

# Definition of an efficient synthetic poly(A) site

N. Levitt, D. Briggs, A. Gil,<sup>1</sup> and N.J. Proudfoot

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

**We constructed and analyzed a synthetic poly(A) (SPA) site that was based on the highly efficient poly(A) signal of the rabbit  $\beta$ -globin gene. By use of the SPA, we demonstrate that the minimum sequences required for efficient polyadenylation are the AATAAA sequence and a GT/T-rich sequence with the correct spacing of 22–23 nucleotides between them. When placed downstream of the poly(A) site of the human  $\alpha 2$ -globin gene, the SPA is used exclusively. We predict that the SPA, with its more extensive GT/T-rich sequence, is a more efficient poly(A) site than  $\alpha$ -globin. Also, we compared the use of the SPA when it is placed either in the exon 3 or intron 2 of the rabbit  $\beta$ -globin gene. When in the exonic position, SPA is used 10-fold more than the regular poly(A) site of rabbit  $\beta$ -globin. In contrast, when it is in the intronic location, no detectable use of SPA is observed; however, the deletion of the donor site of intron 2 reactivates the intronic positioned SPA. These results indicate that the splicing of intron 2 in the rabbit  $\beta$ -globin gene occurs ahead of polyadenylation and have important implications for termination of transcription. Polyadenylation, although required for termination of transcription, is not sufficient; therefore, additional termination signals for RNA polymerase II must exist.**

[*Key Words:* Synthetic poly(A) site; polyadenylation; splicing;  $\beta$ -globin gene; termination signals]

Received October 25, 1988; revised version accepted May 2, 1989.

The identification of the ubiquitous sequence AATAAA at the 3' end of nearly all genes that encode polyadenylated mRNA has long been associated with the 3'-end formation of mRNA (Proudfoot and Brownlee 1976; Fitzgerald and Shenk 1981). Indeed, mutations within this sequence disrupt mRNA synthesis and thereby result in unstable, and often elongated, mRNAs (Higgs et al. 1983; Montell et al. 1983; Orkin et al. 1985); however, AATAAA is insufficient for mRNA polyadenylation because the deletion of sequences immediately 3' to the site of polyadenylation also disrupts mRNA 3'-end formation (for review, see Proudfoot and Whitelaw 1988). No ubiquitous sequence elements are found in these required 3' regions, although a GT-rich or T-rich sequence generally is present (Birnstiel et al. 1985; McLauchlan et al. 1985). Indeed, some confusion has been caused by the apparently nonconserved nature of these GT- or T-rich sequences because it is relatively easy to reconstitute such sequences inadvertently during a deletion analysis of the sequences required for a particular polyadenylation site (Gil and Proudfoot 1987).

We demonstrated previously that the level of stable mRNA produced by a gene is affected critically by the quality of its polyadenylation signals (Whitelaw and Proudfoot 1986). In the case of the rabbit  $\beta$ -globin gene, we demonstrated the presence of both a required GT-

rich and T-rich sequence element immediately downstream of the poly(A) site. Furthermore, we showed that the reconstruction of this poly(A) site, with either the GT- or T-rich element separately, or with altered spacing between these two elements and AATAAA, reduced dramatically the efficiency of the poly(A) site, and thus reduced the level of mRNA (Gil and Proudfoot 1987). From these results it seems reasonable to predict that the poly(A) sites of different genes may vary in efficiency depending on the presence or absence of either a GT- or T-rich downstream element. Genes without a downstream element or with only one downstream element may have evolved to produce less stable mRNA, thus producing less protein product (Proudfoot and Whitelaw 1988).

In this paper we describe the construction of a synthetic poly(A) site (SPA) that is based on the rabbit  $\beta$ -globin gene. We demonstrate that the sequences between the AATAAA and (GT)<sub>n</sub>(T)<sub>n</sub> homologies may be exchanged with random sequences. We define the minimum sequence requirement for an efficient poly(A) site. We describe the use of the SPA as a reagent to investigate at which positions within a gene a poly(A) site can be used. We show that, when placed immediately 3' to the poly(A) site of the human  $\alpha 2$ -globin gene, the SPA is used exclusively, and, interestingly, the poly(A) site of  $\alpha$ -globin possesses a GT-, but no T-rich, downstream sequence. Also, we report that, when placed 5' to the poly(A) site of the parental rabbit  $\beta$ -globin gene, the SPA is used preferentially, which demonstrates that we syn-

<sup>1</sup>Present address: Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA.

Levitt et al.

thesized a fully efficient poly(A) site. Finally, we demonstrate that, when the SPA is placed in the second intron of the rabbit  $\beta$ -globin gene, it is not functional unless splicing signals are deleted. This final result demonstrates that splicing occurs ahead of polyadenylation in the processing of the primary transcript of the  $\beta$ -globin gene.

