The EMBO Journal vol.7 no.7 pp.2107 - 2116, 1988

Splice site selection dominates over poly(A) site choice in RNA production from complex adenovirus transcription units

Guy Adami and Joseph R.Nevins[1]

Howard Hughes Medical Institute, Rockefeller University, New York, NY, USA

[1]Present address: Howard Hughes Medical Institute, Duke University Medical Center, PO Box 3054, Durham, NC 27710, USA

Communicated by N.J.Proudfoot

The adenovirus late genes are organized in a complex transcription unit in which the production of multiple RNAs is controlled by specific RNA processing events. Adding to the complexity is the fact that the early E3 transcription unit is wholly contained within the late transcription unit. Within this overlapping region there are splicing events and poly(A) site choices specific to each transcription unit. Early in infection, there is near exclusive use of the E3 splice sites with little L4 poly(A) site use whereas the reverse is true late in infection. Using a plasmid containing an intact E3 transcription unit, we demonstrate that the L4 poly(A) site, located in the first E3 intron, is not used either in a mock-infected cell or a late infected cell; nor is it used if the same transcription unit is driven by the major late promoter. In addition, the nature of the poly(A) site appeared to be unimportant since the E3 poly(A) site was also not used in the intron if inserted in place of the L4 poly(A) site. Thus, splice site selection appears to dominate over poly(A) site choice. This conclusion was confirmed by the observation that deletion of either the E3 splice donor (5'ss) or the E3 splice acceptor (3'ss) allowed the use of the L4 poly(A) site. Finally, the L4 poly(A) site within the E3 intron was also used when additional sequence, including splicing signals from the L4 region and the tripartite leader, was inserted between the promoter and the E3 processing sites. It appears that splicing is the dominant event in governing production of E3/L4 RNAs.

Key words: adenovirus/complex transcription unit/splice site

Introduction

The biogenesis of an animal cell mRNA is a complex process involving a variety of steps (for review see Nevins, 1983). The initial RNA transcript is capped, polyadenylated, spliced and then transported from the nucleus to the cytoplasm. From the analysis of a variety of cell and viral genes, it is clear that most of these events are subject to regulation so as to alter the final output of the transcription unit. For example, in the cases of complex transcription units which produce multiple mRNAs from a single primary transcription event, alteration in splice site choices or poly(A) site choices can vary the output of mRNA (Nevins, 1983; Leff *et al.*, 1986). Two examples of complex transcription units that have been analyzed in some detail are the calcitonin/CGRP gene and the immunoglobulin μ heavy chain locus. Both specify alternate exons and multiple poly(A) sites with the processing of precursor RNAs differing depending on the cell type. Calcitonin is the product of the calcitonin/CGRP transcription unit when expressed in the thyroid whereas the neuropeptide CGRP is the product of transcription in the brain (Amara *et al.*, 1984). Using transfection assay systems to study the requirements for regulated expression, Leff *et al.* (1987) have shown that differential processing of the calcitonin/CGRP primary transcript in different cells is dependent on RNA splicing choices. In contrast, a similar analysis of the immunoglobulin μ gene, which produces a μ membrane mRNA and protein in B cells and a secreted form in plasma cells (Alt *et al.*, 1980; Early *et al.*, 1980; Rogers *et al.*, 1980), indicated that poly(A) site choice was dominant and determined the output of the RNA dependent on cell type (Galli *et al.*, 1987, 1988).

Perhaps the best studied complex transcription unit is the adenovirus major late. Transcription originates from a single promoter and differential RNA processing, including multiple poly(A) site choices and multiple splicing events, allows for the production of at least 20 mRNAs (for review see Nevins and Chen-Kiang, 1981). Within this 30 kb transcription unit, five poly(A) sites give rise to five families of 3' co-terminal mRNAs. The families of RNAs are generated by alternative splicing of the third leader segment to one or another acceptor site within the body of one of the five groups. The basis for these RNA processing choices is unclear but certainly of great importance for an understanding of mechanisms of post-transcriptional gene control.

In addition to these complex RNA processing choices which take place during expression of the late transcription unit, there is an apparent discrimination between processing signals in a region of overlap between an early transcription unit and the late transcription unit. The early E3 transcription unit is contained within the late transcription unit, overlapping the 3' half of the late L4 sequences including the L4 poly(A) site (see Figure 1). Thus, the same sequences are transcribed into RNA both early and late in infection but there is a dramatic difference in the use of processing signals in these transcripts. Early in infection, E3 transcripts use the E3 splice sites and do not use the L4 poly(A) site whereas late in infection the reverse is true (Bhat and Wold, 1986). We have therefore sought to define the basis for these choices to provide an understanding for the underlying basis of gene control in such complex systems. Our results indicate that in the choice of L4 poly(A) site usage or E3 intron excision, the dominant step is selection of the E3 splice site. The experiments suggest that splicing choice determines poly(A) site choice which has implications for the control of processing of the major late transcripts.

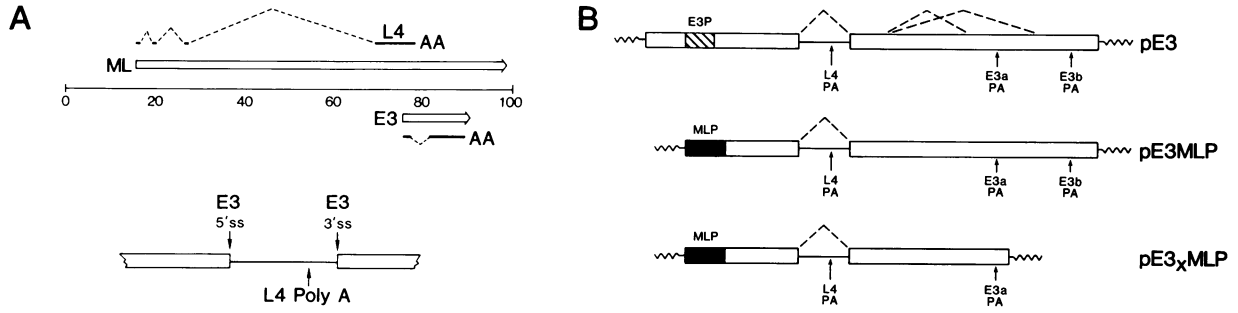

Fig. 1. A. Schematic diagram of the adenovirus genome depicting location of the major late transcription unit (ML), L4 coding sequences and the E3 transcription unit. Numbers 0–100 refer to viral genome map units. The L4 and E3 exons are shown only for general comparison and are not meant to be accurate reflections of all of the mRNA species. For the L4 mRNA the tripartite leader splicing is shown as well as splicing to the L4 mRNA body. E3 splicing is summarized by the inclusion of the major splice present in almost all E3 mRNAs. Shown below is an enlarged view of the first E3 intron shown in Figure 1A. The E3 exons are represented as open boxes with the intron as a line. The L4 poly(A) site is located about midway in this 358 base intron which is spliced out in virtually all E3 messages. **B**. Structure of the pE3 and pE3MLP plasmids. Wavy lines are plasmid sequences, solid or hatched boxes are non-transcribed regions containing the promoter. Arrows indicate poly(A) sites while the major E3 splice (first intron) is shown by the thin line. This splice and the two other E3 splices are represented as dotted lines. pE3 is a *Hin*dIII clone of the entire E3 gene, pE3MLP contains the major late promoter (MLP) substituted for the E3 promoter and pE3$_x$MLP is truncated immediately after the E3a poly(A) site.

