Volume 15 Number 22 1987 Nucleic Acids Research

Identification of a novel sequence that governs both polyadenylation and alternative splicing in region E3 of adenovirus

Helen A.Brady and William S.M.Wold*

Institute for Molecular Virology, St Louis University School of Medicine, 3681 Park Avenue, St Louis, MO 63110, USA

Received June 7, 1987; Revised and Accepted September 21, 1987 Accession no. Y00496

SUMMARY

Region E3 encodes four major overlapping mRNAs with different splicing patterns. There are two poly(A) sites, an upstream site called E3A and a downstream site called E3B. We have analyzed virus mutants with deletions or insertions in E3 in order to identify sequences that function in the alternative processing of E3 pre-mRNAs, and to understand what determines which poly(A) sites and which splice sites are used. In previous studies we established that the 5' boundary of the E3A poly(A) signal is at an ATTAAA sequence. We now show, using viable virus mutants, that the 3' boundary of the E3A signal is located within 47-62 nucleotides (nt) downstream of the ATTAAA (17-32 nt downstream of the last microheterogenous poly(A) addition site). Our data further suggest that the spacing between the ATTAAA, the cleavage sites, and the essential downstream sequences may be important in E3A 3' end formation. Of particular interest, these mutants suggest a novel mechanism for the control of alternative pre-mRNA processing. Mutants which are almost completely defective in E3A 3' end formation display greatly increased use of a 3' splice site located 4 nt upstream of the ATTAAA. The mRNA that uses this 3' splice site is polyadenylated at the E3B poly(A) site. We suggest, for this particular case, that alternative pre-mRNA processing could be determined by a competition between trans-acting factors that function in E3A 3' end formation or in splicing. These factors could compete for overlapping sequences in pre-mRNA.

INTRODUCTION

The formation of a mature mRNA from a primary pre-mRNA transcript generally requires cleavage and polyadenylation to form the 3' end and splicing to join exons together. Much attention has focused on the sequence requirements and the biochemical mechanisms of these two processes. In contrast, very little is known about the sequence requirements and mechanisms which determine alternative pre-mRNA processing in complex transcription units (reviewed in ref. 1).

Investigations into splicing (reviewed in refs. 2, 3) have concluded that consensus sequences at the 5' and 3' splice sites, as well as the lariat branch site are involved. These sequences are required for the formation of a splicing complex which includes small nuclear ribonucleo-

proteins (snRNP) and heterogeneous nuclear RNP core proteins. Some snRNPs are required for splicing and specific snRNPs bind to the consensus sequences.

Regarding 3' end formation, an AATAAA (AAUAAA in RNA) or closely related sequence, located 10-30 nucleotides (nt) upstream of the poly(A) addition (3' end) site, is an essential signal for 3' end formation in most mRNAs (4-7). The AATAAA appears to be required for both polyadenylation and cleavage at the 3' end (8, 9). A factor with the properties of a snRNP has been demonstrated to associate with the AATAAA in vitro (10). Several studies have implicated less conserved sequences downstream of the AATAAA (11-20). The functional role of these downstream sequences remains unclear.

Although minimal sequence requirements have been established for splicing and 3' end formation, it seems unlikely that these sequences alone account for the choice between alternative splice sites and alternative polyadenylation sites in complex transcription units. In particular, it is not known how alternative splice sites and polyadenylation sites are selected, or whether the splicing and polyadenylation signals in complex transcription units differ from those in simple transcription units.

We are using early region E3 of adenovirus as a model system to study pre-mRNA processing in a complex transcription unit (21-28). A schematic of E3 is shown in Figure 1. E3 has a single promoter, and alternative polyadenylation sites termed E3A and E3B. All abundant transcripts have at least one splice (nt 372→768); some messages (mRNA f, h) have a second optional splice with alternative 5' and 3' splice sites. In order to identify cis-acting sequences that function in the processing of E3 pre-mRNA, we constructed small deletion and insertion mutants in viable virus. These mutants were analyzed in virus infected cells to determine how the mutations affect E3 pre-mRNA processing. A major advantage of viable virus mutants is that the mutation can be analyzed under conditions as natural as possible. This contrasts with mutations in plasmid vectors where the transcription unit is out of context. In this paper we describe virus mutants which define sequences essential for the formation of the E3A 3' end. These mutants also suggest a novel mechanism for the control of alternative pre-mRNA processing. Certain mutants defective in E3A 3' end formation display greatly enhanced use of the nearby nt 2157 3' splice site. The E3A polyadenylation site is used to make mRNA a, and the nt 2157 3' splice site is used to make mRNA f (see Fig. 1). With these mutants, mRNA a (40-50% of total E3 mRNAs in wild type) is not made and mRNA f (~15% in wild

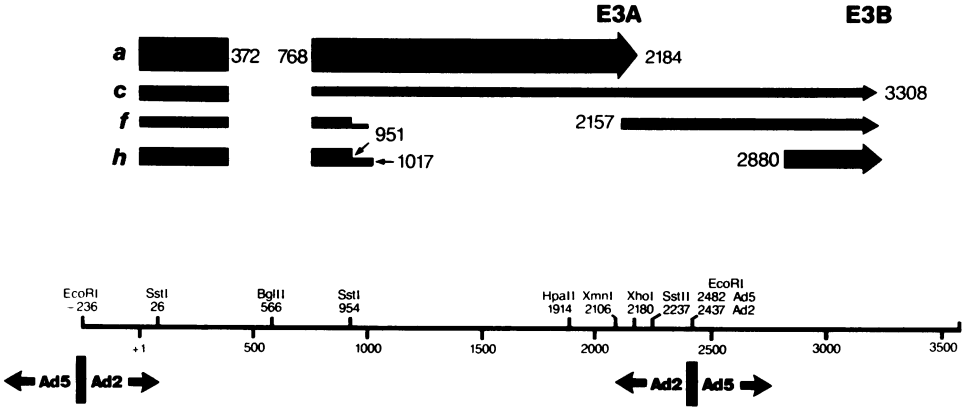

Figure 1. Schematic representation of the E3 transcription unit in rec700. E3 in rec700 has Ad2 sequences between the EcoRI sites at nts -236 and 2437 and Ad5 sequences elsewhere. Nt +1 is the transcription initiation site. Restriction sites are numbered at the 5' base in the recognition sequence. The arrows a, c, f, h indicate the exon structures of the mRNAs with the thickness of the arrow suggesting the abundance of the mRNA. The minor mRNAs b, d, e, g, and i (see reference 27) are not shown. E3A and E3B refer to polyadenylation sites.

type) is made in about 6-fold greater abundance than in wild type, and is virtually the only E3 mRNA made. We suggest for this specific case, that alternative pre-mRNA processing may be determined by a competition between trans-acting factors that function in 3' end formation or in splicing. These factors could compete for overlapping sequences in pre-mRNA.

