5256–5263 Nucleic Acids Research, 1993, Vol. 21, No. 22

Polyadenylation site selection cannot occur *in vivo* after excision of the 3′-terminal intron

Xuedong Liu[1,2] and Janet E.Mertz[1,*]
[1]McArdle Laboratory for Cancer Research and [2]Laboratory of Genetics, 1400 University Avenue, University of Wisconsin, Madison, WI 53706-1599, USA

Received May 19, 1993; Revised and Accepted September 20, 1993

ABSTRACT

Splicing of 3′-terminal introns and polyadenylation of pre-mRNAs can be coupled in an appropriate cell-free system. However, definitive evidence has been lacking as to whether these events are coupled *in vivo* and whether the order of these two processing events is obligatory. Here, we investigated these questions by examining the *in vivo* processing of transcripts that differ solely by the precise insertion of an intron within the first of two polyadenylation signals. Quantitative S1 nuclease mapping and PCR techniques were utilized to analyze the processed RNAs that accumulated in monkey cells transfected with plasmids encoding these transcripts. We found that, whereas all of the primary transcripts that lacked the inserted intron were processed via utilization of the 5′-proximal polyadenylation signal, none of the transcripts initially disrupted in this signal were processed this way even though the disrupting intron had been properly excised and excision sometimes preceded polyadenylation. In addition, deletion of the second polyadenylation signal resulted in failure of spliced transcripts to accumulate. We conclude that selection of, but not necessarily cleavage at the polyadenylation site precedes excision of the 3′-terminal intron *in vivo*; although coupling exists during selection of the sites to be used for polyadenylation and excision of the 3′-terminal intron, the actual order of the subsequent enzymatic reactions is probably simply a reflection of their relative kinetics.

INTRODUCTION

Most pre-mRNAs of higher eukaryotes are processed via numerous steps in which they are capped, polyadenylated, spliced, and then transported from the nucleus to the cytoplasm (1–3). The mechanism of polyadenylation, which involves endonucleolytic cleavage followed by the addition of 50 to 250 adenosine residues to the 3′ end of the RNA, has been intensively studied (4–6). A highly conserved hexanucleotide sequence, AAUAAA, located 10–30 nucleotides 5′ of the poly(A) site, is required for accurate cleavage and polyadenylation both *in vivo* and *in vitro* (5, 6). In addition, a signal referred to as the G/U box, located 3′ of the site of polyadenylation, is also required for efficient poly(A) site utilization (5). Although the G/U box is less well defined in sequence and location, its proximity to the AAUAAA signal is important (7). If the distance between these two sequence elements is increased to beyond 40 nucleotides, poly(A) site usage is lost except in unusual cases in which a highly stable RNA stem–loop structure forms between the AAUAAA and G/U box signals (8, 9).

Specific sequence elements that are required for excision of introns have been determined. These elements include the highly conserved 5′ and 3′ splice sites, the branch point, and polypyrimidine region near the 3′ splice site (1, 3). Mutations at these sites often alter splicing. In addition, splicing requires a minimal intron length (10).

Numerous studies aimed at determining whether the splicing and polyadenylation reactions occur in an obligatory order have yielded contradictory results. Because the kinetics of the polyadenylation reaction is much more rapid than the kinetics of the splicing reaction for most pre-mRNAs (2, 11), polyadenylation usually precedes splicing. In support of this conclusion is the fact that unspliced, polyadenylated transcripts are readily detected both *in vivo* and *in vitro* as intermediates in RNA processing (12 and references therein). However, splicing of 5′-proximal introns usually precedes polyadenylation in very large transcripts because of the larger time interval between synthesis of the 5′- and 3′-ends of the transcripts (13–15 and references therein). Furthermore, results from some experiments designed to examine *in vivo* processing of transcripts that could be alternatively spliced and polyadenylated suggest that splicing precedes polyadenylation (16).

Splicing and polyadenylation reactions can be uncoupled and studied separately *in vitro*. Nevertheless, Niwa and Berget (12, 17, 18) demonstrated that polyadenylation can be stimulated by the presence of an upstream 3′ splice site, and mutations in a polyadenylation signal can inhibit splicing of a 3′-terminal intron in a cell-free processing system under appropriate conditions. These data indicated that splicing of the 3′-terminal intron and polyadenylation may be coupled and provide strong support for the part of their 'exon definition' hypothesis (19) concerned with

* To whom correspondence should be addressed

the 3'-terminal exon, namely, that an early step in pre-mRNA processing involves interaction between the 3'-terminal, 3' splice site and the polyadenylation signal.

In vivo experiments are necessary to establish the physiological relevance of these *in vitro* observations. Numerous studies (20–24 and references therein) have indicated that mutations of a 3' splice site can affect the efficiency of polyadenylation and choice of a polyadenylation site *in vivo*. Unfortunately, all except the most recent of these studies (24) were complicated by the fact that inactivation of the 3' splice site resulted in rapid turnover of the RNA, making careful analysis of its processing unachievable.

In the work described here, we examined whether splicing of a 3'-terminal intron and polyadenylation can occur independently *in vivo* using RNAs designed to accumulate stably in cells regardless of how processed. In strong support of the exon definition hypothesis of Berget and colleagues (19), we show that polyadenylation site selection cannot occur *in vivo* after excision of a 3'-terminal intron. Thus, determination of the endpoints of the 3'-terminal exon occurs *in vivo* as an early step in pre-mRNA processing.

MATERIALS AND METHODS

Cells and transfection

The African green monkey kidney cell line CV-1PD was grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal bovine serum as described previously (25). Co-transfections were performed by a modification of the DEAE–dextran protocol of McCutchan and Pagano (26) essentially as described previously (27). Briefly, 2.0 μg of test plasmid DNA and 1.0 μg of pRSV-Tori, an SV40 T-antigen-encoding plasmid (28), were diluted into 0.5 ml DMEM containing 500 μg of DEAE–dextran (molecular weight 2×10^6, Pharmacia) per ml and 20 mM Hepes (pH 7.25) to a final concentration of 1.0 μg/ml. Confluent monolayers of cells were subcultured 1:5 two days before transfection, washed twice with DMEM containing 20 mM Hepes (pH 7.25), and incubated with 0.5 ml of DNA solution per 100-mm dish for 45 min in a 5% CO_2 incubator at 37°C with rocking every 10 min. The cells were subsequently incubated with DMEM containing 20 mM Hepes (pH 7.25) and 100 mM chloroquine for 4 hr in a 5% CO_2 incubator at 37°C. After chloroquine treatment, the cells were washed twice with DMEM containing 20 mM Hepes (pH 7.25), refed with DMEM containing 2% fetal bovine serum, and incubated at 37°C in a 5% CO_2 incubator for 42–48 h.