Results

Plasmid directed E3 specific processing

To gain insight into the mechanisms which control complex RNA processing events, we have taken advantage of a unique arrangement of transcription units in the adenovirus genome. The major late transcription unit extends from map position 16 to map position 99. Contained within this region and transcribed from the same DNA strand is the early E3 transcription unit (see Figure 1). Although primary transcripts for E3 mRNA and L4 mRNA both contain the L4 poly(A) site, this poly(A) site is only used in the major late transcripts. Likewise, E3-specific splices are only made in E3 transcripts (Bhat and Wold, 1986). How is this decision made? The precursors for these mRNAs differ in several ways. The E3 gene is expressed early after infection and transcription derives from a promoter at 76 map units. The L4 RNA derives from transcription initiated at the major late promoter at map position 16.4 and is only found at late times of infection. Although the late promoter is active early in infection, transcripts terminate prior to the L4 region (Akusjarvi and Persson, 1981; Nevins and Wilson, 1981; Shaw and Ziff, 1981). Thus, one might initially consider two possibilities for the E3-specific processing early and the L4-specific processing late. First, the environment of the cell, in the form of specific *trans*-acting factors, could dictate the observed processing. Perhaps the uninfected cell is incapable of processing at the L4 poly(A) site and to do so requires a factor produced during the infection. In the absence of such a factor, the L4 poly(A) site is not recognized resulting in use of one of the downstream E3 poly(A) sites. Alternatively, the promoter driving transcription might specify the specific processing event in a manner similar to the formation of U1 gene transcripts (Hernandez and Weiner, 1986; Neuman de Vegvar *et al.*, 1986).

To approach these questions, we have constructed a plasmid vector containing the complete E3 gene which thus includes the L4 poly(A) site (Figure 1B). The vector pE3 was introduced into the human 293 cell line by transfection. These cells express the E1A gene (Aiello *et al.*, 1979), allow a normal virus lytic cycle and afford an efficient transfection system (Alwine, 1985). Cytoplasmic RNA was harvested at 50 h post-transfection and analyzed by hybridization to labeled RNA probes followed by T2 digestion and polyacrylamide gel analysis of protected fragments (Favaloro *et al.*, 1980). The normal change in RNA processing in the L4/E3 region is illustrated by analysis of viral RNA isolated from adenovirus infected HeLA cells at 7 and 23 h post-infection as shown in Figure 2. Analysis of RNA with a riboprobe that spans the E3 intron distinguishes between use of the E3 splice and use of the L4 poly(A) site. Use of the L4 poly(A) site will result in a protected band of 420 nt whereas use of the E3 splice sites results in a protected band of 178 nt. At the early time point (7 h), E3 splicing occurs at a high level with only low level use of the L4 poly(A) site whereas at 23 h the opposite result is found with the L4 poly(A) site present in most RNAs measured by this probe. Use of a probe that hybridizes across the L4 poly(A) site and the E3 3′ss gave similar results although the E3 3′ss is used efficiently late since the tripartite leader is frequently spliced to the L5 'y' leader exon which is the same site as the E3 3′ splice site (data not shown). An examination of RNA production from the pE3 plasmid after transfection of 293 cells is shown in Figure 2A. Assays for processing with the 5′ss probe demonstrated that RNA produced from pE3 was efficiently spliced at the E3 splice sites with little evidence for use of the L4 poly(A) site. Thus, the E3 minigene (pE3), expressed in a transfection of 293 cells, appears to be an accurate representation of the expression of the intact viral gene during the early phase of a viral infection.

To test for promoter specificity in directing the subsequent processing, we compared processing of RNA when initiated by the E3 promoter or the major late promoter. The E3 promoter was deleted and replaced with a 300 bp fragment containing the major late promoter (see Materials and methods for details). A 5′ end analysis indicated that RNAs produced from this construct (pE3MLP) initiated at the expected site for major late transcripts (Ziff and Evans, 1978) and transcripts produced from the E3 promoter initiated at the E3 start site (Figure 2B). Analysis of processing of the RNA from these constructs, as shown in Figure 2A, revealed E3 specific splicing, with no difference in the pattern of E3 splice site versus L4 poly(A) site use whether the transcription was driven by the E3 promoter

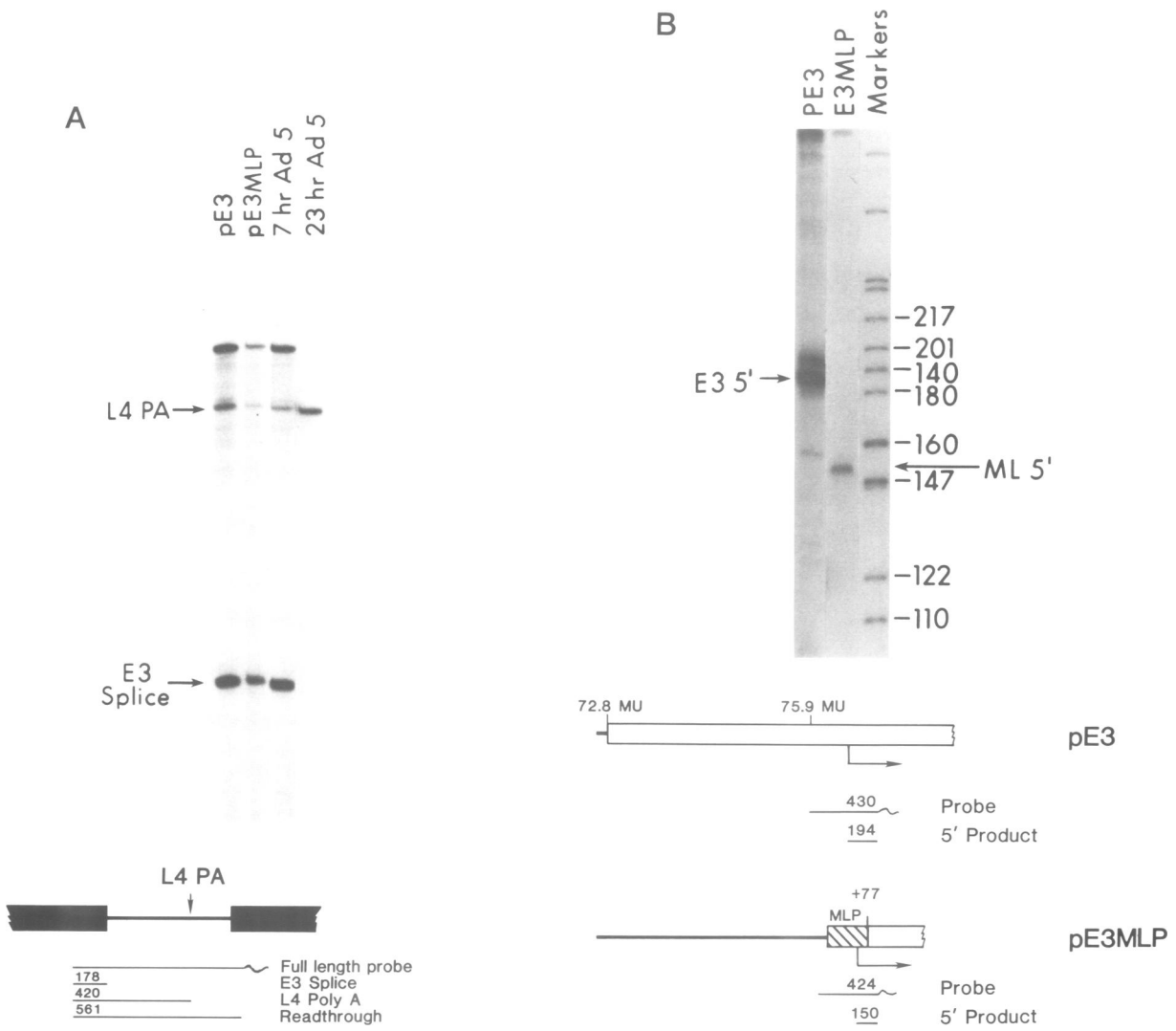

Fig. 2. A. Processing of RNA transcribed from E3 containing plasmids. Cytoplasmic RNA was harvested from 293 cells 50 h after transfection with either pE3 or pE3MLP. For comparison, RNA from Ad5 infected HeLa cells, harvested at either 7 or 23 h, was also analyzed. E3 specific sequences were detected by an RNase protection assay as described in Materials and methods. Shown below is a diagram of the 5'ss probe and the expected protected products. The wavy line at the end of the probe represents plasmid sequence. **B**. Mapping of 5' ends of pE3 and pE3MLP transcripts. Shown at the bottom is a diagram of the promoter region of pE3 and pE3MLP. Open boxes represent viral DNA while lines are plasmid sequence. The MLP as a hatched box is shown fused at nt 77 in the E3 transcriptional unit. A schematic of uniformly labeled RNA probes used in RNase protection analysis is shown as are expected product lengths for correct 5' initiation. The probes for the ML promoter and the E3 promoter are described in Materials and methods (HAP and RAP). Shown above are the results of RNase protection analysis of cytoplasmic RNA isolated from pE3 or pE3MLP transfected 293 cells. Each RNA sample was hybridized to probe prepared from homologous sequence.

or the major late promoter. Thus, it appears that the E3 specific pattern of RNA processing is independent of the promoter which initiates transcription and thus the basis for the switch late in infection to the use of the L4 poly(A) site is not a function of the major late promoter.