## Results

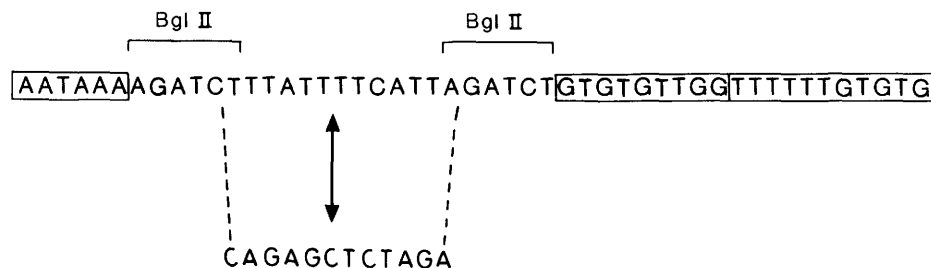
### Construction and analysis of a SPA site

On the basis of our previous studies (Gil and Proudfoot 1987), we reasoned that at least three elements were required to form an efficient poly(A) site: The AATAAA sequence, a GT-rich sequence, and a T-rich sequence. In the case of the rabbit  $\beta$ -globin gene, the GT-rich element is GTGTGTTGG and the T-rich element is T<sub>6</sub>GTGTG. Although these two elements normally are separated by the two nucleotides AA, we demonstrated previously that the  $\beta$ -globin poly(A) site is slightly more efficient when the GT- and T-rich elements are fused together. Therefore, we synthesized the oligonucleotide shown in Figure 1, which contains each of these three essential elements as well as most of the sequences between them. To investigate the possible requirement of the nucleotides between these essential sequence elements, we placed two *Bgl*II restriction sites immediately 3' to AATAAA and 5' to GT-rich sequences. We were careful to maintain correct spacing between the AATAAA and GT-rich sequences because we showed previously that the alteration of their spacing greatly reduces the efficiency of the poly(A) site (Gil and Proudfoot 1987). With these two *Bgl*II sites in place, it was a relatively easy matter to excise the central portion of the synthetic poly(A) site and replace this region with a linker oligonucleotide.

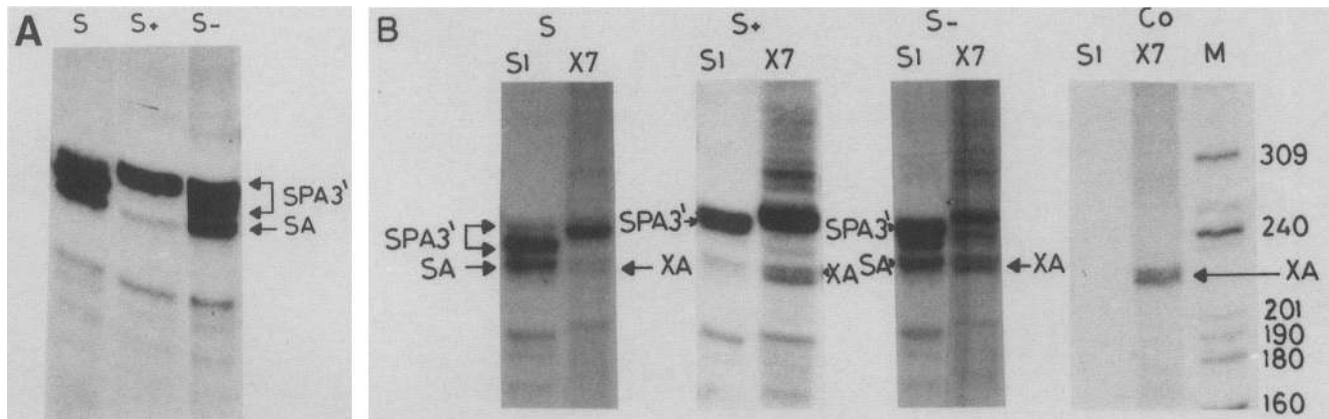
Figure 2 shows our initial data from a test of the functionality of the synthetic poly(A) site. In these experiments, we deleted the polyadenylation signals of the human  $\alpha$ 2-globin gene by use of *Bal*31 exonuclease and replaced them with the SPA (Fig. 3A). An  $\alpha$ -globin gene deletion mutant was selected that lacked the entire 3' sequence from a few nucleotides 5' to the AATAAA sequence of the  $\alpha$  poly(A) site to well within the 3'-flanking sequence (clone S). This  $\alpha$ -globin SPA construct was contained within the transient expression vector

pSVed, which possessed the SV40 replication origin and enhancer sequences (see Materials and methods). Two further constructs were made in which the internal sequence of the SPA between the two *Bgl*II sites (Fig. 1) was replaced with a linker oligonucleotide that was almost the same length as the excised region. Both orientations of the linker construct were obtained (clone S+ and S-). These three  $\alpha$ -globin SPA constructs then were transfected transiently into HeLa cells, and the cytoplasmic RNA obtained was analyzed by a combination of S1 nuclease and exonuclease VII mapping techniques by use of an end-labeled DNA probe (Fig. 3A). It proved necessary to use both of these RNA mapping techniques to establish which bands in the S1 analysis represented authentic mRNA 3' termini.

