MOLECULAR AND CELLULAR BIOLOGY, Aug. 1994, p. 5278–5289
0270-7306/94/$04.00+0

Vol. 14, No. 8

Sequences Homologous to 5′ Splice Sites Are Required for the Inhibitory Activity of Papillomavirus Late 3′ Untranslated Regions

PRISCILLA A. FURTH,† WON-TAEK CHOE,‡ JANELL H. REX, JANET C. BYRNE, AND CARL C. BAKER*

Laboratory of Tumor Virus Biology, National Cancer Institute, Bethesda, Maryland 20892

Received 14 March 1994/Returned for modification 8 May 1994/Accepted 16 May 1994

Expression of bovine papillomavirus type 1 (BPV-1) late genes is limited to terminally differentiated keratinocytes in an infected epithelium. We have previously shown that although the BPV-1 late polyadenylation site is functional in nonpermissive cells, a 53-nucleotide (nt) fragment of the late 3′ untranslated region acts posttranscriptionally to reduce polyadenylated cytoplasmic RNA levels. This 53-nt fragment does not appear to function by destabilizing polyadenylated cytoplasmic RNA (P. A. Furth and C. C. Baker, J. Virol. 65:5806–5812, 1991). In this study, we used site-directed mutagenesis and deletion analysis to demonstrate that the sequence AAG/GUAAGU, which is identical to the consensus 5′ splice site sequence, was both necessary and sufficient for the inhibitory activity of the 53-nt fragment. Furthermore, base pairing between the 5′ end of the U1 small nuclear RNA and this 5′ splice site-like sequence was shown to be required for the inhibitory activity in vivo. We have also further mapped the human papillomavirus type 16 late 3′ inhibitory element (I. M. Kennedy, J. K. Haddow, and J. B. Clements, J. Virol. 65:2093–2097, 1991) to a 51-nt region containing four overlapping sequence motifs with partial homology to 5′ splice sites. Mutation of each of these motifs demonstrated that only one of these motifs is required for the inhibitory activity. However, the presence of the other motifs may contribute to the full inhibitory activity of the element. No BPV-1 or human papillomavirus type 16 mRNAs which are spliced by using the potential 5′ splice sites present in the viral late 3′ untranslated regions have been identified. This suggests that the primary function of these 5′ splice site-like sequences is the inhibition of late gene expression. The most likely mechanism of action of these elements is reduction of polyadenylation efficiency, perhaps through interference with 3′-terminal exon definition.

Most eukaryotic mRNAs are polyadenylated during processing of the pre-mRNA. Cleavage and polyadenylation of a pre-mRNA requires sequence elements both within the 3′ untranslated region (UTR) and downstream of the cleavage site (reviewed in references 38, 61, and 80). The highly conserved AAUAAA sequence (or close variants) is located 10 to 30 nucleotides (nt) upstream of the cleavage site and is required for both cleavage and polyadenylation. Additional U-rich and GU-rich elements, known as downstream elements (DSEs), are located immediately downstream of the cleavage site. These elements determine the stability of the polyadenylation complex and therefore the efficiency of the poly(A) site (31, 79). In several viral systems, including simian virus 40 (SV40), adenovirus type 2, ground squirrel hepatitis virus, cauliflower mosaic virus, and human immunodeficiency virus type 1, U-rich elements (USEs) upstream of the AAUAAA polyadenylation signal increase the efficiency of the poly(A) site and are involved in regulating poly(A) site choice (see references 19, 20, 24, 38, 61, 64, and 74 and references therein). Both USEs and DSEs are U rich, suggesting that they may also be functionally similar (75). In fact, the human immunodeficiency virus USE has been shown to enhance the stability of the polyadenylation complex on the core poly(A) site in vitro, a function reminiscent of the DSEs (30).

There is increasing evidence that splicing and polyadenylation are coupled and can influence each other. The presence of an intron in a pre-mRNA stimulates polyadenylation in vitro (58) and in vivo (37). This suggests that 3′-terminal exons are recognized as a unit and that splicing and polyadenylation factors communicate with each other across the exon (58). Wassarman and Steitz (77) have suggested that coupling is mediated by weak interactions between the U1 small nuclear ribonucleoprotein (snRNP) and the region of the pre-mRNA upstream of the poly(A) site. In contrast, strong interactions between the U1 snRNP and a 5′ splice site in a terminal exon appear to block coupling between splicing and polyadenylation and consequently inhibit polyadenylation (57, 77). These interactions require base pairing between the 5′ end of the U1 snRNA and the pre-mRNA. Elements which influence polyadenylation efficiency have also been shown to interact with snRNP-associated proteins. Lutz and Alwine (50) recently showed that the positively acting USE of the SV40 late poly(A) site specifically binds the U1 snRNP-A (U1A) protein and that this interaction is required for the function of the element. The U1A protein also binds sequences in the 3′ UTR of the U1A pre-mRNA and inhibits polyadenylation of the U1A pre-mRNA (11, 33, 76). Thus, splicing factors can have both positive and negative effects on polyadenylation.