To test for the involvement of virus-specific *trans*-acting factors in the switch in processing and the possibility that promoter specificity might only be seen with the appropriate factors, the pE3MLP plasmid was transfected into 293 cells that were either mock infected or were infected with Ad2 for 19 h. The use of Ad2 as the infecting virus allowed us to take advantage of a sequence divergence between Ad2 and Ad5 (pE3 derives from Ad5) so as to employ an end labeled probe for an S1 assay in which the labeled base was at a position of mismatch (Cladaras and Wold, 1985). S1 analysis of RNA from cells infected with Ad2 virus alone was negative (Figure 3), thus indicating that under these assay conditions, Ad2 viral RNA did not interfere. Examination of RNA derived from the transfected pE3MLP plasmid in the late viral infection demonstrated that the plasmid RNA is processed using the E3 splice with no apparent use of the L4 poly(A) site. Therefore, we conclude that neither the nature of the promoter nor a virus induced *trans*-effect are major factors in the regulation of E3/L4 mRNA processing.

E3 processing choices are unaffected by the nature of the poly(A) site

The results of Figure 3 suggest that the production of E3 transcripts from the E3 transcription unit is not due to regulatory *trans* effects or to the nature of the promoter directing transcription initiation. We thus presume that the processing pattern is directed by *cis*-acting signals in the primary transcripts. If the regulating step in this case was

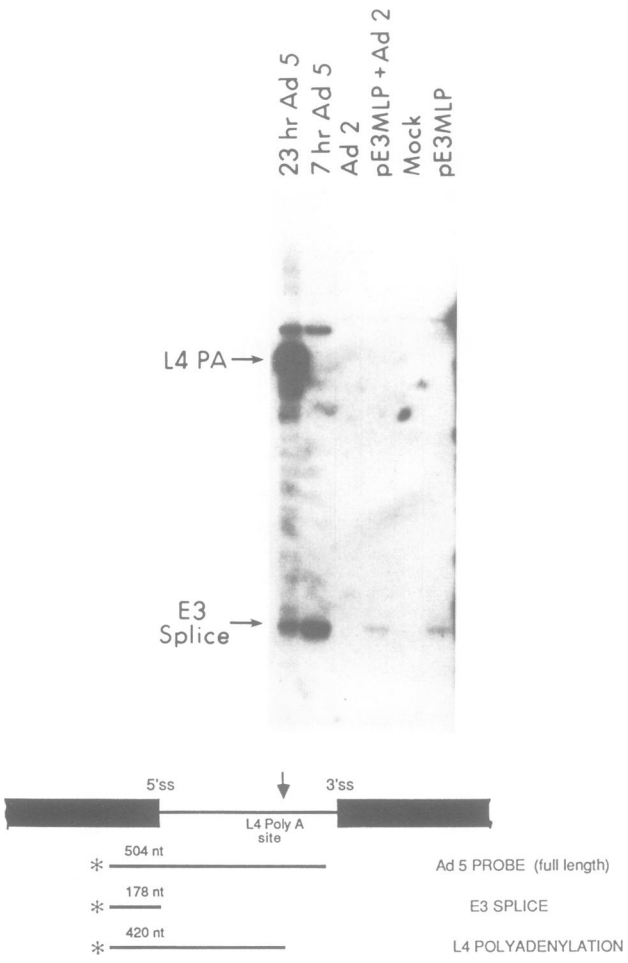

Fig. 3. Effect of viral infection on the processing E3 RNA. 293 cells were transfected with pE3MLP or mock transfected with calf thymus DNA. Transfected cells were either superinfected with Ad2 virus 19 h before harvesting cytoplasmic RNA or were mock infected. S1 analysis with an Ad5 specific end labeled probe, as depicted at the bottom and described in Materials and methods, reveals protected bands corresponding to use of the E3 splice sites or the L4 poly(A) site as labeled. The probe is 3′ end labeled at a position of mismatch between Ad2 and Ad5 and therefore only detects Ad5 RNA under these conditions. **Lane 1**, Ad5 RNA, 23 h post-infection; **lane 2**, Ad5 RNA, 7 h postinfection; **lane 3**, mock transfection, late Ad2 superinfection; **lane 4**, pE3MLP transfection, late Ad2 superinfection; **lane 5**, mock transfection, mock superinfection; **lane 6**, pE3MLP transfection, mock superinfection.

poly(A) site choice, then it would seem likely that there was something special about the L4 or E3 poly(A) sites that allowed their differential usage in the two different primary transcripts. To determine if the E3a poly(A) site was dominant over the L4 poly(A) site in the context of E3 transcription, the L4 poly(A) site was deleted from pE3MLP and replaced by the E3a poly(A) site as depicted in Figure 4 to generate the plasmid pSPA. In addition to the SPA plasmid, several other constructs were made as controls for changes in E3 intron length. RSPA contains the E3a poly(A) site inserted in the E3 intron in the reverse orientation while the plasmid pΔPA simply has deleted the L4 poly(A) site from the intron. As a first step in RNA analysis, the 5′ss probe that is homologous to the wild-type E3 gene was hybridized to poly(A)$^+$ RNA harvested after transfections with each construct shown diagrammatically in Figure 4A. This probe differentiates between transcripts in which the E3 5′ splice has been made and those transcripts which are not processed at the splice site (readthrough). When used to assay the wild-type RNA, this probe also detects use of the L4 poly(A) site. As a control for transfection efficiency, the plasmid pSV-Glo was co-transfected and assayed for expression with a β globin specific probe (Costa *et al.*, 1986). This experiment thus assays for the relative use of the E3 splice site in these various constructions. If a poly(A) site inserted in the intron was efficiently utilized, then we would expect a reduction in the use of the E3 splice site. However, as can be seen in Figure 4B, virtually all RNAs transcribed from each plasmid contained the E3 splice (as measured by 5′ss usage).

An assay was also done with homologous probes (5′ss SPA and 5′ss RSPA) which hybridize like the wild-type 5′ss and through the entire intron of RNA produced from the SPA and RSPA plasmids (Figure 4C). This assay directly measures use of the inserted poly(A) site in each transcript. Like the wild-type 5′ss probe when used with wild-type E3 minigene RNA, three sizes of protected fragments are expected which correspond to RNA with the E3 splice (247 nt), RNA cleaved at the intron poly(A) site (577 nt) and readthrough RNA (700 nt). As depicted in Figure 4D, there was no evidence of a 577 nt band corresponding to use of the E3a poly(A) site when it was inserted in the intron in place of the L4 poly(A) site. Additionally, a band of 125 bases corresponding to the truncated E3a poly(A) site in its wild-type position is expected when the probe with the sense insert is used. It is clear that aside from this band the plasmid with the sense and antisense E3a poly(A) site inserts give identical RNase protection products. This was also true using homologous probes that spanned the 3′ss (Figure 4D). Thus, by each assay there was no evidence for use of the inserted poly(A) site.