MATERIALS AND METHODS

Isolation of Plasmid and Virus Mutants

Deletion mutants were isolated essentially as described (29). Deletions were made in the cloned adenovirus 2 (Ad2) EcoRI-D fragment (nt -236 to 2437) at either the XhoI site at nt 2180 (dl720) or the SstII site at nt 2237 (dl752, dl751, dl750, dl748). All these mutants except dl720 have a BamHI linker (CGGGATCCCG) inserted at the deletion (30). The remaining mutants were constructed from the dl752 plasmid. The dl752 plasmid was linearized at the XhoI site (nt 2180) and then blunted either with mung bean nuclease (dl747) or DNA polymerase Klenow fragment (in721). in722 was constructed by ligating a BamHI linker after the XhoI site was blunted with Klenow. For dl753, the dl752 plasmid was cleaved at the BamHI site, blunt-ended, ligated to an EcoRI linker (CGGAATTCCG), cleaved with EcoRI

(which generated EcoRI sticky ends in the linker and also excised the BamHI-EcoRI fragment (nt 2229 to 2437)), then recircularized. The deletions in all mutants were established by DNA sequence analysis.

The construction of E3 virus mutants from plasmid mutants has been described (29). Briefly the Ad2 EcoRI-D fragment (genomic map positions 76-86) was ligated between the Ad5 EcoRI-A fragment (map position 0-76) and the Ad5 EcoRI-B fragment (map position 86-100). This DNA was used to transfect KB cells and the resulting plaques were screened by agarose gel analysis of the HindIII digested Hirt supernatant.

Nuclease Gel and RNA Blot Analysis

KB cells were infected with mutants and maintained in cycloheximide (25 μg/ml) from 3 to 9 h postinfection. Cytoplasmic and poly(A)$^+$ RNA were prepared (26). Quantitative nuclease gel analysis (31) was carried out with ^{32}P-labeled RNA probes in excess (32), 30 μg of cytoplasmic RNA, and S1 nuclease, as described (21, 28). Probes are described in figure legends. All gels are 6% sequencing gels. ^{32}P-labeled RNA size markers were prepared with SP6 polymerase.

RNA blots were done with poly(A)$^+$ RNA as described (25) using the ^{32}P-labeled RNA probes described in Fig. 7. E3-specific mRNAs in each poly(A)$^+$ RNA preparation were estimated by S1 nuclease analysis of the 1-372 exon (see Results). The same quantity of E3-specific mRNAs were loaded on the RNA blot. Total poly(A)$^+$ RNA was about 1 μg.

RESULTS

Standardization of E3-specific mRNAs in Each Mutant RNA Preparation

Before determining whether the mutations affect processing of individual E3 mRNAs, it was necessary to determine whether the mutations affect total E3 mRNA synthesis. This was done by first standardizing the abundance of adenovirus specific mRNAs synthesized by the various mutants by S1 nuclease analysis of mRNAs from region E1B. The analysis indicated that the ratio of E1B mRNAs to total cytoplasmic RNA was about the same for all mutants; this indicates that the infections were equally efficient. The abundance of total E3 mRNAs was estimated by S1 nuclease analysis of the nt 1-372 exon, with the probe in excess. This exon is present in almost all E3 mRNAs, so its abundance reflects the total number of E3 transcripts present. As shown in Fig. 2, the 1-372 exon, represented by the 372 nt nuclease-protected band, was at about the same abundance with all mutants. This indicates that none of the mutations significantly affect the total accumulation

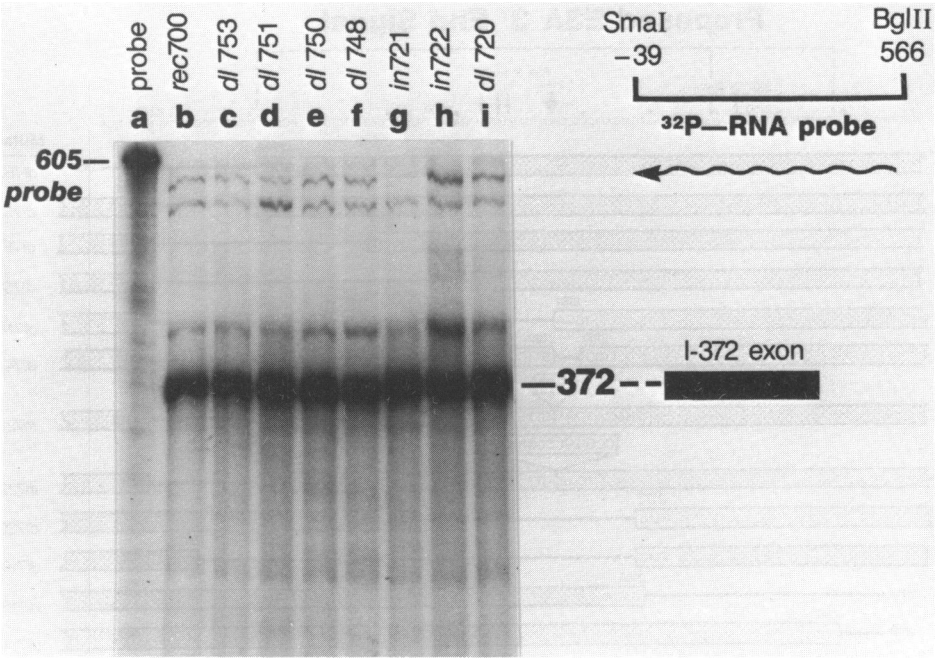

Figure 2. S1 analysis for the 1-372 exon. The ^{32}P-RNA probe, prepared with T7 poly_ merase, consists of 605 nt of rec700 sequences from the BglII (nt 566) to the Sma (nt -39) sites, plus 39 nt of pGem2 sequences at the 5' terminus. In the gel, bold roman type indicates nuclease-protected RNA size markers, italics indicate the probe, and light roman type indicates ^{32}P-labeled RNA size markers. The schematic indicates how the nuclease-resistant fragments correspond to the 1-372 exons in the mRNAs.

of E3 mRNAs. Therefore, any variations in the amount of individual E3 mRNAs, as described below, are due to changes in processing of a common pre-mRNA precursor, and not to effects on transcription, transport, or stability of the abundant mRNAs.

