MOLECULAR AND CELLULAR BIOLOGY, Nov. 1985, p. 2975–2983
0270-7306/85/112975-09$02.00/0

Vol. 5, No. 11

Definition of Essential Sequences and Functional Equivalence of Elements Downstream of the Adenovirus E2A and the Early Simian Virus 40 Polyadenylation Sites

RONALD P. HART, MICHAEL A. McDEVITT, HASEEB ALI, AND JOSEPH R. NEVINS*
The Rockefeller University, New York, New York 10021

Received 24 January 1985/Accepted 26 July 1985

In addition to the highly conserved AATAAA sequence, there is a requirement for specific sequences downstream of polyadenylic acid [poly(A)] cleavage sites to generate correct mRNA 3′ termini. Previous experiments demonstrated that 35 nucleotides downstream of the E2A poly(A) site were sufficient but 20 nucleotides were not. The construction and assay of bidirectional deletion mutants in the adenovirus E2A poly(A) site indicates that there may be redundant multiple sequence elements that affect poly(A) site usage. Sequences between the poly(A) site and 31 nucleotides downstream were not essential for efficient cleavage. Further deletion downstream (3′ to +31) abolished efficient cleavage in certain constructions but not all. Between +20 and +38 the sequence T(A/G)TTTTT was duplicated. Function was retained when one copy of the sequence was present, suggesting that this sequence represents an essential element. There may also be additional sequences distal to +43 that can function. To establish common features of poly(A) sites, we also analyzed the early simian virus 40 (SV40) poly(A) site for essential sequences. An SV40 poly(A) site deletion that retained 18 nucleotides downstream of the cleavage site was fully functional while one that retained 5 nucleotides downstream was not, thus defining sequences required for cleavage. Comparison of the SV40 sequences with those from E2A did not reveal significant homologies. Nevertheless, normal cleavage and polyadenylation could be restored at the early SV40 poly(A) site by the addition of downstream sequences from the adenovirus E2A poly(A) site to the SV40 +5 mutant. The same sequences that were required in the E2A site for efficient cleavage also restored activity to the SV40 poly(A) site.

The mechanism for formation of the polyadenylic acid [poly(A)] site involves a posttranscriptional RNA processing step (19, 20). The primary transcript is cleaved, probably through an endonucleolytic process (21), very rapidly after the poly(A) site is transcribed. The precise mechanism for the recognition of a single poly(A) site, or for the choice among multiple possible sites, is unknown. The elucidation of this mechanism has become particularly important after the discovery of genes that contain multiple poly(A) sites and the fact that the selection of a given poly(A) site can be a regulatory event. For example, the calcitonin transcription unit specifies both the calcitonin mRNA and the calcitonin gene-related peptide mRNA (25). These transcripts are apparently regulated in a tissue-specific manner by poly(A) site selection (1). Also, the immunoglobulin mu and delta genes are part of the same transcription unit. For these genes, the decision as to which mRNA is to be produced involves a choice between poly(A) sites (13, 30). Finally, the selection of the poly(A) sites in the adenovirus late transcription unit can vary depending upon the time of infection (21, 22). Clearly, the mechanism of selection of sites for poly(A) is crucial for a functional gene, and in addition it provides a step at which gene regulation may occur (20).

The sequence AAUAAA can be found within 10 to 25 nucleotides upstream of the cleavage site of transcripts which are polyadenylated [poly(A)$^+$] (23), and this sequence is necessary for formation of the mRNA 3′ terminus (6, 11, 17, 28). However, this sequence is also found within the transcribed region of many genes at sites where there is no processing of the mRNA (such as the early simian virus 40 [SV40] transcription unit), clearly suggesting the need for additional information in the selection of a poly(A) site. We recently demonstrated that sequences downstream of the adenovirus E2A poly(A) site were essential for efficient cleavage (15). Deletion of the downstream sequence, leaving an intact AATAAA and cleavage site, resulted in the production of RNA transcripts that were not cleaved at the poly(A) site. A similar result has also been found in several other genes (8, 16, 26, 27), suggesting a common organization of poly(A) sites. In this study, we took two approaches to further define poly(A) site function. First, sequences in the E2A poly(A) site were narrowed down to those critical for function. Second, to provide a basis for generalization, we examined sequences in a different poly(A) site, the early SV40 site, for those that might be essential for poly(A) site formation. Finally, the same E2A sequences which are essential for efficient function of the E2A poly(A) site can be used to rescue deletion mutants of the SV40 poly(A) site, suggesting that these two downstream elements are functionally related.

MATERIALS AND METHODS

Cells. Human 293 cells (10), grown in Dulbecco modified Eagle medium with 10% fetal calf serum, were used throughout.

Transfection of plasmids and preparation of RNA. Plasmid DNA concentration was determined by the diphenylamine assay, and plasmids were transfected into 293 cells as previously described (12, 15). RNA was prepared from whole cells 48 to 72 h after transfection by the guanidine thiocyanate lysis method (5), followed by pelleting through a cesium cushion. After pelleting, RNA was reprecipitated from 2 M ammonium acetate with 2.5 volumes of ethanol,

* Corresponding author.

washed, and dried in a vacuum. The RNA was suspended in water, and its A_{260} was measured.

S1 nuclease assays. The −2/−1 and −2/+11 *Ava*I-*Hae*II probes were prepared from the pE2(−2/−1) and pE2(−2/+11) plasmids respectively. An *Ava*I fragment spanning the poly(A) site was recovered by agarose gel electrophoresis and elution.

This fragment was end labeled with [α-^{32}P]TTP and digested with *Hae*II, which cleaves 58 nucleotides downstream of the poly(A) site. The labeled single-strand fragment was separated from the unlabeled fragment by electrophoresis in a 13% acrylamide–urea sequencing gel and eluted (14).

S1 mapping experiments were performed essentially as before (15), except samples were incubated at 37°C for 20 min with pancreatic RNase instead of using NaOH to remove excess RNA.

