The EMBO Journal vol.10 no.7 pp.1833-1842, 1991

A pause site for RNA polymerase II is associated with termination of transcription

Pita Enriquez-Harris, Nicola Levitt, Diana Briggs and Nick J.Proudfoot

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Communicated by N.J.Proudfoot

Termination of transcription by RNA polymerase II has been postulated to involve a pausing process. We have identified such a pause signal, 350 bp into the 3′ flanking region of the human α2 globin gene at a position where termination is thought to occur. We show that this pause signal enhances the utilization of an upstream poly(A) site which is otherwise out-competed by a stronger downstream poly(A) site. We also demonstrate that the pause site rescues a poly(A) site that is inactive due to its location within an intron. Using nuclear run-on analysis we show that elongating RNA polymerase II molecules accumulate over this pause signal. Furthermore we show that when the pause site is positioned immediately downstream of a strong poly(A) signal, significant levels of transcription termination take place.
Key words: pause signal/poly(A) site/RNA polymerase II/transcription termination

Introduction

In higher eukaryotes, the 3′ ends of mRNA molecules do not appear to correspond to the position at which the RNA polymerase disengages from the DNA template (as is the case in most prokaryotic RNA molecules). Instead, these 3′ ends are formed by processing events which specifically cleave the transcript and add a poly(A) tail. Both *cis*- and *trans*-acting factors are required for these reactions and are now relatively well characterized (Manley, 1988; Proudfoot and Whitelaw, 1988). Transcription termination has been defined as the cessation of RNA chain elongation followed by release of the RNA polymerase from the DNA template (Holmes *et al.*, 1983). Recently, a series of studies has enabled the beginning of an understanding of the molecular events which lead to transcription termination by RNA polymerase II. It is known that this enzyme is capable of transcribing a gene well beyond the poly(A) signal at least in the cases so far studied. Termination regions mapped for genes transcribed by RNA polymerase II (polII genes) lie between 100 and 4000 bp downstream of the poly(A) site (reviewed by Proudfoot, 1989). The RNA polymerase II therefore would appear to have a surprising capacity for remaining actively bound to the transcription unit. Such a property of the enzyme would be of great advantage during the transcription of very large genes such as the dystrophin gene (2 Mb). On the other hand, it makes it difficult to envisage how or why RNA polymerase II transcription should terminate without the aid of *cis*-acting sequences which would indicate to the enzyme that the 3′ end of a transcription unit had been reached. Indeed in the case of snRNA genes transcribed by RNA polymerase II, a specific 3′ terminal sequence is known to elicit transcriptional termination (Kunkel and Pederson, 1985; Hernandez and Weiner, 1986; Neuman de Vegvar *et al.*, 1986).

A requirement for a functional poly(A) site in the termination of transcription by RNA polymerase II has been shown in a number of different polII genes: human α2 globin, mouse β globin, adenovirus major late and polyoma virus late (Whitelaw and Proudfoot, 1986; Logan *et al.*, 1987; Connelly and Manley, 1988, 1989a; Lanoix and Acheson, 1988). However, it is also likely that other sequences, hundreds or even thousands of base pairs downstream of the poly(A) signal, play a role in termination (Citron *et al.*, 1984; Hagenbuchle *et al.*, 1984; LeMeur *et al.*, 1984; Logan *et al.*, 1987). In the adenovirus major late promoter, an inverted CCAAT box sequence functions as a terminator, preventing readthrough from an upstream gene, into the major late promoter (Connelly and Manley, 1989a). The orientation dependence of this element and deleterious effects of single base substitutions throughout suggest that a CCAAT box binding protein is required (Connelly and Manley, 1989b). The adenovirus major late promoter is reminiscent of RNA polymerase I genes whose promoter regions are shown to include termination signals that act on transcription from upstream polI genes (Grummt *et al.*, 1986; Henderson and Sollner-Webb, 1986; McStay and Reeder, 1986). However, it is unclear if the concept of transcription initiation factors which double as termination factors for polII genes is generally applicable. Certainly in chromosomal polII genes, all cases so far investigated indicate that termination occurs in the 3′ flanking region some distance upstream of the next polII promoter (Proudfoot, 1989).

The discovery that termination of transcription by RNA polymerase II requires a functional poly(A) site has led to a number of different models for this process (see Discussion). One model has suggested that the slowing down or pausing of the elongating RNA polymerase beyond the 3′ processing signals of the gene is a key component of the termination process (Proudfoot, 1989). In keeping with this prediction, we present data which provide evidence from two separate studies for such a pause signal in the 3′ flanking region of the human α2 globin gene. Using assays in which RNA processing factors are competed for by two closely positioned processing signals, we have previously shown that (i) a stronger downstream poly(A) site out-competes a weaker upstream poly(A) site, and (ii) splicing out-competes polyadenylation even at a strong poly(A) site (Levitt *et al.*, 1989). As described in these studies, the presence between two such competing RNA processing signals of a 92 bp sequence from the 3′ flanking region of the human α2 globin gene is found in each case to encourage utilization of the