Mapping of Sequences Required for mRNA 3' End Formation at the E3A Site

Figure 3 illustrates the sequence at the Ad2 E3A polyadenylation site, as well as sequences deleted in the virus mutants. Sequence analysis of cDNA clones established that mRNA 3' ends form in an AT-rich region at nt 2191, 2193, and 2196 (thin arrows in Fig. 3; 33, 34). S1 nuclease mapping data indicate that 3' ends also form at nt ~2184 (thick arrow in Fig. 3). We presume that cleavage occurs after the A residue at nt 2184, based on the marked preference for cleavage after A's in many 3' end sites (35-37). In

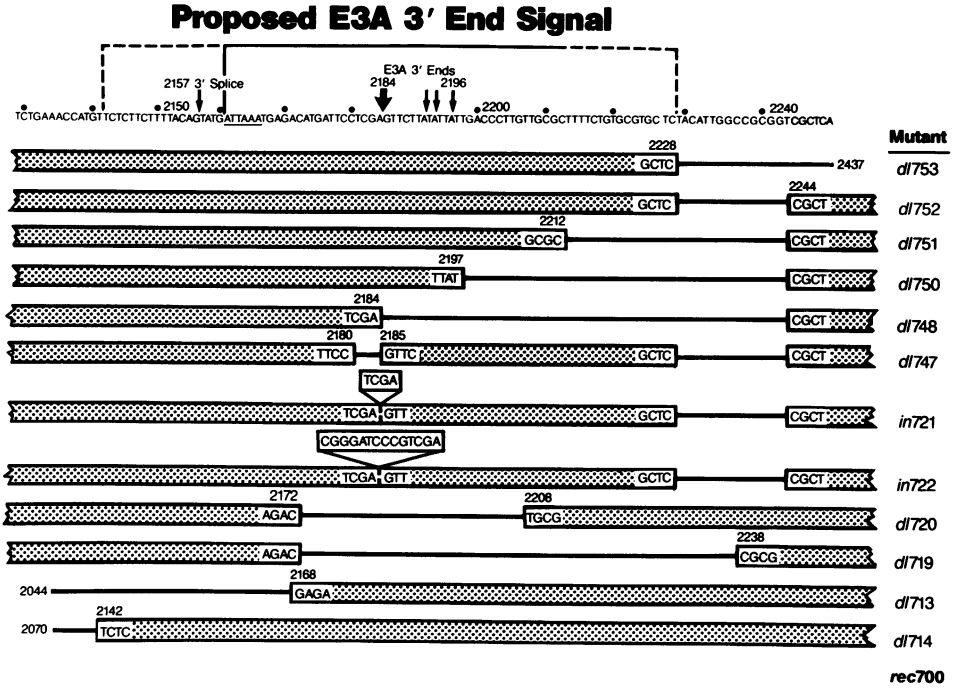

Figure 3. DNA sequence of the E3A polyadenylation site, of the 2157 3' splice site, and of sequences deleted in the virus mutants. The proposed E3A 3' end signal is shown; the dashed lines indicate regions of uncertainty. The sequences deleted in the mutants are as follows: dl753 (Δ2229-2436), dl752 (Δ2229-2243), dl751 (Δ2213-2243), dl750 (Δ2198-2243), dl748 (Δ2185-2243), dl747 (Δ2181-2184), dl720 (Δ2173-2207), dl719 (Δ2173-2237), dl713 (Δ2044-2167), dl714 (Δ2070-2141). dl752, dl751, dl750, and dl748 have a BamHI linker (CGGGATCCCG) inserted in the deletion. dl747, in721, and in722 were all constructed from dl752, so they have the same 14 bp deletion plus a 10 bp BamHI linker as in dl752.

this paper we will assume that nt 2184 is a bona fide 3' end site, even though it could be an artifact due to nibbling by S1 nuclease into the hybrid from the 3' ends formed in the AT-rich region.

Previous analyses of deletion mutants have identified the 5' boundary of the sequences required for E3A 3' end formation (Fig. 3; 21). dl713 (Δ2044-2167) is defective whereas dl714 (Δ2070-2141) is wild type: thus, the 5' boundary of the 3' end signal lies between nt 2142 and 2168. The ATTAAA at nt 2161-2166 is undoubtedly an essential component of the signal. dl719 (Δ2173-2237) is defective, indicating that sequences downstream of the ATTAAA and including the poly(A) addition sites are required.

We now describe mutants constructed to define more fully the E3A 3' end

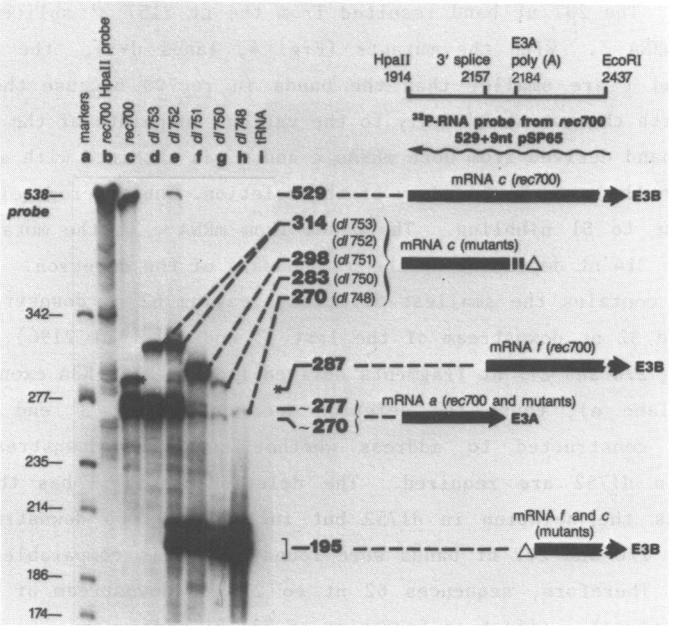

Figure 4. S1 analysis for E3A 3' end formation by mutants whose deletions approach the 3' end sites from the downstream side. The ^{32}P-RNA probe, prepared with SP6 polymerase, consists of 529 nt of rec700 sequences from the EcoRI (nt 2437) to the HpaII (nt 1914) sites, plus 9 nt of pSP65 sequences at the 5' terminus. The triangles in the schematic indicate discontinuity between the probe and the mutant mRNAs. See legend to Fig. 2 for details.

signal. dl748, dl750, dl751, and dl752 have deletions that leave the 3' end sites intact but remove downstream sequences. mRNA from cells infected by these mutants was analyzed by the nuclease gel procedure. The ^{32}P-labeled RNA probe contained Ad2 sequences from the EcoRI site (nt 2437) to the HpaII site (nt 1914) and was transcribed using SP6 polymerase (see the schematic in Fig. 4). Several fragments were observed with rec700 (wild type) RNA (Fig. 4, lane c). The 270 and 277 nt bands represent cleavage at the E3A 3' end sites, i.e. these bands are the portion of the probe protected by the 768-E3A exon of mRNA a. The exact position of 3' end formation cannot be ascertained because of the tendency of S1 nuclease to nibble into the frayed ends of hybrids. Most likely the 270 nt band corresponds to 3' ends at nt 2184, and the 277 nt band to 3' ends at several positions downstream in the AT-rich region (nt 2191-2196) (Fig. 3). The other rec700 bands derive from mRNAs c and f. The 529 nt band corresponds to mRNA c which is colinear with

the probe. The 287 nt band resulted from the nt 2157 3' splice (2157-E3B exon) in mRNA f. With the mutants (Fig. 4, lanes d-i), the bands from mRNAs c and f are smaller than the bands in rec700 because the probe is colinear with the mutant RNA only to the various endpoints of the deletions. A ~195 nt band derived from both mRNAs c and f was obtained with all mutants which share the same 3' boundary of the deletion. The heterogeneity of this band is due to S1 nibbling. The bands from mRNA c in the mutants varied from 270 to 314 nt depending on the 5' boundary of the deletion.