RNase protection assay. A probe for the E2A poly(A) site was prepared by cloning a *Nar*I fragment containing the E2 poly(A) site from plasmid pE2(+35) into the *Acc*I site of pGEM1 (Promega Biotec). This clone contains wild-type E2 sequence from 240 nucleotides upstream of the poly(A) site to +35 nucleotides downstream, as well as approximately 30 nucleotides further downstream from the pE2(+35) plasmid. Labeled SV40 probe RNA was synthesized from a clone containing a 237-base-pair (bp) *Sau*3A fragment of SV40 DNA in the vector pSP64 (Promega Biotec). Each plasmid was linearized with a restriction endonuclease and used as a template in a reaction medium consisting of 40 mM Tris (pH 7.5), 10 mM NaCl, 6 mM $MgCl_2$, 10 mM spermidine, 20 mM dithiothreitol, 2 U of placental RNase inhibitor (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) per μl, 0.5 mM of each of three cold nucleotide triphosphates, 15 μM GTP, 10 μM [α-^{32}P]GTP (700 Ci/mmol), 100 ng of DNA per ml, and 500 U of SP6 RNA polymerase (New England Nuclear Corp., Boston, Mass.) per ml. Synthesis proceeded for 1 h at 40°C, followed by the addition of 20 μg of RNase-free DNase I (Bethesda Research Laboratories) per ml and further incubation at 37°C for 15 min. The synthetic RNA was then phenol-chloroform extracted and ethanol precipitated from 2 M ammonium acetate. Typically, about 50 to 80% of the labeled nucleotide was incorporated, yielding a probe with a specific activity of about 5×10^8 cpm/μg. Cell RNA (5 to 20 μg) was coprecipitated with about 5 ng of probe RNA and suspended with 20 μl of 80% redistilled formamide–40 mM PIPES (piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (pH 6.7)–0.4 M NaCl–1 mM EDTA. The mixture was heated briefly to 70°C and hybridized at 55 to 57.5°C overnight. A 300-μl portion of an ice-cold solution consisting of 50 mM sodium acetate (pH 4.5), 2 mM EDTA, 100 mM NaCl, and 33 U of T2 RNase (Bethesda Research Laboratories) per ml was added and allowed to digest at 30°C for 1 h. The RNA was then extracted, precipitated, and run on an 8% polyacrylamide sequencing gel (14).

Construction of E2A bidirectional deletion mutants. To prepare the E2A sequence deleted in a 5′ to 3′ direction, wild-type pE2 was digested with *Ava*I, and the 1,905-nucleotide fragment spanning the E2A poly(A) site was isolated. The purified DNA was then digested with BAL 31, and *Bcl*I linkers were added. The DNA was cut with *Bcl*I and *Sal*I, and DNA of the appropriate size was purified in an agarose gel. These fragments were then ligated to a *Bam*HI-*Sal*I fragment isolated from pE2(+20), pE2(+12), and pE2(−2). Recombinants were selected for ampicillin resistance, and DNA was prepared from individual colonies. The +12/+43 plasmid was constructed by preparing a *Sau*3A fragment from the pE2(+20/+43) recombinant. As described above, the original series of mutants were formed by ligation of a *Bam* site with a *Bcl* site. This leaves a *Sau*3A site at the site of recombination (i.e., +43 in the +20/+43 mutant). This *Sau*3A fragment was then inserted at the *Bam* site of the +12 mutant to generate the +12/+43 recombinant. The +20/+32 recombinant was likewise generated by using the *Sau*3A fragment from the original pE2(+12/+32) mutant. The nature of the recombinants was determined by restriction mapping and finally by DNA sequencing.

Construction of SV40 poly(A) site deletions. Plasmid pEC (12a) was linearized with *Bam*HI and digested with BAL 31 as previously described (15), and the deleted gene portion was isolated by digestion with *Xba*I. This fragment was recloned into the *Bam*HI-*Xba*I vector sequences from pEC. Resulting clones were screened and sequenced.

Recombination with E2 downstream sequences. Clones containing deletions downstream of the E2 poly(A) site were digested with *Sau*3A and run on a acrylamide gel. Fragments corresponding to sequences downstream of the cleavage site were gel purified (14) and recloned into pEC(+5) linearized with *Bam*HI. Clones resistant to *Bam*HI were screened and sequenced.

RESULTS

Construction of additional E2A poly(A) site mutants. Our previous studies established that there is a requirement for sequences immediately downstream of the adenovirus E2A poly(A) site for efficient cleavage at the poly(A) site to occur (15). The sequence at the E2A poly(A) site and the original 3′ deletion mutants are depicted in Fig. 1A. The previous experiments employing these mutants demonstrated that the pE2(+35) plasmid was fully functional whereas the pE2(+20) plasmid was not. Thus, the 3′ border of a critical sequence must lie somewhere between 20 and 35 nucleotides downstream of the cleavage site. Those results demonstrated that an essential sequence exists, but they did not allow a precise identification of the critical sequence. Furthermore, if essential sequences were repeated in the downstream region, then such analyses might only yield information about the minimal requirements. To define the downstream sequence requirements more precisely, we constructed additional mutants, in the form of small internal bidirectional deletions. Adenovirus E2A DNA was deleted in a 5′ to 3′ direction starting at a point just upstream of the AATAAA sequence. The deleted DNA was then recombined with three of the original E2A poly(A) site deletion mutants: pE2(−2), pE2(+12), and pE2(+20). The result is a nearly wild-type sequence interrupted by small deletions (or small duplications) and the insertion of a linker molecule. The structure of these constructions is depicted in Fig. 1B.

The E2A downstream region contains multiple elements. To assay for function in these various mutants, human 293 cells were transfected with each of the plasmids, and RNA was prepared 48 h later. For most of the experiments reported here, cleavage at the poly(A) site was then determined by an RNase protection assay with an SP6-generated RNA probe that spanned the poly(A) site. Our previous assays utilized a 3′ S1 assay (15). However, we have recently changed to the SP6 assay for two reasons. First, there is greater sensitivity with the synthetic, single-stranded RNA probe. Second, our S1 probe protected only a very short fragment which often gave rise to considerable heterogeneity of the protected fragment, making quantitation difficult. SP6 RNA protection assays of the original +35 and +12 plasmids were very similar to previous results (16) (Fig. 2A). E2A RNA cleaved

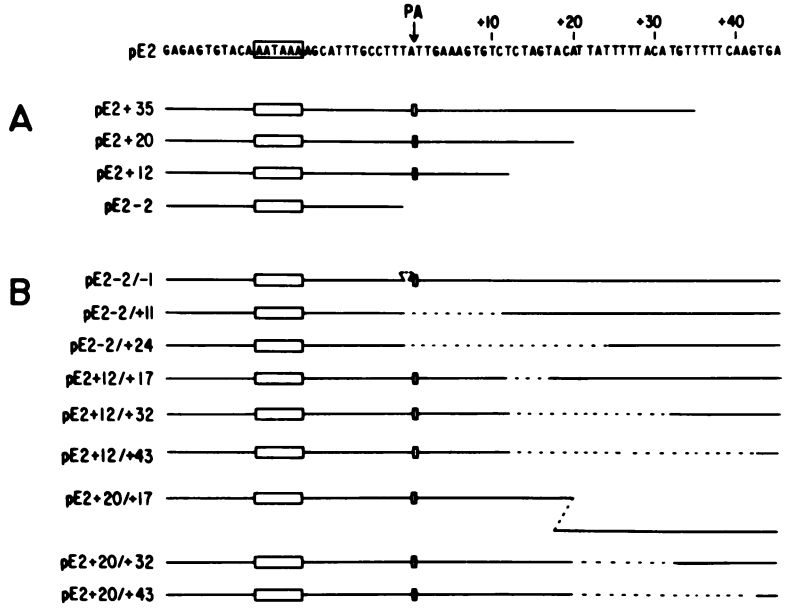

FIG. 1. Schematic diagram of the E2A poly(A) site and structure of the mutants. (A) DNA sequence at the E2A poly(A) site depicting the AATAAA and the major poly(A) site (PA). Below is shown a depiction of the previously described 3′ deletion mutants (15). The mutants are labeled according to the endpoint of the deletion relative to the poly(A) site. (B) Schematic diagram of the bidirectional deletion mutants. The solid lines indicate E2A sequence, and the dotted lines indicate the presence of a linker sequence (CCGGATCAG) and deleted E2A sequence.