It was apparent that the E3a poly(A) site in the E3 intron was not functioning. To ensure that we had cloned the entire poly(A) site and that it had the potential to function, we inserted this fragment into a different region of the body of the pE3$_x$MLP minigene. This construct pExSPA contains the duplicated E3a poly(A) site in an exon ~300 nt downstream of the first E3 intron with the L4 poly(A) site still present in the E3 intron (Figure 5A). We assayed RNA from this construct with a probe that would differentiate the original E3a poly(A) site from the duplicated E3a poly(A) site in order to determine the efficiency of use of the E3a poly(A) site inserted into the E3 exon. Hybridization to an RNA probe followed by T2 digestion revealed a protected fragment of 350 nt corresponding to the original E3a site (used or unused) which is present in all clones (Figure 5B). A band of 214 nt, seen only in the ExSPA lane, is due to readthrough of the inserted poly(A) site while a band at 127 nt is a measure of the use of the inserted poly(A) site. From these results, it is evident that the E3a poly(A) site is used ~30% of the time which is the same level of usage as the original E3a site (data not shown). Thus, these results indicate that the E3a poly(A) site element as we define it is functional. However, it is not used when inserted in the E3 intron but rather the E3 splice is favored. Thus, poly(A) site selection appears to be secondary to E3 splice site selection.

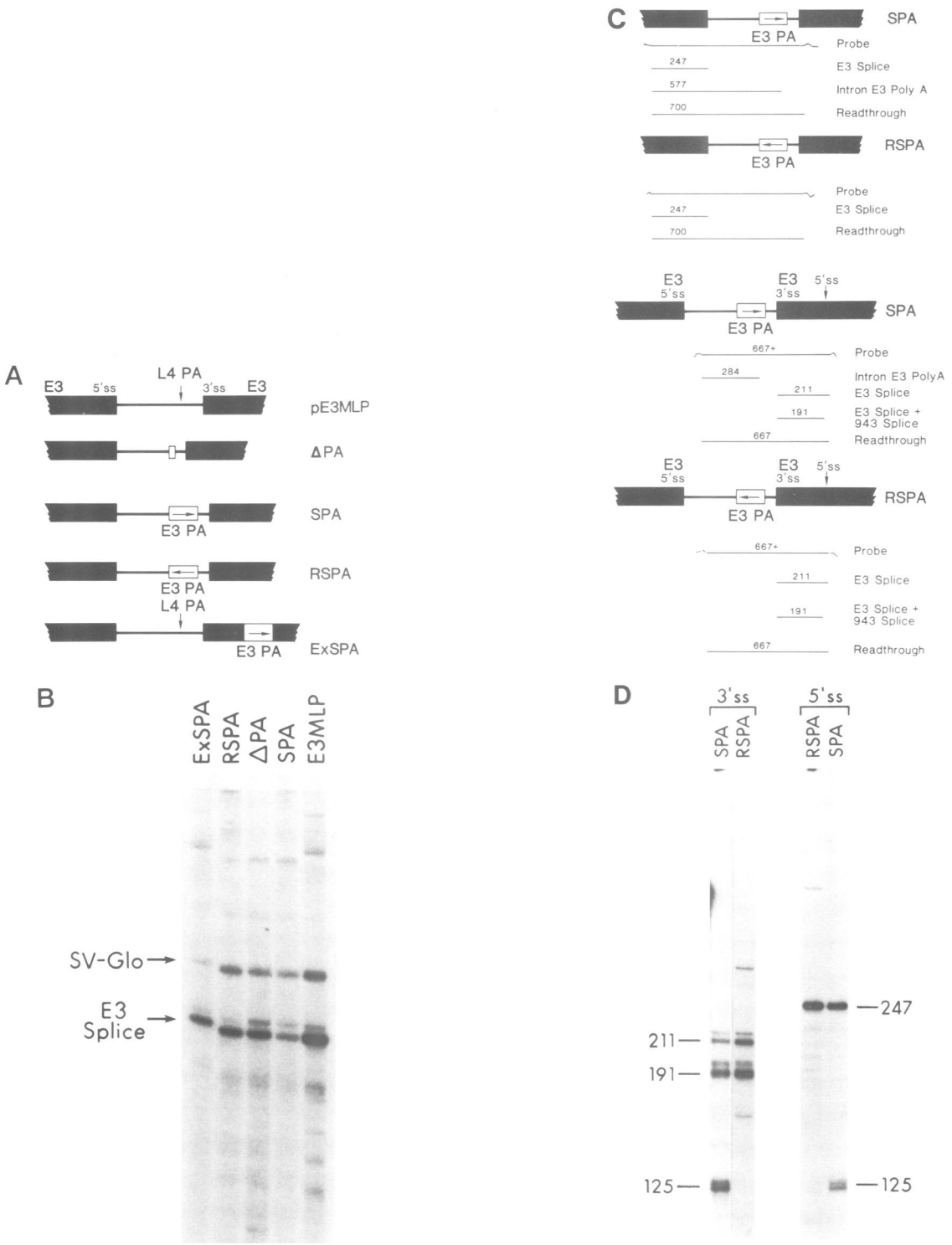

Fig. 4. A. Schematic diagram of the structure of the poly(A) site substitution plasmids. The diagram is a close up of the first E3 intron. Exon sequences are represented as black boxes while the intron is a line. ΔPA is a deletion of the L4 poly(A) site with a *SalI* linker inserted at the site. SPA contains the E3a poly(A) site inserted in place of the L4 poly(A) site. RSPA contains the E3a poly(A) site in place of the L4 poly(A) site but inserted in the reverse orientation. ExSPA contains the E3a poly(A) site inserted at an *XbaI* site in the E3 exon and retains the L4 poly(A) site. To simplify cloning, the vector used in these experiments was a pE3MLP minigene truncated just past the E3a poly(A) site. This plasmid ($pE3_xMLP$) produces transcripts which do not use the L4 poly(A) site while the E3a polyadenylated RNA is present at a high level, similar to the parent plasmid. **B**. RNase protection analysis of poly(A)$^+$ RNA from cells transfected with the substituted poly(A) site variants. As shown below, the probe is the 5′ss probe from the WT pE3 sequence. Assay of β-globin RNA from pSV-Glo in the same samples serves as an internal control for transfection efficiency. The SV-Glo probe has been described (Costa *et al.*, 1986). **C**. Schematic diagram of probes used to assay the two intron poly(A) site substitution mutants SPA and RSPA which contain the E3A poly(A) site in the sense and antisense orientation. Homologous probes which hybridize across both the 5′ss and the intron and the 3′ss and intron are used in the RNase protection assay. These are shown diagrammatically along with the expected protected probe fragments. The probes are described in Materials and methods. **D**. Results of RNase protection analysis using homologous probes described above. On the left are results with the 3′ splice site probes and on the right those with the 5′ splice site probes. The 125 nt protection product is representative of use of the normal E3a poly(A) site resulting from hybridization of the probe to sequences at the 3′ end of the transcript.

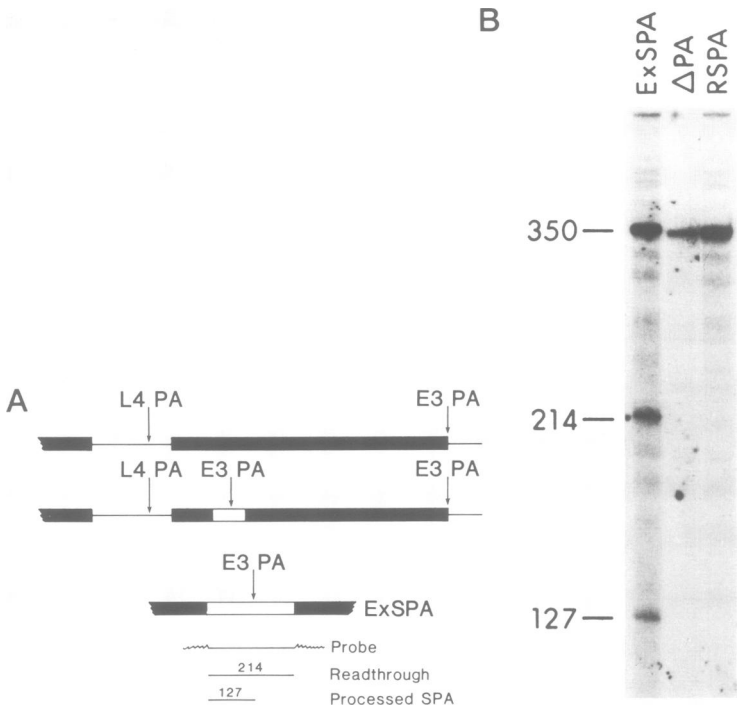

Fig. 5. Use of the E2a poly(A) site when inserted in an exon downstream of the E3 splice/L4 poly(A) site. **A**. Schematic diagram of probe used to assay poly(A) site use from pExSPA. The probe, pSP, is homologous to the wild-type E3A poly(A) site and so potentially can hybridize to two places on one molecule. Shown are the sizes of protected probe fragments depending on poly(A) site usage of the exon inserted E3A poly(A) site. **B**. RNase protection analysis to determine usage of the extra E3A poly(A) site inserted into the E3 exon of EXSPA.

