver. Finally, α and β globins are expressed in adult bone marrow and form adult haemoglobin. From this pattern of globin gene expression during the development of erythroid tissues, it is inferred that there is one switch in expression of the human α-like globin gene cluster from the ζ globin gene to α globin genes at an early stage of foetal development (for review, see Maniatis *et al.*, 1980).

As a direct approach to study the control of α-like globin gene transcription, the procedure of gene transfer or DNA-mediated transfection of tissue culture cells may be utilised, in particular the transient expression system. Using this procedure, it has been shown that both α1 and α2 globin genes are efficiently transcribed in a variety of mammalian tissue culture cells such as human HeLa or monkey Cos7 cells (Mellon *et al.*, 1981). Similarly, ψα globin gene displays significant but low transcriptional activity when transfected into Cos7 cells (Whitelaw and Proudfoot, 1983). Here we describe a systematic attempt to express ζ and ψζ genes in HeLa and Cos7 cells. Essentially no transcription was detected from either the ζ or ψζ promoters in a variety of expression vectors. Only when the SV40 transcriptional enhancer (Banerji *et al.*, 1981) was placed in close proximity to the ζ or ψζ promoter could low level transcription from ζ and ψζ be detected.

As an alternative approach, we have used the procedure of micro-injection into the nuclei of *Xenopus laevis* oocytes (Gurdon and Melton, 1981). This system has been extensively used to study the transcription of RNA polymerase III genes such as tRNA and 5S (Cortese *et al.*, 1978; Kressman *et al.*, 1978; Korn, 1982). Furthermore, some RNA polymerase II transcribed genes such as histone (e.g., Probst *et al.*, 1979; Hentschel *et al.*, 1980), viral thymidine kinase (McKnight *et al.*, 1981), SV40 and adenovirus RNA (Mertz and Gurdon, 1977; Wickens and Gurdon, 1983; Asselberg *et al.*, 1983), are transcribed with efficient initiation and processing in oocytes. However, when other RNA polymerase II genes, such as rabbit β globin and chick ovalbumin, whose expression is specific to differentiated cells, have been injected into oocytes, no correctly initiated transcripts were seen (Rungger *et al.*, 1981; Wickens *et al.*, 1980). Here we report that both the ζ and ψζ globin gene promoters function efficiently when injected into *Xenopus* oocytes, in contrast to their inactivity in tissue culture cells. We postulate that *Xenopus* oocytes possess specific factors that allow the ζ and ψζ promoters to function.

Results

Transient expression of ζ and ψζ genes in human HeLa tissue culture

Two transient expression systems have been successfully utilised to study the transcription of globin genes: the HeLa cell system described in this section and the Cos7 cell system described later. Both make use of the replication and/or promoter enhancer features of SV40. Banerji *et al.* (1981) have demonstrated the efficient transcription of the rabbit β globin gene in an expression vector containing the SV40 origin, enhancer and T antigen gene. This plasmid, following transfection into HeLa cells, replicates in the cell nucleus, and the β gene within this plasmid is efficiently transcribed due to the presence of the SV40 transcriptional enhancer. We wished to investigate the transcriptional features of the ζ and ψζ globin genes using the HeLa cell transient expression system. To this end, we subcloned the ζ and ψζ genes together with a hybrid ζ/ψζ gene construct into the expression vector SVpBR328 as

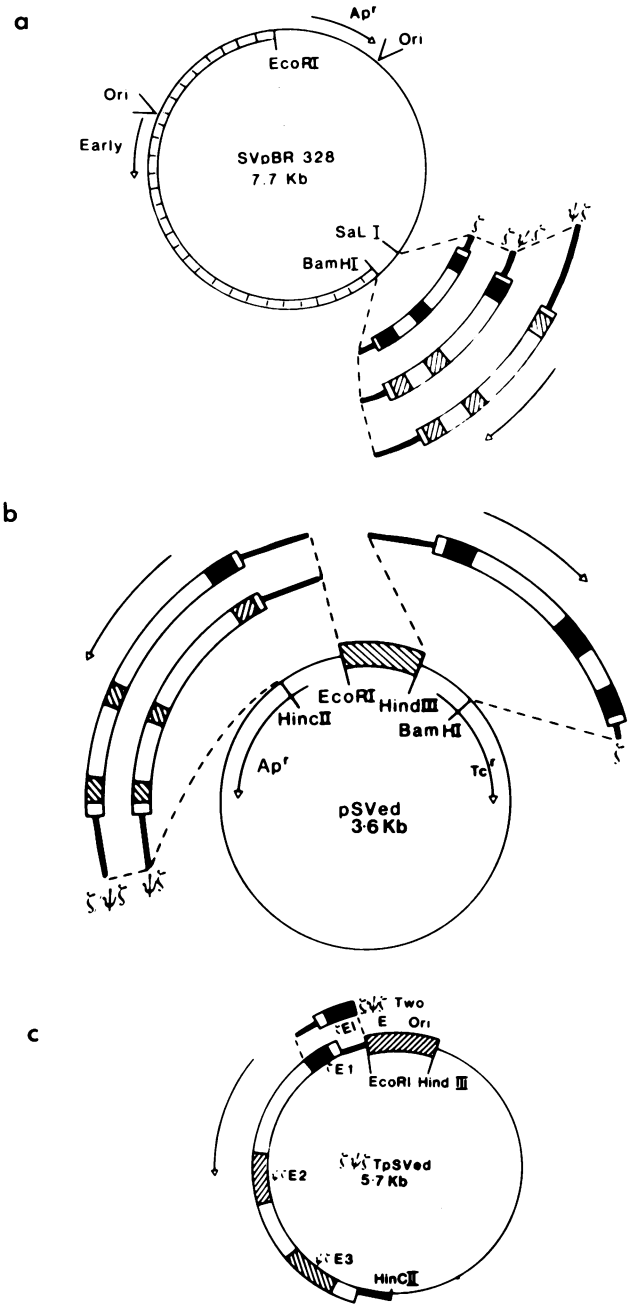

Fig. 1. (1a) Diagram of SVpBR328 constructs. Hatched area, SV40; lined area, pBR328; thick lines 5′ and 3′ flanking sequence of ζ and ψζ genes; open boxes, introns; filled in boxes, ζ exons; hatched boxes, ψζ exons. Arrows denote direction of transcription of adjacent gene. Restriction sites at which SV40, ζ or ψζ sequence is joined to pBR328 are indicated (see Materials and methods). **(1b)** Diagram of constructs based on pSVed. Hatched box between *Eco*RI and *Hind*III is SV40 enhancer plus origin sequence (not drawn to scale). Otherwise, legend as for Figure 1a. **(1c)**ζ/ψζ T pSVed construction. Again, hatched box between *Eco*RI and *Hind*III is SV40 enhancer plus origin sequence (not drawn to scale). Otherwise, legend as for Figure 1a. See Materials and methods for details of hybrid gene constructions.