Figure 2A demonstrates that each  $\alpha$ -SPA construct produced significant amounts of S1 signal that corresponded to mRNA that used the SPA. Furthermore, the internal sequence was clearly inessential, as it could be replaced by two random sequences (S+ and S-) without affecting the levels of signal obtained. Yet, the spacing between the AATAAA and (GT)<sub>n</sub> elements had to be maintained close to the normal gene sequence because Gil and Proudfoot (1987) showed previously that an SPA with a 5-nucleotide-deficient internal sequence is wholly nonfunctional as a poly(A) site. Interestingly, the exact site of polyadenylation varied between S, S+, and S-. As shown in Figure 2, A and B, S had two authentic 3' termini, represented by a minor top band and major center band. These products were confirmed by the use of the alternative exonuclease VII (X7) mapping procedure. The products of X7 are always a few nucleotides larger than those of S1, reflecting slight steric differences between the two nucleases. The bottom band in lane S (SA), was absent from the X7 lane, which indicates that it was caused by AT-rich sequences that destabilized the RNA-DNA duplex. Such structures are sensitive to the single-strand-specific endonuclease activity of S1. Both S+ and S- RNA had slightly different 3' termini. Lane S+ contained just one strong band, whereas lane S- displayed a triplet band pattern. The X7 data confirmed the S1 nuclease data for sequences S+ and S-, although the smallest S1 band for S- might have been another S1 artifact because it is not visible clearly in the X7 lane. Finally, it should be mentioned that all of the X7 lanes showed an artifact band (XA) that was present in the tRNA control lane (Co).



**Figure 1.** Sequence of the synthetic poly(A) site. The essential AATAAA and (GT)<sub>n</sub>(T)<sub>n</sub> are boxed in. The two internal *Bgl*II sites are indicated by horizontal brackets. The position and sequence of oligonucleotide C are indicated also.



**Figure 2.** RNA mapping of clones S, S+, and S-  $\alpha$ 2W3'PSPVed. (A) S1-digestion products of S, S+, and S- by use of the DNA probe indicated in Fig. 3A. S and S- give doublet bands, whereas S+ gives a single band (SPA3'). (SA) Artifact bands caused by the AT-rich nature of the SPA sequence (see B). (B) Comparison of S1 and exonuclease VII (X7) digestion products for RNA from S, S+, and S- transfections that were probed with homologous DNA probes as indicated in Fig. 3A. (Lane Co) Control digests with tRNA instead of transfected HeLa cell cytoplasmic RNA.

In conclusion, we demonstrated that the SPA is an active poly(A) site and, furthermore, we have demonstrated that the three elements AATAAA, (GT)<sub>n</sub>, and (T)<sub>n</sub> were sufficient to form this RNA-processing signal. As noted by others, the sequences between AATAAA and (GT)<sub>n</sub>, although unimportant for the efficiency of the poly(A) site, did influence the exact position at which poly(A) was added (Fitzgerald and Shenk 1981; Mason et al. 1986).

#### *Placing the SPA 3' to the poly(A) site of the human $\alpha$ -globin gene*

Figure 3A shows the nucleotide sequence for the poly(A) site region of the human  $\alpha$ 2-globin gene. The AATAAA and a potential downstream GT-rich sequence are boxed in. To define the sequence requirements for the  $\alpha$ -globin poly(A) site, as well as to test its activity in competition with the SPA, we constructed *Bal31* exonuclease deletion mutants as indicated in Figure 3. At each deletion end point, the SPA was inserted in the correct orientation. Furthermore, the same 3'-flanking sequence was positioned against the SPA in each clone, so that the 3'-end formation activity of each construct was dependent only on the deleted sequence and not the sequence that was placed against the SPA. Each construct was transfected into HeLa cells, and cytoplasmic RNA was purified.

Figure 3B presents the S1 analysis for  $\Delta$ 1- $\Delta$ 5 RNAs. For each S1 reaction, a homologous DNA probe, labeled at the *Bst*EII site in exon 3 of the  $\alpha$ -globin gene was used. Surprisingly, the SPA was used exclusively in  $\Delta$ 1- $\Delta$ 3. No band was detectable at the  $\alpha$ -poly(A) site position (see  $\alpha$  lane); only a doublet band of decreasing size corresponding to the SPA at the 3' deletion end points could be seen.  $\Delta$ 3C is a control to  $\Delta$ 3, in which no SPA was added to the deletion end point. In this case, the  $\alpha$ -globin poly(A) site was fully functional. This demonstrated that the exclusive use of SPA in  $\Delta$ 1- $\Delta$ 3 was not caused by the

deletion of sequences required for the  $\alpha$ -poly(A) site to work.

Surprisingly, the next deletion,  $\Delta$ 4, completely switched poly(A) site selection to  $\alpha$ . In this clone, the  $\alpha$  and SPA sites abutted one another directly (see Fig. 3A). Apparently this arrangement favored the weaker 5'  $\alpha$ -poly(A) site. The next deletion,  $\Delta$ 5, lacked the GT-rich box of the  $\alpha$ -poly(A) site so that, as predicted, the  $\alpha$ -poly(A) site was inactivated, and the SPA was used once again.