Recombinant plasmids

Standard recombinant DNA techniques were used in the construction of all plasmids (29). The starting plasmids psPA and psPΔA, kindly provided by B.Cullen, have been described in detail previously (9). The plasmid psPA contains the cytomegalovirus immediate early (CMV-IE) promoter 5' of sequences derived primarily from the U3 region of the HIV-1 LTR. These are followed by a synthetic poly(A) site located 204 bp 5' of a natural poly(A) site derived from the rat preproinsulin II gene (Figure 1). The plasmid psPΔA is identical to psPA except for a mutation of the AATAAA signal in the synthetic poly(A) to AAGAAG. psPA+4, which contains a 4 bp insertion (5'-CAGG-3') between the *Hin*dIII and *Tth*111I sites, was made by insertional PCR-based mutagenesis. The plasmid psPA+IVS2 differs from psPA+4 by the insertion in the sense orientation of the second intron of the human β-globin gene; it was constructed by insertion of a PCR-generated DNA fragment consisting of the human β-globin second intron (IVS2) sequence with attached *Hin*dIII and *Tth*111I ends into the *Hin*dIII–*Tth*111I site of psPA. The plasmid psPA+IVS2mini differs from psPA+IVS2 in that it lacks much of the sequence derived from the human β-globin IVS2 (i.e., nucleotides 514–1178 relative to the transcription initiation site in the human β-globin gene); it was constructed in a similar manner as was psPA+IVS2, except that the template used in the PCR reaction was pβIVSΔP/ΔS, a plasmid containing an IVS2-deleted variant of the human β-globin gene (23). The plasmid psPA+IVS2ΔIIA, lacking 282 bps including the preproinsulin polyadenylation signal, was made by deletion of the smaller *Bst*EII–*Eco*RI fragment from psPA+IVS2.

Purification and structural analysis of RNA

Nuclear and cytoplasmic RNAs were purified from transfected monkey cells as described previously (27). The *Bsp*MII-cut, 3' end-labeled probes shown schematically in Figure 3B were used for quantitative S1 nuclease mapping as described previously (30) to determine the polyadenylation sites used in 3'-end formation. DNA sequence analysis of the product of reverse-PCR was used to determine precisely the sequence of the processed RNA synthesized from psPA+IVS2 in the region surrounding the excised intron. In brief, primer PA2 (5'-GGATCCACG-ATGCCGCGCTTCT-3'), complementary to a sequence 3' of the intron (see Figure 4C for location), was mixed with cytoplasmic RNA in 1× PCR buffer (Promega), heated at 80°C for 1 min, annealed at 62°C for 15 min, and extended by incubation at 48°C for 1.5 h with AMV reverse transcriptase (Life Sciences). PCR-based amplification of the cDNA was performed by addition of primer PA51 (5'-TCGAGCCCTCAG-ATCCTGCATA-3'), homologous to sequence 5' of the intron (see Figure 4C for location), and *Taq*I DNA polymerase followed by 35 cycles of amplification. The profile involved denaturation at 94°C for 30 sec, annealing at 55°C for 30 sec, and extension at 72°C for 2 min in a Perkin-Elmer Cetus thermal cycler. DNA sequence analysis of the resulting PCR product was performed by the dideoxy chain termination method (29) with Sequenase 2.0 (United States Biochemicals).

Reverse-PCR techniques were also used to identify transcripts of psPA+IVS2 that had been spliced, but not yet cleaved for polyadenylation. This analysis was performed essentially as described above except that (i) primer PA3 (5'-ACTCCAGCCG-ACCCCACTCATGTT-3') (see Figure 4C for location) was used in place of primer PA2 for amplification, and (ii) random primer (Pharmacia) was used to prime the reverse transcription.

RESULTS

Experimental design

Proudfoot and colleagues (16) constructed a consensus synthetic poly(A) site and used it to demonstrate that the sequence required for efficient polyadenylation is AAUAAA plus a G/U box with 22–23 nucleotides between them. Brown *et al.* (9) confirmed and extended this observation using an expression vector, psPA, which contains a consensus synthetic poly(A) site 5' of a poly(A) site derived from the rat preproinsulin gene and sequence from the U3 and R regions of the LTR of HIV 5' to the synthetic poly(A) site (Figure 1). As is true for lentiviruses and retroviruses, the transcripts made from this plasmid can be

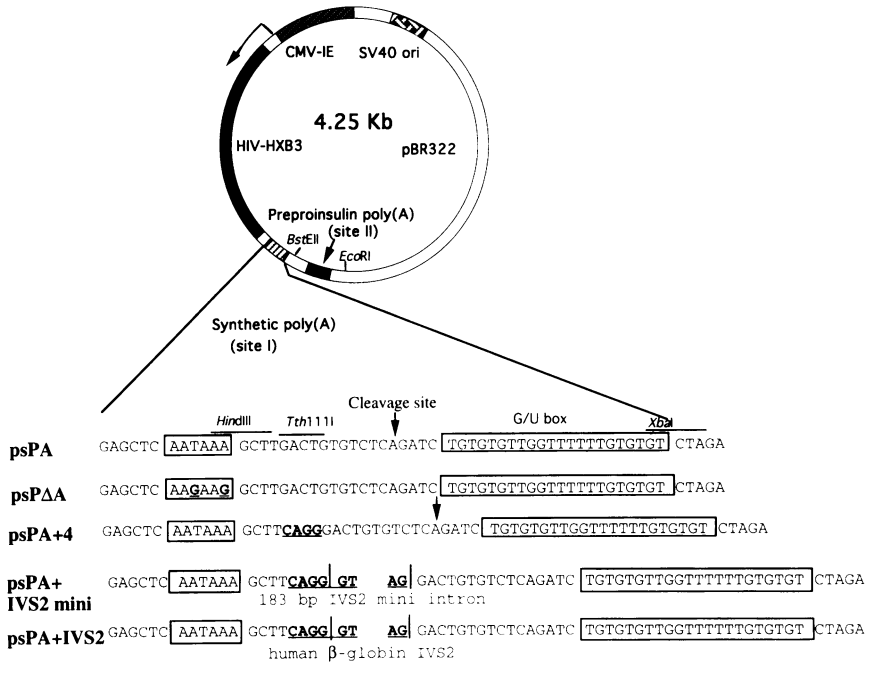


Figure 1. Schematic diagram of the constructs used in this study. Bold-faced letters indicate the alterations in sequence from psPA. The boxes indicate the locations of the AAUAAA and G/U box signals.

polyadenylated in the absence of splicing (9). They are processed *in vivo* exclusively by utilization of the synthetic poly(A) site; however, both mutations that inactivate the synthetic poly(A) signal, such as the one in psPΔA (Figure 1), and insertions that disturb the spacing between the AAUAAA and G/U box of the synthetic poly(A) result in the utilization, instead, of the preproinsulin poly(A) site (9; see below).