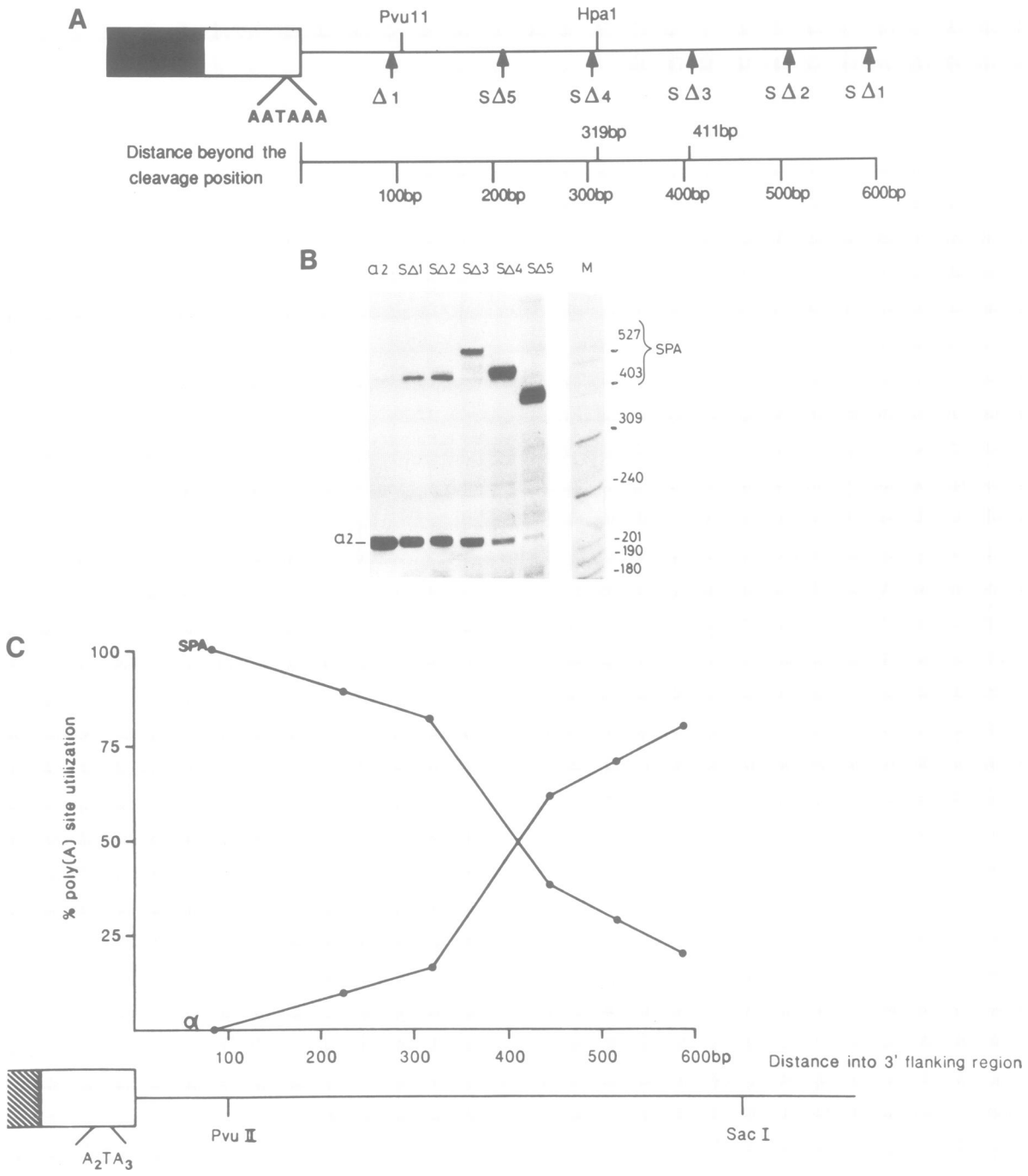

Fig. 1. A. Bal31 deletion mutants of the α2 globin 3′ flanking region. Arrows denote deletion endpoints. Exons (solid box), non-coding sequences (open box) and 3′ flanking region (line) are shown. Bal31 exonuclease was used to delete sequences 5′ of the *Sac*I positioned ~665 bp downstream of the α2 poly(A) site. The SPA was inserted at each endpoint. Δ1 was generated by Bal31 exonuclease deletion from the *Pvu*II site and has been described previously (Levitt *et al.*, 1989). **B.** S1 nuclease analysis of the (*Sac*I) Bal31 deletion mutants (SΔ1 to SΔ5). The α2 control plasmid gives a band at ~220 nt which corresponds to the use of the α2 poly(A) site. Larger bands correspond to the use of the SPA. A trend in the differential usage of the two poly(A) sites may be observed. **C.** Plot of percentage poly(A) site utilization against distance in the 3′ flanking region, as determined by densitometric analysis of the data shown in B.

upstream signal. Furthermore, when this sequence is placed in the antisense orientation, acting merely as a spacer of the two processing signals, it has a negligible effect. From this we conclude that the 92 bp region acts in an orientation dependent manner to decrease the rate at which the downstream, stronger processing signal is reached by the RNA polymerase. We also provide direct evidence from nascent chain analysis for a significantly increased polymerase density over the 92 bp region. Such an accumulation of polymerase molecules would be expected in a region of transcriptional pausing. When the 92 bp region is placed 3′ to an efficient poly(A) site, we not only detect transcriptional pausing but also significant levels of termination beyond the pause signal. Thus the pause site appears to interact with the poly(A) site to facilitate transcriptional termination.

Results

Specific sequence in the 3′ flanking region of the α2 globin gene enhances utilization of an upstream poly(A) site

An indirect method of assaying for pausing or termination was devised. It has previously been shown that when a strong, synthetic poly(A) site (SPA) is placed immediately downstream of the relatively weak α2 globin poly(A) site, the SPA out-competes the α2 poly(A) site (Levitt *et al.*, 1989). We decided to investigate whether the positioning of the SPA further downstream would alter its usage. A series of Bal31 exonuclease deletion mutants in the 3′ flanking region of the α2 gene was constructed, as represented by Figure 1A. The SPA was inserted in the correct orientation at each deletion end-point with the same 3′ flanking sequence positioned against the SPA in each clone. In this way, it was possible to investigate the effects of moving the SPA further downstream from the α2 poly(A) site. Whilst the sequence between the α2 poly(A) site and the SPA was altered by the deletions, the sequence downstream of the SPA was kept constant. Any change in the ratio of usage of the two competing poly(A) sites could then be considered due to either a distance effect, or to a specific sequence between them.

Figure 1B shows the S1 analysis of cytoplasmic RNA from cells transfected with SΔ1 to SΔ5. A homologous DNA probe labelled at the *Bst*EII site in exon 3 of the α2 globin gene (Figure 2A) was used for analysis of SΔ3, SΔ4 and SΔ5. RNAs from SΔ1 and SΔ2 were analysed with a heterologous probe made from SΔ4 DNA. The resulting autoradiograph was then scanned by a densitometer in order to calculate the percentages of usage of the α2 and synthetic poly(A) sites. These data were then plotted as seen in Figure 1C. The α2 control illustrates the wild-type situation, in which only the α2 poly(A) site is utilized. Between SΔ5 and SΔ4, a small decline in the use of the more distant SPA is observed. Accompanying this trend is a compensating increase in the usage in the weaker α2 poly(A) site. Similarly, between SΔ3 and SΔ1 the same phenomenon occurs—as the stronger SPA is moved further away from the α2 poly(A) site, its usage declines, whilst that of the α2 poly(A) site increases. The interesting observation from these data is that between SΔ3 and SΔ4, there is a significant deviation from the otherwise gradual trend. The 92 bp sequence lying between the deletion endpoints in SΔ3 and SΔ4 would appear to have the property of increasing the use of the upstream α2 poly(A) site. Were one to extrapolate the trend observed between SΔ5 and SΔ4, one would expect that for SΔ3, the α poly(A) site would be used in ~22% of the transcripts. Instead, it is used in 58% of the transcripts. Again, merely by extrapolation, such a rate of usage of the α2 poly(A) site might not be expected until the SPA had been moved ~1100 bp downstream of the α2 poly(A) site.

Our interpretation of these data is as follows: once a poly(A) site has been revealed in the transcript, it may readily be used. If an alternative poly(A) site is transcribed shortly afterwards, the processing factors may preferentially cleave and process one of the two sites. Should the upstream signal be the stronger it will always be more efficiently processed, being the first polyadenylation substrate available as transcription proceeds, as well as more effectively competing for the processing factors. If, however, the stronger poly(A) signal lies downstream, then the weaker upstream poly(A) site will be processed as a function of the distance lying between the two sites. The further away the strong poly(A) signal, the greater the opportunity for processing factors to recognize the weak upstream site.

The 92 bp sequence between the deletion endpoints in SΔ3 and SΔ4 behaves as might a spacing of 1100 bp in that it increases the usage of the upstream α poly(A) site to the same degree expected of a 1.1 kb spacer (based on extrapolation of the graph in Figure 1C). This could conceivably be due to some degree of transcriptional termination occurring in this 92 bp region. Alternatively, the polII may not actually dissociate from the template. Instead, it may merely transcribe the 92 bp at the same rate as it would transcribe a 'neutral' 1.1 kbp spacer, i.e. a spacer lacking termination, attenuation or 'pause' sequences. The 92 bp region could then be described as a 'pause' site since it appears to have a kinetic effect on transcription.

Transcriptional analysis of the α2 3′ flanking region

To demonstrate directly that the 92 bp sequence in the α2 3′ flanking region is a transcriptional pause site, we made a more direct investigation of the transcriptional state of the α2 gene and its 3′ flanking region. Nascent RNA chains from the isolated nuclei of HeLa cells presented with the human α2 globin expression plasmid, pα2W3′PS (Figure 2A), were elongated in the presence of [α-^{32}P]UTP (nuclear run-on analysis). Under the conditions used, transcription initiation and RNA processing do not occur. Preinitiated RNA chains are elongated by ~100 nt (Weber *et al.*, 1977; see also Bentley and Groudine, 1988; Connelly and Manley, 1989; and references therein). This nuclear run-on assay, then, measures the density of polymerase molecules engaged in transcription along the gene transcription unit. Labelled RNA extracted from these transfected cell nuclei was hybridized to single-stranded M13 DNA containing contiguous fragments from the α2 globin gene and its 3′ flanking region (see Figure 2B). The single-stranded M13 DNA probes were either run on an agarose gel and blotted onto nylon membranes (Figure 2B and C) or immobilized on nylon membrane to which they were applied using a slot-blot apparatus (Figure 2D). Any variations in the signal intensity between the probes are accounted for by (i) differences in the UTP content of the transcript at a particular stage and (ii) variations in the density of polymerase molecules over any region.