at the poly(A) site protects 240 nucleotides of the SP6 probe RNA. In addition to this major cleavage site, there is a minor cleavage site four to six nucleotides downstream yielding a slightly larger protected product. RNA produced from the +12 plasmid was inefficiently cleaved when compared with 3′ end formation from RNA produced from the +35 plasmid (about 15% of the +35). In these assays, poly(A) site function is measured by the amount of the correctly cleaved band irrespective of the amount of readthrough transcript. Through the course of a number of analyses, we found that the stability and thus recovery of the poly(A)$^-$ readthrough transcript can vary. Thus, we cannot be certain of the

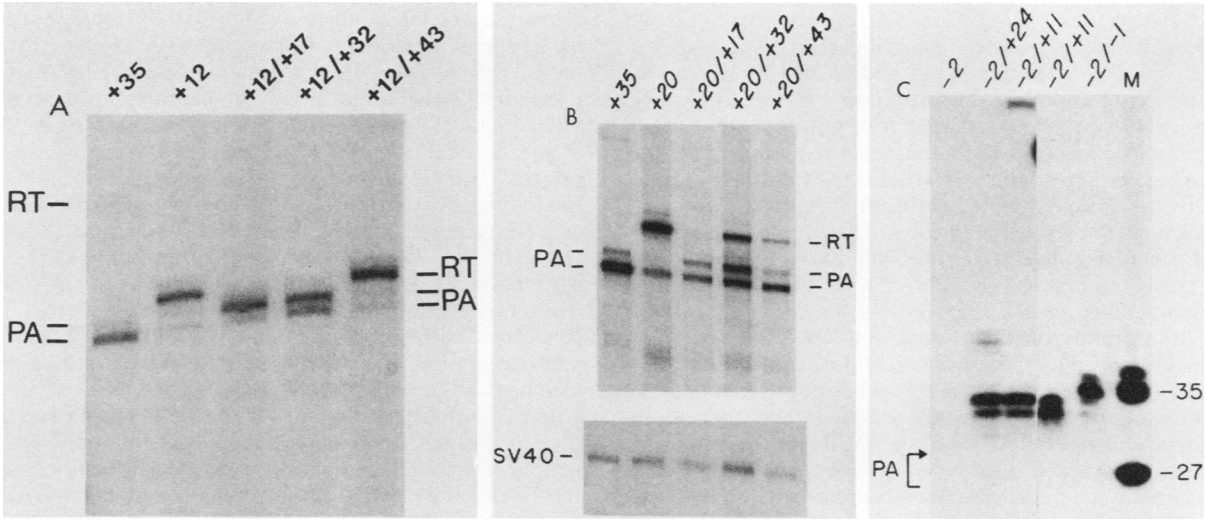

FIG. 2. RNase protection assay of E2A poly(A) site bidirectional deletion mutants. (A) Assays of +12-series mutants. Human 293 cells were transfected with the various plasmids, and total cell RNA was assayed with an SP6 RNA that spanned the poly(A) site. The major poly(A) site is indicated by PA and represents a major site and a minor site (see text). The uncleaved RNA from the +12 mutant is indicated on the right. Readthrough (RT) RNA from the +35 mutant is indicated on the left. (B) Assays of the +20-series mutants. Assays were performed as above. Shown below this figure are internal controls with the SV40 poly(A) site. The pEC plasmid was cotransfected with the +20 mutants, and samples of the RNAs were assayed for SV40 poly(A) site formation as described in the legend to Fig. 3. (C) Assays of the −2-series mutants. 3′ S1 assays of pE2(−2), pE2(−2/+24), pE2(−2/+11), and pE2(−2/−1). RNA from transfections was hybridized with a 3′ labeled probe prepared from pE2(−2/+11) except for the sample from the −2/−1 transfection, which was hybridized to a probe prepared from the homologous plasmid. The size of the normal poly(A) site band is indicated (PA). Assays from two separate experiments are shown (left three lanes show one experiment, and right two lanes show another). M, DNA markers of 37 and 25 nucleotides.

efficiency with which the transcripts are cleaved. The only way of determining such a value with no ambiguity would be to measure processing of pulse-labeled RNA, an assay not feasible here.

In contrast to the results with the +12 plasmid, the two internal deletion mutants derived from the +12 plasmid that restore downstream sequences from +17 or +32 produced RNA that was efficiently cleaved, equal to that of the +35 plasmid. The construction that deletes sequence between +13 and +42 (+12/+43), however, produced RNA that was inefficiently cleaved at the poly(A) site, no better than the original +12 mutant. Thus, sequences between +13 and +31 are not required for efficient cleavage, but further deletion to +42 markedly affects cleavage. Sequences within the +12/+17 region may be important for the site of cleavage since for both the +12/+17 and the +12/+32 recombinants the cleavage frequently takes place at the minor poly(A) site. Alternatively, the alteration in site of cleavage may be due to the alteration of the normal spacing of sequences in this region. This alternate cleavage site corresponds to a series of A residues at position +4 to +6 (Fig. 1). This site is actually an authentic poly(A) site since RNA with this 3′ terminus is found in the poly(A)$^+$ fraction (data not shown). Furthermore, this does not appear to be an aberrant process since this cleavage site is in fact seen with the +35 plasmid (Fig. 2A) as well as in adenovirus-infected cells (data not shown), although much less frequently than the normal site.

We next assayed the +20 series of bidirectional deletion mutants (Fig. 2B). Once again, assay of the original pE2(+20) plasmid gave results similar to previous data indicating inefficient cleavage; in this experiment, the production of RNA from the +20 plasmid cleaved at the poly(A) site was about 15% of that from the +35 plasmid. Replacement of downstream sequence with either the +17 fragment or the +32 fragment indeed improved activity although not to the full +35 level (about 80% when the internal control is taken into account). In contrast to the results with the +12 series, however, the +43 construction also improved cleavage. Thus, we see no sharp distinction in this series as we observed with the other mutants and as will be described below for a heterologous construction utilizing the early SV40 poly(A) site. We must therefore conclude that there are sequences distal to +43 that are able to restore cleavage function but are dependent on the nature of the construction.