shown in Figure 1a. This vector, previously described by Grosveld *et al.* (1982), contains most of the plasmid pBR328 and virus SV40 joined at their *Bam* and *Eco*RI sites. The SV40 gene functions, origin of replication, enhancer and T antigen gene are all intact, so that this plasmid replicates in HeLa cell nuclei and enhances the transcription of genes cloned into it. As shown in Figure 2, analysis of mRNA 3′ ends using S1 nuclease reveals the presence of ζ globin mRNA 3′ ends. Similarly with the hybrid ζ/ψζ gene (Figure 1a), ψζ mRNA 3′ ends are detectable. However, the intact ψζ gene gave no detectable 3′-end signal. This negative result cannot be accounted for by a failed transfection of the HeLa cells since the human α1 globin gene in α1 pSVod (Mellon *et al.*, 1981) co-transfected with all three gene-containing plasmids, gave a similar 3′-end signal (Figure 2) in each transfection, proving that all three transfections were equally successful. These results at first sight suggested that ζ but not ψζ globin gene is transcriptionally active in HeLa cells. Figure 3 shows a primer extension experiment (Proudfoot *et al.*, 1980), with a 3′ end-labelled primer isolated from ζ globin exon 1 extended with reverse transcriptase on the three RNAs, together with K562 RNA as a positive control for ζ globin mRNA. K562 is a ζ globin-expressing human erythroid cell line that expresses up to 10 pg/cell of embryonic and foetal globins following induction with haemin (Rutherford *et al.*, 1979). As indicated, although cDNA transcripts of ζ and ζ/ψζ are detectable, they do not terminate at the correct ζ *Cap* site as indicated by the K562 control, but rather extend beyond this position to beyond the 5′ extremity of the ζ gene insert in SVpBR328. It seems likely that these transcripts initiate from an SV40 promoter. The absence of these upstream transcripts in ψζ may reflect the larger portion of 5′ flanking sequence in the ψζ gene clone as compared with the ζ gene clone, 2.0 kb rather than 0.6 kb (Figure 1a). Possibly the SV40 transcripts terminate within the ψζ 5′ flanking sequence and so do not read through into the ψζ gene. These data demonstrate the total transcriptional inactivity of the ζ and ψζ gene promoters in this system.

Transient expression of ζ and ψζ genes in monkey Cos7 tissue culture cells

Mellon *et al.* (1981) have described the transfection of human α globin-containing plasmids into monkey Cos7 cells. This cell line is stably transformed with origin-defective SV40 and expresses endogenous SV40 T antigen (Gluzman, 1981). The plasmid used in these experiments, pSVod, contains the SV40 origin sequence, so that replication occurs to high copy numbers as endogenous T antigen is available. High levels of α globin mRNA were detected in these experiments. No SV40 enhancer was used in the expression vector since α globin is efficiently transcribed in the absence of enhancer elements. Since the ζ and ψζ globin genes were not transcriptionally active in the HeLa cell transient expression system, we wished to test their transcriptional activity in the alternative Cos7 system. One possible explanation for the ζ and ψζ globin genes inactivity in the HeLa cell experiments could relate to their requirement for an enhancer element close to their promoter sequences. In the SVpBR328 constructs, the ζ and ψζ gene promoters were several kilobases away from the SV40 enhancer. Such a requirement has been demonstrated with the mouse immunoglobin lambda light chain gene which is only transcribed in the HeLa cell system when the SV40 enhancer is within a few hundred bases of its promoter (Picard and Schaffner, 1983).

As shown in Figure 1b, we modified the pSVod plasmid to pSVed by replacing the SV40 origin fragment with a larger piece of SV40 DNA including both the enhancer and 72-bp repeat sequence, as well as the origin sequence. We first subcloned the intact ζ gene with ~500 bp of 5′ flanking sequence into pSVed on the origin side of the SV40 fragment (ζ pSVed). We then subcloned the intact ψζ gene with ~300 bp