To study the relationship between polyadenylation and splicing *in vivo*, we constructed several derivatives of psPA. Plasmids psPA+IVS2 and psPA+IVS2mini contain the human β-globin IVS2 and a 183 bp deleted variant of it, respectively, inserted between the AAUAAA and G/U box of the synthetic poly(A) site present in psPA (Figure 1). By disturbing the spacing between the AAUAAA and G/U box, these intronic insertions should prevent recognition of the synthetic poly(A) site prior to their excision. As a positive control, plasmid psPA+4 was also constructed (Figure 1). The transcripts synthesized from this plasmid are identical in sequence to those present in cells after excision of the intron from transcripts of psPA+IVS2 and psPA+IVS2mini. If poly(A) site selection occurs after splicing, as suggested by the experiments of Levitt *et al.* (16), we would expect most, if not all of the transcripts synthesized from psPA+IVS2 and psPA+IVS2mini to be processed via utilization of the synthetic poly(A) site — i.e., to RNA species D in Figure 2. On the other hand, if poly(A) site selection cannot occur after excision of the introns, all of the transcripts should be processed, instead, via utilization of the preproinsulin poly(A) site — i.e., to RNA species E in Figure 2. To distinguish between the aforementioned hypotheses, monkey cells were transfected with each of these plasmids and pRSV-Tori. The latter plasmid, encoding SV40 T-antigen, allows replication of the test plasmids to high copy number, which results in high levels of transcription. Forty-eight hours after transfection, the cells were harvested, and

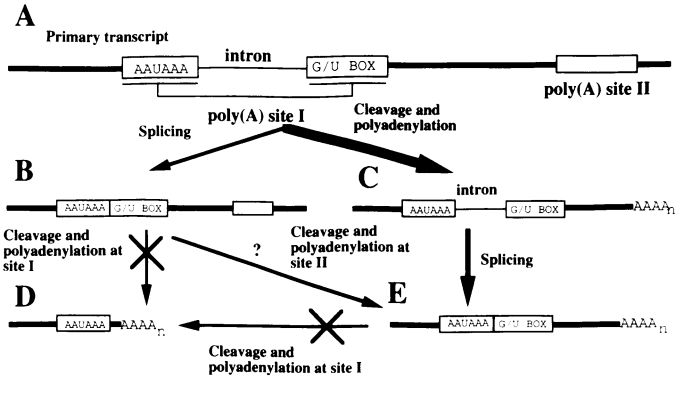


Figure 2. Summary of pathways utilized in the processing of transcripts synthesized from psPA+IVS2 and psPA+IVS2mini. The boxes, not drawn to scale, indicate the locations of the AAUAAA and G/U box signals. The thick solid arrows indicate the predominant pathway utilized in processing of these transcripts based upon the data presented here. The thin arrows indicate other potential pathways: those crossed out were not detectably used; the A → B pathway was used at low levels; and the usage of the B → E pathway was not determined.

nuclear and cytoplasmic RNAs were isolated. These RNAs were analyzed by quantitative S1 nuclease mapping with *Bsp*MII-digested, 3′ end-labeled probes homologous to the cDNA versions of each of the primary transcripts (Figure 3).

Polyadenylation site selection cannot occur after excision of the 3′-terminal intron

In confirmation of the finding of Brown *et al.* (9), RNAs synthesized from psPA were processed exclusively via utilization of the synthetic poly(A) site (site I), while RNAs synthesized from psPΔA were processed exclusively via utilization of the

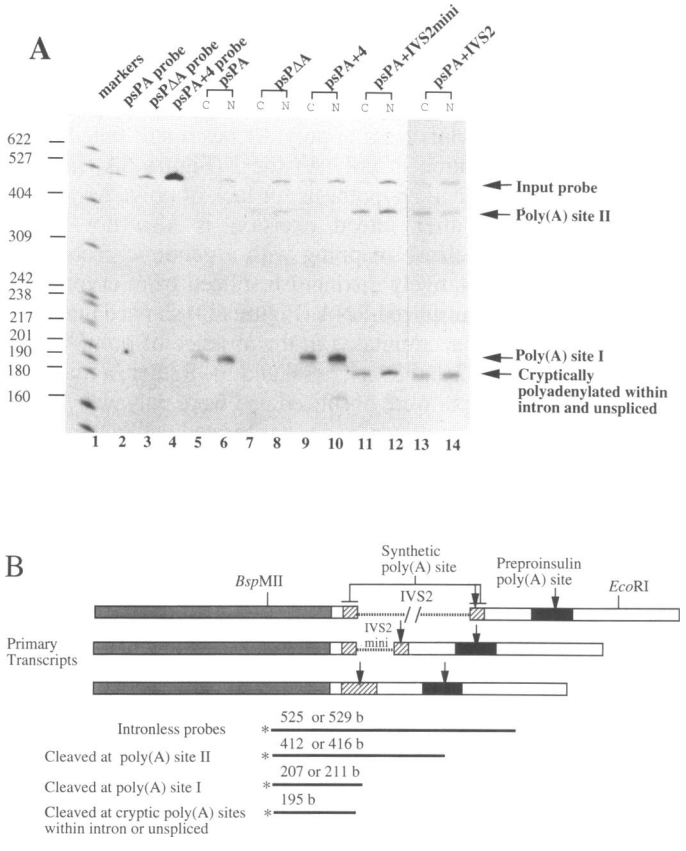


Figure 3. Poly(A) site I is not used in processing of the intron-containing primary transcripts. (A) 3′-end analysis of the RNAs accumulated in CV-1PD cells cotransfected with each of the plasmid DNAs indicated in Figure 1 and pRSV-Tori at a 2:1 molar ratio. Forty-eight hours after transfection, nuclear (N) and cytoplasmic (C) RNAs were isolated and analyzed by S1 nuclease mapping. The probes used, shown schematically in panel B, were *Bsp*MII-cut, 3′ end-labeled DNA fragments homologous to the cDNA version of each of the primary transcripts. After hybridization at 50°C and treatment with S1 nuclease at 25°C, the resulting protected fragments were electrophoresed in 5% polyacrylamide gels containing 8 M urea. A band approximately 170 b in length was detected nonspecifically in variable, low amounts in all RNA samples; its origin remains unclear. The size markers were *Msp*I-cut pBR322 DNA. (B) Schematic diagram of the 3′ end-labeled, cDNA probes made from psPA, psPΔA, and psPA+4 that were used in the 3′-end mapping experiment in panel A and the fragments protected by hybridization with the indicated RNAs. The textures in the transcripts correspond to those used in Figure 1 to indicate various DNA sequences. The downward arrows indicate the major sites of cleavage for polyadenylation. The various segments of the RNAs are not drawn to scale.