dl752 contains the smallest deletion, leaving 62 nt downstream of the ATTAAA, and 32 nt downstream of the last 3' end site (nt 2196). dl752 RNA gave strong 270 and 277 nt fragments derived from the 768-E3A exon of mRNA a (Fig. 4, lane e); thus, the deletion does not affect 3' end formation. dl753 was constructed to address whether sequences downstream of the deletion in dl752 are required. The deletion in dl753 has the same 5' boundary as the deletion in dl752 but includes 209 bp downstream. With dl753, the 270 and 277 nt bands were found at levels comparable to rec700 (lane d). Therefore, sequences 62 nt to 271 nt downstream of the ATTAAA have no discernable effect on formation of E3A 3' ends.

The further deletion of sequences approaching the E3A poly(A) addition sites did alter 3' end processing. Deletion of 16 additional nucleotides in dl751, to within 16 nt of the last (nt 2196) 3' site, markedly decreased 3' end formation as indicated by the reduction in the 277 and 270 nt bands (Fig. 4, lane f). The deletions in dl750 and dl748 had even more dramatic effects. The deletion in dl750 leaves 1 nt downstream of the last AT-rich 3' site: the 277 nt band cannot be seen and the 270 band is barely visible (lane g). The deletion in dl748 removes all the 3' sites in the AT-rich region, and all the sequences downstream of the nt 2184 site: 3' end formation is completely abolished (lane h). The faint ~270 nt band seen with dl748 is due to mRNA c, i.e. this is the protected fragment which extends from the 5' boundary of the deletion to the 3' end of the probe (this has been confirmed by use of a probe transcribed from a dl748 plasmid clone, data not shown). We conclude that sequences downstream of the E3A 3' sites are necessary for full efficiency of 3' end processing. Furthermore, deletion of these essential downstream sequences did not alter the position of 3' end formation. Since dl753 and dl752 have wild type phenotype, at most 32 nt of downstream sequences are sufficient for full efficiency of cleavage at all, or almost all, the 3' end sites. Since dl751 is markedly deficient in processing at all the 3' sites, the sequences deleted in dl751

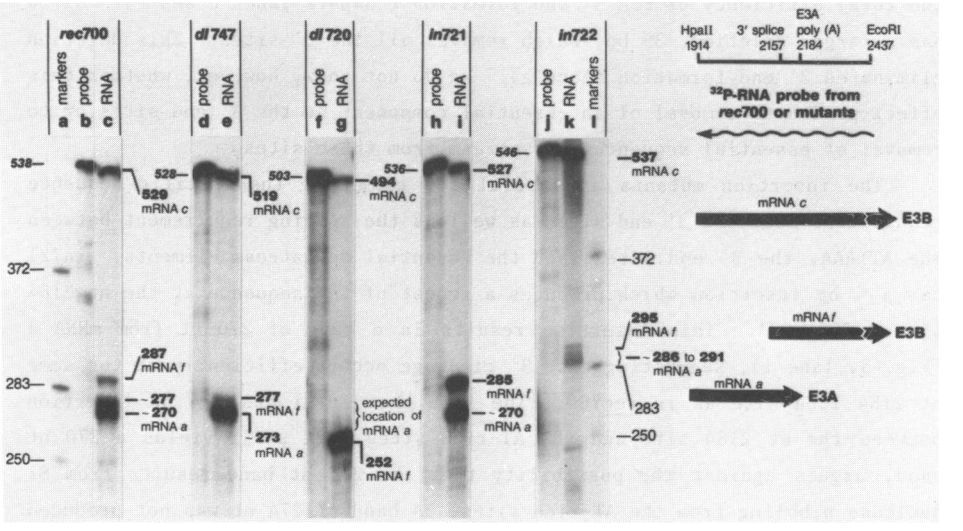

Figure 5. S1 analysis for E3A 3' end formation and for splicing at the 2157 3' splice site, by mutants with deletions or insertions near the 3' cleavage sites. Each mutant was analyzed with a probe prepared from its own DNA. See the legend to Fig. 2 for details.

and retained in dl753 must define the 3' boundary of the E3A 3' end signal; the boundary is 17-32 nt downstream of the last 3' end site.

The second series of mutants contain insertions (in721, in722) or deletions (dl747, dl720) at or near the 3' sites. in721, in722, and dl747 were constructed from the dl752 plasmid so they contain the 14 bp deletion (plus 10 bp BamHI linker) present in dl752; as discussed above, the dl752 deletion does not affect 3' end processing. These mutants were analyzed to determine whether specific sequences at the 3' sites are required. (Some of these mutants also affected the synthesis of mRNAs f and c; this will be discussed in a later section). The probe is the same as in Fig. 4 except that the probe for each mutant was transcribed from the corresponding mutant plasmid. Thus, the probes and protected fragments vary in size according to the size of the deletion or insertion.

dl747 contains a 4 bp deletion which removes the sequence TCGA which includes the nt 2184 3' site (Fig. 3). This deletion eliminated the 270 nt band and resulted in a band (or doublet) of ~273 nt (Fig. 5, lane e). Apparently the 3' ends for mRNA a form predominately in the AT-rich region, since this would yield a ~273 nt band. This deletion has little effect on

the total efficiency of E3A 3' end formation (compare lanes e and c). dl720 has a larger deletion, 35 bp, which removes all the 3' sites. This deletion eliminated 3' end formation (lane g). We do not know, however, whether this effect is due to removal of an essential component at the 3' end sites or to removal of essential sequences downstream from these sites.

The insertion mutants are useful for examining the specific sequence requirement near the 3' end sites as well as the spacing requirement between the ATTAAA, the 3' end sites, and the essential downstream elements. in721 has a 4 bp insertion which produces a repeat of the sequence at the nt 2184 site: TCGATCGA. This insertion results in a band of 270 nt from mRNA a (Fig. 5, lane i), suggesting that 3' cleavage occurs efficiently at the same nt 2184 TCGA site as in rec700. The fact that in721 has a 4 bp insertion between the nt 2184 site and the AT-rich sites, yet still yields a 270 nt band, argues against the possibility that the 270 nt band results from S1 nuclease nibbling from the AT-rich sites. A band of 274 nt was not produced in in721, indicating that cleavage did not occur at the repeated downstream TCGA. We cannot tell if cleavage occurred at the AT-rich sites, because the resulting bands of ∿281-286 nt would comigrate with the 285 nt band from mRNA f.