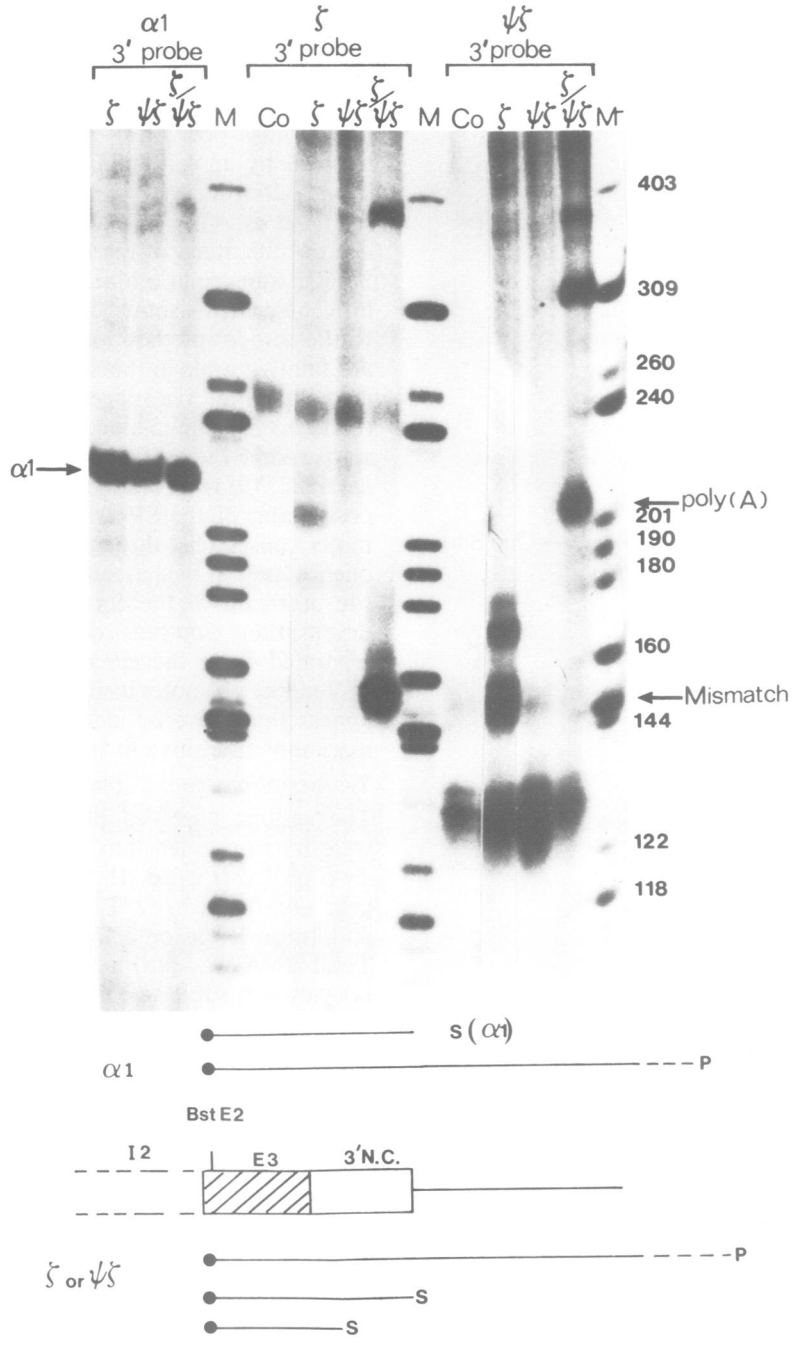

Fig. 2. 3′ S1 analysis of HeLa cell total RNA transformed with ζ, ψζ and ζ/ψζ SVpBR328 plasmids. Co, minus RNA control; M, size markers. Position of α1 signal is indicated. Position of expected ζ or ψζ 3′ ends indicated by poly(A). Position of sequence divergence between ζ and ψζ mRNA 3′ ends indicated by mismatch. The details of the probes (P) used and the signals (S) obtained from this S1 mapping experiment are shown below and in Materials and methods.

of 5′ flanking sequence into pSVed but on the enhancer side of the SV40 fragment (ψζ pSVed). ψζ pSVed was then modified to ζ/ψζ pSVed by replacing the ψζ 5′ flanking sequence with an equivalent piece of ζ 5′ flanking sequence. Finally, as shown in Figure 1c, we deleted most of the ζ gene 5′ flanking sequence so that the SV40 enhancer is within 50 bp of the ζ CCAAT box promoter sequence (ζ/ψζ T pSVed). As a by-product of the ζ/ψζ T pSVed construct, we obtained an equivalent plasmid with the ζ promoter in the opposite orientation (ζ/ψζ Two pSVed).

Figure 4 shows S1 analysis with a ζ gene 5′ probe, of RNA obtained from Cos7 cells transfected with the above constructs. Essentially no correct ζ mRNA 5′ ends (*Cap* site) or upstream transcription were detected with any of the ζ pSVed constructs except for ζ/ψζ T pSVed. These same results have been obtained in different transfection experiments. Furthermore, co-transfection of the ζ gene pSVed constructs with human α-globin pSVed in each case revealed equivalent levels of globin mRNA (data not shown). This indicates that the different Cos7 transfections were equally efficient. A faint mismatch band is present in the ψζ lane corresponding to the divergence position between the ζ probe and the ψζ 5′ flanking sequence. ζ/ψζ pSVod, an analogous construct to ζ pSVed, but lacking the SV40 enhancer, again gave no tran-

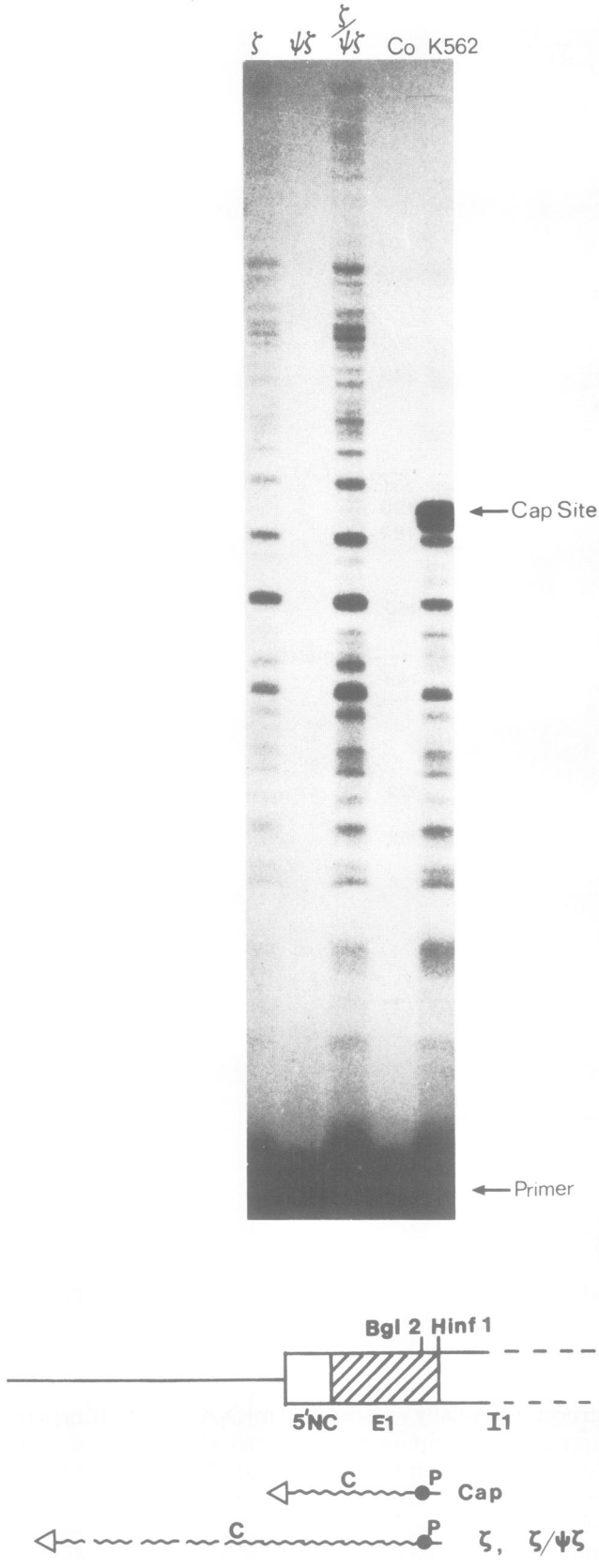

Fig. 3. 5′ primer extensions on HeLa cell total RNA transformed with ζ, ψζ and ζ/ψζ SVpBR328 plasmids. Co, minus RNA control; K562, positive control with ζ globin mRNA containing K562 total RNA. Position of primer and ζ *Cap* site are indicated. A diagram showing the positions of the primer (P) and cDNA transcripts (C) with respect to the ζ gene structure is provided. The labelled 3′ end of the primer is indicated by a dot.