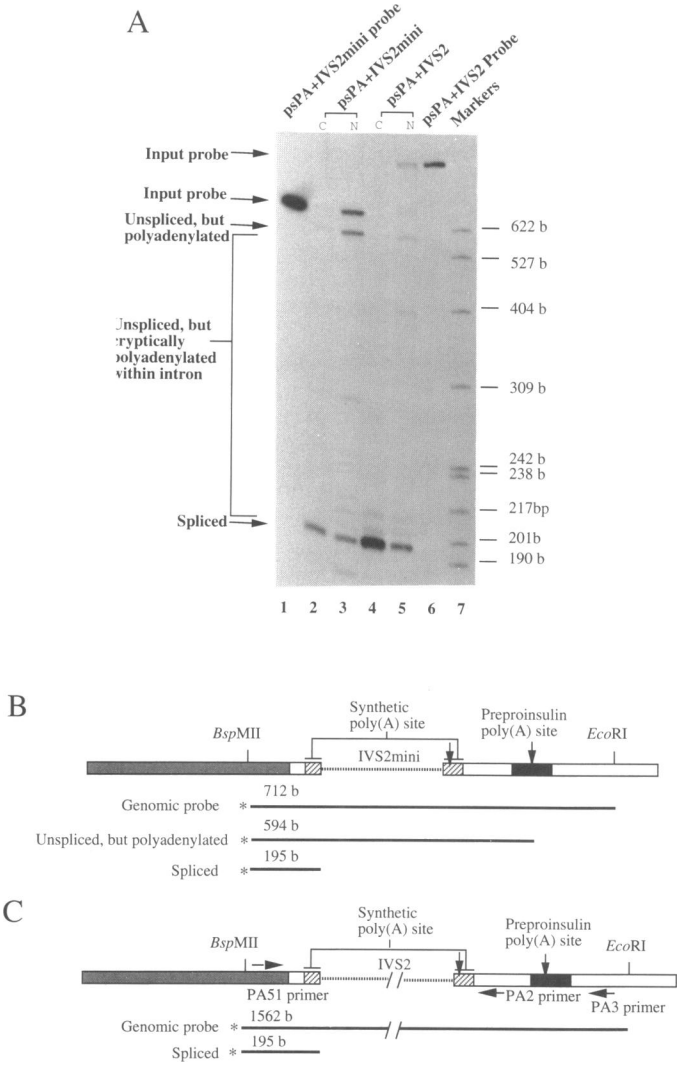


Figure 4. The introns are excised from the transcripts of psPA+IVS2 and PsPA+IVS2mini. (A) 5′ splice site usage determined by S1 nuclease mapping with the *Bsp*MII-cut, 3′ end-labeled probes shown schematically in panels B and C. The RNA samples were the same as those analyzed in the experiment shown in Figure 3A, lanes 11−14. Shown is an autoradiogram of the S1 nuclease-protected DNAs electrophoresed in a 5% polyacrylamide gel containing 8 M urea. (B and C) Schematic diagrams of the 3′ end-labeled, genomic probes used in the 5′ splice site mapping of the RNAs synthesized from psPA+IVS2mini and psPA+IVS2, respectively, and the fragments protected by hybridization with the indicated RNAs. The various segments of the RNAs are not drawn to scale.

preproinsulin poly(A) site (site II) (Figure 3A, lanes 5−8). Transcripts made from psPA+4 were also processed predominantly by cleavage at poly(A) site I (Figure 3A, lanes 9 and 10). Therefore, the presence of the 4 additional bases, CAGG, in the transcripts made from psPA+4 relative to those made from psPA does not significantly affect the selection of the poly(A) site used in 3′-end formation. In contrast, transcripts synthesized from psPA+IVS2 and psPA+IVS2mini were never processed via utilization of poly(A) site I (Figure 3A, lanes 11−14) even though these RNAs were expected to have been identical in primary sequence to those made from psPA+4 after excision of their introns (Figure 1). Long exposures of these gels also failed to show any protected DNA fragments of the size expected for cleavage of these latter RNAs at poly(A) site I (data not shown).

The cryptically polyadenylated RNAs indicated in Figure 3A probably arose from cleavage and polyadenylation within the inserted introns. The appearance of numerous minor bands when probes complementary to the primary transcripts were used for S1 mapping (Figure 4; see also Figure 6 below) provides support for this supposition. These cryptic sites of polyadenylation probably arose because: (i) numerous sequences similar to G/U- or U-rich elements can weakly function as downstream elements for polyadenylation (7); (ii) polyadenylation can still occur if the sequences that disrupt the appropriate spacing between the AAUAAA and G/U signals can form a stable secondary structure (9); and (iii) non-consensus AAUAAAs can function weakly as upstream elements for polyadenylation (5). Since polyadenylation occurred within the intron in these cryptically polyadenylated

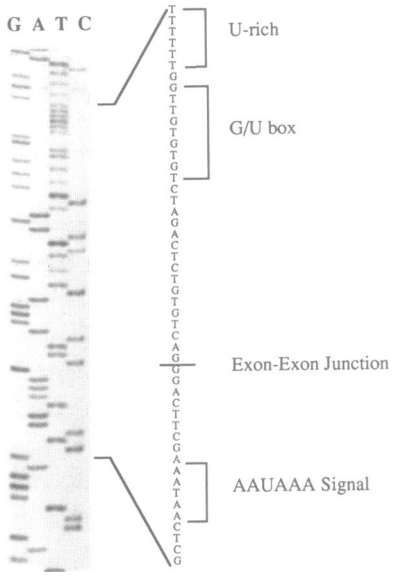


Figure 5. Transcripts of psPA+IVS2 are processed *in vivo* by precise excision of the intron to generate cytoplasmic mRNAs containing an unused polyadenylation signal. A portion of the RNA sample analyzed in the experiment shown in Figure 3, lane 13 was reverse-PCR amplified in the region surrounding poly(A) site I as described in Materials and Methods. Afterward, DNA sequence analysis of the amplified product was performed. Shown here is an autoradiogram of the resulting sequencing gel.