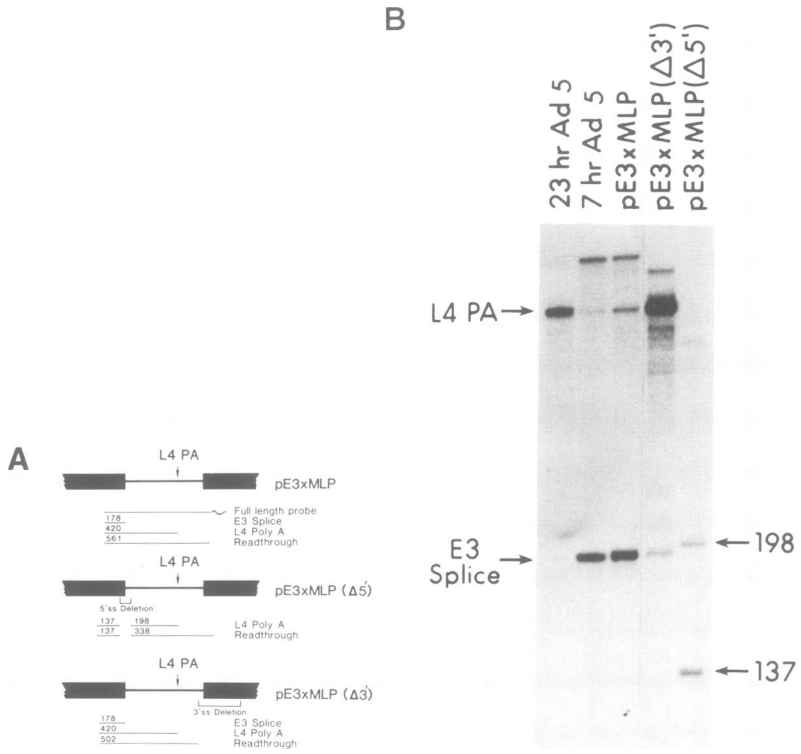

Fig. 6. Alteration of E3 splicing allows use of the L4 poly(A) site. **A**. Close-up of the first E3 intron in plasmid pE3xMLP and also the variants with the 5′ss deleted (Δ5′) and the 3′ss deleted (Δ3′). RNase protection analysis was performed with the same 5′ss probe used in experiments outlined in Figure 2A. Sizes of protected fragments and the RNAs to which they correspond are shown. **B**. RNase protection analysis of poly(A)$^+$ cytoplasmic RNA from transfection with pE3$_x$MLP, pE3$_x$MLP (Δ3′) and pE3$_x$MLP (Δ5′). Shown for comparison are assays of early and late Ad5 RNA.

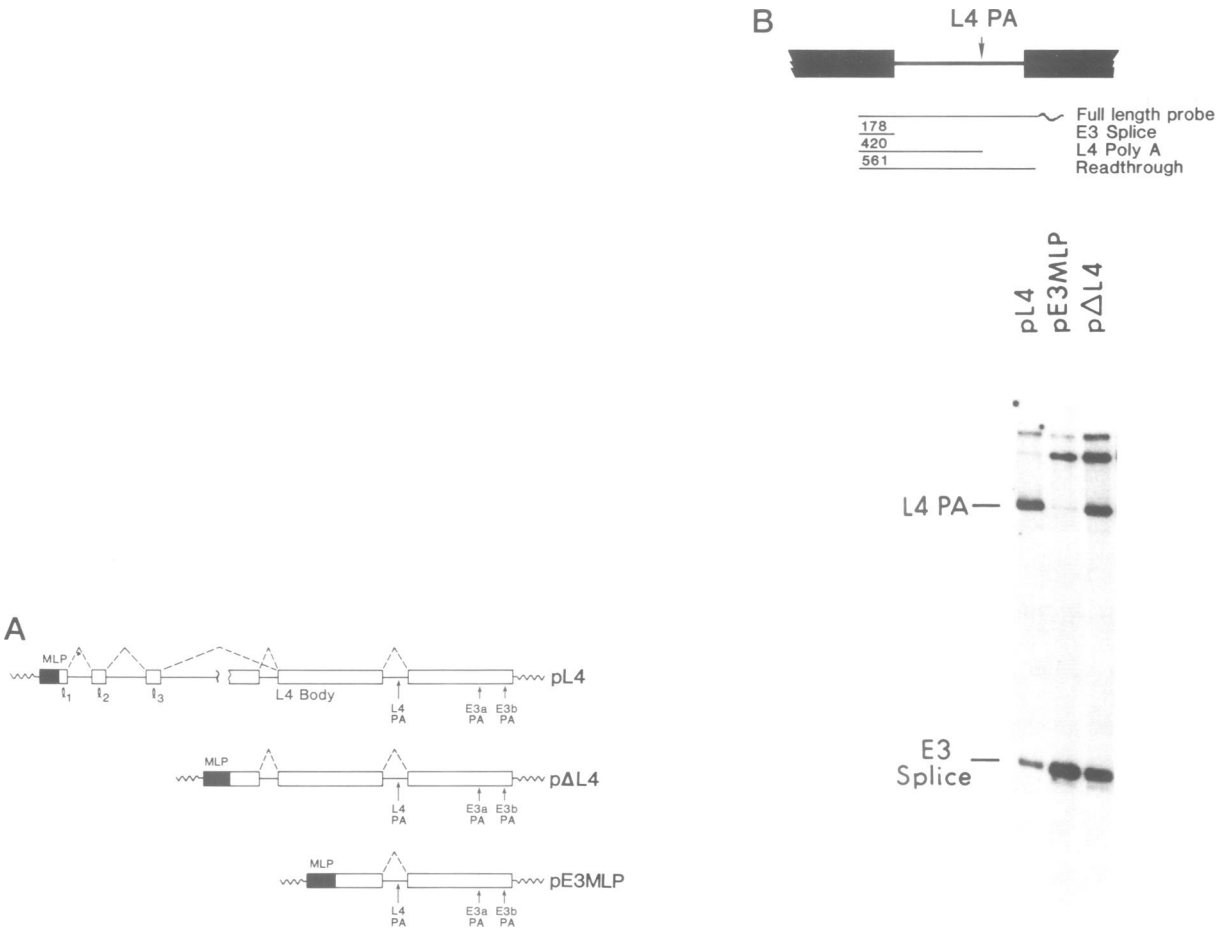

Fig. 7. Effect of L4 splicing signals on use of the L4 poly(A) site. **A**. Schematic diagram of the pL4 and pΔL4 plasmids as compared to pE3MLP. The filled boxes represent the major late promoter, open boxes labeled as l1, l2 and l3 are the tripartite leader sequences and dashed lines are splicing pathways. Introns are denoted as thin lines and exons as open boxes. Wavy lines are plasmid sequences. In the pL4 plasmid, the viral genome sequences between a *Bst*XI site at 27 map units and the *Hin*dIII site at 73 map units have been deleted (indicated by the break between the intron after the 13 sequence and the L4 body). For each plasmid, the E3 promoter has been deleted. **B**. RNase protection assays for processing of pL4, pE3MLP or pΔL4 transcripts at the E3 splice sites and the L4 poly(A) site using the same 5′ss probe described in Figure 2B.