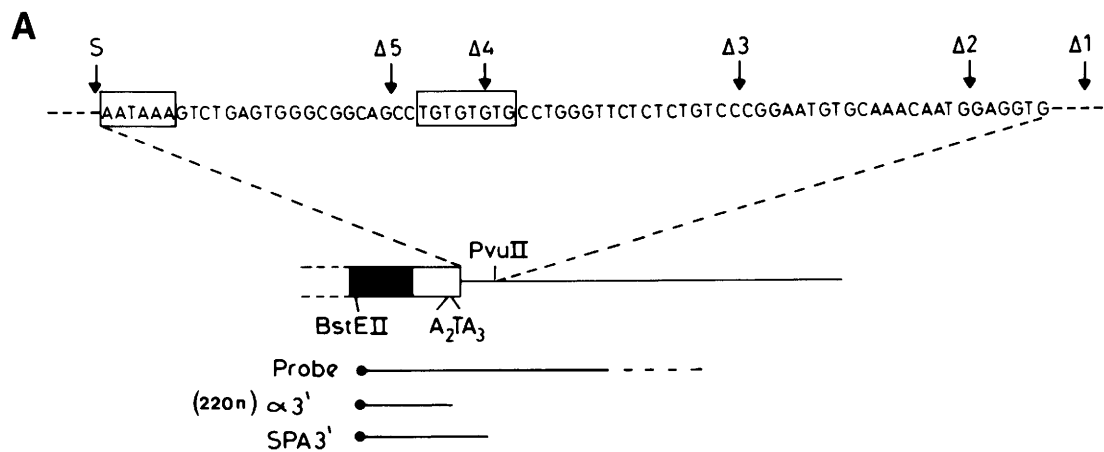
The above results allowed us to draw a number of interesting conclusions. First, the SPA outcompeted the  $\alpha$ -poly(A) site completely. Because both poly(A) sites possessed AATAAA signals, it followed that their different downstream sequences caused this difference. Clearly, the fact that SPA possessed both a GT-rich and T-rich element, whereas  $\alpha$  possessed only a GT-rich element, is likely to be the explanation of this effect because we demonstrated previously that the rabbit  $\beta$ -poly(A) site is weakened substantially when its T-rich element is deleted (Gil and Proudfoot 1987). Second, we showed that when both poly(A) sites are directly adjacent to one another ( $\Delta$ 4), the weaker  $\alpha$ -poly(A) site is used. This may reflect some steric preference for the 5'-positioned poly(A) site. Third, the fact that the next deletion ( $\Delta$ 5) switches back to SPA suggests that the GT-rich box in the  $\alpha$ -gene sequence is indeed part of the  $\alpha$ -poly(A) site.

#### *The SPA functions efficiently in the exon but not in the intron of the rabbit $\beta$ -globin gene*

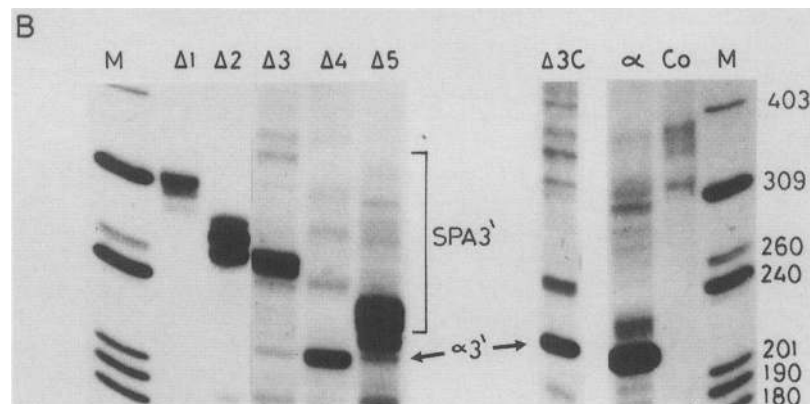
In the previous section, we tested the effect of the placement of the SPA to a location 3' to a natural poly(A) site. In these final experiments, we wished to address the question of how the SPA would behave when 5' to an efficient poly(A) site, either in the exon or in the intron of a gene. We reasoned that it would be useful to use the rabbit  $\beta$ -globin gene as our test system because the SPA was derived from the gene, and because its second intron



Levitt et al.



**Figure 3.** (A) Diagram of the 3' end of the human  $\alpha 2$ -globin gene. The sequence around the  $\alpha$ -poly(A) site is shown, as is the precise position of each deletion end point. The AATAAA and  $(GT)_n$  sequences of  $\alpha$ -globin are boxed. Finally, the positions of the DNA probe and S1 or exoVII products are indicated. (B) S1 analysis of RNA from HeLa cells transfected with  $\Delta 1$ – $\Delta 5$  plasmids as well as  $\Delta 3C$  and the original  $\alpha 2W3'$  PS pSVed plasmid ( $\alpha$ ). (Lane Co) tRNA control experiments; (lane M) DNA markers. Homologous DNA probes were used for each RNA labeled at the *BstEII* site in  $\alpha$  globin (Fig. 3A).