Finally, we assayed the −2 series of mutants for processing efficiency (Fig. 2C). For these assays we used the 3′ S1 method rather than the SP6 protection assay. In this experiment, similar to previous results (15), we did not detect an RNA produced from the pE2(−2) plasmid. This appears to be due to instability of the unprocessed RNA since in other experiments we detected such a readthrough RNA. However, the addition of the normal downstream sequences to this mutant does indeed restore RNA 3′ end formation. Addition of sequences starting from +24 downstream or +11 downstream produced efficient 3′ end formation (Fig. 2C). As indicated in the figure, the 3′ ends are downstream of the normal site, which is deleted in these constructions. In fact, the cleavage in each of these constructions takes place at a C-A sequence in the linker. Thus, the normal poly(A) site is not absolutely required, a result consistent with previous data derived from mutation at the late SV40 poly(A) site (6). The results obtained with the −2/−1 construction in which the normal poly(A) site has been returned are also shown in Fig. 2C. In this case, cleavage now occurs at a G-A residue in the linker, two nucleotides downstream from the C-A site. We conclude that the poly(A) site is not essential and that relative spacing plays an important role as evidenced by the shift in sites within the linker.

From the data presented above we draw the following conclusions. From our previous work and data presented here it is clear that the minimal requirement for function is 35 nucleotides downstream of the poly(A) site; 20 nucleotides of downstream sequence are not sufficient. Thus, there must be critical sequences between 20 and 35 nucleotides downstream of the poly(A) site. Sequences between −2 of the poly(A) site and +31 are not essential for efficient cleavage. Finally, there may be a functional sequence distal to +43. These sequences clearly are not critical when the sequences up to +35 are present since the pE2(+35) plasmid functions. Possibly there is repetition of functional sequences in the downstream region. In addition, although there appears to be activity in the sequence downstream of +43, this is only true in certain contexts. A likely explanation for such a result is a requirement for proper spacing between the cleavage site and the downstream sequence or possibly the generation of an active sequence due to the nature of the recombination.

Assay for early SV40 poly(A) site formation. To investigate the sequences in a different poly(A) site that might be necessary for cleavage, we made use of a plasmid (depicted in Fig. 3A) containing the early SV40 poly(A) site. The plasmid pEC was constructed previously as an assay system for the adenovirus E2 promoter (12a) and consists of the E2 promoter fused to the chloramphenicol acetyl transferase coding sequence, an SV40 splice site, and the early SV40 poly(A) site. The construction leaves 53 nucleotides of SV40 sequence downstream of the major cleavage site. As will be described elsewhere, there are minor cleavage sites (indicated by the arrows in Fig. 3A) apparently directly by the upstream AATAAA sequence. To assay for poly(A) site function, we transfected the pEC plasmid into human 293 cells where the expression of this plasmid directed by the E1A-inducible E2 promoter is high. Total cell RNA was extracted 48 h later and assayed for SV40 3′ termini by using an SP6-generated RNA probe (Fig. 3B). The probe originates at the *Bam*HI site and terminates at a *Sau*3A site. The total size of the probe is 277 nucleotides and includes 34 nucleotides of SP6 sequence at the 5′ end and 6 nucleotides at the 3′ end. Thus, SV40 transcripts not cleaved at the poly(A) site (readthrough) would protect 242 nucleotides (including the entire *Bam*HI site sequence), and RNA cleaved at the major poly(A) site would protect 189 nucleotides. In Fig. 3B, the site of polyadenylation in the clone pEC is compared with that of wild-type early SV40 RNA from various sources. The size of the band protected by RNA from early SV40-infected CV-1 cells is 189 nucleotides and agrees with the previously determined site of poly(A) cleavage in the early region (24). RNA from cells transfected with pEC yielded the same protected band as well as a small quantity of 242-nucleotide band due to uncleaved transcripts. This readthrough transcript is apparently confined to the nucleus as there was no evidence of the RNA in the cytoplasmic sample (the band in the cytoplasmic sample just below the readthrough band is an incomplete RNase digestion product). Fractionation of the transfection RNA by oligo(dT)-cellulose chromatography indeed demonstrated that the 189-nucleotide band is poly(A)$^+$ and the 242-nucleotide band is poly(A)$^-$ (data not shown). In addition, there is a series of bands which represent minor cleavage sites upstream of the major site. These minor cleavage sites are apparently directed by the AATAAA sequence located 23 bp upstream of the major AATAAA element. Finally, a