Figure 2B shows an example of such a nuclear run-on experiment using α2W3′PS transfected nuclei. When antisense (−) M13 probes are used, there is clearly a greater signal from probe 6 including the putative pause site region [+319 to +429, where 0 marks the position of the poly(A) site] than from anywhere else in the α2 globin transcription unit. Table I lists the U content of each portion of the α transcript (detected by the − sense probes). Taking into account the slightly higher proportion of U residues in the probe 6 transcript, it is still evident that the probe 6 signal is 3- to 4-fold greater than the average signal over the α coding region (based on data shown as well as numerous unpublished experiments). It should also be noted in Figure 2B that there is significant transcriptional signal in the sense (+) M13 probes which must reflect transcription off the other strand of the α2W3′PS plasmid. Presumably these transcripts initiate from the SV40 early promoter present in the SV40 origin region (see hatched box in Figure 2A). Figure 2B allows us to conclude that the highest density of

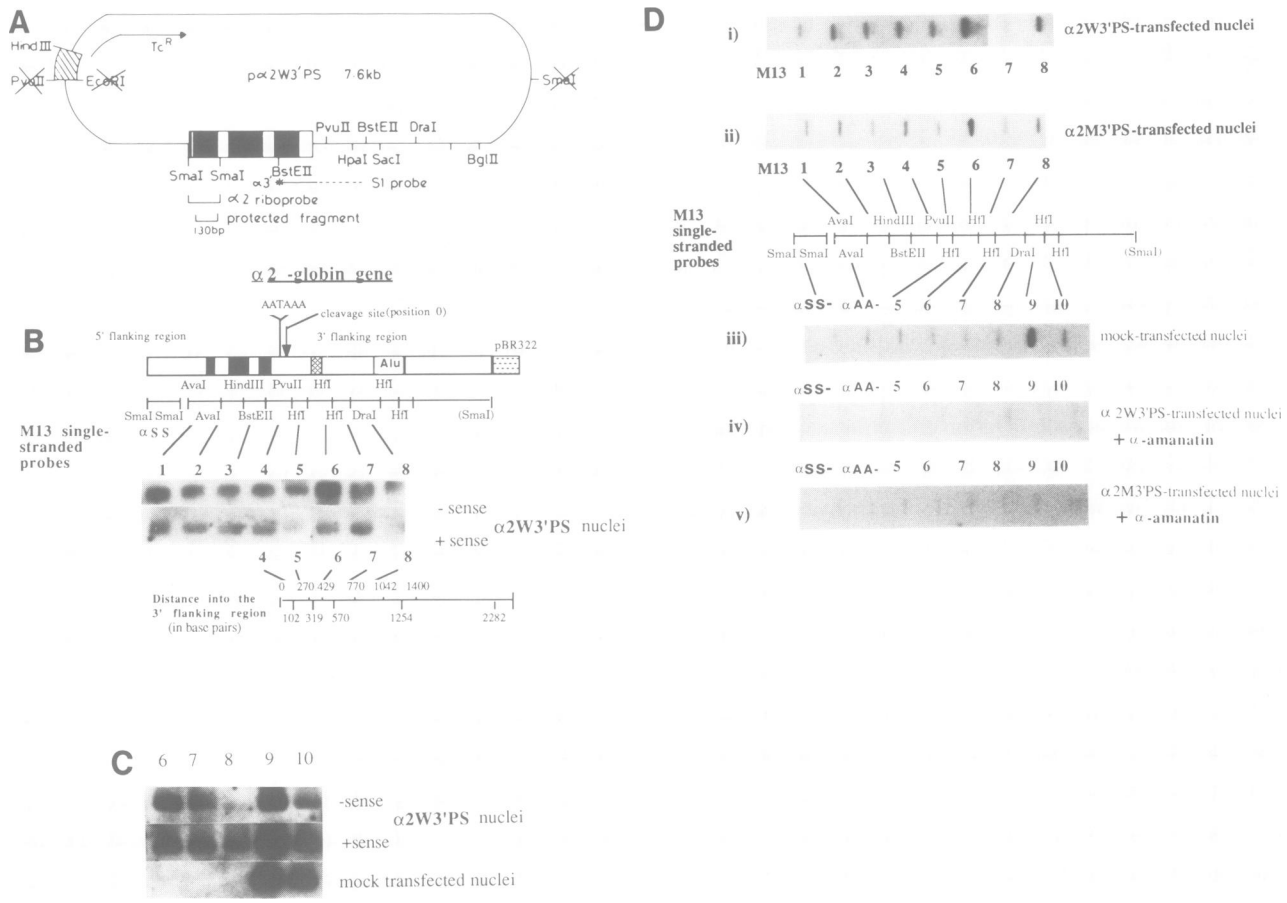

Fig. 2. A. Schematic representation of plasmid pα2W3′PS, the construction of which is described in Whitelaw and Proudfoot (1986). Exons (solid boxes), introns and non-coding sequence (open boxes) and SV40 origin sequence (hatched box) are shown. The fragment from the α2 globin exon 1 which was used to construct the α2 riboprobe is also shown, as is the protected fragment generated by RNase digestion of probe:α2 transcript hybrids. **B.** Nuclear run-on analysis of pα2W3′PS transfected nuclei. Both sense and antisense probes were used. A diagram showing the positions of these single-stranded M13 DNA probes on the α2 globin gene map is shown. Probes containing the antisense or sense fragments were immobilized on nylon filters and used to detect radio-labelled sense or antisense transcripts, respectively, from pα2W3′PS-transfected nuclei. Exons (solid boxes), non-coding (open boxes), pBR322 (striped box), Alu repeats (labelled box) and the pause site sequence (hatched box) are all shown. **C.** Nuclear run-on analysis of pα2W3′PS transcription downstream of the pause site. The probes used are described in B. The high intensity of signal over clone 6 (containing the pause site) and clone 9 (containing the Alu repeat) may clearly be seen. As in B, filters containing both sense and antisense probes are shown. Also shown is a filter from a nuclear run-on assay performed on mock-transfected nuclei. **D.** Nuclear run-on analysis of pα2W3′PS compared to pα2M3′PS transfected nuclei. A different method of filter preparation and nuclear run-on analysis was employed (see Materials and methods). Also shown is the effect of α-amanitin on the nuclear run-on signal as well as another mock transfection experiment.

polymerase molecules anywhere in the α2 globin transcription unit is observed between +270 and +570. This is the region encompassed by the α 3′ flanking sequence in M13 clone 6, and includes the 92 bp sequence implicated as a pause site in the previous section. It is unlikely that the high transcription signal in clone 6 is due to newly initiated transcripts, since under the run-on conditions used, initiation does not occur and in any case no cryptic promoter sequences are known to exist between +270 and +570. We therefore propose that polymerase molecules are being paused or stalled as they pass through this part of the DNA template. Such an effect of polymerase build-up over a region shortly downstream of a poly(A) site has also been observed in the rabbit α globin gene (R.Hardison, personal communication).

Further downstream of the putative pause region is a repetitive Alu repeat sequence (Shen and Maniatis, 1982). Alu transcripts from the entire endogenous genome will therefore hybridize to M13 DNA probes containing fragments which include such sequences. In Figure 2C, it can be seen that clones 9 and 10 include sequences which in both sense and antisense orientations, strongly hybridize

Table I.