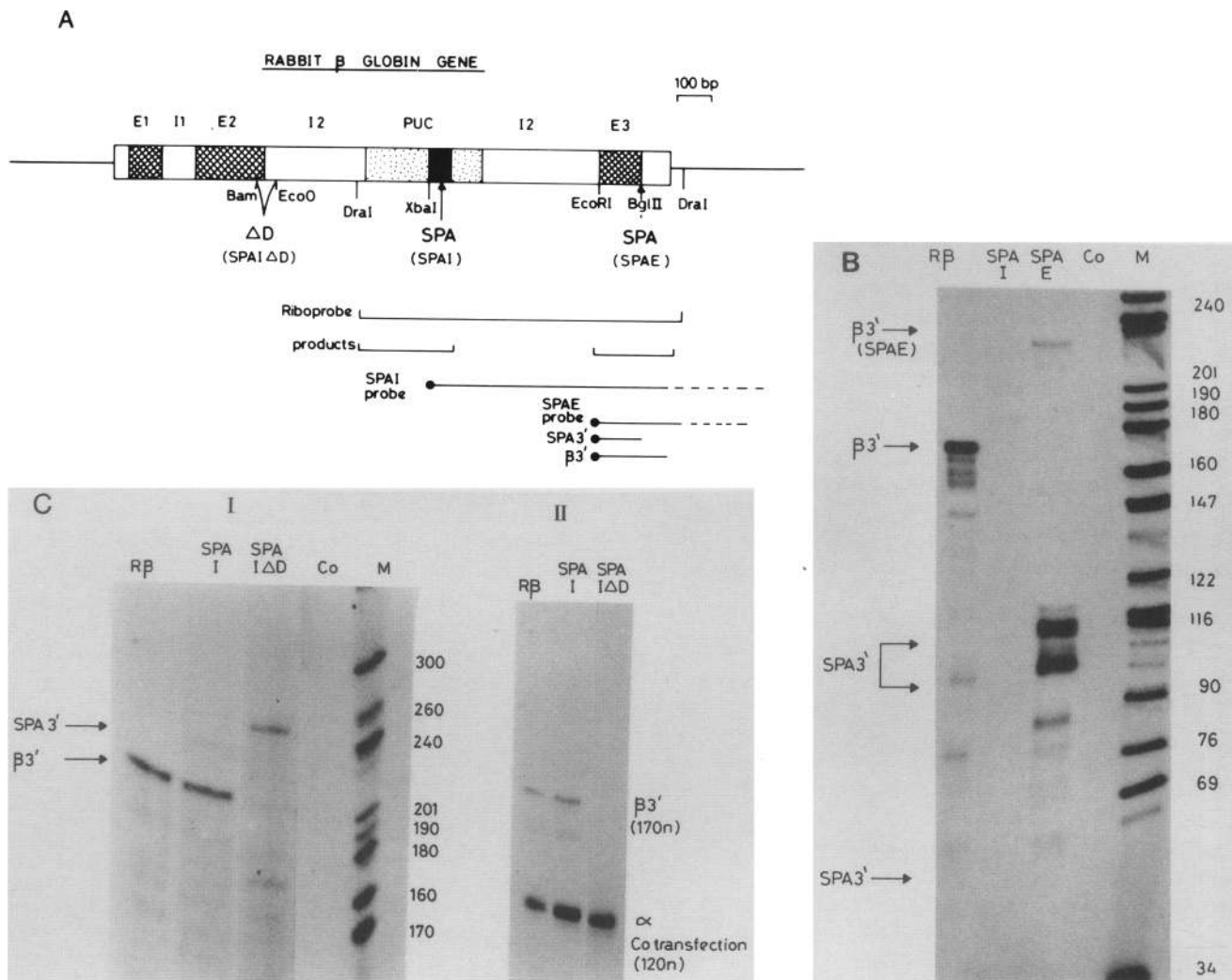


is over 600 bp, considerably larger than the  $\sim 100$ -bp introns of  $\alpha$ -globin. The cloning of inserts into the middle of this larger intron, therefore, should not disrupt the splicing signals at either end of the intron. Figure 4A shows a diagram of the various R $\beta$ –SPA constructs as well as the positions of the S1 nuclease probes and signals. The SPA was cloned into a polylinker sequence that was placed in the middle of the second intron of the rabbit  $\beta$ -globin gene (clone SPAI). The SPA was cloned also into a unique *BglIII* site at the termination codon of exon 3 (clone SPAE).

Figure 4B shows the S1 analysis of cytoplasmic RNA that was obtained from HeLa cells transfected with the original R $\beta$  plasmid or the R $\beta$ –SPA constructs SPAE and SPAI. With the SPA cloned into exon 3 of the rabbit  $\beta$  globin gene (Fig. 4B, lane SPAE), the SPA was used 10 times as efficiently as the normal 3'-positioned poly(A) site. The faint  $\beta 3'$  signal was shifted to correspond to a fragment 54 nucleotides larger than the R $\beta$  control lane  $\beta 3'$  signal, the size of the inserted SPA oligonucleotide. We demonstrated previously that when two copies of the rabbit  $\beta$ -globin poly(A) site were placed in tandem, the first poly(A) site was used exclusively. However, if the first poly(A) site was weakened by the alteration of its downstream sequence, then proportionally more of the mRNA used the 3'-positioned poly(A) site (Gil and Proudfoot 1987). Therefore, the SPA can be judged to be almost as efficient as the normal rabbit  $\beta$ -

globin gene poly(A) site. Surprisingly, when the SPA was placed 400 bp further 5' in the middle of intron 2, the SPA was not used. As shown in lane SPAI, no signals were detectable at the expected size of 50 nucleotides.

The fact that the SPA appeared to be nonfunctional when placed in the middle of intron 2 can be explained two ways. Either intron 2 was spliced efficiently, thereby preventing polyadenylation of the SPA, or the SPA was positioned inadvertently in a region of sequence that somehow prevented its use. To distinguish between these two possibilities, we made another construct SPAI  $\Delta D$  (Fig. 4A), in which the donor site of intron 2 of the rabbit  $\beta$ -globin gene was deleted (a deletion of 24 bp that included the obligatory GT sequence). As shown in Figure 4A, an antisense RNA probe was made by the subcloning of a *DraI* fragment that contained part of intron 2 including the SPA, through to just beyond the poly(A) site of the rabbit  $\beta$ -globin gene, and into the plasmid SP65 (Melton et al. 1984). Cytoplasmic RNA from HeLa cells that were transfected with plasmids R $\beta$ , R $\beta$ SPAI, or R $\beta$ SPAI $\Delta D$  was hybridized to this *DraI* riboprobe. The RNase digestion products obtained are shown in Figure 4C. R $\beta$  and SPAI both gave a band of 220 nucleotides that corresponded in size to exon 3. In contrast, SPAI $\Delta D$  gave virtually no exon 3 product, but instead formed a new band of 270 nucleotides that corresponded in size to an unspliced mRNA that uses the SPA (see Fig. 4A for positions of RNA products). These