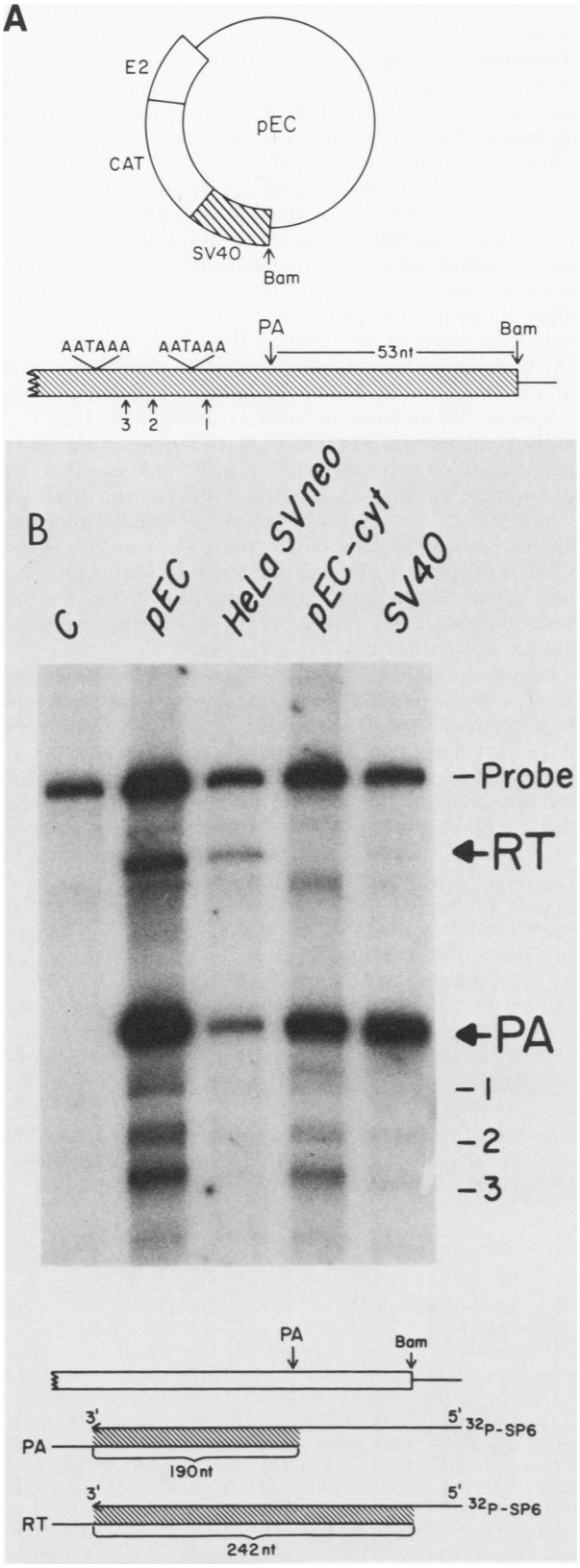

FIG. 3. Schematic diagram of the pEC plasmid and RNase protection assay of the early SV40 poly(A) site. (A) The plasmid pEC (12a) consists of an adenovirus E2 promoter, the chloramphenicol acetyl transferase-coding sequence (9), and SV40 sequences including a splice site and the early SV40 poly(A) site. There are two AATAAA sequences in the early SV40 poly(A) site separated by 23 bp. The location of the major early 3′ end is indicated by PA. In this particular construction, there are 53 nucleotides of SV40 sequence downstream from the poly(A) site. Minor cleavage sites are indicated by the arrows and 1, 2, 3. (B) RNase protection assay. Human 293 cells were transfected with pEC, and total cell RNA was prepared. The RNA was then assayed with an SP6-generated RNA probe depicted at the bottom of the figure. SV40 transcripts that extend through the poly(A) site (RT) will protect an RNA of 242 nucleotides in length, while a transcript cleaved at the poly(A) site (PA) will protect an RNA of 189 nucleotides in length. The figure depicts such an assay with RNA prepared from mock-transfected cells (C), pEC-transfected cells (pEC), RNA from a stable HeLa cell line expressing a neomycin-resistant gene utilizing the early SV40 poly(A) site (HeLa SVneo), cytoplasmic RNA from cells transfected with pEC (pEC-cyt), and RNA from an early infection of CV-1 cells with SV40 virus (SV40). nt, Nucleotides.

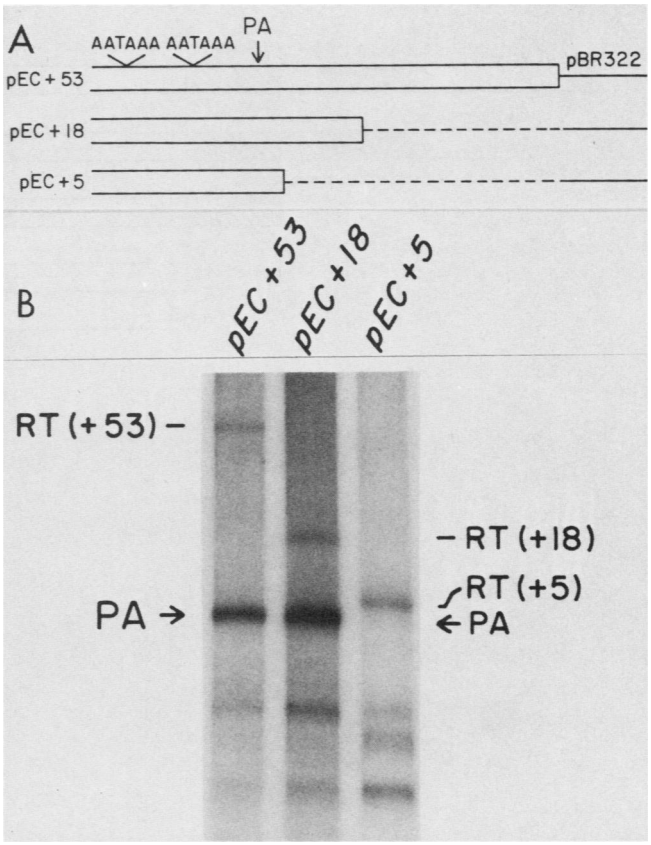

FIG. 4. Assay of SV40 poly(A) site deletion mutants. (A) Schematic diagram depicting the structure of the starting pEC plasmid [pEC(+53)] and the 3′ deletion mutants. The mutants are labeled according to the number of nucleotides of SV40 sequence remaining downstream of the poly(A) site (PA). (B) RNase protection assay for 3′ end formation from the deletions. Assays were as described in the legend to Fig. 3 with the same SP6 probe. The readthrough (RT) transcripts derived from the various mutants are indicated.

HeLa cell line containing the pSV2neo gene which confers resistance to the drug G 418 and which contains the same SV40 poly(A) site displays the same pattern of 3′ end formation. The minor cleavage sites appear to be utilized more frequently from the pEC plasmid, either in trans-