scription from the ζ promoter. ζ/ψζ T pSVed gave a significant signal at the ζ *Cap* site position together with strong signals 5′ to the ζ *Cap* site, in part from beyond the 5′ end of the ζ gene insert, demonstrated by the mismatch band, and in part from a position ~ 90 bp upstream from the *Cap* site denoted by an X. The closest band to the *Cap* site corresponds to the AT rich ATA box sequence TATATAA (denoted by ATA box). Presumably this band is an artefact due to 'breathing' of the RNA/DNA duplex at this position. Similar transcripts can be detected at a very low level *in vivo* in K562 cells, as shown in the K562 lane of Figure 4 (T.R. Rutherford, in preparation). Both the X and ATA box bands are faintly visible in this lane in addition to an over-exposed *Cap* site band. We suggest that this upstream initiation site (band X) may be analogous to the upstream transcriptional starts recently identified in the human ε globin gene (Allan *et al.*, 1982, 1983). From the above data, it would seem that the positioning of the SV40 enhancer close to the ζ gene promoter causes read-through transcription from the SV40 sequence itself as well as stimulation of an inefficient initiation site upstream of the correct ζ *Cap* site. Under these conditions, the ζ *Cap* site is only slightly activated. When further separated from the enhancer sequence, as in ψζ or ζ/ψζ pSVed, the promoter displays no detectable activity. It would appear that the ζ or identical ψζ promoter sequence is remarkably insensitive to transcriptional enhancement.

Transcription from ζ and ψζ promoters in Xenopus oocytes

The various ζ pSVed plasmid constructs described above were micro-injected into oocytes: ζ, ψζ, ζ/ψζ T and ζ/ψζ Two pSVed (Figures 1b and c). In addition, two deleted ζ gene constructs ζ/ψζ T△P and △PB, which delete exon 1 and intron 1 or exon 2 and intron 2 respectively, were tested. Total RNA was purified from the different micro-injected oocytes and subjected to 5′-end analysis using S1 nuclease with a ζ globin gene 5′ probe.

As shown in Figure 5, the S1 protected DNA obtained with these different RNA samples in each case gave strong signals corresponding to the ζ mRNA *Cap* site as indicated by the K562 RNA control. Weaker signals were detected in ζ gene 5′ flanking sequence upstream of the *Cap* site. In the ψζ sample, there is a major signal at the mismatch point between the ψζ gene and the ζ probe. This represents the sum of all transcripts starting in the ψζ flanking sequences and in the vector.

When most of the ζ gene flanking sequences are deleted (ζ/ψζ T), or when the ζ promoter sequence region is inverted with respect to the rest of the gene (ζ/ψζ Two), there is still a significant signal from the *Cap* site position. The fainter signals obtained with these two RNA samples as compared to the others is due to a combination of a less efficient microinjection and the use of less RNA in the S1 experiment. However, ratio of upstream to correctly initiated transcripts is nearly equal. The mismatch band represents transcripts originating in the vector, possibly from the SV40 late promoter, which is known to be functional in oocytes ((Wickens and Gurdon, 1983; Miller *et al.*, 1982). Interestingly, ζ/ψζT, but not the opposite promoter orientation ζ/ψζ Two construct, gave a strong band X at one of the upstream start positions found at low levels in K562 RNA, and high levels in Cos7 cells transformed with ζ/ψζ T pSVed. Again, when most of the internal sequences are deleted (△P deletes all of exon 1 and most of intron 1 while △PB deletes the 3′ end of intron 1, exon 2 and most of intron 2 in the ζ/ψζ T construct), there is still a strong signal from the major *in vivo Cap* site. Thus,