RNAs, polyadenylation superseded intron excision. Thus, selection of poly(A) site I to yield stable mRNAs did not occur, regardless of whether the interrupting intron present in these transcripts was excised.

Two experiments were performed to eliminate the possibility that the failure to process the psPA+IVS2 and psPA+IVS2mini transcripts via utilization of poly(A) site I was a consequence of their introns not being excised as expected. In the first experiment, S1 nuclease mapping analysis of the RNAs was performed with *Bsp*MII-cut, 3′ end-labeled probes complementary to the primary transcripts (Figure 4). As expected, most of the RNA that accumulated in the cytoplasm was discontinuous with the probes at the location of the 5′ splice site of the inserted human β-globin IVS2 (Figure 4A, lanes 2 and 4).

In the second experiment, we determined the precise sequence in the region surrounding poly(A) site I of the spliced RNA accumulated in cells transfected with psPA+IVS2. This was accomplished as follows: First, cDNA of the cytoplasmic RNA was synthesized by reverse transcription with a primer, PA2, which is complementary to sequences 90−110 bases 3′ of the inserted intron (Figure 4C). Second, the poly(A) site I-containing region of the cDNA was amplified with PCR primers PA51, which is homologous to sequences 98−78 bases 5′ of the intron, and PA2 (Figure 4C). Lastly, the resulting amplification product was gel-purified and sequenced. The data from this experiment (Figure 5) indicated definitively that the excision of IVS2 was precise, with the processed, cytoplasmic RNA containing the exact sequence of the synthetic poly(A) site. Therefore, these RNAs failed to be processed via utilization of poly(A) site I even though this poly(A) site was regenerated by excision of the intron.

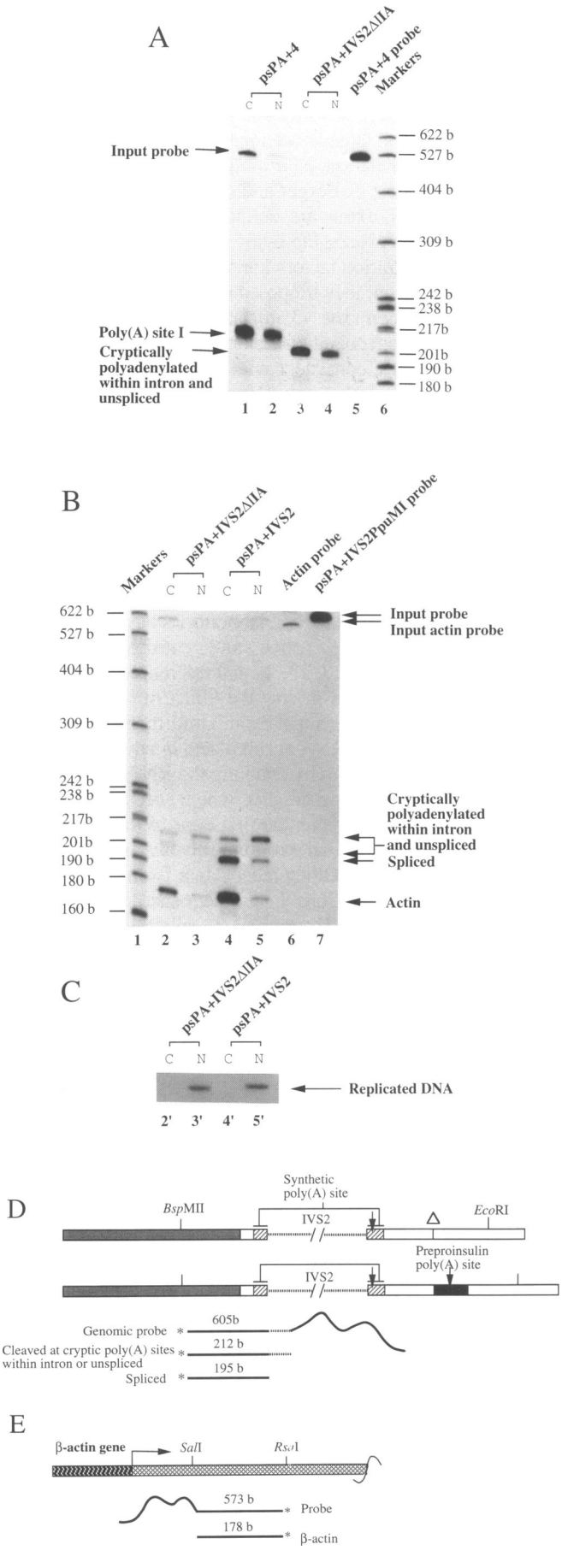
Another plausible hypothesis is that selection of poly(A) site I after intron excision was not observed because it is kinetically unfavorable when poly(A) site II is present. To test this hypothesis, the plasmid psPA+IVS2ΔIIA was constructed. This plasmid differs from psPA+IVS2 by deletion of poly(A) site II. S1 nuclease mapping analysis with cDNA probes (Figure 3B) indicated that transcripts lacking poly(A) site II still failed to be processed via utilization of poly(A) site I (Figure 6A, lanes 3 and 4). Thus, this kinetic hypothesis for lack of polyadenylation at poly(A) site I after intron excision is also not valid. Interestingly, S1 nuclease mapping with a genomic probe that enabled us to quantitatively distinguish spliced from cryptically polyadenylated and unspliced RNA (Figure 6D) showed that little if any spliced RNA accumulated in the absence of poly(A) site II (Figure 6B, lanes 2 and 3 vs lanes 4 and 5). Rather, a majority of the stable transcripts were unspliced and were polyadenylated, instead, at the weak, cryptic poly(A) sites located within the intron (Figure 6B, lanes 2 and 3). Normalization to the relative amount of cellular actin RNA (Figure 6B) and replicated plasmid DNA (Figure 6C) present in these samples to correct for differences in RNA recoveries and transfection efficiencies indicated that (i) spliced RNA lacking poly(A) site II accumulated to at most 1/10th of the level it did in its presence, and (ii) unspliced RNA overaccumulated 1.5- to 2-fold. Thus, intron excision probably also occurred little, if at all, in the absence of poly(A) site II.