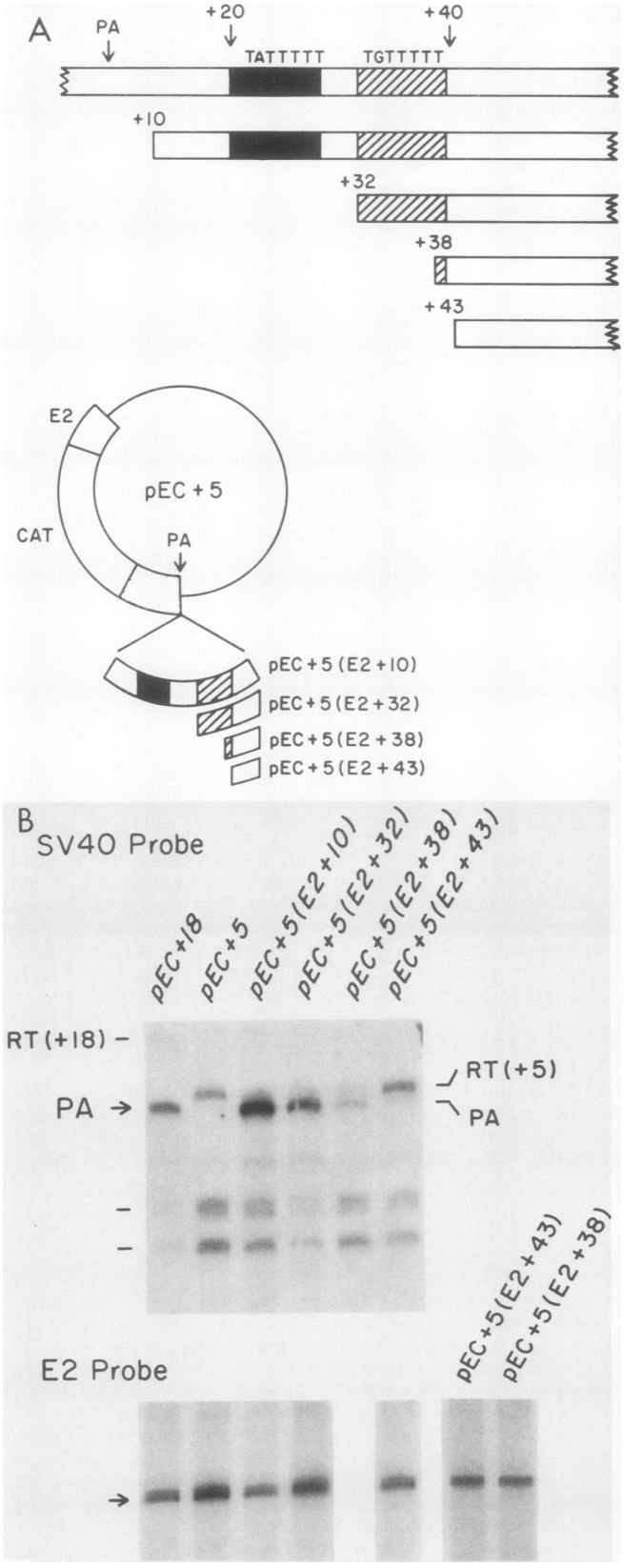

FIG. 5. Restoration of SV40 poly(A) site function with adenovirus E2A downstream sequence. (A) Schematic diagram depicting the downstream region of the E2A poly(A) site and the repeated element T(A/G)TTTTT. The two elements are shown in the schematic as hatched and solid blocks to distinguish them. The endpoints of E2A mutants at +10, +32, +38, and +43 are indicated. Each of these endpoints represents a *Sau*3A site from the linker sequence. There is a natural *Sau*3A site 216 nucleotides downstream of the poly(A) site in the wild-type E2A sequence. Below is shown the pEC(+5) deletion mutant. The *Sau*3A fragments from the E2A bidirectional deletion mutants were inserted into pEC(+5) at the *Bam* site (+5 position). PA, Major poly(A) site. (B) Assay for poly(A) site formation in the hybrid plasmids. RNA was prepared from transfections with the various plasmids. In addition, the pE2 plasmid was included in each transfection to serve as an internal control. Assays for SV40 poly(A) site formation utilized the SP6 probe as described in the legend to Fig. 3B. Shown below are assays of E2A as internal controls. In the initial assay, the (E2+38) sample was lost. The assay of this RNA was repeated (shown on the right) along with the (E2+43) sample to allow comparison with the other original assays. RT, Readthrough.

fections or in stable cell lines, than from the native SV40 genome.

Definition of sequences essential for cleavage. A series of deletions of pEC were produced by digestion of *Bam*HI-linearized plasmid with BAL 31 and recloning into vector sequences. These mutants (Fig. 4A) are indicated by the number of nucleotides remaining relative to the major poly(A) site. Plasmids containing these deletions were transfected into 293 cells, and whole-cell RNA was prepared and probed by an RNase protection assay (Fig. 4B). In each case, a readthrough transcript can be detected (labeled RT in the figure) the size of which decreases as the homologous coding sequence decreases. A deletion leaving only 18 bp beyond the poly(A) site [pEC(+18)] allows normal 3′ end formation, equal to or somewhat better than that obtained with the +53 plasmid. However, removal of an additional 13 bp, leaving 5 bp beyond the poly(A) addition site [pEC(+5)], greatly reduced the frequency of cleavage at the normal poly(A) addition site [about 5% of pEC(+53) or pEC(+18)]. The decrease in usage of the normal poly(A) site often led to an increase in the levels of poly(A)$^+$ cleavage products mapping to the upstream (minor) poly(A) sites. The essential sequence for the use of these poly(A) sites is currently being investigated. These results thus define the presence of sequences downstream of the early SV40 poly(A) cleavage site that are required for efficient cleavage.

Complementation of SV40 downstream deletions by insertion of adenovirus E2A downstream sequences. Sequences downstream of the adenovirus E2A poly(A) site are essential, as shown in this paper as well as in previous work (15). Indeed, the focus of the present work is to begin to determine the common elements of the poly(A) site. An examination of the two downstream sequences (E2A and early SV40), however, did not reveal significant homologies. Possibly, the two are functionally the same even though homologies are not readily detected. Alternatively, the two poly(A) sites may operate quite differently. To probe this question, we made hybrid poly(A) sites using the early SV40 AATAAA element and the adenovirus E2A downstream element. To determine if the downstream element of SV40 can be complemented by the apparently nonhomologous E2 downstream element, we cloned E2 fragments into the *Bam*HI linker site of pEC(+5). Plasmids from a series of bidirectional deletions near the poly(A) site of E2A (Fig. 1B) contained a *Bam*HI linker sequence fused with a *Bcl*I linker at the deletion site, leaving a *Sau*3A site. These clones were digested with *Sau*3A to yield fragments beginning at +10,

+32, +38, or +43 bp and extending to a natural *Sau*3A site 216 bp beyond the E2A poly(A) site (Fig. 5A). The fragments were cloned into the *Bam*HI site of pEC(+5), thus at the +5 position with an 8-bp linker intervening, and recombinants were screened and sequenced. Four clones were isolated which contained a fragment of E2 in the proper orientation. All four clones were transfected into 293 cells along with pEC(+5) and pEC(+18). RNA was prepared and assayed by RNase protection (Fig. 5B).

Two results are clear from this experiment. First, the adenovirus E2A downstream sequences can indeed restore function to the impaired SV40 poly(A) site. For instance, compare the poly(A) site band (Fig. 5, PA) obtained with pEC(+5) to that obtained with pEC(+5)[E2(+10)]. There appears to be complete restoration of function. In fact, the usage of the SV40 poly(A) site in this hybrid construction is consistently higher than that seen with the pEC(+18). The second point evident from this experiment concerns the sequences in the E2A poly(A) site that are required for restoration of cleavage at the SV40 poly(A) site. The E2A fragment with an endpoint at +32 is nearly as effective as the fragment beginning at +10. Furthermore, compare the formation of the PA band with pEC(+5)[E2(+32)] to that with pEC(+5)[E2(+38)] or pEC(+5)[E2(+43)]. Clearly, there is a loss of function when the E2A sequences between +32 and +43 are removed. Thus, the results are similar to those obtained with the +12 series of E2 deletion mutants.