Probe	Length (nucleotides)	Number of U residues in α transcript
αSS	470	80
1	240	34
2	270	37
3	200	34
4	300	71
5	160	26
6	300	55
7	200	51
8	270	63
9	210	35
10	150	31

to labelled, nascent RNA. When mock-transfected nuclei were used, only clones 9 and 10 showed significant hybridization to the labelled transcripts. These results are consistent with the interpretation that any signal observed over clones 9 and 10 is due to transcription of endogenous Alu repeats. The presence of the Alu sequence at this precise location, therefore, makes it difficult to study transcriptional termination of the α2 globin gene beyond the pause site.

Figure 2D shows a nuclear run-on experiment which compares nuclei of cells transfected with pα2M3′PS with pα2W3′PS (Whitelaw and Proudfoot, 1986). pα2M3′PS is a plasmid identical to pα2W3′PS except for a thalassemic mutation in the poly(A) site. It can be seen that the distribution of polymerases over the α2 globin gene with the mutant poly(A) site bears a close resemblance to the situation for pα2W3′PS. In particular a strong signal is observed for probe 6. We would expect that in the α2 globin gene with the mutant poly(A) site, polymerases may still accumulate over a pause site, but without the associated poly(A) signal no termination would occur. These results also confirm that there is an accumulation of polymerases over probe 6 as in Figure 2B. However, we notice in this experiment that for both transfections there is an immediate drop in signal with probe 7 in contrast to Figure 2B where probe 7 has a higher signal (although still 3-fold lower than probe 6). The nuclear run-on protocol was somewhat modified in Figure 2D as compared with Figure 2B and C (see Materials and methods). In particular the incubation time of nuclei in *in vitro* transcription was 15 min longer in Figure 2B and C which may result in the more spread out transcription signal peaks, i.e. higher signals in probes 7 and 10, following the pause site and Alu repetitive transcripts.

Figure 2D also shows three control nuclear run-on assays. Mock-transfected nuclei give no significant hybridization signal except for over clone 9, which contains the largest portion of the Alu repeat. The absence of any signal over probe 6 in this and in the mock-transfected experiment shown in Figure 2C rules out the possibility that the high signal obtained might derive from endogenous transcripts cross hybridizing to probe 6. In the presence of 5 μg/ml α-amanitin, nuclei transfected with either pα2W3′PS or pα2M3′PS give no significant hybridization signal in the examined part of the 3′ flanking region of the α2 gene. These results clearly demonstrate that (i) there is little background hybridization, except in the case of Alu repeat-containing fragments; (ii) polymerase accumulation over clone 6 is sensitive to low concentrations of α-amanitin; and (iii) transcripts which hybridize to the Alu repeat are also sensitive to low concentrations of α-amanitin. Although isolated Alu repeats are transcribed by RNA polymerase III (Shen and Maniatis, 1982), many Alu repeats (such as this one) lie within polII transcription units and will therefore be transcribed by RNA polymerase II. It is therefore not surprising that a reduced Alu repeat signal is observed in the presence of 5 μg/ml α-amanitin.

Activation of a poly(A) site positioned in an intron by the α2 pause site

We have previously demonstrated that an efficient poly(A) site will not function if placed within an intron (Levitt *et al.*, 1989). This is thought to be due to the predominance of splicing over polyadenylation (Adami and Nevins, 1988). We have suggested that the kinetics of splicing are such that an intronic poly(A) signal will be used only slightly, if at all, presumably because polyadenylation is the slower of the two processes (Levitt *et al.*, 1989). If this is so, we would expect that by allowing more time for the utilization of an intronic poly(A) signal, the efficiency of polyadenylation at this position might be increased. The 92 bp sequence from the α2−α1 globin intergenic region is predicted, from our previous two experiments, to have the property of kinetically altering, or pausing, transcription in its vicinity. We constructed plasmid molecules in which this 92 bp sequence was positioned in both orientations directly downstream of the synthetic poly(A) site (SPA) in the second intron of the rabbit β globin gene (see Levitt *et al.*, 1989, for details of construction of the parent plasmid pSPAI). These constructs were named pSPAIP1+ and pSPAIP1− and are represented in Figure 3A. If the putative pause site does indeed have a kinetic effect on transcription such that polymerase molecules are paused as they transcribe this region, then we would expect the intronic poly(A) site to be utilized.

An antisense RNA probe was made by subcloning a fragment containing the SPA and some pUC119 polylinker sequence from pSPAI into plasmid pSP65 (Melton *et al.*, 1984). Cytoplasmic RNA samples from HeLa cells transfected with constructs derived from RβSVpBR328 were hybridized to this probe (pySPA$^-$). Figure 3A shows the extents of probe pySPA$^-$ and of the protected RNA obtained when processing and polyadenylation have occurred at the intronic SPA. Figure 3B shows an autoradiograph of an RNase protection assay on cytoplasmic RNA from HeLa cells transfected with SPAIP1+ and SPAIP1−. As controls, RNA samples from HeLa cells transfected with rabbit β globin, SPAI and SPAIΔD, were also analysed. In SPAIΔD, the splice donor site has been deleted and so the intronic SPA is efficiently utilized (Levitt *et al.*, 1989). This, then, controls for the maximum possible use of the intronic SPA, whilst SPAI controls for the basal utilization of the SPA in the intron. Each construct was co-transfected with the α2 globin plasmid pα2W3′PS (see Figure 2A) to control for transfection efficiencies. Thus, each RNA sample was also analysed with an α2 globin riboprobe. This antisense probe was constructed by subcloning a *Sma*I−*Sma*I fragment from exon 1 of the α2 globin gene into pSP65 (Figure 2A).

The autoradiograph shown in Figure 3B was analysed by denistometry to allow comparison of the levels of protected RNA between the experiments in which transfection efficiencies were variable. Table II shows the relative utilization of the SPA in HeLa cells transfected with the various constructs. Since Rβ possessed no SPA in the second intron, any signal on the autoradiograph at this position was assumed to be the background due to probe degradation. Such signal was therefore subtracted from the corresponding signal for all other constructs. The α globin co-transfection signal was equalized in all cases, in order that all experiments might be directly compared.

From Table II and Figure 3B it can be seen that there is a negligible use of the SPA in cells transfected with SPAI or SPAIP1−. However, positioning the pause site immediately downstream of the SPA, in the correct orientation (SPAIP1+), increases the utilization of the intronic SPA to 25% of that seen when the splice donor site is deleted. Since the splice site deletion will prevent splicing, it effectively removes the SPA from the intron and allows 3′ end processing or polyadenylation to occur at this position.