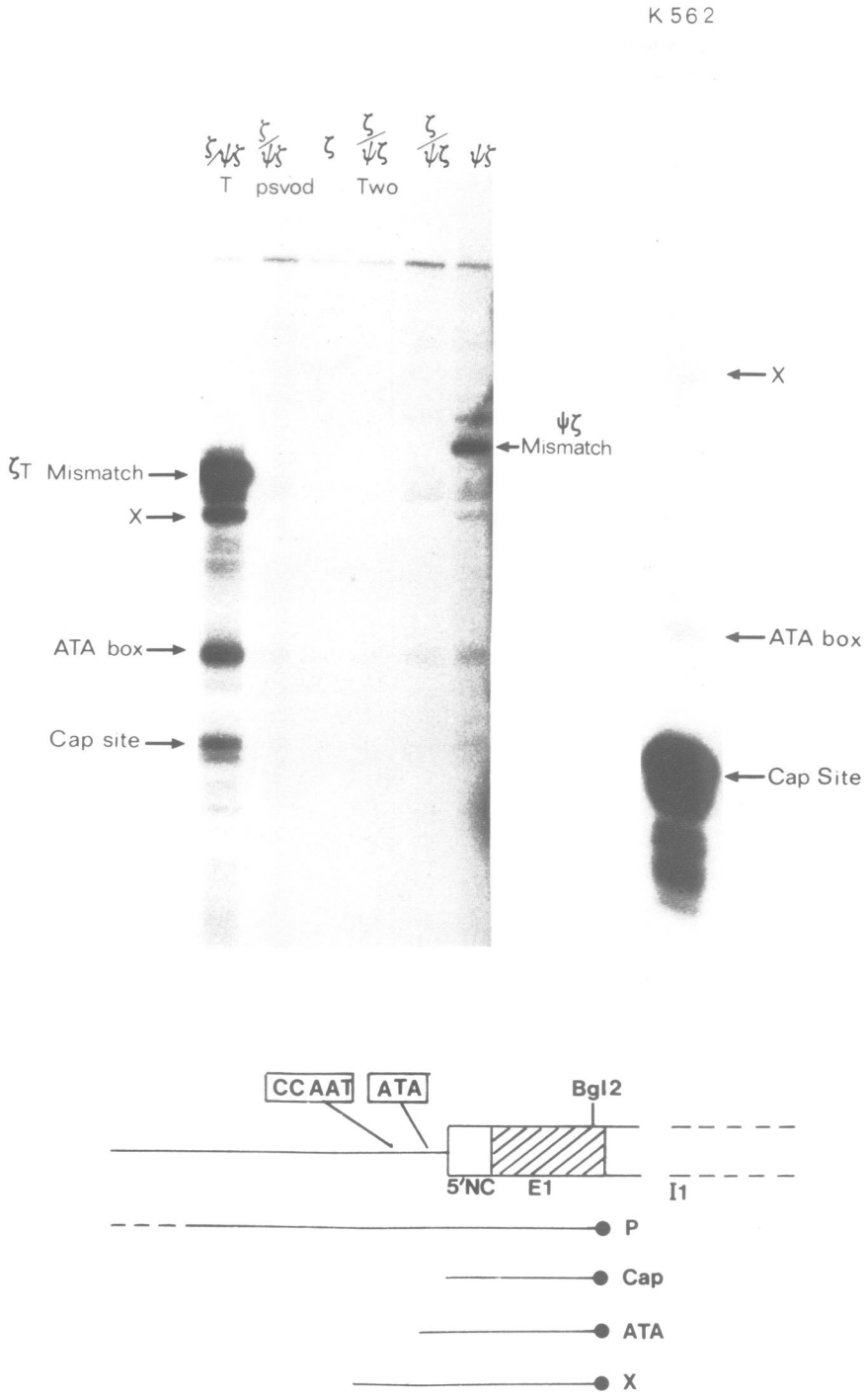

Fig. 4. 5′ end S1 analysis of Cos7 cell total RNA transformed with ζ/ψζ T pSVed, ζ/ψζ pSVod, ζ, ζ/ψζ Two, ζ/ψζ and ψζ pSVed. See the diagram below for details of the positions of the probe(s) and signals (*Cap* site, ATA box and band X) with respect to the ζ gene structure. ζ/ψζ T mismatch and ψζ mismatch indicates the point of sequence divergence between the *Eco*RI-*Bgl*II ζ 5′ probe and ζ/ψζ T or ψζ pSVed. K562 lane is an over-exposed radioautograph of 5′-end S1 analysis with K562 RNA. The faint X, ATA box and over-exposed *Cap* site bands are indicated (see text).

both ζ and ψζ promoters are functional in *Xenopus* oocytes and only relatively short sequences adjacent to the promoter (−111 to +38 bp) are required for correct transcription. The observation that the ψζ promoter functions as well as the ζ promoter in *Xenopus* oocytes is therefore not surprising, since the ζ and ψζ promoter sequences are identical in the −111 to +38 bp region.

We have recently shown that transcription from the ζ promoter in oocytes is inhibited by a low dose of α amanitin, sufficient to inhibit RNA polymerase II but not polymerase III (data not shown). Furthermore, Northern blot analysis of the RNA from oocytes injected with the ζ or ψζ constructs demonstrates the presence of high levels of discrete ζ specific bands, but not all of these may be correctly processed (G.A. Partington, unpublished results).

Discussion

We have observed that the ζ and ψζ genes are transcriptionally inactive when transfected into two mammalian tissue culture cell lines (monkey Cos7 and human HeLa). This result