Alteration of E3 splicing enables use of the L4 poly(A) site

The above results indicate that the location of the L4 poly(A) site is important in its regulation in the context of E3 transcription and suggests that E3 splicing dominates over the use of the poly(A) site. To determine more directly the influence of the E3 splice sites on use of the L4 poly(A) site, we assayed RNA transcripts from plasmids in which the 5′ss or the 3′ss of the E3 intron was deleted (Figure 6A). These deletions were 83 nt for the 5′ss deletion and 332 nt for the 3′ss deletion. RNA analysis after transfection of these plasmids into 293 cells revealed that deletion of either splice site from the $pE3_xMLP$ minigene resulted in inefficient use of the companion splice site. No evidence of protected bands corresponding to messages using the remaining E3 5′ss or 3′ss in RNA from the splice site deletion mutants was seen; cryptic splice sites were apparently used only at very low levels if at all. As depicted in Figure 6B, the inactivation of E3 splicing resulted in the production of transcripts that used the L4 poly(A) site. This was true for either the 5′ splice site deletion or the 3′ splice site deletion. RNA bands of 198 nt for the 5′ss deletion and 420 nt for the 3′ss deletion were the prominent RNase protection products, indicative of efficient L4 poly(A) site usage. Thus the L4 poly(A) site in the wild-type pE3 transcription unit is functional but appears to be inefficiently used due to its location.

Transcripts derived from a major late minigene use the L4 poly(A) site

Although the L4 poly(A) site is not utilized efficiently in the context of E3 transcription, it is used in a late viral infection and dominates over the use of the E3 splice sites (see Figure 2). Neither the environment of the late infected cell nor the identity of the promoter appears to influence this selection, as seen from the results of Figure 3. The other major difference between early and late infection is the nature of the L4 and E3 primary transcripts. The L4 specific RNA initiates at map position 16.4 and thus includes ~20 kb of sequence 5′ to the E3/L4 sequences. Perhaps this additional sequence influences use of E3 splice sites and the L4 poly(A) site. To pursue this possibility, an L4 minigene was constructed comprising a viral DNA clone of the major late promoter and tripartite leader with introns fused to a region containing one out of the four possible intact L4 mRNA

bodies. The E3 promoter is deleted to insure that all RNA precursors initiate at the major late promoter. The transcripts from this clone can theoretically make leader to leader exon splices, leader 3 to L4 body splices, and leader 3 to L5 body splices (Sharp, 1984). In addition, there is an intron within the L4 body region which could be utilized in this construct. 293 cells were transfected with this L4 minigene (pL4) and with the E3 minigene (pE3MLP). Analysis of splicing of the transcripts generated by this L4 minigene revealed that leader 3 was spliced to an upstream exon (probably leader 2) quite efficiently, leader 3 was spliced to a downstream exon (L4 and/or L5 exons) less efficiently and splicing within the L4 exon region was also apparent in most transcripts, both from pL4 as well as late viral RNA (data not shown).

As shown in Figure 7, RNA analysis with the E3 5′ss probe revealed that the L4 poly(A) site was used quite efficiently in transcripts derived from the L4 minigene. This is in sharp contrast to the processing of these sequences derived from pE3MLP where the L4 poly(A) site is used infrequently. Although the ratio of L4 poly(A) site usage to E3 splicing in the L4 minigene was not identical to late viral RNA, it was much higher than for the E3 minigene. Thus, the addition of the tripartite leader and L4 exon sequences apparently was able to influence the use of the L4 poly(A) site despite the presence of the E3 splicing signals.

One additional construct was made which deletes the tripartite leader exons but left the L4 exon region (pΔL4). The major late promoter was fused to the *Hin*dIII site at 72.8 map units. This construct would allow synthesis of an L4 RNA lacking the tripartite leader and thus devoid of a leader−L4 body splice. However, the splice in the L4 exon is still made in >90% of the transcripts from this construct. RNA analysis using the E3 5′ss probe revealed that although the E3 splice/L4 poly(A) choice is increased in favor of the E3, the L4 poly(A) site was still used about half the time (Figure 7). Thus, the absence of splicing of the tripartite leader sequences appears to diminish the use of the L4 poly(A) site, but processing of these transcripts, compared to pE3 transcripts, still favored the L4 poly(A) site over the E3 splice sites. Although the transcripts from these plasmids differ in ways other than splicing choices, we feel the most likely explanation for the enhanced use of the L4 poly(A) site is the ability of a splice to be made prior to the E3 intron.

Discussion

The expression of complex transcription units is characterized by the potential for alternative splicing choices and alternative poly(A) site choices (see Nevins, 1983 and Leff *et al.*, 1986 for reviews). When the final output of such transcription units can be subject to regulation, it becomes of critical importance to define the exact event of mRNA processing at which the regulatory control is directed. It is clear from studies of a variety of genes that 3′ end formation and poly(A) addition can precede splicing since poly(A)-containing unspliced RNAs can be detected (Nevins, 1984 for review). A more definitive study, employing the adenovirus major late transcription unit, was possible since the rate of synthesis was sufficiently high to allow analysis of very briefly labeled RNA. It was found that newly synthesized poly(A) containing RNA was largely unspliced (Nevins and Darnell, 1978; Nevins, 1979). Thus, it would appear that 3′ end formation and poly(A) addition are early events in mRNA processing with splicing generally occurring later. These results would suggest that in those cases where alternative splicing is coupled with alternative poly(A) site selection that the choice of the poly(A) site would be the critical event. This indeed appears to be true for the immunoglobulin μ heavy chain transcription unit. The $\mu_{secreted}$ (μ_s) and $\mu_{membrane}$ (μ_m) mRNAs are characterized by distinct poly(A) sites and distinct splicing patterns. Analysis of mutations that affect the ratio of the μ_s to μ_m RNA indicated that the poly(A) sites were the primary determinants of the final mRNA levels (Galli *et al.*, 1988).

However, despite the fact that 3′ end formation and poly(A) addition precede splicing, it is also possible to imagine that a splice site choice could dictate poly(A) site selection even though the actual splicing event occurs later. For instance, it is known that *in vitro* spliceosome assembly is quite rapid and occurs before the actual splicing event (Padgett *et al.*, 1986 for review). Also, electron microscope analyses have shown *in vivo* complex formation on nascent mRNAs (Osheim *et al.*, 1985). These results suggest that unspliced non-polyadenylated RNA could be complexed with the splicing machinery prior to polyadenylation. Thus, one might speculate that in some cases a splicing choice is made before the poly(A) site is utilized but the actual splicing event (cleavage and ligation) occurs only later. The results presented here suggest that for the adenovirus E3 and L4 region just that sort of situation does occur. The mere presence of the L4 poly(A) site in a transcript did not insure its use, even when included in an RNA initiated at the major late promoter and synthesized in the environment of a late infected cell. Rather, it appears likely that the use of the L4 poly(A) site apparently requires the presence and use of L4 splicing signals upstream.

The possibility of a splice site choice dominating over poly(A) site selection has been demonstrated in the calcitonin/CGRP genes. Although in that system polyadenylation of the primary transcript occurs prior to the splicing event, a commitment to make a certain splice occurs first and is dominant. For the calcitonin/CGRP system it is easy to envisage how choosing one splicing path can rule out selection of a poly(A) site since the regulated splice site and the poly(A) site are virtually overlapping. Thus, once the splicing machinery forms at the splice site, cleavage and polyadenylation at the calcitonin mRNA poly(A) site could be physically blocked. Like the calcitonin/CGRP system, kinetic data show that the E3 transcripts are polyadenylated prior to executing the major E3 splice (Sittler *et al.*, 1986). However, unlike the calcitonin/CGRP system, the L4 poly(A) site does not overlap with the E3 splice site and thus one must presume that once spliceosome formation has occurred, the intron is committed to be spliced out and the poly(A) site is unavailable. Therefore, our finding that a primary transcript containing upstream L4 sequences does not make the E3 splice efficiently suggests that splicing decisions in an RNA might be interactive, with the commitment to make one splice having an effect on whether another splice is made.