The larger insertion in in722 showed a considerably stronger effect. in722 contains a 14 bp insertion at the same position as the insertion in in721. The insertion leaves the TCGA at the nt 2184 site as it occurs in rec700, and results in a repeat of this TCGA 11 bp downstream. This 14 bp insertion resulted in a group of bands of ∿286-291 nt from mRNA a (Fig. 5, lane k). No fragment of 270 nt was produced, indicating that in contrast to in721, the TCGA at nt 2184 is not used. Also, the downstream repeated TCGA is not used, as this would yield a band of 283 nt. Thus, in722 apparently uses the sites in the AT-rich region, which would yield bands of ∿291 nt. The efficiency of 3' end formation is reduced to less than half that of rec700. Thus, the insertion in in722 abolishes 3' end formation at the nt 2184 site, and reduces the efficiency at the AT-rich sites.

Mutations That Abolish E3A 3' End Formation Result in Enhanced Use of the 3' Splice Site at Nucleotide 2157

During the analysis for E3A 3' end formation, we observed that several mutants had increased amounts of the 2157-E3B exon from mRNA f; this would result from enhanced use of the nt 2157 3' splice site. This is shown in Fig. 5 for some of the mutants. About the same quantity of the 2157-E3B exon was formed by rec700 (287 nt band in lane c), by dl747 (277 nt band in

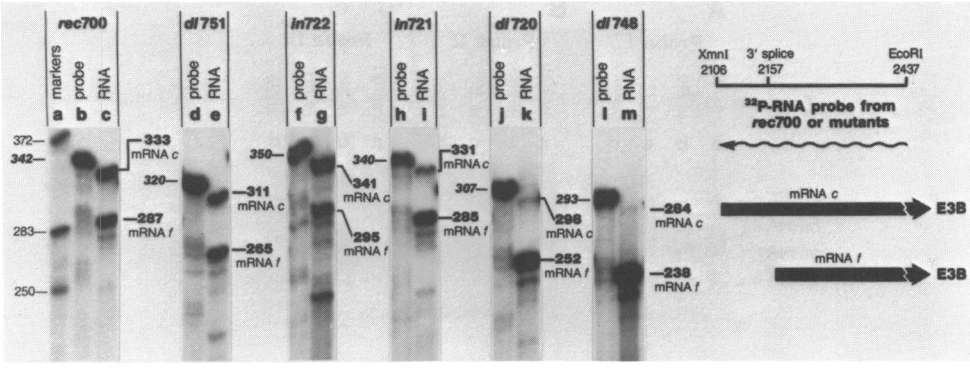

Figure 6. S1 analysis for splicing at the 2157 3' splice site. Each mutant was analyzed with a probe prepared from its own DNA. See the legend to Fig. 2 for details.

lane e), and by in722 (295 nt band in lane k). The 2157-E3B exon is dramatically increased in dl720 (252 nt band in lane g). The exon appears to be elevated in in721 (285 nt band in lane i), although this cannot be ascertained by this gel because the bands for the 2157-E3B exon of mRNA f and those for the 768-E3A exon of mRNA a would comigrate. To circumvent this comigration problem we used a shorter probe, extending from the EcoRI site (nt 2437) to the XmnI site (nt 2106). This probe overlaps fully with mRNA c, almost fully with mRNA f, but only slightly with mRNA a (Fig. 6). The probe for each mutant was transcribed from the corresponding plasmid, so the lengths of the probes and protected fragments vary, but in all cases the larger fragment represents mRNA c and the smaller fragment represents mRNA f.

As expected for rec700, protected fragments of 333 nt (mRNA c) and 287 nt (2157-E3B exon of mRNA f) were formed (Fig. 6, lane c). In this experiment the frequency of the 2157 3' splice is best illustrated by the ratio of mRNA f to that of the readthrough mRNA c. The relative frequency of the 2157 3' splice among the mutants may also be roughly estimated from the intensity of the bands corresponding to the 2157-E3B exon; however, this must be viewed with caution because we do not know that the specific activities of the different probes are the same. For rec700, note that the mRNA c and f fragments are at about equal abundance. These mRNA c and f fragments are also at the same abundance with dl751 (lane e) and in722 (lane g), as well as with dl747 and dl752 (data not shown). In contrast, with in721 the abundance of mRNA f is increased relative to mRNA c as would be expected

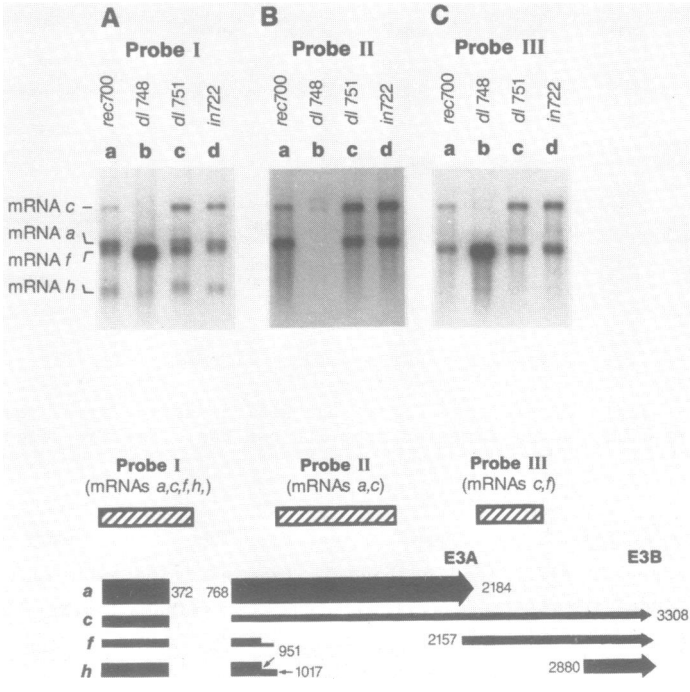

Figure 7. RNA blot analysis for synthesis of E3 mRNAs. Probes were ^{32}P-labeled SP6 RNAs. Probe I consists of rec700 sequences from nt 27 to 560, probe II of nt 959 to 1694, and probe III of nt 2181 to 2442.

from increased use of the 2157 3' splice site (lane i). The effect is much more obvious with dl720 (lane k), dl748 (lane m), and dl750 (data not shown), where the band for mRNA f is very abundant and that for mRNA c is barely visible. Also, for these mutants the band for mRNA f is much more intense than the band for mRNA f obtained with rec700, dl747, and in722. We conclude that in addition to abolishing E3A 3' end formation, the deletions in dl720, dl750, and dl748 result in a dramatic increase in splicing at the nearby nt 2157 3' splice site. The 4 bp insert in in721 somewhat (~2 fold) increases use of the 2157 3' splice.