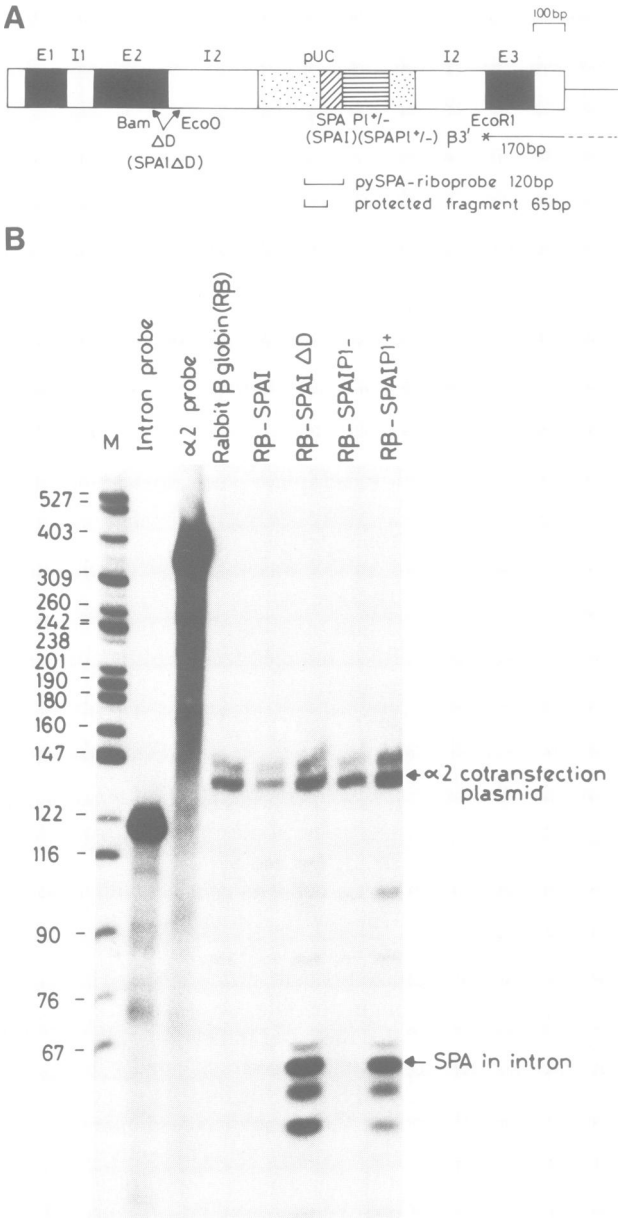

Fig. 3. A. Representation of rabbit β globin gene derivatives SPAIP1+ and SPAIP1−. Exons (solid boxes), introns and non-coding sequences (open boxes), pUC119 polylinker sequence (stippled boxes), the synthetic poly(A) site (diagonally striped box) and either orientation of the pause site sequence (horizontally striped box) are shown. Both probes used to detect transcripts from these plasmid constructions are shown. The *Eco*RI site was used to end-label the S1 DNA probe, and this probe was therefore made from the wild-type rabbit β globin plasmid RβSVpBR328 in which there is the pUC119-derived *Eco*RI site. The riboprobe construct pySPA− was made using a *Hin*dIII−*Eco*RI fragment from plasmid SPAI (see Levitt *et al.*, 1989) containing pUC119 and synthetic poly(A) site sequence. **B.** RNase protection analysis of RNA from HeLa cells transfected with RβSVpBR328 and derivatives thereof. Aliquots of each probe used in the hybridization are also shown, as before hybridization and RNase treatment. The α2 co-transfection plasmid transcripts are represented by the band at 130 bp. The SPA-usage transcripts are represented by the band at 65 bp. The fainter bands below the 65 bp band are probably over-digestion products caused by 'breathing' of the AT-rich hybrids. In order to ensure that a negative result in the RNase protection assay was not due to a lack of transcript from the Rβ-derived constructs, a further control experiment was performed. All the RNA samples described were also analysed by S1 nuclease. The DNA probe used to detect the α2 globin transcripts was the end-labelled *Bst*EII-digested construct, pα2W3′PS. The probe used to detect rabbit β globin transcripts was the end-labelled *Eco*RI-digested pRβ. Both DNA probes will hybridize to transcripts from the third exon of the α2 and rabbit β-globin genes, respectively (see A). It was indeed found by such an analysis that both α and β third exon transcripts were present in all RNA samples except those from SPAIΔD transfections (data not shown). Although α transcripts were present in the latter, no β third exon transcripts were detected, as expected, and consistent with our previous results (Levitt *et al.*, 1989).

Table II.

	Relative utilization of SPA in transiently expressed constructs				
	Rβ	SPAI	SPAIΔD	SPAIP1−	SPAIP1+
α2 globin[a]	1	1	1	1	1
SPA in Rβ intron 2	0	0.07	2.77	0.02	0.69
% utilization of SPA relative to SPAIΔD	0%	2.5%	100%	0.7%	25%

[a]α2 globin (co-transfection) signals on autoradiographs are normalized in order to allow a direct comparison of SPA usage in the various constructs.

In SPAIP1+, however, there is no such deletion. Any recognition of the SPA by polyadenylation factors can only occur by inhibition of splicing by another mechanism. The pause site rescues the upstream intronic SPA by favouring polyadenylation over splicing in 25% of transcripts. We propose that this is achieved by the pausing of polymerase molecules immediately after transcription of the SPA. In this way, the normally less efficient polyadenylation reactions are given an opportunity to take place before the more efficient splicing reaction removes the SPA from the transcript, thereby preventing its utilization.

The pause site co-operates with the synthetic poly(A) site to elicit termination of transcription

The previous section demonstrates that although the SPA is inactive when positioned in a functional intron, it can be 'rescued' when the 92 bp pause sequence is placed immediately downstream. These results suggest that the pause site slows down the polymerase in time to allow polyadenylation to occur at the SPA, before the acceptor site of the intron is transcribed. However, it is equally possible that the pause site together with SPA elicit termination as suggested by the model of Proudfoot (1989). Both such processes may occur to some extent. However, a third interpretation could be that the pause site itself contains a cryptic donor splice site which may compete with the rabbit β donor site. If the cryptic donor site was active, the SPA would then be in the exon, and its use should be unconstrained. To test these possibilities directly, we carried out nuclear run-on analysis across the SPA/pause site construct SPAIP1+, in which the pause site is placed closely 3′ to the intronic SPA.

Figure 4A shows a diagram of SPAIP1+ together with the positions of the four different M13 probes subcloned from this region (BH−, HR−, R_1S_1−, S_2R_2−). αSS+ and αAA− are control probes that give a measure of background transcription and transfection efficiency based on the co-transfected pα2W3′PS plasmid respectively (see legend to Figure 4). Figure 4B then shows the data obtained from