DISCUSSION

From the results presented here as well as those presented in a previous study (15), it is now clear that the functional element of a poly(A) site contains more than just the AATAAA sequence element. In two cases, sequences immediately downstream of the cleavage site are essential for efficient processing of the primary RNA transcript. Furthermore, similar results have now been obtained for several other poly(A) sites (8, 16, 26, 27). The significance of the findings reported here are twofold. First, the studies with the E2A deletion mutants as well as the reconstruction of the SV40 poly(A) site have allowed a more precise definition of essential sequences. From the initial 3′ deletion mutants, we were able to demonstrate that 35 nucleotides were sufficient for efficient cleavage. The assays of the +12 bidirectional deletion mutant series showed that sequences up to +31 can be deleted; the same was also true with the heterologous construction involving the early SV40 poly(A) site. Taking these two results together leaves the sequence between +32 and +35, TGTT. Although it is possible that this is the entirety of an essential element, we feel that it is likely to encompass more. If one examines the DNA sequence downstream of the cleavage site one finds that the sequence T(A/G)TTTTT is repeated at +22 to +28 and at +32 to +38. Examination of the mutant assays reveals that activity remains as long as one copy of this sequence is present. We suggest that the essential element may be this repeated heptanucleotide sequence T(A/G)TTTTT. In addition, it is possible that there are even more elements since in one series of constructions, sequences distal to +43 could restore activity. The context of the construction, possibly the relative spacing of the functional signals, may influence the activity of a given construction. A final definition of any given sequence as being critical and sufficient will require the isolation of the element and assay under controlled circumstances as well as point mutagenesis within this region, and these experiments are in progress.

Second, the requirement for specific sequences downstream of the poly(A) cleavage site is now not a unique feature of the adenovirus E2A poly(A) site. The data presented here clearly show that such is also the case for the early SV40 poly(A) site, and similar studies have been carried out for several other poly(A) sites (8, 16, 26, 27). The definition of additional sequences that are involved in the function of a poly(A) site allows us to begin to examine the sequences for potential common features. As we suggested above, a likely candidate sequence in the E2A poly(A) site is the T(A/G)TTTTT element. The critical sequences in the SV40 poly(A) site must be within 18 nucleotides of the cleavage site (Fig. 6). Within this region there is no sequence precisely homologous to the E2A element, despite the fact that the two appear to be functionally homologous. In general, the SV40 sequence in the essential region is GT-rich. An examination of the other poly(A) sites where essential sequences have been defined reveals the presence of both types of sequence, of (A/G)TTTTT, thus quite similar to the postulated E2A sequence, and GT-rich sequences similar to the SV40 element. The relative location of such sequences, the fact that they are in the essential region, and the homology to the E2A sequence and the SV40 downstream region suggest that these may be a common functional element. It seems possible that there are at least two classes of downstream sequence, one representing SV40 homology (and surveyed extensively in reference 16) and the other representing E2A homology. Furthermore, the possibility that there may be functional sequences in the E2A gene distal to +43 that do not appear to be homologous to either of the above sequences suggests a potential for even greater complexity.

In addition to the participation of downstream sequences in efficient processing of mRNA molecules at a poly(A) site, our data suggest that sequences just downstream of the poly(A) site may be important in the selection of poly(A) sites. This may not be a result of actual specific sequences since similar phenomena were observed in many of the constructions described in this work. Instead, spacing between critical elements such as the AAUAAA and the downstream sequence may be the determining factor. Similar results have been obtained previously with the late SV40 poly(A) site (6) and the bovine growth hormone poly(A) site (29).

Several recent results suggest that a small RNA could be involved in the formation of the poly(A) site. First, there is a requirement for the U1 ribonucleoprotein in the recently described in vitro coupled transcription-cleavage reaction (18). The addition of antisera specific to the U1 ribonucleoprotein blocked the generation of the 3′ terminus. Second, studies on the generation of the 3′ terminus of histone H3 RNA have demonstrated the involvement of a small RNA in the cleavage process (4, 7). Although the histone RNAs are not poly(A)$^+$ and do not possess an AAUAAA sequence, the process of 3′ end formation nevertheless has similarities to the formation of a poly(A) site. Generation of the histone 3′ terminus involves an RNA chain cleavage, and there is a requirement for sequences downstream of the cleavage site (3, 4). Finally, there has been a suggestion that the U4 RNA might be involved in poly(A) site selection (2). Analysis of poly(A) site sequences suggests a possible interaction of U4 RNA with the AAUAAA element and a downstream sequence (consensus of CAPyUG). Certainly, these potential U4 homologies cannot be sufficient since the SV40 +5 mutant retains the sequence (CACUG) but is not functional. Of course, this does not rule out the potential involvement of such a sequence, only that

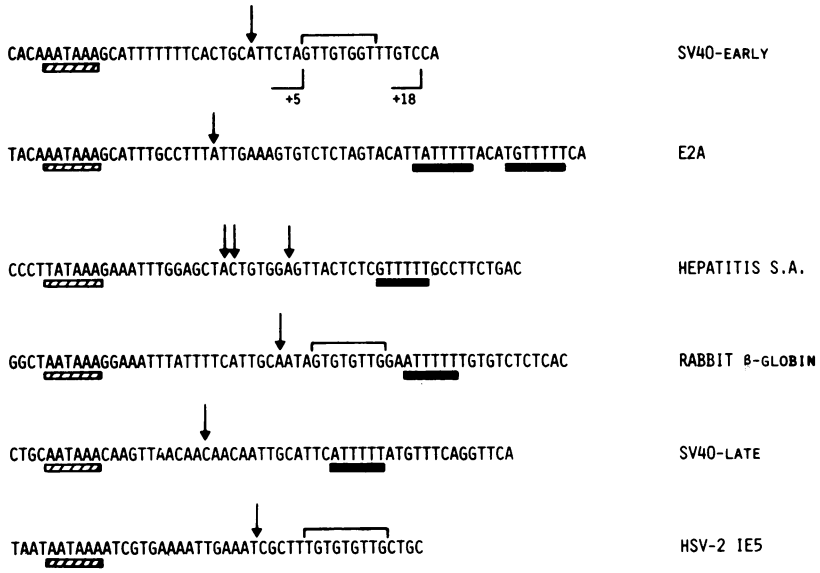

FIG. 6. Sequences of essential regions downstream of poly(A) sites. DNA sequence from the AATAAA element (underscored with a hatched box) to the 3′ border of sequences essential for poly(A) site formation is shown for five transcription units. The sites of cleavage are indicated by the arrows. The two deletions that define the SV40 essential sequences are indicated below the sequence (i.e., +18 is functional and +5 is not). The T(A/G)TTTTT sequences in the E2A poly(A) site and the homologous sequences in the other sites are underscored by a solid box. The sequence homologous to the SV40 early downstream element is marked above the sequence with a bracket. Data for the hepatitis surface antigen (SA) poly(A) site are from Simonsen and Levinson (27); data for the rabbit β-globin site are from Gil and Proudfoot (8); data for the SV40 late site are from Sadofsky and Alwine (26); and data for the herpesvirus immediate-early 5 gene are from McLauchlan et al. (16).

the sequence is not the sole requirement in addition to the AATAAA. However, the sequences that are postulated to interact with the U4 RNA in the E2A poly(A) site can be deleted (the −2/+11 or the −2/+24 recombinants) with function still retained. It is still possible that an additional U4-interacting site is brought into proximity in these constructions so as to allow an interaction. Clearly, an in vitro system is required to allow a decision to be made about the components necessary for poly(A) site formation, either proteins or ribonucleoprotein particles.