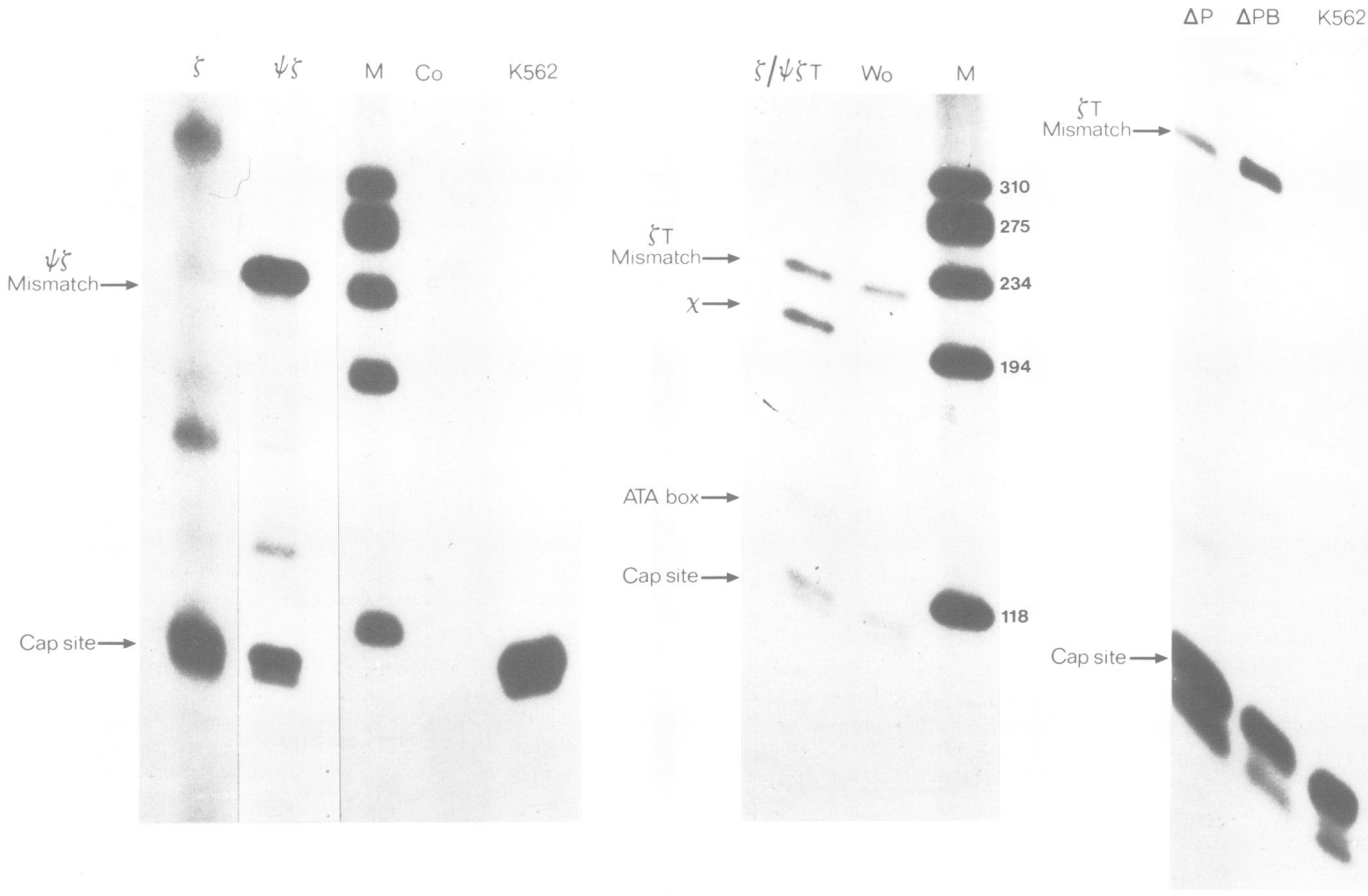

Fig. 5. 5′-end S1 mapping of RNA purified from *Xenopus* oocytes injected with ζ, ψζ, ζ/ψζ T, ζ/ψζ Two, ζ/ψζ TΔP and ζ/ψζ TΔPB pSVed as compared to K562 RNA and a minus RNA control (Co). M, size markers. Details of probe and signals obtained with respect to ζ gene sequence are demonstrated in Figure 4. Lanes ΔP, ΔPB and adjacent K562 used the different *Eco*RI-*Pvu*II ζ 5′ probe since the *Bgl*II site used for the other probe is deleted in ζ/ψζ TΔP pSVed.

has been independently confirmed by Lau and Kan (1983). These workers transfected an SV40 cosmid vector containing the entire human α-like globin gene cluster into monkey Cos7 cells and mouse L cells. In both transfected cells, high levels of α but very little ζ globin mRNAs were detected. Our results demonstrate that this inactivity of the ζ and ψζ genes can be partially alleviated when the SV40 enhancer is placed very close to the ζ or ψζ promoter (ζ/ψζ T pSVed). However, it is clear that the ζ promoter is extremely refractory to transcriptional enhancement since an upstream transcriptional initiation site, barely detectable in erythroid RNA, is stimulated to the same level as the normal ζ *Cap* site by the SV40 enhancer. Other high level transcripts are activated from the SV40 origin region when the SV40 enhancer is placed close to the ζ gene promoter.

The inactivity of the ζ globin promoter in mammalian tissue culture cells is reminiscent of the mouse immunoglobulin genes. In a similar fashion to ζ, the mouse kappa light-chain gene promoter has been carefully analysed for transcriptional activity in Cos7 cells and shown to display no transcriptional activity (Stafford and Queen, 1983). Similar results have been obtained with mouse kappa and lambda light-chain and heavy-chain genes transfected into mouse L cells (Gillies and Tonegawa, 1983). More interestingly, both the mouse heavy-chain and kappa light-chain gene promoters only function efficiently when transformed into immunoglobulin producing mouse myeloma cell lines. This promoter activity is entirely dependent on an internal intronic promoter enhancer sequence (Banerji *et al.*, 1983; Gillies *et al.*, 1983; Queen and Baltimore, 1983). However, unlike the SV40 enhancer sequence, the immunoglobulin gene enhancers are all highly cell-specific. It remains an intriguing possibility that ζ and possibly ψζ globin genes similarly possess a cell specific enhancer.

A quite different pattern of promoter activity is demonstrated for ζ and ψζ genes when injected into *Xenopus* oocytes; both promoters allow efficient transcription from the correct major *in vivo Cap* site. This efficient recognition of the ζ and ψζ gene promoters in oocytes is in marked contrast to other RNA polymerase II genes. Although several viral genes have been shown to function efficiently in *Xenopus* oocytes, generally other RNA polymerase II genes, which are transcribed in a tissue-specific manner *in vivo*, do not initiate transcription efficiently or accurately in oocytes (Gurdon and Melton, 1981). In particular, transcription of a cloned rabbit β globin gene in oocytes was reported to be entirely due to spillover transcription from the vector sequences (Rungger *et al.*, 1981). It is therefore surprising to find accurate and efficient transcription from the ζ promoters in oocytes.

Correct transcription from the construct ζ/ψζ T shows that sequences upstream of −111 are not required for pro-

moter function. Similarly the constructs △P and △PB show that internal sequences beyond +38 are not required. Even the ζ construct with the promoter region inverted with respect to the rest of the gene (ζ/ψζ Two) gives transcripts initiating at the ζ *Cap* site. Since the ζ and ψζ promoters are identical between positions −111 and +38, it is not surprising that both are functional. As described above in the *Xenopus* oocyte transcription experiments, some weaker ζ gene transcripts were observed, coming from 5′ flanking sequences upstream of the major *Cap* site. These are not simply artefactual bands produced in the oocyte system, because in fact transcripts from the same sites, but at very much lower levels, can be observed *in vivo* (T.R.Rutherford, in preparation). Similar minor upstream transcripts have been observed *in vivo* for the ε (Allan *et al.*, 1982) and β (Ley and Nienhuis, 1983; Carlson and Ross, 1983) globin genes, but their function is unknown. In the construct ζ/ψζ T, one of the upstream sites (band X) was activated to become a major initiation site. This same upstream site was demonstrated to be a major initiation site when ζ/ψζ T pSVed was transfected into Cos7 cells.

Since all the constructs described above include SV40 enhancer sequences, it might be thought that this is responsible for efficient ζ transcription in oocytes. In fact, the host range of the SV40 enhancer does not extend to *Xenopus* (de Villiers and Schaffner, 1981), and in general both SV40 and plasmid sequences have been observed to have an inhibitory effect on correct transcription in *Xenopus* cells (Rusconi and Schaffner, 1981; Gurdon and Melton, 1981). Furthermore, we have observed correct and efficient ζ gene transcription in a construct that lacks the SV40 enhancer (data not shown).

Our results show that cloned human ζ genes are not transcribed in HeLa or Cos7 tissue culture cells under conditions where other human globin genes are transcribed. In contrast, *Xenopus* oocytes, which do not correctly transcribe other globin genes, will transcribe the cloned ζ genes efficiently and accurately. We speculate that specific embryonic factors are required for ζ gene transcription, that these factors are absent from HeLa and Cos cells but that homologous factors may be present in *Xenopus* oocytes.