RNA blot analysis was used to confirm the nuclease gel results. The ^{32}P-labeled RNA probes used are described in Fig. 7. Probe I is specific to the first exon and therefore will hybridize to all E3 mRNAs. This probe shows that mRNA a is very abundant in rec700, representing about half of the total E3 mRNA, whereas mRNAs c, f, and h are less abundant (Fig. 7A, lane a). The identity of the bands as the mRNAs shown in Fig. 7 was established

previously (25). (The faint band above mRNA c is the scarce unspliced mRNA i which extends from the mRNA cap site to the E3B polyadenylation site (25)). As expected from the nuclease gel data, mRNA a is absent in dl748 (lane b), and is reduced in dl751 (lane c) and in722 (lane d). mRNA f is very abundant in dl748, and mRNAs c and h are very scarce.

Since mRNAs a and f nearly comigrate, additional probes were used to distinguish them on the blot. Probe II, which hybridizes only to mRNAs a, c, and i, confirms that mRNA a in dl751 (Fig. 7B, lane c) and in722 (lane d) is reduced in comparison to rec700 (lane a). In dl748 (lane b), mRNA a was not detected, and mRNA c (and i) was much reduced. Probe III, which hybridizes only to mRNAs c, f, and i, establishes that the abundant mRNA in dl748 is mRNA f (Fig. 7C, lane b). We conclude that the RNA blots confirm the results of the nuclease gel analysis. In mutants dl751 and in722 which decrease the efficiency of 3' end formation (but don't abolish it), mRNA a is reduced but the other E3 mRNAs continue to be made (in slightly increased abundance compared to rec700). In dl748 which abolishes E3A 3' end formation, only mRNA f is made in abundance. The same result was found with dl720, dl750, and dl719 which are also E3A 3' end mutants (data not shown). In the RNA blot analysis of dl751, mRNA a and mRNA c appear to be about equally abundant (Fig. 7, lanes c) in contrast to the nuclease gel results where mRNA c appears to be more abundant (Fig. 4, lane f). With nuclease gel analysis, an mRNA which protects a greater length of probe results in a stronger band, and heterogeneity due to S1 nibbling makes other bands appear fainter. Thus, RNA blot analysis is more representative of the relative abundances of all E3 mRNAs for a particular mutant than is nuclease gel analysis.

DISCUSSION

E3A 3' End Processing Signal

A previous study established that the 5' boundary of the E3A 3' end signal is located between nt 2142 and nt 2168 (see dl714 and dl713 in Fig. 3), and includes the ATTAAA at nt 2161-2166 (21). We have now mapped the 3' boundary of the E3A 3' end signal. It is located somewhere between nt 2213 and 2228, i.e. the sequences deleted in dl751 (partially defective) but retained in dl752 and dl753 (both wild type). The 3' boundary is 47-62 nt downstream of the ATTAAA, and 17-32 nt downstream of the last 3' end site at nt 2196.

A requirement has been reported for sequences downstream of other 3'

end processing sites both in vivo with expression vectors (11-13, 15, 16, 18, 20, 38, 39) and in vitro (14, 15). The essential downstream element is roughly 20 nt long and is located a few nucleotides downstream from the 3' end site. Our results with E3A are in agreement with these conclusions; this is significant because the E3A site and our method of analysis differ in several ways from most other studies. First, E3A has ATTAAA instead of AATAAA. Second, our study was done with virus mutants which are transcribed in a "natural" context, rather than with plasmid expression vectors. In another study with virus mutants, Sadofsky and Alwine (17) reported that a 39 bp deletion located 8 bp downstream from the SV40 late poly(A) addition site reduced the efficiency of 3' end formation at the normal site. Third, and of particular interest, E3A is the "upstream" polyadenylation site in a complex transcription unit with two alternative polyadenylation sites. Nevertheless, the E3A signal appears to be similar to other polyadenylation signals. We note that deletion of sequences 62 to 271 nt (dl753) downstream from the ATTAAA does not affect E3A 3' end formation, indicating that this region does not determine the choice between the E3A and E3B sites. The Ad2 E2A polyadenylation site, which has been analyzed in an expression vector (13, 16) and in vitro (14, 15), is also an upstream polyadenylation site analogous to E3A. The E2A site seems similar to E3A, although the essential downstream element differs in sequence (see below). Also, the E2A site apparently has redundant downstream elements which do not seem to be present in E3A.

Other workers have reported that deletions downstream from the 3' end site change its position but have little effect on its efficiency (19, 36). Our results differ from these reports. Mason et al. (36) suggested that the negative effect of deletions might be to bring nonspecific "poison" downstream sequences into proximity with the AATAAA and the 3' end site. We cannot exclude this possibility, even though our downstream deletions have the same 3' boundary.

Sequences downstream from 3' end sites are not highly conserved, although a loose consensus sequence, YGTGTTYY, was found ~30 nt downstream of AATAAA in 67% of a large number of cases examined (39). Recent studies suggest that the requirement is merely for a GT-rich element; deletion of this downstream element in the SV40 early and HSV-1 TK polyadenylation sites prevents 3' end formation (13, 20, 38). Studies on the Ad2 E2A polyadenylation signal indicate that a second type of element may exist at the same position as the GT-rich element, but one that is T-rich (13, 15). Sequences

of the form $(T)_n$ $(A)_p$ $(T)_q$ (40) and CAYTG (35) have also been noted at polyadenylation sites.

Examples of these various consensus sequences are found in E3A. CGAGTTCT, which conforms to the YGTGTTYY consensus, is at nt 2182-2189 which includes the 3' end site at nt 2184. This sequence is partially deleted in dl747, and it is interrupted by the 4 bp insertion in in721, both with little effect. Thus, either this sequence is not important or redundant elements are present. A region TTCTTATATTATT at nt 2186 to 2198 is both T-rich and has the form $(T)_n$ $(A)_p$ $(T)_q$. This region is deleted in dl720 where E3A 3' end formation is absent. However, this cannot be the entire signal since dl751 which has this sequence intact is markedly defective. There is a ~33 bp GT-rich region (nt 2197-2229) downstream from the last 3' end site at nt 2196, and which is retained in dl752 and dl753 which form 3' ends normally. This GT-rich region must be important because deletion of 16 nt (dl751) of it from the downstream side reduces 3' end formation very significantly, and deletion of the entire region (dl750) reduces 3' end formation to a trace. However, dl750 does have residual 3' end activity, so this GT-rich region cannot be completely obligatory. Only with dl748, whose deletion removes the 3' end sites in the AT-rich region as well as all sequences downstream from the nt 2184 site, were we unable to detect 3' end formation. This suggests that the AT-rich region (nt 2186-2198) downstream from the nt 2184 site is of some importance, at least for the 2184 site.