How do these results bear on the control of viral gene expression during an early and late infection? The E3 transcription unit, which includes the L4 poly(A) site, is transcribed early but not late. The results presented here indicate that the exclusive production of E3 mRNA and the

failure to utilize the L4 poly(A) site is due to initiation at the E3 promoter which excludes L4 splice signals. Thus, the E3 splice sites are recognized, and the downstream E3 poly(A) sites are used. The L4 poly(A) site in the intron is not used and is eventually spliced out. Transcripts do initiate at the major late promoter during the early phase of infection, but they terminate prior to reaching the L4 poly(A) site (Nevins and Wilson, 1981). Later in infection, there is a large increase in transcription from the major late promoter and transcription proceeds to near the end of the genome (Fraser *et al.*, 1979). At that time, recognition of the L4 splice sites leads to use of the L4 poly(A) site precluding use of the E3 splice sites. The important suggestion from this work is that the use of the L4 poly(A) site may depend on recognition of the L4 splice sites. If this is true for each of the major late transcripts, then this would imply that splicing choices dictate late gene expression. Poly(A) sites in introns that were committed to be spliced out would not be seen in that particular pre-mRNA molecule. In actuality such a scenario would simplify the process of late gene expression. If poly(A) site choice was the primary event, then one might imagine the need for two elements of selection, first the poly(A) site and then the splice site to determine the mRNA in the 3′ co-terminal group. If, however, splice site choice dictated the poly(A) site, then this one selection would be sufficient. Once the splice site choice was made the first poly(A) site in the downstream exon would be used.

Finally, it appears that such processing selections are subject to regulation. As stated above, the major late transcription unit is expressed early although transcription only includes the L1, L2 and L3 region (Nevins and Wilson, 1981). However, despite the transcription of L1−L3, a single L1 mRNA is the predominant product, representing >50% of the nuclear transcripts (Chow *et al.*, 1979; Shaw and Ziff, 1980; Akusjarvi and Persson, 1981; Nevins and Wilson, 1981). Late in infection, the relative production of the L1 RNA declines significantly, representing only a small percentage of the primary transcripts. One could envisage several scenarios to account for these changes. For instance, this could reflect the appearance of new factors late in infection that favor the downstream splice sites. Alternatively, there could be a single factor for which all splice sites compete. Perhaps early in infection L1 predominates because the factor is not limiting and L1 is seen first. Late in infection, the large number of transcripts may render such a factor limiting allowing the other sites to compete. Regardless of the actual mechanism, the real value of the experiments described here is to focus the thinking with respect to such regulation on splice site choices.

Materials and methods

Plasmids

Restriction sites in Ad5 sequences are identified by genomic map position or nucleotide number in the E3 gene. pE3, the E3 minigene, was constructed by subcloning an Ad5 *Hin*dIII fragment (map unit 73.6−85.0) into the *Hin*dIII site of pUC8 such that the cloning site would be upstream of the E3 gene. pE3MLP, the E3 minigene with the major late promoter (MLP) substituted for the E3 promoter, was constructed in three steps. A *Hin*dIII/*Xho*I fragment comprising the left half of the E3 gene was cloned into the *Hin*dIII/*Xho*I site of pUC8-X to form pE3$_x$. pUC8-X is a derivative of pUC8 that contains an *Eco*RI site converted to an *Xho*I site (gift of B.Cohen). The MLP was obtained as a 306 base long *Xho*I−*Rsa*I Ad5 fragment cloned into the *Sal*I−*Hin*cII sites of pUC13 (gift of G.Galli). The MLP, as a *Hin*dIII−*Pvu*II fragment, was inserted into the *Hin*dIII−*Nae*I (+77) sites of pE3$_x$t. This MLP fragment devoid of splice sites replaces the E3 promoter and further upstream sequences of pE3$_x$. This clone pE3$_x$MLP is made into a full length E3 gene analogue by inserting the *Hin*dIII−*Xho*I fragment from this plasmid, filled in at the *Hin*dIII site, into the *Bam*HI−*Xho*I sites in the pE3 plasmid (the *Bam*HI site being blunted). This replaces the wild-type 5′ half of the pE3 clone with the MLP driven version of the E3 gene. This plasmid is designated pE3MLP.

L4 poly(A) site substitutions were all made in pE3$_x$MLP derivatives. To form pΔPA, pE3$_x$MLP was digested with *Bgl*II (+569) and *Mbo*II (+646), blunted, and a *Sal*I linker inserted prior to self ligation. This plasmid, pΔPA, contains no L4 poly(A) site and instead has a *Sal*I linker at the site of the deletion. From pE3, a blunted fragment *Fnu*DII (+2099)−*Acc*I (+2303) containing the E3a poly(A) site was cloned into the filled in *Sal*I site of pΔPA. This substitutes the E3a poly(A) site for the L4 poly(A) site in both orientations (pSPA: sense and pRSPA: antisense). The control plasmid, pExSPA, with the E3a poly(A) site inserted into an E3 exon, contains the blunted *Fnu*DII−*Acc*I fragment inserted into the filled in *Xba*I (+1029) site in pE3$_x$MLP.

E3 splice site mutants were both made in pE3$_x$MLP. pE3$_x$MLP (Δ5′) was constructed in two steps. First a *Pst*I (+121)/*Bst*NI (+330) fragment from pE3$_x$MLP, filled in at the *Bst*NI site, was inserted into the *Pst*I (+121) and *Apa*I site (+414) sites (the *Apa*I site blunted) of pE3$_x$MLP. This effectively deletes the sequence between *Bst*NI (+330)−*Apa*I (+414) which includes the E3 5′ splice site and also deletes the *Pst*I fragment containing the MLP. The MLP was restored by inserting a *Hin*dIII−*Tth*III (+125) fragment from pE3$_x$MLP blunted at the *Tth*III (+125) site into the *Hin*dIII and blunted *Tth*III (+125) sites of this clone to create pE3$_x$MLP (Δ5′). pE3$_x$MLP (Δ3′) was formed by inserting the *Hin*dIII−*Rsa*I (+695) fragment of pE3$_x$MLP into the *Hin*dIII and blunted *Xba*I (+1028) sites of pE3$_x$MLP. This deletes the sequences between the *Rsa*I−*Xba*I sites which includes the E3 3′ splice site and 290 bases downstream.

The L4 minigene was constructed in two steps. First the E3 promoter from pE3 was deleted by removing a *Sma*I/*Sma*I fragment (map units 75.9−76.8) of pE3. Next, an *Sph*I (map unit 14.7)−*Bst*XI (map unit 27.0) fragment was derived from the Ad5 subclone 46, an Ad5 clone of map unit 0−42 in pBR322 (gift of P.Freimuth). The fragment was blunted and cloned into the filled-in *Bam*HI site in the polylinker upstream of the E3 gene in pE3. This creates an L4 minigene (pL4) with the MLP, all tripartite leader exons and introns, and the entire body of at least one out of four L4 RNAs. There is an 11 kb deletion of the L4 intron in this plasmid which removes completely the bodies of the L1, L2 and L3 mRNAs.

To construct an L4 minigene with no leader splicing nor leader exon-mRNA body splicing, the *Hin*dIII−*Pvu*II fragment containing the MLP was inserted into the blunted *Bam*HI site of the polylinker in pE3. This effectively deletes the second and third leader exons and all leader splice sites. This clone (pΔL4) contains a gene with the MLP fused to the body of one L4 gene.