Taking these data together, we conclude that selection of the poly(A) site I for production of stable RNA can not occur after excision of the interrupting intron in these transcripts.

Excision of a 3′-terminal intron can precede polyadenylation *in vivo*

The data in Figure 4A, lane 3 also indicated the presence in the nuclei of the cells transfected with psPA+IVS2mini of RNAs that had been cleaved at poly(A) site II, but not spliced — i.e., RNAs corresponding to species C in Figure 2. Presumably, these RNAs were intermediates in processing from RNA species A→E in Figure 2. Their presence at fairly high levels is consistent with polyadenylation usually or always preceding intron excision in the processing of these transcripts. [Our failure to detect significant quantities of unspliced, polyadenylated RNA in the nuclei of the cells transfected with psPA+IVS2 may be a consequence of either (i) S1 nuclease artifactually cleaving the expected, fairly long, protected DNA fragment to heterogeneous sizes and/or (ii) the rate of excision of the full-length human β-globin IVS2 being faster than that of the mini-intron variant of it (31)]. Therefore, another explanation for the failure to detect RNAs of psPA+IVS2 and psPA+IVS2mini processed via utilization of poly(A) site I is that polyadenylation of the transcripts might always precede excision of their introns.

To test the validity of this latter hypothesis, we examined the nuclei of cells transfected with psPA+IVS2 for the presence of RNA that had been spliced, but not yet polyadenylated at either poly(A) site I or II. This was accomplished by reverse transcription with random hexanucleotide $p(N)_6$ primer, followed by PCR amplification with primers PA51 and PA3, the latter being complementary to the sequence 86 bases 3′ of the AATAAA of poly(A) site II (Figure 4C). RNA from cells transfected with psPA+4 was reverse transcribed and amplified in parallel as a control. Size analysis of the amplified products indicated the presence of spliced RNA not cleaved for polyadenylation at poly(A) site II (Figure 7, lane 3). Thus, intron excision can occur in the absence of prior polyadenylation at poly(A) site II. The above-noted absence of RNA species D in Figure 2 in cells transfected with psPA+IVS2mini and



psPA+IVS2 (Figure 3A, lanes 11−14) indicated that these spliced RNAs were not subsequently processed by cleavage and polyadenylation at poly(A) site I to yield stable mRNAs. The failure of these spliced RNAs to be processed in this manner was probably a consequence of their having been already committed to cleavage and polyadenylation at site II even though these actual enzymatic reactions had not yet occurred on these RNA molecules. Alternatively, the spliced RNA detected in Figure 7, lane 3 represented (i) RNA polyadenylated at cryptic polyadenylation sites situated 3′ of the region to which the primer PA3 was annealed, or (ii) dead-end products that could not function as intermediates in the synthesis of either RNA species D or E. Although the experiments presented here do not enable us to distinguish among these three possibilities, we consider the commitment hypothesis to be the one most likely to be valid for the reasons discussed below. In conclusion, polyadenylation site selection cannot occur after splicing even though the excision of the intron probably can precede the cleavage and polyadenylation reaction.

DISCUSSION

To determine whether selection of a polyadenylation site can occur after excision of a 3′-terminal intron, we constructed genes in which the first of two polyadenylation signals exists only after excision of the introns present in the transcripts synthesized from these genes (Figures 1 and 2). The RNAs which accumulated in cells transfected with these genes failed to be polyadenylated at poly(A) site I (Figure 3) even though (i) the introns had been properly excised to regenerate poly(A) site I (Figures 4 and 5), (ii) poly(A) site I was utilized efficiently in the processing of transcripts made from cDNA versions of these genes (Figure 3), and (iii) excision of the introns sometimes preceded cleavage for polyadenylation at poly(A) site II (Figure 7). Deletion of the second, potentially competing polyadenylation signal also did not result in poly(A) site I utilization; rather, spliced transcripts failed to accumulate and, likely, were never made (Figure 6). Therefore, poly(A) site selection could not occur in these

Figure 6. Poly(A) site I is not used even in the absence of poly(A) site II. (A) 3′-end analysis of the RNAs accumulated in CV-1PD cells cotransfected with psP+IVS2ΔIIA and pRSV-Tori. The RNAs were isolated as described in the legend to Figure 3A and analyzed by S1 nuclease mapping with the cDNA probes shown schematically in Figure 3B. (B) 5′ splice site usage determined by S1 nuclease mapping with the *Bsp*MII-cut, 3′ end-labeled, genomic probes shown schematically in panel D. Note that this particular probe is discontinuous with the primary transcripts starting 17 bases 3′ of the 5′ splice site. As an internal control for RNA recovery and nuclear/cytoplasmic fractionation, the relative amounts of cellular β-actin RNA present in each sample were quantified concurrently by inclusion in each S1 mapping reaction of the β-actin probe shown schematically in panel E. (C) Relative transfection efficiencies as determined by Southern blot analysis of replicated plasmid DNA. A portion of each nucleic acid sample analyzed in panel B was removed prior to treatment with DNase I, digested with *Dpn*I and *Hin*dIII, electrophoresed in a 1% agarose gel, transferred to a Nytran membrane, and probed with radiolabeled plasmid DNA. (D) Schematic diagram of the probe used in the 5′ splice site usage experiment shown in panel B and the fragments protected by hybridization with the indicated RNAs. The wavy line indicates the region of the probe non-homologous with the primary transcripts. (E) Schematic diagram of the cellular β-actin probe included in the S1 mapping experiment shown in panel B and the fragment protected by hybridization with cellular β-actin RNA. The wavy line indicates the region of the probe non-homologous with the RNA.

transcripts *in vivo* after excision of their 3′-terminal intron. Thus, splicing and polyadenylation are not independent events *in vivo*.

A likely explanation for this finding is that factors involved in splicing and polyadenylation interact with each other to define the 3′-terminal exon, with the processing sites at both ends of the 3′-terminal exon being selected by the binding of appropriate factors before cleavage occurs at either end (Figure 8). After selection of the 3′ splice site of the 3′-terminal intron and the poly(A) site, polyadenylation may obligatorily precede excision of the 3′-terminal intron as the next step in an ordered reaction. More likely, excision of the 3′-terminal intron preceding or following polyadenylation may simply be a reflection of the relative average rates of these two enzymatic reactions, with polyadenylation usually occurring first because it is a faster reaction. Even if splicing can functionally precede polyadenylation, the polyadenylation signal regenerated by excision of the disrupting intron could not be utilized in our experiments because a commitment to utilize poly(A) site II for 3′-end formation had already been made and could not be redone.