Materials and methods

Construction of ζ gene expression vectors

SVpBR328 constructs (Figure 1a). (ζ) The ζ gene, purified from CosHGα1 (Proudfoot *et al.*, 1982) *Eco*RI (filled in) − *Bam*HI was ligated into the transient expression vector SVpBR328 (Grosveld *et al.*, 1982) cut with *Bam*HI and *Sal*I (filled in). (ψζ) The ψζ gene purified from λHGα1 (Lauer, 1981) *Eco*RI (filled in) − *Bam*HI was ligated into SVpBR328 as with ζ. (ζ/ψζ) The ζ promoter fragment *Eco*RI (filled in) − *Bgl*II was ligated to the ψζ gene fragment *Bgl*II - *Bam*HI and the hybrid ζ/ψζ was ligated into SVpBR328 as above. The unique *Bgl*II site in ζ and ψζ is at the 3′ end of exon 1 (Proudfoot *et al.*, 1982).

The above ligation reactions were transformed into competent *Escherchia coli* (MC1061) and clones selected on ampicillin containing L agar plates. Mini-plasmid preparations were tested by restriction enzyme analysis and the correct recombinant plasmids were grown up in large scale in MC1061 and purified on CsCl gradients (Maniatis *et al.*, 1982).

pSVed constructs (Figure 1b and 1c). The transient expression vector pSVod containing the SV40 origin fragment (Mellon *et al.*, 1981) was modified to contain a larger piece of SV40 including both origin and enhancer sequences (pSVed). Thus the SV40 origin region and 72-bp repeat enhancer *Pvu*II(270)-*Hind*III(5171) was ligated into pSVod *Eco*RI (filled in)-*Hind*III. The *Eco*RI to *Pvu*II ligation reformed the *Eco*RI site. (ζ) The ζ gene *Eco*RI (filled in)-*Bam*HI fragment (purified as before) was ligated into pSVed cut with *Hind*III (filled in)-*Bam*HI. (ζ/ψζ) The ζ/ψζ gene made as described above (except that the *Eco*RI end was left as a sticky end and the *Bam*HI end was filled in) was ligated into pSVed, cut with *Eco*RI and partially cut with *Hinc*II (so that only the *Hinc*II site in the Ap^r gene of pSVed was cut). (ψζ) The ψζ gene contained within the DNA fragment *Bst*EII (filled in)-*Bam*HI (filled in) (*Bst*EII cuts ~300 bp 5′ to the ψζ *Cap* site as well as at the 5′ end of exon 3) was ligated into pSVed cut with *Eco*RI (filled in) and *Hinc*II as above. (ζ/ψζ T) The 5′ flanking region of ζ/ψζ pSVed was deleted bringing the SV40 enhancer close to the ζ *Cap* site as follows. A *Sau*3A-*Bgl*II fragment (both ends filled in) was purified from the ζ gene that includes ~100 bp of 5′ flanking sequence plus the 5′ non-coding region and exon 1 of ζ. ζ/ψζ pSVed was cut with *Eco*RI and *Bgl*II and both ends were filled in. The ζ promoter fragment was ligated into this vector. The correct transcriptional orientation is called ζ/ψζ T pSVed (T stands for truncated promoter); while the opposite orientation insert with the ζ promoter facing in the opposite direction to the rest of the gene is called ζ/ψζ Two (wo stands for wrong orientation). (ζ/ψζ T △P and △PB) The ζ gene sequence in ζ/ψζ T pSVed was partially deleted by excising a *Pvu*II fragment that includes all exon 1 and most of intron 1 (△P) or a *Pvu*II to *Bst*EII fragment that includes the 3′ end of intron 1, exon 2 and most of intron 2 (△PB) [see Proudfoot *et al.* (1982) for precise positions of these restriction enzyme sites in the gene sequence].

Transcription systems

Transient expression. HeLa cells or Cos7 cells grown in DMEM, 10% new born calf serum, were transformed with SVpBR328 plasmids (HeLa) or pSVed plasmids (Cos7). DNAs (20 μg/90 mm plate) were precipitated with calcium phosphate and added to subconfluent HeLa or Cos7 cells. After 12 h, the medium was changed and the cells were incubated for a further 48 h. Cells were then harvested and total RNA purified by proteinase K digestion, phenol extraction, ethanol precipitation and finally, DNase I digestion (Maniatis *et al.*, 1982). The efficiency of HeLa and Cos7 cell transfection (% cells expressing transfected genes) and the average number of plasmid copies per cell were not directly measured in these experiments, but have been described by others (Mellon *et al.*, 1981; Grosveld *et al.*, 1982).

Micro-injection of ooctyes. The micro-injection technique and oocyte culture conditions have been described previously (Mertz and Gurdon, 1977). Stage V-VI oocytes (Dumont, 1972) were taken from female *X. laevis* caught in the wild, weighing 100 g or less. Each oocyte was injected with 50 nl containing 25 ng of plasmid. The oocytes were then incubated at 19°C for 2 days. Oocytes usually in groups of 20−30 were homogenised quickly in 1 ml of buffer containing 0.3 M sodium acetate pH 5, 0.8% SDS and 4 μg/ml polyvinylsulphate. The homogenate was then diluted to 5 ml with the same buffer and phenol-chloroform extracted. The organic phase was re-extracted with 5 ml 0.2 M NaCl 50 mM Tris-HCl pH 9 and 0.8% SDS. The combined aqueous phases were re-extracted with first pheno-chloroform and then chloroform. Finally, total RNA was ethanol precipitated from the aqueous phase.

RNA mapping

S1 mapping (Berk and Sharp, 1977; Weaver and Weissman, 1979). ψζ, ζ and α globin transcripts were mapped using a double strand DNA probe 3′ end-labelled (filled in with [α-^{32}P]dGTP and Klenow DNA polymerase I) at the *Bst*EII site present at the 5′ end of exon 3 in each gene (Figure 2). The probe DNA was annealed to the different RNAs in 30 μl of 80% formamide, 0.04 M Pipes pH 6.8, 0.4 M NaCl, 0.1 mM EDTA by denaturing at 80°C for 10 min, then 53°C overnight. 300 μl of ice cold S1 buffer (0.25 M NaCl, 0.03 M NaOAC pH 4.6, 2 mM $ZnSO_4$, 50 μg/ml denatured sonicated carrier DNA) plus S1 (3000 units) was quickly added to each hybridisation and incubated for 1 h at 30°C. S1 reactions were ethanol precipitated and fractionated by electrophoresis on denaturing, 7 M urea polyacrylamide gels. 5′-end analysis of ζ and ψζ globin transcripts was carried out using a single-stranded DNA probe. Either a 650-bp *Eco*RI-*Bgl*II (Figure 4) or a 600-bp *Eco*RI-*Pvu*II ζ promoter fragment was treated with calf intestinal phosphatase and kinase-labelled with [γ-^{32}P]ATP. The DNA was then denatured at 95°C in 50% DMSO and strand-separated by electrophoresis on a native 5% polyacrylamide gel. The antisense strand was then annealed to RNA as above except that incubations were carried out at 30°C rather than 53°C. S1 digestion and fractionation was carried out as before.