**Figure 4.** (A) Diagram of rabbit  $\beta$ -globin gene. Exons (hatched boxes); noncoding and intron sequences (open boxes); pUC119 sequence (stippled boxes); pUC119 polylinker sequence (solid boxes). The positions at which SPA has been inserted are indicated. The DNA and RNA probes used in S1 or RNase analysis and the products obtained are indicated. The *Xba*I and R1 restriction sites shown in the figure were the sites of end-labeling for the SPAI and SPAE probes, whereas the two *Dra*I sites define the region of the gene construct used in the riboprobe. Finally, the *Bam*HI and *Eco*O 109 are indicated, deletion between which resulted in the SPAI $\Delta$ D construct. (B) S1 analysis of RNA from HeLa cells transfected with R $\beta$ , SPAI, or SPAE R $\beta$ SVpBR328. The positions of the  $\beta$ 3' and SPA 3' signals are indicated. (Lane Co) tRNA control S1 reaction. (C) Analysis of RNA from HeLa cells transfected with the SPAI $\Delta$ D construct using RNase protection (I) and S1 analysis (II) compared with R $\beta$  and SPAI. RNase-protection products obtained (I);  $\beta$ 3' signals for the three samples (II). The  $\alpha$ -globin cotransfection control signal is indicated also. In fact, this band of 120 nucleotides was obtained as a result of a mismatch between mRNA transcribed from the  $\alpha$ 2 3' PSpSVed transfected plasmid and an  $\alpha$ 1-globin probe labeled at its *Bst*EII site. The 3'-noncoding regions of human  $\alpha$ 2- and  $\alpha$ 1-globin genes diverge in sequence at this position (Michelson and Orkin 1980).

results demonstrate clearly that, when the SPA is placed in intron 2 of the rabbit  $\beta$ -globin gene, it is nonfunctional because of the splicing of the intron. Deletion of the donor site of intron 2 inactivates splicing and thereby reactivates the SPA. As a final confirmation of these results, Figure 4C shows S1 analysis of the R $\beta$ , SPAI, and SPAI $\Delta$ D RNAs using a DNA probe labeled at the R1 site in the rabbit  $\beta$ -globin gene (together with a probe for the  $\alpha$  globin cotransfection control). Both R $\beta$  and SPAI gave normal R $\beta$  3'-end signals, whereas SPAI $\Delta$ D gave no R $\beta$  3'-end signals because the 3' end of its mRNA used the SPA in intron 2.

## Discussion

These studies describe the construction of an efficient synthetic poly(A) site. We used this duplex oligonucleotide as a reagent to investigate a number of parameters associated with poly(A)-site use.

First, we are able to demonstrate unambiguously that the minimum sequence elements AATAAA, GT rich, and T rich are sufficient to constitute a poly(A) site, providing that correct spacing is maintained between them. Although a substantial number of poly(A) site signals have been investigated, the minimum sequences re-

quired for polyadenylation were not established previously. In particular, the possible role of a sequence between the A<sub>2</sub>UA<sub>3</sub> and the GT-rich element was controversial. Although our studies rule out specific sequence requirements for this region, we, as others, find that the different sequences promote slight variations in the precise position at which poly(A) is added (Fitzgerald and Shenk 1981; Mason et al. 1986).

The exclusive use of the SPA, when placed 3' to the poly(A) site of the human  $\alpha$ -globin gene, was surprising. The fact that the SPA possesses more extensive downstream sequences (both a GT-rich and a T-rich element) than  $\alpha$ -globin is likely to be the explanation for this observation; however, it is still surprising that the  $\alpha$ -globin poly(A) site is completely inactive when placed adjacent to the SPA. Presumably, the 3'-processing factors that interact with the poly(A) sites must have a considerably higher affinity for SPA than  $\alpha$ . Once the SPA is selected, processing at the  $\alpha$ -poly(A) site must be excluded mutually.

Another unexpected result of the experiments in which SPA is placed 3' to the  $\alpha$ -poly(A) site is that when the two poly(A) sites are brought within a few nucleotides of each other,  $\alpha$  is used rather than SPA. We suggest that steric factors introduced by the abnormally close positions of the two sites may favor the 5'-positioned poly(A) site even though its signal is the weaker of the two signals. Clearly, the results described here have some implication for the differential use of a particular poly(A) site when it is present in the 3'-untranslated region of a gene with multiple poly(A) sites.

When placed 5' to the rabbit  $\beta$ -globin poly(A) site in exon 3, the SPA is used at a 90% level. We showed previously that when two copies of the poly(A) site of the rabbit  $\beta$ -globin gene are placed in tandem, the 5' poly(A) site is used exclusively, unless the 5' poly(A) site is inactivated partly, in which case an increase in the 3' poly(A) site occurs (Gil and Proudfoot 1984, 1987). Therefore, it follows that the SPA is very nearly as efficient as the original rabbit  $\beta$ -globin poly(A) site. However, when the SPA is placed in intron 2 of the  $\beta$ -globin gene, it is no longer used at detectable levels unless the splicing signals of this intron are disrupted. These results demonstrate that polyadenylation must occur after splicing has taken place, which raises some interesting mechanistic aspects of the overall transcription and processing of a primary *poIII* transcript.