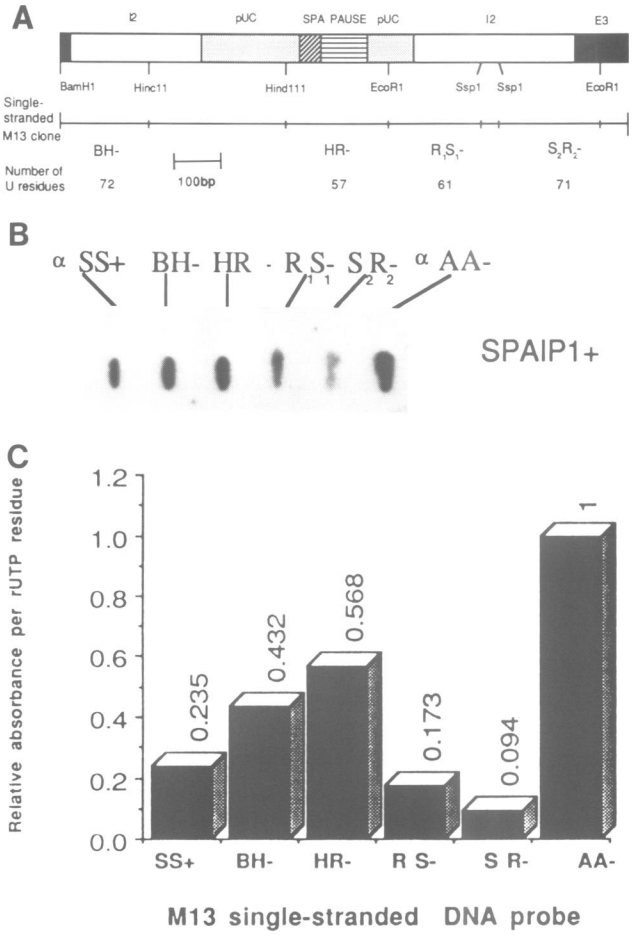

Fig. 4. A. Diagram of rabbit β globin derivatives SPAIP1+ showing fragments used in the construction of M13 clones for use as probes in nuclear run-off assays. Exons (solid boxes), introns (open boxes), pUC119 sequences (shaded boxes), SPA (diagonally striped box) and pause site sequences (horizontally striped box) are indicated. The number of UTP residues in each fragment is also shown.
B. Autoradiographs of filters containing M13 single-stranded DNA probes (as shown in A) used to detect SPAIP1+ transcripts generated in nuclear run-on assays. A control M13 probe, αAA− (also called 1−), containing a fragment from exon 1 of plasmid pα2W3′PS (see Figure 2B) was used to measure co-transfection efficiency. In these experiments we might have expected high background hybridization levels, since the plasmid SPAIP1+ contains sequences from pUC119 which are complementary to M13 sequences when the SPAIP1+ plasmid is transcribed in the antisense. In order to control for this background hybridization, we used an M13 probe, αSS+ containing the α globin promoter fragment in the sense orientation. The insert disrupts the pUC119-homologous region in the M13 DNA, and in this way provides a more accurate control than would M13mp18/19, since all the M13 clones used in the experiments possess inserts of a similar size. Levels of pα2W3′PS fragment αAA transcription are high in this assay. **C.** Bar chart showing quantification of hybridization signals in Figure 2B.

nascent transcription across the SPA region in SPAIP1+ while Figure 4C shows quantification of these nuclear run-on data. The 30% increase in signal from probes BH− to HR− indicates a small pausing effect. Probes R_1S_1- and S_2R_2- then give progressively lower signals both well below BH−, the transcription signal before the pause site and, in fact, below the background signal detected by αSS+. The fact that S_2R_2- gives a signal 5-fold lower than BH− suggests an 80% level of transcription termination following the pause site in SPAIP1+. These results also argue against the possibility that the pause site activates the SPA in the intron of the rabbit β globin gene by functioning as a cryptic donor site. If this were the case, nearly equal signals would be predicted for each of the nuclear run-on probes.

Discussion

The experiments described in this paper define a novel transcriptional control element in the 3′ flanking region of the human α2 globin gene, which functions as a pause signal to elongating RNA polymerase II. Our evidence for the existence of this signal is both through direct analysis of nascent RNA transcription and by the effect that this element has on enhancing the activity of an upstream poly(A) signal. This poly(A) site is otherwise inactive due to competition with either a stronger downstream poly(A) site or an overriding RNA splicing mechanism. In both of these situations the pause site appears to have a kinetic effect in allowing more time for the weaker RNA processing signal to operate. We have also demonstrated that when the highly efficient, synthetic poly(A) site is placed next to the pause signal, the combined elements elicit significant levels of transcriptional termination.

It is interesting to consider the possible role played by this RNA polymerase II pause site in the normal process of transcriptional termination by the human α2 globin gene. It has been previously demonstrated (Whitelaw and Proudfoot, 1986) that in erythroid cells the α2 globin gene terminates transcription at least to some extent a few hundred nucleotides past the poly(A) addition signals (close to the pause site) rather than several kilobases further 3′ as has been demonstrated for mouse β globin (Citron *et al.*, 1984) and α amylase (Hagenbuchle *et al.*, 1984). These results are entirely consistent with a role for the pause site in the termination of transcription of the α2 globin gene. However, the nuclear run-off analysis data (Figure 2) on transcription of the α2 globin gene when transfected into non-erythroid HeLa cells show no significant levels of termination beyond the pause site region. Interpretation of these results is made difficult by the presence of repetitive Alu repeat sequences soon after the pause site, which would effectively obscure small levels of termination in this region. Also the transient expression system used in these studies, in which high copy number plasmids are transcribed separately from the HeLa cell chromosomal genes, may give high levels of non-specific transcription all over the plasmid template. In spite of these considerations, it is clear that the pause site when placed downstream of the stronger synthetic poly(A) site does elicit significant levels of termination in a transient expression system. It therefore follows that the termination process of the α2 globin gene may be intentionally weak due to its less efficient poly(A) site. Consistent with this suggestion it has been demonstrated that in the DNA tumour virus, polyoma termination of transcription of the late transcripts occurs only weakly but is greatly enhanced by the insertion of a stronger poly(A) signal (Lanoix and Acheson, 1988).

It has been demonstrated that the α2 gene encodes a 2- to 3-fold higher steady state level of mRNA than the downstream α1 gene (Orkin and Goff, 1981; Liebhaber and Kan, 1983; Liebhaber *et al.*, 1985). These levels are also reflected in the protein products of the two genes (Liebhaber *et al.*, 1986; Shakin and Liebhaber, 1986). The promoters of both α globin genes are identical and their polyadenylation sites are not significantly dissimilar. Since the α2 globin gene

Fig. 5. Nucleotide sequence of the 92 nt pause site. The repeated CA_4 sequence is indicated.

lies upstream of α1, it is possible that the observed inefficiency of termination by α2 may permit transcriptional interference with α1, thereby down-regulating levels of α1 globin. Intriguingly, thalassemic mutations in the α2 gene have a more severe phenotype than those in the α1 gene. Since the α2 gene is required to produce 66−75% of all α globin, such a phenotype is to be expected. Also, failure of the cell to sufficiently transcribe or polyadenylate α2 transcripts would not only lead to a lack of α2 mRNA, but might also prevent any termination between the two α globin genes [since a functional poly(A) site is required for termination]. This would then cause interference with the α1 gene, as shown in a model system by Proudfoot (1986). Both types of mutation and their phenotypes, then, are consistent with the finding that α2 globin termination may be inefficient.

Two models have been proposed to account for the interconnection between polyadenylation and termination of transcription demonstrated in these studies. First, Logan *et al.* (1987) have postulated the interaction of the elongating RNA polymerase with an 'anti-terminator' factor. This factor would allow readthrough of premature termination sites, such as that present in the c-*myc* gene (Bentley and Groudine, 1988; Kerppola and Kane, 1988; Wright and Bishop, 1989). Upon arrival of the transcription complex at the poly(A) site, the anti-terminator would be released. Once this has occurred the transcription complex would be released at termination signals as found in *in vitro* systems by various groups (Baek *et al.*, 1986; Dedrick *et al.*, 1987; Kerpolla and Kane, 1988; Resnekov and Aloni, 1989) and *in vivo* by others (Frayne and Kellems, 1986; Logan *et al.*, 1987; Connelly and Manley, 1989a; Dressler and Fraser, 1989). Although this model does not discount the involvement of a pause signal in the termination process, a second model specifically postulates the existence of such a pause site, positioned at some distance downstream from the poly(A) signal (Connelly and Manley, 1988, 1989a; Proudfoot, 1989). In this model, after cleavage at the poly(A) signal, an uncapped RNA is generated. Such an RNA sequence would be extremely unstable, and would be either degraded by a 5′ to 3′ exonuclease activity or unwound from the DNA template by a helicase activity, until the elongating polymerase were reached. At this point, polymerase release (termination) would occur. By pausing the polymerase at some distance from the cleavage site, the efficiency of the termination process would be increased.

It is interesting to speculate how pausing of the RNA polymerase II complex could be achieved. One possible mechanism might be by direct obstruction through the binding of a sequence specific protein: a so-called 'road block'. Indeed, Connelly and Manley (1989a,b) have presented evidence for a CCAAT box binding protein that functions as a terminator. Clearly this process is highly specific since other DNA binding proteins such as SP1 had no effect in their system. Furthermore, the CCAAT box termination effect was shown to be orientation specific arguing that the precise mode of binding of the protein to the CCAAT box was also critical. We as yet have no information whether or not a protein interacts with the α pause site either at the DNA or RNA level.