Primer extension (Proudfoot *et al.*, 1980; Devine *et al.*, 1982). ζ Globin transcript 5′ ends were mapped as shown in Figure 3 using a single strand antisense DNA fragment *Bgl*II-*Hinf*I 3′ end-labelled with [α-^{32}P]dGTP and Klenow DNA polymerase I. DNA primers and RNA were annealed in 10 μl of 10 mM Pipes pH 6.4, 0.4 M NaCl, 80°C 10 min, 63°C overnight. 50 μl of reverse transcriptase buffer (50 mM Tris pH 8.2, 10 mM DTT, 6 mM $MgCl_2$, 0.5 mM dATP, dCTP, dGTP, dTTP) plus reverse transcriptase (5 units) were added to hybridisations and incubated at 42°C for 1 h. The reactions were ethanol precipitated and fractionated by electrophoresis on denaturing 7 M urea polyacrylamide gels.

Acknowledgements

We gratefully acknowledge Nicola Yarwood for technical assistance. We thank Professor D.J.Weatherall for encouragement throughout these studies. T.R.R. is supported by the British Cancer Research campaign. N.J.P. is supported by a Medical Research Council project grant no. G8108316CB.

References

Allan,M., Grindlay,G.J., Stefani,L. and Paul,J. (1982) *Nucleic Acids Res.*, **10**, 5133-5144.
Allan,M., Lanyon,W.G. and Paul,J. (1983) *Cell*, **35**, 187-197.
Asselberg,F.A.M., Smart,J.E. and Mathews,M.B. (1983) *J. Mol. Biol.*, **163**, 209-238.
Banerji,J., Rusconi,S. and Schaffner,W. (1981) *Cell*, **27**, 299-308.
Banerji,J., Olson,L. and Schaffner,W. (1983) *Cell*, **33**, 729-740.
Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Carlson,D.P. and Ross,J. (1983) *Cell*, **34**, 857-864.
Cortese,R., Melton,D.A., Tranquilla,T. and Smith,J.D. (1978) *Nucleic Acids Res.*, **5**, 4593-4611.
Deisseroth,A., Nienhuis,A., Turner,P., Velez,R., Anderson,W.F., Lawrence,J., Creagan,R. and Kucheriapati,R. (1977) *Cell*, **12**, 205-218.
de Villiers,J. and Schaffner,W. (1981) *Experentia*, **37**, 642.
Devine,J.M., Tsang,A.S. and Williams,J.G. (1982) *Cell*, **28**, 793-800.
Dumont,J.N. (1972) *J. Morphol.*, **136**, 153-180.
Gillies,S.D. and Tonegawa,S. (1983) *Nucleic Acids Res.*, **11**, 7981-7997.
Gillies,S.D., Morrison,S.L., Oi,V.T. and Tonegawa,S. (1983) *Cell*, **33**, 717-728.
Gluzman,Y. (1981) *Cell*, **23**, 175-182.
Grosveld,G.C., De Boer,E., Shewmaker,C.K. and Flavell,R.A. (1982) *Nature*, **295**, 120-126.
Gurdon,J.B. and Melton,D.A. (1981) *Annu. Rev. Genet.*, **15**, 189-218.
Hentschel,C., Probst,E. and Birnstiel,M.L. (1980) *Nature*, **288**, 100-102.
Huehns,E.R., Flynn,F.U., Butler,E.A. and Beaven,G.H. (1961) *Nature*, **189**, 496-497.
Korn,L.J. (1982) *Nature*, **295**, 101-105.
Kressman,A., Clarkson,S.L.G., Pirrotta,V. and Birnstiel,M.L. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1176-1180.
Lau,Y-F. and Kan,Y.W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5225-5229.
Lauer,J. (1981) Ph.D. Thesis, Caltech.
Lauer,J., Shen,C-K.J. and Maniatis,T. (1980) *Cell*, **20**, 119-130.
Ley,T.J. and Nienhuis,A.W. (1983) *Biochem. Biophys. Res. Commun.*, **112**, 1941-1048.
Liebhaber,S.A., Goosens,M., Poon,R. and Kan,Y.W. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 7054-7058.
Maniatis,T., Fritsch,E.F., Lauer,J. and Lawn,R.M. (1980) *Annu. Rev. Genet.*, **14**, 145-178.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, a Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
McKnight,S.L., Gavis,E.R., Kingsbury,R. and Axel,R. (1981) *Cell*, **25**, 385-398.
Mellon,P., Parker,V., Gluzman,Y. and Maniatis,T. (1981) *Cell*, **27**, 279-288.
Mertz,J.E. and Gurdon,J.B. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 1502-1506.
Michelson,A.M. and Orkin,S.H. (1980) *Cell*, **22**, 371-377.
Miller,T.J., Stephens,D.L. and Mertz,J.E. (1982) *Mol. Cell. Biol.*, **2**, 1581-1594.
Picard,D. and Schaffner,W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 417-421.
Probst,E., Kressman,A. and Birnstiel,M.L. (1979) *J. Mol. Biol.*,, **135**, 709-732.
Proudfoot,N.J. and Maniatis,T. (1980) *Cell*, **21**, 537-544.
Proudfoot,N.J., Gil,A. and Maniatis,T. (1982) *Cell*, **31**, 553-563.
Proudfoot,N.J., Shander,M.H.M., Manley,J.L., Gefter,M.L. and Maniatis, T. (1980) *Science (Wash.)*, **209**, 1329-1336.
Queen,C. and Baltimore,D. (1983) *Cell*, **33**, 741-748.
Rungger,D., Matthias,P.D. and Huber,J.P. (1981) in Schweiger,H.G. (ed.), *International Cell Biology*, 1980-1981, Springer-Verlag, Berlin, pp. 28-32.
Rusconi,S. and Schaffner,W. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 5051-5055.
Rutherford,T.R., Clegg,J.B. and Weatherall,D.J. (1979) *Nature*, **280**, 164-165.
Stafford,J. and Queen,C. (1983) *Nature*, **306**, 77-79.
Weaver,R.F. and Weissman,C. (1979) *Nucleic Acids Res.*, **6**, 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.

Received on 5 March 1984; revised on 9 April 1984