Our results with dl747, in721, and in722 address the role of sequences at the 3' cleavage sites as well as the spacing between the various elements of the E3A 3' end signal. The 4 bp deletion in dl747, which removes the site at nt 2184, results in the alternative use of the other sites in the AT-rich region. The 4 bp insertion in in721 does not prevent use of the nt 2184 site. The efficiency of 3' end formation is not significantly affected by these two mutations. The 35 bp deletion in dl720, which removes all the 3' end sites, destroys 3' end formation. However, this could be due to partial removal of the essential downstream elements. The 14 bp insertion in in722 reduces the efficiency of 3' end formation to less than 50% of wild type. Of interest, 3' end formation at the nt 2184 site is abolished. One explanation for this result is that the insert inhibits the binding of trans-acting factors required for cleavage. An alternate explanation is that the insert changes the spacing between different elements of the E3A signal. For example, the insert increases the distance between the nt 2184 site and the essential downstream element, and this could eliminate use of

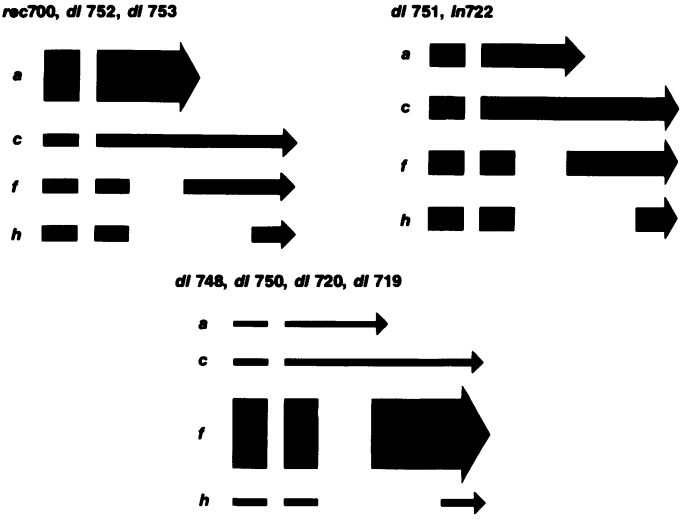

Figure 8. Schematic illustrating the abundances of E3 mRNAs in three phenotypic groups of mutants. The thickness of the arrows represents the relative abundance of the mRNA.

the nt 2184 site. Also, the insert increases the distance between the ATTAAA and the AT-rich sites, and this could reduce the efficiency of these sites. Studies on other polyadenylation sites, Ad2 E2A (15) and HSV-1 TK (20), have also indicated that the spacing between the AATAAA and the essential downstream element are important in the efficiency of 3' end formation.

Our view of the E3A 3' end signal is illustrated in Fig. 9. We conclude that there are at least two critical sequence elements for E3A 3' end formation, the ATTAAA and a short region located immediately downstream of the poly(A) addition sites. Sequences at the actual cleavage site may play some role although not a crucial role. The result with in722 suggests that spacing may be important. The spacing constraint implies an interaction between the required sequence elements, either directly through RNA secondary structure (e.g. see refs. 16, 17, 19) or indirectly or through trans-acting factors. If different factors (or complexes of factors) associate with the E3A signal elements, a change in the spacing between these elements might reduce the efficiency of such an interaction. However, the spacing cannot be stringent since the 4 bp insertion and deletion had no major effects.

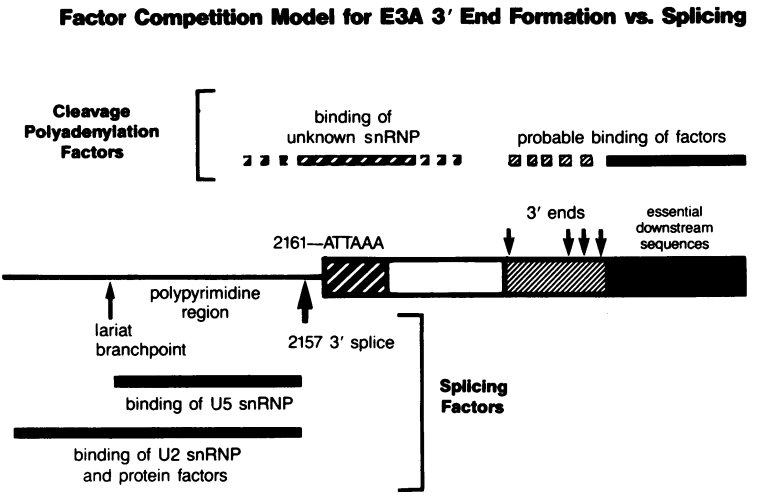

Figure 9. Schematic illustrating the E3A 3' end signal as well as the factor competition model for E3A 3' end formation versus splicing. The thick bar indicates the E3A 3' end signal as well as its presumed different functional elements. The black area indicates the essential downstream element (which may extend further to the left than is shown), the light hatched area indicates the 3' cleavage sites (nts 2184, 2191, 2193, 2196) and the heavy hatched area indicates the ATTAAA sequence which begins at nt 2161. The 2157 3' splice (used by mRNA f) is shown with its essential polypyrimidine region (24) and lariat branchpoint. The binding regions for snRNPs are based on data from references 10, 41, 44, 45.

Control of Alternative Pre-mRNA Processing

We have identified with dl748, dl750, and dl720 a region which affects two aspects in pre-mRNA processing. The region is part of the E3A 3' end signal, and it plays a major role in splicing at the nt 2157 3' splice site. The deletion of exon sequences in these mutants results in dramatically enhanced splicing at the 2157 3' splice site. The effects of the various mutants on E3A 3' end formation and splicing at the 2157 3' splice site are summarized in Fig. 8. Our working model on how this region may function in alternative pre-mRNA processing is illustrated in Fig. 9. The key feature is that splicing factors and cleavage/polyadenylation factors compete for binding to overlapping sequences in pre-mRNA. That such competition can occur is strongly suggested by the fact that the 2157 3' splice site is only 4 nt upstream of the ATTAAA element of the E3A signal. Splicing factors would be the snRNP, probably U5 snRNP, which binds near the 3' splice site (41), and U2 snRNP and proteins which bind to the lariat branchpoint site and splice site (42-44). A cleavagepolyadenylation factor would be the