Cell culture transfections

293 cells were grown in Dulbecco's Modified Eagle's Medium with 10% fetal bovine serum. Cells were plated at 5 × 10^6 cells/100 mm plate 12−36 h prior to transfection. Monolayer cultures were transfected with 20 μg plasmid DNA (plus 1 μg control DNA when included) using the $CaPO_4$ coprecipitation method (Graham and van der Eb, 1973), followed at 4−5 h by a glycerol shock (Parker and Stark, 1979). Cytoplasmic RNA was harvested 48−50 h post-transfection essentially by the method of Favaloro *et al.* (1980). When indicated, poly(A)$^+$ RNA was selected by oligo(dT) cellulose chromatography. For superinfection experiments, 293 cells were transfected as described above and 31 h post-transfection, the cells were infected with 20 p.f.u./cell of Ad2. Cytoplasmic RNA was harvested at 50 h.

Nuclease S1 analysis

The 3′ end labeled double stranded probe for examination of E3 5′ss usage and L4 poly(A) site usage in infected cells (Figure 3) was prepared as follows. An *Acc*I (+191)−*Rsa*I (+695) fragment from pE3 was gel isolated and then 3′ end labeled with [α-^{32}P]dGTP (Amersham) plus cold dATP and the Klenow fragment of *Escherichia coli* DNA polymerase (Boehringer-Mannheim). Hybridization to cytoplasmic RNA was performed at two concentrations of probe giving identical results to insure that there was an excess of probe. Hybridization was overnight at 62°C with S1 digestion at 35°C (Favaloro *et al.*, 1980). Protected fragments were separated on a 6% denaturing polyacrylamide gel. The 3′ labeled nucleotides are Ad5 specific and are lost in hybrids with Ad2 when digested under these conditions.

RNase protection assays

Riboprobe assays followed the method of Hart *et al.* (1985). All probes were cloned in BlueScript (Stratagene) and were synthesized *in vitro* using [α-^{32}P]UTP and T7 or T3 polymerase accordingly. Probes contain additional sequences from that shown in figures (usually 10–40 bases from the cloning region).

Two 5′ end probes were made. To make RAP the *Eco*RI (−237)−*Pst*I (+121) fragment from pE3 was cloned into the *Eco*RI and *Pst*I sites of BlueScript. This clone was cut with *Xho*I prior to probe synthesis. HAP was constructed by inserting the *Hin*dIII−*Pst*I (+121) fragment from pE3MLP into the *Hin*dIII and *Pst*I sites of BlueScript. This clone was also digested with *Xho*I prior to probe synthesis.

The E3 5′ss and L4 poly(A) site probe was made from a plasmid containing the *Acc*I (+193)−*Acc*I (+752) fragment from pE3, blunted and cloned into the *Sma*I site of BlueScript. The plasmid was digested with *Xho*I prior to probe synthesis.

Plasmids to make probes to examine intron poly(A) site usage in E3 splice intron mutants were synthesized from the parent plasmids. SPA-3′ contains an *Apa*I (+412)−*Mst*I (+942) fragment from pSPA inserted into the *Apa*I−*Sma*I sites of BlueScript. This plasmid was cut with *Pvu*II prior to probe synthesis to give a template without a 3′ overhang. SPA-5′ contains the *Hin*dIII−*Rsa*I (+695) fragment from pSPA inserted into the *Hin*dIII−*Rsa*I sites of BlueScript. The plasmid was cut with *Tth*III (+115) and the end filled in prior to SPA-5′ probe synthesis. Probes for pSPA-R were synthesized analogously.

The probe to examine the E3a poly(A) site usage in pExSPA was from a plasmid pSP containing the *Sac*II (+1881)−*Pst*I (+2372) fragment from pE3 cloned into BlueScribe. Digestion was with *Pvu*II prior to probe synthesis.

Acknowledgments

G.A. is an ACS postdoctoral fellow. This work was supported by a grant from the NIH (GM-35894).

References

Aiello,L., Guiloyle,R., Huebner,K. and Weinmann,R. (1979) *Virology*, **94**, 460–469.
Akusjarvi,G. and Persson,H. (1981) *Nature*, **292**, 420–426.
Alt,F.W., Bothwell,A.L.M., Knapp,M., Siden,E., Mather,E., Koshland,M. and Baltimore,D. (1980) *Cell*, **20**, 293–301.
Alwine,J.C. (1985) *Mol. Cell. Biol.*, **5**, 1034–1042.
Amara,S.G., Evans,R.M. and Rosenfeld,M.G. (1984) *Mol. Cell. Biol.*, **4**, 2151–2160.
Bhat,B.H. and Wold,W.S. (1986) *J. Virol.*, **60**, 54–63.
Chow,L.T., Broker,T.R. and Lewis,J.B. (1979) *J. Mol. Biol.*, **134**, 265–303.
Cladaras,C. and Wold,W.S. (1985) *Virology*, **140**, 28–43.
Costa,R., Lai,E. and Darnell,J. (1986) *Mol. Cell. Biol.*, **6**, 4697–4708.
Early,P., Rogers,J., Davis,M., Calame,K., Bond,M., Wall,R. and Hood,L. (1980) *Cell*, **20**, 313–319.
Favaloro,J., Treisman,R. and Kamen,R. (1980) *Methods Enzymol.*, **65**, 718–749.
Fraser,N.W., Nevins,J.R., Ziff,E. and Darnell,J.E. (1979) *J. Mol. Biol.*, **129**, 643–656.
Galli,G., Guise,J.W., McDevitt,M.A., Tucker,P.W. and Nevins,J.R. (1987) *Genes Dev.*, **1**, 471–481.
Galli,G., Guise,J., Tucker,P.W. and Nevins,J.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2439–2443.
Graham,F.L. and Van der Eb,A.J. (1973) *Virology*, **52**, 456–467.
Hart,R.P., McDevitt,M.A., Ali,H. and Nevins,J.R. (1985) *Mol. Cell. Biol.*, **5**, 2975–2983.
Hernandez,N. and Weiner,A.M. (1986) *Cell*, **47**, 249–258.
Leff,S.E., Rosenfeld,M.G. and Evans,R.M. (1986) *Annu. Rev. Biochem.*, **55**, 1091–1117.
Leff,S.E., Evans,R.M. and Rosenfeld,M.G. (1987) *Cell*, **48**, 517–524.
Neuman de Vegvar,H.E., Lund,E. and Dahlberg,J.E. (1986) *Cell*, **47**, 259–266.
Nevins,J.R. (1979) *J. Mol. Biol.*, **130**, 493–506.
Nevins,J.R. (1983) *Annu. Rev. Biochem.*, **52**, 441–466.
Nevins,J.R. (1984) In Apirion,D. (ed.), *Processing of RNA*. CRC Press, Boca Raton, Florida, pp. 133–150.
Nevins,J.R. and Chen-Kiang,S. (1981) *Adv. Virus Res.*, **26**, 1–35.
Nevins,J.R. and Darnell,J.E. (1978) *Cell*, **15**, 1477–1493.
Nevins,J.R. and Wilson,M.C. (1981) *Nature*, **290**, 113–118.
Osheim,Y.N., Miller,O.L.,Jr and Beyer,A.L. (1985) *Cell*, **43**, 143–151.
Padgett,R.A., Grabowski,P.J., Konarska,M.M., Seiler,S. and Sharp,P.A. (1986) *Annu. Rev. Biochem.*, **55**, 1119–1150.
Parker,G. and Stark,G. (1979) *J. Virol.*, **31**, 360–369.
Rogers,J., Early,P., Carter,C., Calame,K., Bond,M., Hood,L. and Wall,R. (1980) *Cell*, **20**, 303–312.
Sharp,P.A. (1984) In Ginsberg,H. (ed.), *The Adenoviruses*. Plenum Press, New York.
Shaw,A.R. and Ziff,E.B. (1980) *Cell*, **22**, 905–916.
Sittler,A., Gallinaro,H. and Jacob,M. (1986) *Nucleic Acids Res.*, **14**, 1187–1207.
Ziff,E.B. and Evans,R.M. (1978) *Cell*, **15**, 1463–1475.

Received on February 29, 1988; revised on April 5, 1988