First, it is clear that an efficient poly(A) site is not sufficient for termination of transcription. A number of studies have demonstrated that termination of transcription by RNA polymerase II requires active polyadenylation (Whitelaw and Proudfoot 1986; Logan et al. 1987; Connelly and Manley 1988; Lanoix and Acheson 1988) and possibly no other specific sequences. However, in the case of the mouse  $\beta$ -globin gene, it has been demonstrated that a region in its 3'-flanking sequence is required in addition to the poly(A) site of the gene for efficient termination of transcription (Falck-Pedersen et al. 1985; Logan et al. 1987). The results described here demonstrate also that an efficient poly(A) site is not suf-

ficient to prevent transcription from reading several hundred base pairs beyond the poly(A) site to the end of the intron, thus allowing preferential splicing instead of polyadenylation. Second, the fact that splicing occurs ahead of polyadenylation suggests that the formation of spliceosomes must exclude factors involved in 3' processing and polyadenylation. Two examples of the predominance of splicing over polyadenylation have been reported previously. First, in the case of the calcitonin gene, its poly(A) site is prevented from working if alternative splicing occurs to form the related calcitonin gene-related peptide (CGRP). In cells making CGRP, splicing excludes the use of the intronic calcitonin poly(A) site (Leff et al. 1987). Similarly, in adenovirus transcription, the Late 4 (L4) poly(A) site is within the intron of the overlapping Early 3 (E3) transcription unit. Again, the L4 poly(A) site is not used when E3 is spliced (Adami and Nevins 1988). Clearly, our observations here on the inactivity of the SPA, when inserted into the intron of the rabbit  $\beta$ -globin gene, are consistent entirely with those two natural examples. The notion that splicing predominates over polyadenylation may, therefore, be a general feature of RNA processing.

## Materials and methods

### SPA oligonucleotide constructions

The SPA oligonucleotides shown in Figure 1 were synthesized on an Applied Biosystems DNA Synthesizer. The actual oligonucleotides made were: (1) GATCCAATA<sub>4</sub>GATCT<sub>3</sub>AT<sub>4</sub>CATTAGATCTGTGTGTTGGT<sub>6</sub>GTGTG; and (2) GATCCACACA<sub>6</sub>CCAACACACAGATCTAATGA<sub>4</sub>TA<sub>3</sub>GATCTTTTATTG. These formed the duplex SPA with 5' overhangs of GATC to allow ligation into *Bam*HI or *Bgl*II restriction sites. A third oligonucleotide (3) was synthesized and was a direct complement to the first oligonucleotide. Oligonucleotides (1) and (3), on annealing, formed a flush-ended SPA. The linker oligonucleotide was synthesized previously [oligonucleotide C (Gil and Proudfoot 1987)].

### Human $\alpha 2$ -globin gene constructions

The plasmid  $\alpha 2W3$ 'PS pSVed (Whitelaw and Proudfoot 1986) contained the human  $\alpha 2$ -globin gene with extensive 5'- and 3'-flanking regions placed in a transient expression vector (pSVed) that has the SV40 origin of replication and enhancer sequences (Proudfoot et al. 1984). The plasmid was linearized with *Pvu*II and subjected to *Bal*31 exonuclease digestion. Following ligation in the presence of *Bgl*II oligonucleotide linkers, deletion mutants  $\Delta 1$ – $\Delta 5$ , and S were selected. These six deletion mutants were digested with *Bgl*II, which excises a fragment with one end at the deletion end point and the other end at a constant position approximately 2 kb into the 3'-flanking region of the  $\alpha 2$ -globin gene. After ligation in the presence of SPA, the six final constructs were obtained that had SPA cloned into each deletion end point in the correct orientation with the same 3'-flanking region sequence 3' to the SPA. Then clone S was digested with *Bgl*II to remove the middle portion of the SPA and oligonucleotide C was ligated in its place in both orientations (S+ and S-).  $\Delta 3C$  lacked the SPA sequence but was otherwise identical to  $\Delta 3$ .



*Rabbit  $\beta$ -globin gene constructions*

The plasmid pBSVpBR328 (Grosveld et al. 1982) contained the whole rabbit  $\beta$ -globin gene in a plasmid in which SV40 and pBR328 were fused between their *EcoRI* and *BamHI* sites. The SV40 T-antigen gene was left intact to allow the replication of SV40-origin-containing plasmids in transient expression experiments.

A 322-bp *PvuII* fragment from the plasmid pUC119, which contained a polylinker sequence, was inserted into the middle of  $\beta$ -globin intron 2 at a *ScaI* site. SPA was inserted into this polylinker at the *SmaI* site (SPA1) or into a *BglII* site at the termination codon in exon 3 (SPA2). Finally, SPA1 was deleted at the donor site of intron 2 between a *BamHI* site just within exon 2 and an *EcoO* 109 site 10 bp into the intron (SPA1 $\Delta$ D). The SP6 riboprobe was constructed by the insertion of a *DraI* 1.0-kb fragment from intron 2 and exon 3 of the rabbit  $\beta$ -globin SPA1 gene construct into the polylinker *SmaI* site of pSP65 in the reverse orientation.