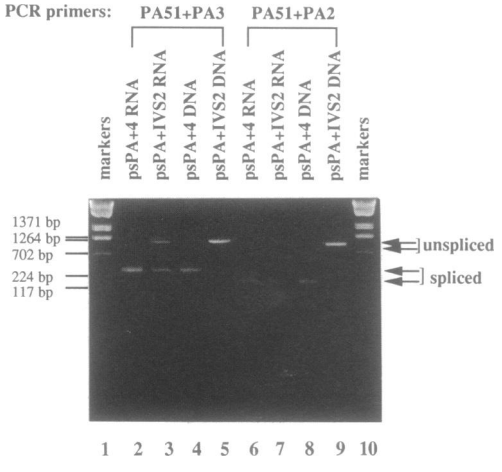


Figure 7. Intron excision can precede cleavage at poly(A) site II. Portions of the RNA samples analyzed in the experiment shown in Figure 3, lanes 10 and 14 were reverse transcribed with random primer and 20 units of AMV reverse transcriptase at 40°C for 75 min. Afterward, the resulting cDNAs were amplified by PCR as described in Materials and Methods with primers PA51 and PA3 (lanes 2–5, see Figure 4C) or primers PA51 and PA2 (lanes 6–9). As controls, plasmid DNAs psPA+4 and psPA+IVS2 were PCR-amplified in parallel. Shown here is a photograph of the PCR products sized by electrophoresis in a 1.0% agarose gel stained with ethidium bromide. Lanes 1 and 10, *Bst*EII-cut λ DNA; lanes 2 and 6, psPA+4 nuclear RNA; lanes 3 and 7, psPA+IVS2 nuclear RNA; lanes 4 and 8, psPA+4 DNA; and lanes 5 and 9, psPA+IVS2 DNA.

The above explanation is a simple extension of the exon definition hypothesis of Berget and colleagues (19). According to their hypothesis, exons are defined as units during an early step in pre-mRNA processing by 3′ splice site and downstream 5′ splice site recognition factors interacting with each other. For the 3′-terminal intron, they proposed that exon definition requires interaction between the 3′-terminal 3′ splice site and polyadenylation site recognition factors. Several observations with cell-free systems are consistent with this proposal. First, Niwa *et al.* (17) have shown that polyadenylation is stimulated *in vitro* by placement of an intron upstream of the polyadenylation signal; however, mutation of the 3′ splice site eliminates the stimulatory effect. Second, mutation of the polyadenylation signal depresses *in vitro* splicing of proximal, but not distal introns (18). Third, insertion of a 5′ splice site within the 3′-terminal exon depresses polyadenylation *in vitro* (32).

To demonstrate the physiological relevance of the above-mentioned findings, they need to be validated with *in vivo* approaches. Studies from several laboratories have suggested that inactivation of the 3′ splice site causes inhibition of polyadenylation *in vivo* (20, 23, 24 and references therein). Niwa *et al.* (32) also examined *in vivo* the effect of insertion of a 5′ splice site within the 3′-terminal exon. Unfortunately, in all these experiments, little or no RNA accumulated *in vivo*. Our approach circumvented this problem by enabling the RNAs to accumulate *in vivo* via alternative processing when necessary. The data presented here provide among the first clear *in vivo* evidence that definition of the 3′-terminal exon is probably an early event in pre-mRNA processing. Other recent data from our laboratory (33) indicate that alterations in the sequence of the (i) poly-pyrimidine tract of the 3′-terminal 3′ splice site and (ii) the 3′-terminal exon can affect the efficiency of both intron excision and polyadenylation *in vivo*. Luo *et al.* (21, 22) have shown that poly(A) site choice is related to 3′ splice site choice in processing of polyoma virus late transcripts. Taken together, these data

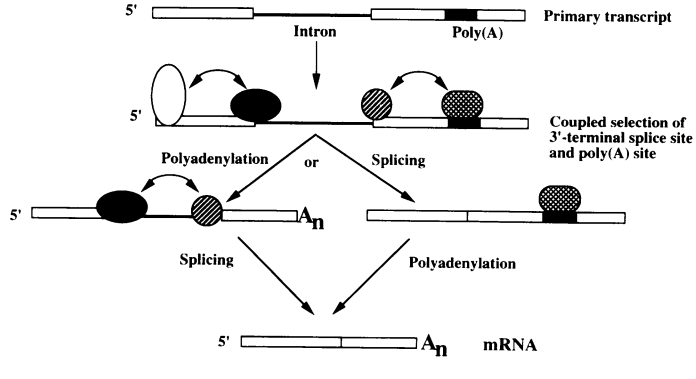


Figure 8. Model for processing of intron-containing pre-mRNAs. The open rectangles and lines designate exons and introns, respectively. The cross-hatched ellipses and filled rectangles represent polyadenylation machinery and poly(A) signals. The hatched circles and the stippled ellipses represent 3′ splice site and 5′ splice site recognition factors, respectively. The cap recognition factor is designated by the open oval. The curved arrows indicate interactions between these factors. The poly(A) tail is represented by A_n.

provide strong evidence that interactions between splicing and polyadenylation have physiological relevance.

Another plausible interpretation of our data is that the splicing and polyadenylation reactions are compartmentalized *in vivo*, with transcripts no longer being accessible to polyadenylation factors after splicing. This hypothesis and the 3′-terminal exon definition hypothesis are not mutually exclusive. In fact, both may very well be the case *in vivo*, with compartmentalization being part of the reason that exon definition needs to be an early event in pre-mRNA processing. Another plausible hypothesis is that proteins or RNPs present upon removal of the intron make poly(A) site I inaccessible to the polyadenylation machinery. Finally, polyadenylation may always precede splicing of the 3′-terminal intron *in vivo*. However, the finding that polyadenylation signals fail to be used when placed within introns (16, 34, 35) argues against the validity of the last of these hypotheses. Whatever the reasons, it is clear that polyadenylation site selection cannot occur after excision of the 3′-terminal intron both *in vivo* (this paper) and *in vitro* (12).