unidentified snRNP which binds to the AATAAA (10). Presumably other unknown factors bind to the essential downstream element and the poly(A) addition sites. Binding of factors to the downstream elements and their interaction with the snRNP which binds to the ATTAAA would tend to block the binding of splicing factors to the 2157 3' splice site. This would limit use of the 2157 3' splice site. Deletion of the essential downstream element as has occurred in dl748, dl750, dl720, and dl719 would prevent this blocking effect by eliminating the binding of the snRNP to the ATTAAA. Thus, with these mutants, the 2157 3' splice is used with very high efficiency. With dl751 and in722, which are markedly defective in E3A 3' end formation yet do not exhibit specifically enhanced use of the 2157 3' splice site, we presume that they retain the capacity to bind cleavage/polyadenylation factors that block the 2157 3' splice site. We note that binding of factors to the lariat branchpoint and 3' splice sites is thought to be among the earliest events in splicing and in splicing complex formation (41-44). Thus, blockage of the 2157 site, as proposed above, should limit the assembly of the 2157 site into the splicing complex in the correct configuration for the 2157 splice. Our model does not exclude the possibility that alternative pre-mRNA secondary structures are important in the choice between the E3A 3' end site and the 2157 3' splice site. For example, the cleavage/polyadenylation factors could stabilize the pre-mRNA in the configuration for 3' end formation rather than splicing.

ACKNOWLEDGEMENTS

This work was supported by grant DMB 8605560 from the National Science Foundation and by grant CA24710 from the Public Health Service.

*To whom correspondence should be addressed

REFERENCES

1. Leff, S.E., and Rosenfeld, M.G. (1986) Ann. Rev. Biochem. 55, 1091-1117.
2. Green, M.R. (1986) Ann. Rev. Genet. 20, 671-708.
3. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., and Sharp, P.A. (1986) Ann. Rev. Biochem. 55, 1119-1150.
4. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature (London) 263, 211-214.
5. Fitzgerald, M., and Shenk, T. (1981) Cell 24, 251-260.
6. Montell, C., Fisher, E.F., Caruthers, M.H., and Berk, A.J. (1983) Nature (London) 305, 600-606.
7. Wickens, M., and Stephenson, P. (1984) Science 226, 1045-1051.

8. Manley, J. L., Yua, H., and Ryner, L. (1985) Mol. Cell. Biol. 5, 373-379.
9. Zarkower, D., Stephenson, P., Sheets, M., and Wickens, M. (1986) Mol. Cell. Biol. 5, 2975-2983.
10. Hashimoto, C., and Steitz, J.A. (1986) Cell 45, 581-591.
11. Conway, L., and Wickens, M. (1985) Proc. Natl. Acad. Sci. USA 82, 3949-3953.
12. Gil, A., and Proudfoot, N.J. (1984) Nature (London) 312, 473-474.
13. Hart, R.P., McDevitt, M.H., Ali, H., and Nevins, J.R. (1985) Mol. Cell. Biol. 5, 2975-2983.
14. Hart, R.P., McDevitt, M.H., and Nevins, J.R. (1985) Cell 43, 677-683.
15. McDevitt, M.A., Hart, R.P., Wong, W.W., and Nevins, J.R. (1986). EMBO J. 5, 2907-2913.
16. McDevitt, M.A., Imperiale, M.J., Ali, H., and Nevins, J.A. (1984) Cell 37, 993-999.
17. Sadofsky, M., and Alwine, J.C. (1984) Mol. Cell. Biol. 4, 1460-1468.
18. Simonsen, C.C., and Levinson, A.D. (1983) Mol. Cell. Biol. 3, 2250-2258.
19. Woychik, R.P., Lyons, R.H., Post, L., and Rottman, R.M. (1984) Proc. Natl. Acad. Sci. USA 81, 3944-3948.
20. Zhang, F., Denome, R.M., and Cole, C.N. (1986) Mol. Cell. Biol. 5, 2975-2983.
21. Bhat, B.M., and Wold, W.S.M. (1985) Mol. Cell. Biol. 5, 3183-3193.
22. Bhat, B.M., and Wold, W.S.M. (1986) J. Virol. 57, 1155-1158.
23. Bhat, B.M., and Wold, W.S.M. (1986) J. Virol. 60, 54-63.
24. Bhat, B.M., Brady, H.A., and Wold, W.S.M. (1985) Mol. Cell. Biol. 5, 2405-2413.
25. Bhat, B.M., Brady, H.A., Pursley, M.H., and Wold, W.S.M. (1986) J. Mol. Biol. 190, 543-557.
26. Cladaras, C., Bhat, B., and Wold, W.S.M. (1985) Virology 140, 44-54.
27. Cladaras, C., and Wold, W.S.M. (1985) Virology 140, 28-43.
28. Deutscher, S.L., Bhat, B.M., Pursley, M.H., Cladaras, C., and Wold, W.S.M. (1985) Nucleic Acids Res. 13, 5771-5788.
29. Wold, W.S.M., Deutscher, S.L., Takemori, N., Bhat, B.M., and Magie, S.C. (1986) Virology 148, 168-180.
30. Lathe, R., Kieny, M.P., Skory, S., and Lecocq, J.P. (1984) DNA 3, 173-180.
31. Berk, A.J., and Sharp, P.A. (1978) Cell 12, 721-732.
32. Melton, D.A., Kreig, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984) Nucleic Acids Res. 12, 7035-7056.
33. Ahmed, C.M.I., Chanda, R., Stow, N., and Zain, B.S. (1982) Gene 19, 297-301.
34. Stalhandske, P., Persson, H., Perricaudet, M., Philipson, L., and Pettersson, U. (1983) Gene 22, 157-165.
35. Berget, S.M. (1984) Nature (London) 309, 179-182.
36. Mason, P.J., Elkington, J.A., Lloyd, M.M., Jones, M.B., and Williams, J.G. (1986) Cell 46, 263-270.
37. Moore, C.L., Skolnik-David, H., and Sharp, P.A. (1986) EMBO J. 5, 1929-1938.
38. Cole, C.N., and Stacy, T.P. (1985) Mol. Cell. Biol. 5, 2104-2113.
39. McLauchlin, J., Gaffney, D., Whitton, J.L., and Clements, J.B. (1985) Nucleic Acids Res. 13, 1347-1368.
40. LeMoullec, J.M., Akŭsjarvi, G., Stalhandske, P., Pettersson, U., Chambrand, B., Gilardi, P., Nasri, M., and Perricaudet, M. (1983) J. Virol. 48, 127-134.
41. Chabot, B., Black, D.L., LeMaster, D.M., and Steitz, J.A. (1985) Science 230, 1344-1349.

42. Black, D.L., Chabot, B., and Steitz, J.A. (1985) Cell 42, 737-750.
43. Ruskin, B., Greene, J.M., and Green, M.R. (1985) Cell 41, 833-844.
44. Konarska, M.M., and Sharp, P.A. (1986) Cell 46, 845-855.
45. Black. D.L., and Steitz, J.A. (1986) Cell 46, 694-704.