Figure 5 shows the nucleotide sequence of the 92 nt pause site. As indicated, part of this sequence is highly A rich and contains a nearly perfect $(CA_4)_6$ repeat sequence. The 3′ half of the pause site has no obvious protein binding sites nor can it be folded into any significantly stable RNA structure. One possibility is that the $(CA_4)_6$ repeat sequence confers some specific structure on the DNA template and thereby slows down elongating RNA polymerases. Alternatively, the $(CA_4)_6$ repeat might be intentionally unfolded, to permit access of a protein factor to the exposed RNA primary structure.

Our results with the positioning of the pause site in between two RNA processing signals have interesting implications. Similar sequences might be employed between multiple poly(A) sites or splice sites for the purpose of regulating usage of the adjacent processing signals. This has been suggested to be the case for selection of poly(A) sites in immunoglobulin genes, with a potential regulatory role for interleukin-6 (Raynal *et al.*, 1989). Transcriptional termination or pausing has also been postulated to play a part in the selection of early or late polyadenylation sites in bovine papillomavirus type 1 (BPV-1) (Baker and Noe, 1989). The involvement of pausing in transcriptional termination may be optional. Where a high efficiency of termination is not required, no pause sites need be employed, and termination would occur over a long stretch of the template (several kilobases). We would predict that pause sites would exist predominantly between closely positioned genes. In such positions these signals might be required to increase termination efficiency, in order to protect the downstream gene from transcriptional inteference. We are currently investigating the existence of such pause sites between closely placed genes in the HLA locus.

Materials and methods

Plasmid constructions

The plasmid pα2W3′PS (Figure 2A) consists of the three exons of the α2 globin gene with 2.0 kb of 5′ flanking sequence and 2.3 kb of 3′ flanking region. Details of the construction of pα2W3′PS have been published previously (Whitelaw and Proudfoot, 1986). We have renamed the plasmid pα2W3′PS instead of α2W3′PSpSVed. The *Sac*I deletion series of plasmid constructions (SΔ1 to SΔ5, as in Figure 1) were made by linearizing pα2W3′PS with *Sac*I, and using Bal31 exonuclease digestion to delete sequences in both directions. *Bgl*II oligonucleotide linkers were cloned in at each endpoint. By subsequently digesting with *Bgl*II, all sequences originally situated between the *Sac*I and *Bgl*II sites were removed (see Figure 2A). The duplex SPA oligonucleotide with 5′ overhangs of GATC (see Levitt *et al.*, 1989) was ligated into the *Bgl*II restriction sites of the pα2W3′PS *Sac*I deletion series of vectors. In this way, derivatives were generated of pα2W3′PS in which the SPA was inserted ~200 bp and 600 bp downstream of the wild-type α2 globin poly(A) site. Sequences downstream of the SPA were identical in all five derivatives selected and comprise the sequence 3′ to the *Bgl*II site in the original pα2W3′PS.

In order to generate the M13 clones for use as single-stranded probes in the nuclear run-on analysis of pα2W3′PS-transfected nuclei, M13mp19 was linearized with *Sma*I to use as a vector. All restriction fragments to

be inserted were blunt-ended with Klenow enzyme and ligated into the *Sma*I site of M13mp19. The *Sma*I−*Sma*I (SS), *Ava*I−*Ava*I (AA or 1), *Ava*I−*Hin*dIIII (2), *Hin*dIII−*Bst*EII (3) and *Bst*EII−*Pvu*II (4) fragments were all gel-purified following digestion of pα2W3′PS with *Ava*I, *Hin*dIII, *Bst*EII and *Pvu*II restriction enzymes. To generate the remaining fragments from the 3′ flanking region, a *Pvu*II−*Dra*I fragment was purified and digested with *Hin*fI.

Most of the plasmids used in the intronic poly(A) site activation assay have been described previously by Levitt *et al.* (1989). To generate the plasmids pSPAIP1+/−, *Hpa*I−*Bgl*II fragments containing the pause site were purified from clone SΔ3, in which the pause site lies between the unique *Hpa*I site and the SPA (which contains two *Bgl*II sites, one close to its 5′ end). The *Hpa*I−*Bgl*II fragments were blunt-ended using Klenow enzyme, then cloned in both orientations into the unique *Sac*I site in the recombinant plasmid pSPAI. The *Sac*I 3′ overhangs were blunted using T4 DNA polymerase. To construct the riboprobe plasmids, an ~100 bp *Hin*dIII−*Eco*RI fragment from pSPAI containing the SPA in the pUC119 polylinker was purified and ligated into the *Hin*dIII−*Eco*RI sites of pSP65 making the directional clone, pySPA−. The α2 globin co-transfection control riboprobe construct, pαSS−, was made by purification of a 240 bp *Sma*I−*Sma*I fragment from pα2W3′PS and its ligation into the *Sma*I site of pSP65.

The M13 clones used as probes in the nuclear run-off analysis of pSPAIP1+ transfected nuclei all generated by directional cloning into the appropriate M13 vector. The *Bam*HI−*Hin*cII fragment was purified from RβSVpBR328, and ligated into the *Bam*HI and *Hin*cIII sites of M13mp9. The *Hin*dIII−*Eco*RI fragment was purified from pSPAIP1+ and ligated into the *Hin*dIII−*Eco*Ri sites of M13mp18 to construct HR−. The *Eco*RI−*Ssp*I, and *Ssp*I−*Eco*RI fragments were co-purified from pSPAI and ligated into the *Eco*RI and *Hin*cII sites of M13mp19 and M13mp18 respectively, generating R_1S_1- and S_2R_2-. These M13 clones were analysed by dideoxy sequencing.

Unless otherwise stated, in all cases of plasmid constructions described here, the orientation of any inserts was determined by restriction enzyme analysis.

Cell transfection and RNA analysis

The methods used for growth and transfection of HeLa cells, isolation of cytoplasmic RNA and S1 nuclease analysis have been described previously (Johnson *et al.*, 1986; Levitt *et al.*, 1989). *In vitro* RNA synthesis and generation of radio-labelled probes was performed as described in Melton *et al.* (1984). RNase protection analysis of cytoplasmic RNA was carried out as described by Melton *et al.* (1984), with the following modification of conditions: hybridizations were at 51°C, and RNase A and T_1 digestions were performed at 30°C. All S1 nuclease or RNase digestion products were electrophoresed through 4−8% polyacrylamide−8 M urea sequencing gels (Maxam and Gilbert, 1980), the actual percentage of polyacrylamide used depending on the size range of products expected. Size markers used were generated by *Msp*I digestion of pϵ0.7 (Baralle *et al.*, 1980) followed by end-labelling. Autoradiographic exposures shown were obtained with intensifying screens.

In vitro nascent chain analysis

Cells were harvested by scraping into PBS 36−48 h after transfection, depending on cell confluency. The cells were spun down at 3000 r.p.m. (Beckman bench centrifuge) at 4°C for 5 min and resuspended in 10 ml of cold cell lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 1 mM EDTA, 0.5% NP-40). Following 10 min incubation on ice, the nuclei were spun down at 3K for 10 min at 4°C. In the first method used (as described in Figure 2A−C), the nuclei were resuspended in 200 μl 10 mM Tris, pH 7.8, 5 mM $MgCl_2$, 90 mM KCl, 5 mM dithiothreitol, 50% glycerol. Elongation was carried out in this solution plus 0.05 mM rATP, rGTP, rCTP and 100 μCi [α-^{32}P]rUTP (Amersham). Reactions were terminated by the addition of ~30 U of RNase-free DNase (Miles or Boehringer Mannheim) and incubation at 37°C for 5 min. Deproteinization was achieved by addition of 200 μl 0.2 M Tris, pH 7.5, 0.3 M NaCl, 2% SDS, 25 mM EDTA and ~100 μg proteinase K. The samples were incubated for 30 min at 37°C, followed by phenol−chloroform extraction and ethanol precipitation. The pellet was resuspended in 300 μl 10 mM Tris, pH 7.5, 10 mM $MgCl_2$ and 30 U DNase and incubated for 30 min at 37°C. The nuclear RNA was then worked up and resuspended in 100 μl of hybridization solution (4 × SSC, 10 × Denhardt's, 200 μg/ml sonicated herring DNA, 0.2% SDS, 50% formamide, 10% dextran sulphate). The labelled RNA was heated at 85°C for 5 min before being added to the hybridization bag. The latter contained single-stranded M13 DNA bound to nylon filters (Hybond-N, Amersham). The M13 single-stranded DNAs were either purified by agarose electrophoresis then Southern blotted onto nylon filters (Figure 2B and C), or else bound to the nylon filters using a 'hybrislot' manifold (BRL) (Figures 2D and 4B), essentially as previously described (Whitelaw and Proudfoot, 1986). Prehybridization, hybridization and washing of filters were also performed as described in the above reference, except that filters were finally washed in 0.5 × SSC, 0.1% SDS at 65°C for 60 min. Filters were exposed to sensitive X-ray film, with intensifying screen, for ~48 h.