Finally, at the heart of these studies is the question of regulation of poly(A) site formation. We and others have argued that the highly conserved AATAAA element does not allow the flexibility essential for a regulated selection, just as the TATA element alone cannot provide specificity in promoter selection. Whether the essential sequences located downstream of the cleavage site are the site of action of regulatory events must await an analysis of a poly(A) site that is subject to regulation. The observation that there may be multiple classes of sequences in the downstream region that are essential for activity leaves the possibility open.

ACKNOWLEDGMENTS

This work was supported by a grant from the American Cancer Society (MV141B). R.P.H. is a National Institutes of Health postdoctoral fellow, and M.A.M. is supported by a Medical Scientist Training Program grant from the National Institutes of Health (T32-GM07739). J.R.N. is the recipient of a Research Career Development Award (CA00666).

LITERATURE CITED

1. **Amara, S. G., R. M. Evans, and M. G. Rosenfeld.** 1984. Calcitonin/calcitonin gene-related peptide transcription unit: tissue-specific expression involves selective use of alternative polyadenylation sites. Mol. Cell. Biol. **4**:2151–2160.
2. **Berget, S. M.** 1984. Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature (London) **309**:179–182.
3. **Birchmeier, C., R. Grosschedl, and M. L. Birnstiel.** 1982. Generation of authentic 3′ termini of an H2A mRNA in vivo is dependent on a short inverted DNA repeat and spacer sequences. Cell **28**:739–745.
4. **Birchmeier, C., D. Schumperli, G. Sconzo, and M. L. Birnstiel.** 1984. 3′ editing of mRNAs: sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. Proc. Natl. Acad. Sci. USA **81**:1057–1061.
5. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
6. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 RNAs. Cell **24**:251–260.
7. **Galli, G., H. Hofstetter, H. G. Stunnenberg, and M. L. Birnstiel.** 1983. Biochemical complementation with RNA in the *Xenopus* oocyte: a small RNA is required for the generation of 3′ histone mRNA termini. Cell **34**:823–828.
8. **Gil, A., and N. J. Proudfoot.** 1984. A sequence downstream of AAUAAA is required for rabbit β-globin 3′ end formation. Nature (London) **312**:473–474.
9. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
10. **Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn.** 1977. Characteristics of a human cell line transformed by DNA from human adenovirus 5. J. Gen. Virol. **36**:59–72.
11. **Higgs, D. R., S. E. Y. Goodbourn, J. Lamb, J. E. Clegg, D. J. Weatherall, and N. J. Proudfoot.** 1983. α-Thalassaemia caused by a polyadenylation signal mutation. Nature (London) **306**:398–400.
12. **Imperiale, M. J., L. T. Feldman, and J. R. Nevins.** 1983. Activation of gene expression by adenovirus and herpesvirus regulatory genes acting in trans and by a cis-acting adenovirus enhancer element. Cell **35**:127–136.

12a. **Imperiale, M. J., R. P. Hart, and J. R. Nevins.** 1985. An

enhancer-like element in the adenovirus E2 promoter contains sequences essential for uninduced and E1A-induced transcription. Proc. Natl. Acad. Sci. USA **82**:381–385.

13. **Mather, E. L., K. J. Nelson, J. Haimovich, and R. P. Perry.** 1984. Mode of regulation of immunoglobulin mu and delta chain expression varies during B-lymphocyte maturation. Cell **36**:329–338.
14. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
15. **McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins.** 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. Cell **37**:993–999.
16. **McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucleic Acids Res. **13**:1347–1368.
17. **Montell, C., E. F. Fisher, M. H. Caruthers, and A. J. Berk.** 1983. Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3′ consensus sequence AAUAAA. Nature (London) **305**:600–629.
18. **Moore, C. L., and P. A. Sharp.** 1984. Site-specific polyadenylation in a cell-free reaction. Cell **36**:581–591.
19. **Nevins, J. R.** 1983. The pathway of eukaryotic mRNA formation. Annu. Rev. Biochem. **52**:441–466.
20. **Nevins, J. R.** 1984. Poly(A) in eukaryotic mRNA, p. 133–150. *In* D. Apirion (ed.), Processing of RNA. CRC Press, Boca Raton, Fla.
21. **Nevins, J. R., and J. E. Darnell.** 1978. Steps in the processing of Ad-2 mRNA: poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell **15**:1477–1493.
22. **Nevins, J. R., and M. C. Wilson.** 1981. Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. Nature (London) **290**:113–119.
23. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
24. **Reddy, V. B., P. K. Ghosh, P. Lebowitz, M. Piatek, and S. Weissman.** 1979. Simian virus 40 early mRNAs. I. Genomic localization of 3′ and 5′ termini and two major splices in mRNA from transformed and lytically infected cells. J. Virol. **30**:279–296.
25. **Rosenfeld, M. G., J. J. Mermod, S. G. Amara, L. W. Swanson, P. E. Sawchenko, J. Rivier, W. W. Vale, and R. M. Evans.** 1983. Production of a novel neuropeptide encoded by the calcitonin gene via tissue-specific RNA processing. Nature (London) **304**:129–135.
26. **Sadofsky, M., and J. C. Alwine** 1984. Sequences on the 3′ side of hexanucleotide AAUAAA affect efficiency of cleavage at the polyadenylation site. Mol. Cell. Biol. **4**:1460–1468.
27. **Simonsen, C. C., and A. D. Levinson.** 1983. Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. Mol. Cell. Biol. **3**:2250–2258.
28. **Wickens, M., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3′ end formation. Science **226**:1045–1051.
29. **Woychik, R. P., R. H. Lyons, L. Post, and F. M. Rottmann.** 1984. Requirement for the 3′ flanking region of the bovine growth hormone gene for accurate polyadenylation. Proc. Natl. Acad. Sci. USA **81**:3944–3948.
30. **Yuan, D., and P. W. Tucker.** 1984. Transcriptional regulation of the μ-6 heavy chain locus in normal murine B lymphocytes. J. Exp. Med. **160**:564–583.