Our conclusion that selection of the 3′-terminal 3′ splice site and polyadenylation site occurs concomitantly as an early step in processing, with the actual order of the subsequent reactions depending upon the relative kinetics of each reaction, leads us to reinterpret several earlier experiments. First, Zeevi *et al.* (36) concluded that polyadenylation and splicing are not coupled *in vivo* based upon their observation that inhibition of polyadenylation with cordycepin did not abolish splicing *in vivo*. However, cordycepin prevents poly(A) tail addition, but not the cleavage reaction and earlier steps in poly(A) site selection. Therefore, exon definition would still be free to occur in their experiment. Second, a number of studies have indicated that both natural and synthetic polyadenylation signals (16, 34, 35) failed to be used when positioned within introns. These results have been interpreted as evidence that splicing precedes polyadenylation. However, given the findings presented here and recently by Niwa *et al.* (32), a more likely explanation for these data is that these intron-encoded polyadenylation signals were ignored during exon definition because an adjacent, upstream, 3′ splice site needed to define the 3′-terminal exon was lacking.

Finally, the interactions that occur between 3′-terminal 3′ splice sites and polyadenylation sites can be subject to regulation. For example, in the processing of *Drosophila doublesex* (*dsx*) mRNA, use of the intrinsically weak 3′ splice site in the female-specific intron is coupled with selection of the female-specific polyadenylation site, with *trans*-acting factors *tra* and *tra*-2 mediating this coupling (37 and references therein). In the processing of immunoglobulin Mμ heavy chain transcripts, the μ_s polyadenylation site is situated within an intron downstream of a weak 5′ splice site, with the weakness of this 5′ splice site being critical for utilization of the μ_s polyadenylation site (38). Quite likely, this 5′ splice site needs to be weak to insure that its presence does not prevent the interactions between the μ_s polyadenylation signal and upstream 3′ splice site necessary for processing to the μ_s mRNA species. Therefore, differential gene expression may be achieved in some cases by regulating definition of the 3′-terminal exon used in pre-mRNA processing.

ACKNOWLEDGMENTS

We thank Bryan Cullen for kindly providing plasmids psPA and psPΔA, Robin Haugh and Qun Fang Hou for technical assistance, and D.Brow, J.Dahlberg, T.Eisenbraun, P.Lambert, Y.Li, R.Spritz, M.Wickens, and members of our laboratory for discussions and helpful comments on the manuscript. This work was supported by US Public Health Service grants CA07175 and CA22443 from the National Institutes of Health.

REFERENCES

1. Green,M.R. (1986) *Annu. Rev. Genet.* **20**, 671–708.
2. Leff,S.E., Rosenfeld,M.G. and Evans,R.M. (1986) *Annu. Rev. Biochem.* **55**, 1091–1117.
3. Padgett,R.A., Grabowski,P.J., Konarska,M.M., Seiler,S. and Sharp,P.A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150.
4. Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell* **41**, 349–359.
5. Wahle,E. and Keller,W. (1992) *Annu. Rev. Biochem.* **61**, 419–440.
6. Wickens,M. (1990) *Trends Biochem.* **15**, 277–281.
7. Gill,A. and Proudfoot,N.J. (1987) *Cell* **49**, 399–406.
8. Ahmed,Y.F., Gilmartin,G.M., Hanly,S.M., Nevins,J.R. and Greene,W.C.(1991) *Cell* **64**, 727–737.
9. Brown,P.H., Tiley,L.S. and Cullen,B.R. (1991) *Genes Devel.* **5**, 1277–1284.
10. Wieringa,B., Hofer,E. and Weissmann,C. (1984) *Cell* **37**, 915–925.
11. Nevins,J.R. and Darnell,J.E.Jr (1978) *Cell* **15**, 1477–1493.
12. Niwa,M. and Berget,S.M. (1991) *Gene Expression* **1**, 5–14.
13. LeMaire,M.F. and Thummel,C.S. (1990) *Mol. Cell. Biol.* **10**, 6059–6063.
14. Ohno,M., Sakamoto,H. and Shimura,Y. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5187–5191.
15. Beyer,A.L. and Osheim,Y.N. (1988) *Genes Devel.* **2**, 754–765.
16. Levitt,N., Briggs,D., Gil,A. and Proudfoot,N.J. (1989) *Genes Devel.* **3**, 1019–1025.
17. Niwa,M., Rose,S.D. and Berget,S.M. (1990) *Genes Devel.* **4**, 1552–1559.
18. Niwa,M. and Berget,S.M. (1991) *Genes Devel.* **5**, 2086–2095.
19. Robberson,B.L., Cote,G.J. and Berget,S.M. (1990) *Mol. Cell. Biol.* **10**, 84–94.
20. Chiou,H.C., Dabrowski,C. and Alwine,J.C. (1991) *J. Virol.* **65**, 6677–6685.
21. Luo,Y. and Carmichael,G.G. (1991) *Mol. Cell. Biol.* **11**, 5291–5300.
22. Luo,Y. and Carmichael,G.G. (1991) *J. Virol.* **65**, 6637–6644.
23. Ryu,W.-S., Gelembiuk,G., Liu,X. and Mertz,J.E., manuscript in preparation.
24. Nesic,D., Cheng,J. and Maquat,L.E. (1993) *Mol. Cell. Biol.* **13**, 3359–3369.
25. Good,P.J., Welch,R.C., Ryu,W.-S. and Mertz,J.E. (1988) *J. Virol.* **62**, 563–571.
26. McCutchan,J.H. and Pagano,J.S. (1968) *J. Natl. Cancer Inst.* **41**, 351–357.
27. Ryu,W.-S. and Mertz,J.E. (1989) *J. Virol.* **63**, 4386–4394.
28. Yu,X.-M., Gelembiuk,G.W., Wang,C.-Y., Ryu,W.-S. and Mertz,J.E. (1991) *Nucleic Acids Res.* **19**, 7231–7234.
29. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
30. Barkan,A. and Mertz,J.E. (1984) *Mol. Cell. Biol.* **4**, 813–816.
31. Lang,K.M., van Santen,V.L. and Spritz,R.A. (1985) *EMBO J.* **4**, 1991–1996.
32. Niwa,M., MacDonald,C.C. and Berget,S.M. (1992) *Nature* **360**, 277–280.
33. Liu,X. and Mertz,J.E., manuscript in preparation.
34. Adami,G. and Nevins,J.R. (1988) *EMBO J.* **7**, 2107–2116.
35. Brady,H.A. and Wold,W.S.M. (1988) *Mol. Cell. Biol.* **8**, 3291–3297.
36. Zeevi,M., Nevins,J.R. and Darnell,J.E.Jr (1981) *Cell* **26**, 39–46.
37. Hedley,M.L. and Maniatis,T. (1991) *Cell* **65**, 579–586.
38. Peterson,M.L. and Perry,R.P. (1989) *Mol. Cell. Biol.* **9**, 726–738.