*Analysis of human  $\alpha$ 2-globin and rabbit  $\beta$ -globin gene constructs by transient expression in HeLa cells*

Subconfluent HeLa cells, on 90-cm tissue culture plates, were transfected with 15  $\mu$ g of test plasmid and 5  $\mu$ g of control plasmid using standard calcium phosphate precipitation techniques. After 2 days, cytoplasmic RNA was purified and subjected to S1 nuclease or exonuclease VII digestion. All of these procedures have been described previously by Johnson et al. (1986).

**References**

- Adami, G. and J.R. Nevins. 1988. Splice site selection dominates over poly(A) site choice in RNA produced from complex adenovirus transcription units. *EMBO J.* 7: 2107–2116.
- Birnstiel, M.K., M. Busslinger, and K. Strub. 1985. Transcription termination and 3' processing: The end is in site. *Cell* 41: 349–359.
- Connelly, S. and T. Manley. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* 2: 440–452.
- Falck-Pedersen, E., J. Logan, T. Shenk, and J.E. Darnell. 1985. Transcription termination within the E1A gene of adenovirus induced by insertion of the mouse  $\beta$  major globin terminator element. *Cell* 40: 897–905.
- Fitzgerald, M. and T. Shenk. 1981. The sequence of 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24: 251–260.
- Gil, A. and N.J. Proudfoot. 1984. A sequence downstream of AAUAAA is required for rabbit  $\beta$ -globin mRNA 3' end formation. *Nature* 312: 473–474.
- . 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit  $\beta$  globin mRNA 3' end formation. *Cell* 49: 399–406.
- Grosveld, G.C., E. de Boer, C.K. Shewmaker, and R.A. Flavell. 1982. DNA sequences necessary for transcription of the rabbit  $\beta$ -globin gene *in vivo*. *Nature* 295: 120–126.
- Higgs, D.R., S.E.Y. Goodbourn, J. Lamb, J.B. Clegg, D.J. Weatherall, and N.J. Proudfoot. 1983.  $\alpha$ -Thalassaemia caused by a polyadenylation signal mutation. *Nature* 306: 398–400.
- Johnson, M., C. Norman, M.A. Reeve, J. Scully, and N.J. Proudfoot. 1986. Tripartite sequences within and 3' to the sea urchin H2A histone gene display properties associated with a transcriptional termination process. *Mol. Cell. Biol.* 6: 4008–4018.
- Lanoix, J. and N.H. Acheson. 1988. A rabbit  $\beta$ -globin polyadenylation signal directs efficient termination of transcription of polyoma virus DNA. *EMBO J.* 7: 2515–2522.
- Leff, S.E., R.M. Evans, and M.G. Rosenfeld. 1987. Splice commitment dictates neuron-specific alternative RNA processing in calcitonin/CGRP gene expression. *Cell* 48: 517–524.
- Logan, J., E. Falck-Pedersen, J.E. Darnell, and T. Shenk. 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse  $\beta$  globin gene. *Proc. Natl. Acad. Sci.* 84: 8306–8310.
- Mason, P.J., J.A. Malgorzata, M.B. Jones, and J.B. Williams. 1986. Mutations downstream of the polyadenylation site of a *Xenopus*  $\beta$ -globin mRNA affect the position but not the efficiency of 3' processing. *Cell* 46: 263–270.
- McLauchlan, J., D. Gaffney, J.L. Whitton, and J.B. Clements. 1985. The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res.* 13: 1347–1368.
- Melton, D.A., P.A. Krieg, M.R. Rebagliati, T. Maniatis, K. Zinn, and M.R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* 12: 7035–7056.
- Michelson, A.M. and S.H. Orkin. 1980. The 3' untranslated regions of the duplicated human  $\alpha$ -like globin genes are unexpectedly divergent. *Cell* 22: 371–377.
- Montell, C., E.F. Fisher, M.H. Caruthers, and A.J. Berk. 1983. Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' consensus sequence AAUAAA. *Nature* 305: 600–605.
- Orkin, S.H., T.-C. Cheng, S.E. Antonarakis, and H.H. Kazazian. 1985. Thalassaemia due to a mutation in the cleavage-polyadenylation signal of the human  $\beta$ -globin gene. *EMBO J.* 4: 453–456.
- Proudfoot, N.J. and G.G. Brownlee. 1976. The 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* 263: 211–214.
- Proudfoot, N.J., T. Rutherford, and G. Partington. 1984. Transcriptional analysis of human  $\zeta$  globin genes. *EMBO J.* 3: 1533–1540.
- Proudfoot, N.J. and E. Whitelaw. 1988. In *Frontiers in molecular biology* (ed. D.B. Hames and D. Glover), pp. 97–129. IRL Press, Oxford.
- Whitelaw, E. and N.J. Proudfoot. 1986.  $\alpha$ -Thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human  $\zeta$  globin gene. *EMBO J.* 5: 2915–2922.



## Definition of an efficient synthetic poly(A) site.

N Levitt, D Briggs, A Gil, et al.

*Genes Dev.* 1989, **3**:

Access the most recent version at doi:[10.1101/gad.3.7.1019](https://doi.org/10.1101/gad.3.7.1019)

---

### References

This article cites 22 articles, 3 of which can be accessed free at:  
<http://genesdev.cshlp.org/content/3/7/1019.full.html#ref-list-1>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