In the method subsequently used (as described for Figure 2D), the nuclei were then resuspended in 210 μl of 50 mM Tris, pH 8.3, 5 mM $MgCl_2$, 0.1 mM EDTA, 40% glycerol. Elongation was carried out at 30°C for 15 min in the presence of 55 mM Tris, pH 8.3, 7.8 mM $MgCl_2$, 0.1 mM EDTA, 167 mM KCl, 0.28 mM rATP, rGTP, rCTP and 100 μCi [α-^{32}P]rUTP, 40% glycerol. Reactions were terminated by the addition of ~30 U of RNase-free DNase, and incubation at 37°C for a further 30 min. The reactions were deproteinized by addition of 290 μl 0.1 M Tris, pH 7.5, 0.15 M NaCl, 1% SDS, 10 mM EDTA with ~100 μg of proteinase K per ml, incubation at 37°C for 1 h and subsequent phenol−chloroform extraction and isopropanol precipitation in the presence of 2.5 M NH_4Ac. The nuclear RNA pellet was suspended in 50 μl STE (10 mM NaCl, 10 mM Tris, pH 7.5, 1 mM EDTA) and then 50 μl hybridization solution (as described previously). The rest of the protocol was as described for the first method. The key differences between the two protocols lie in the length of time allowed for elongation, the concentrations of ribonucleotides used and the use of two separate DNase I treatments in the first method.

Acknowledgements

We are grateful to the other members of our laboratory for discussion and encouragement throughout these studies. We also thank Drs Sheila Connelly, David Bentley and Stephanie Wright for help and advice on the intricacies of nuclear run-off analysis. Finally, we thank the Wellcome Trust for financial support (project grant no. 16354).

References

Adami,G. and Nevins,J.R. (1988) *EMBO J.*, **7**, 2107−2116.
Baek,K.-H., Sato,K., Ito,R. and Agerwal,K. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7623−7627.
Baker,C.C. and Noe,J.S. (1989) *J. Virol.*, **63**, 3529−3535.
Baralle,F.E., Shoulders,C. and Proudfoot,N.J. (1980) *Cell*, **21**, 621−626.
Bentley,D.L. and Groudine,M. (1988) *Cell*, **53**, 245−256.
Citron,B., Falck-Pederson,E., Salditt-Georgieff,M. and Darnell,J.E., Jr (1984) *Nucleic Acids Res.*, **12**, 8723−8731.
Connelly,S. and Manley,J. (1988) *Genes Dev.*, **2**, 440−452.
Connelly,S. and Manley,J. (1989a) *Cell*, **57**, 561−571.
Connelly,S. and Manley,J. (1989b) *Mol. Cell. Biol.*, **9**, 5254−5259.
Dedrick,R.L., Kane,C.M. and Chamberlin,M.J. (1987) *J. Biol. Chem.*, **262**, 9098−9108.
Dressler,G.R. and Fraser,N.W. (1989) *J. Gen. Virol.*, **70**, 1337−1345.
Frayne,E.G. and Kellems,R.E. (1986) *Nucleic Acids Res.*, **14**, 4113−4125.
Grummt,I., Kuhn,A., Bartsch,I. and Rosenbauer,H. (1986) *Cell*, **47**, 901−911.
Hagenbuchle,O., Wellauer,P.K., Cribbs,D.L. and Schibler,U. (1984) *Cell*, **38**, 737−744.
Henderson,S. and Sollner-Webb,B. (1986) *Cell*, **47**, 891−900.
Hernandez,N. and Weiner,A.M. (1986) *Cell*, **47**, 249−258.
Holmes,W.M., Platt,T. and Rosenberg,M. (1983) *Cell*, **32**, 1029−1032.
Johnson,M.R., Norman,C., Reeve,M.A., Scully,J. and Proudfoot,N.J. (1986) *Mol. Cell. Biol.*, **6**, 4008−4018.
Kerppola,T.K. and Kane,C.M. (1988) *Mol. Cell. Biol.*, **8**, 4389−4394.
Kunkel,G.R. and Pederson,T. (1985) *Mol. Cell. Biol.*, **5**, 2332−2340.
Lanoix,J. and Acheson,N.H. (1988) *EMBO J.*, **7**, 2515−2522.
LeMeur,M.A., Galliot,B. and Gerlinger,P. (1984) *EMBO J.*, **3**, 2779−2786.
Levitt,N., Briggs,D., Gil,A. and Proudfoot,N.J. (1989) *Genes Dev.*, **3**, 1019−1025.
Leibhaber,S.A. and Kan,Y.W. (1983) *J. Clin. Invest.*, **71**, 461−466.
Liebhaber,S.A., Cash,F.E. and Main,D.M. (1985) *J. Clin. Invest.*, **76**, 1057−1064.
Liebhaber,S.A., Cash,F.E. and Ballas,S.K. (1986) *J. Biol. Chem.*, **261**, 15327−15333.
Logan,J., Falck-Pederson,E., Darnell,J.E., Jr and Shenk,T. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8306−8310.
McStay,B. and Reeder,R.H. (1986) *Cell*, **47**, 913−920.
Manley,J.L. (1988) *Biochim. Biophys. Acta*, **950**, 1−12.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499−560.

Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Neuman de Vegvar,H.E., Lund,E. and Dahlberg,J.E. (1986) *Cell*, **47**, 259–266.
Orkin,S.H. and Goff,S.C. (1981) *Cell*, **24**, 345–351.
Proudfoot,N.J. (1986) *Nature*, **322**, 562–565.
Proudfoot,N.J. (1989) *Trends Biochem. Sci.*, **14**, 105–110.
Proudfoot,N.J. and Whitelaw,E. (1988) In Hames,B.D. and Glover,D.M. (eds), *Transcription and Splicing* (Frontiers in Molecular Biology). Oxford University Press, Oxford, pp. 97–129.
Raynal,M.-C., Liu,Z., Hirano,T., Mayer,L., Kishimoto,T. and Chen-Kiang,S. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8024–8028.
Resnekov,O. and Aloni,Y. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 12–16.
Shakin,S.A. and Liebhaber,S.A. (1986) *J. Clin. Invest.*, **78**, 1125–1129.
Shen,C.-K.J. and Maniatis,T. (1982) *J. Mol. Appl. Genet.*, **1**, 343–360.
Weber,J., Jelinek,W. and Darnell,J.E., Jr (1977) *Cell*, **10**, 611–616.
Whitelaw,E. and Proudfoot,N.J. (1986) *EMBO J.*, **5**, 2915–2922.
Wright,S. and Bishop,J.H. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 505–509.

Received on February 13, 1991