We have used the papillomaviruses as a model system for studying posttranscriptional regulation of gene expression during cellular differentiation. Bovine papillomavirus type 1 (BPV-1), which transforms rodent cells in culture and induces fibropapillomas or warts in cattle, is the best studied of the

* Corresponding author. Mailing address: Building 41, Room C111, National Institutes of Health, Bethesda, MD 20892. Phone: (301) 496-2078. Fax: (301) 402-0055.

† Permanent address: Division of Infectious Diseases, Department of Medicine, University of Maryland Medical School, Baltimore, MD 21201. (On a leave of absence during this study.)

‡ Present address: Harvard College, Cambridge, MA 02138.

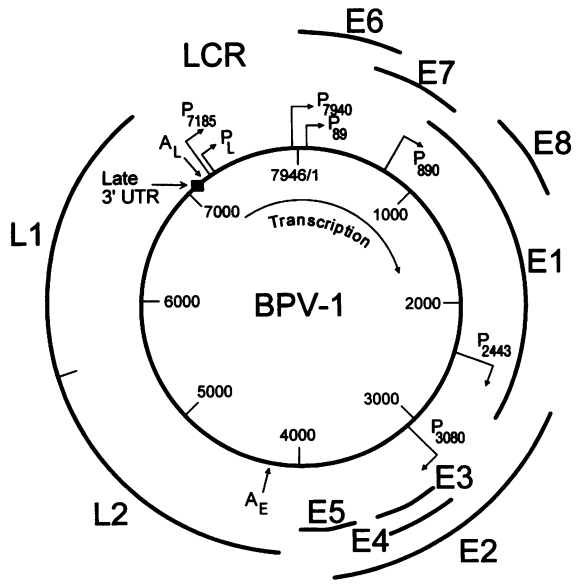

FIG. 1. BPV-1 genomic map. Numbers inside the circle indicate nucleotide positions. The arcs represent the ORFs which lie on only one strand. ORFs in the early region are labeled E; late-region ORFs are labeled L. Early promoters are indicated by an arrow labeled P_n, where n is the approximate nucleotide position of the RNA initiation site. P_L is the late promoter whose initiation site maps between nt 7214 and 7256. The early (A_E) and late (A_L) poly(A) sites are located at nt 4203 and 7175, respectively. LCR, long control region.

papillomaviruses and has served as the prototype for molecular studies of this group of viruses. BPV-1 DNA is maintained extrachromosomally in both transformed cells and productively infected fibropapillomas (44, 45). Both early and late genes are transcribed from a single complex transcription unit which occupies one entire strand of the 8-kb circular DNA viral genome (Fig. 1) (for a review, see reference 4). The complex pattern of splicing and polyadenylation observed in BPV-1 mRNAs suggests the importance of posttranscriptional regulation in BPV-1 gene expression. In the transformed cell, early-region mRNAs are transcribed from multiple promoters but are polyadenylated at a single poly(A) site (A_E) located at the 3′ end of the early region (4). Late-region mRNAs, which encode the viral capsid proteins, are expressed only in the fully differentiated keratinocytes of a productively infected fibropapilloma (70). These mRNAs are transcribed from the late promoter (P_L) which is active only in differentiated keratinocytes and are polyadenylated at a fibropapilloma-specific poly(A) site (A_L) located at the 3′ end of the late region (7, 9). Although no cytoplasmic mRNAs are expressed from the late region in BPV-1-transformed murine C127 cells, nuclear runoff analysis indicates that there is significant transcription of the late region in these cells (8). Thus, early in infection and in BPV-1-transformed cells, there must be mechanisms which either prevent pre-mRNAs transcribed from early promoters from being polyadenylated at the late poly(A) site or destabilize any late mRNAs which are made. In contrast, late in infection, pre-mRNAs transcribed from P_L are polyadenylated at both poly(A) sites and are relatively stable (7, 29). Transcription termination, or perhaps more likely transcriptional pausing, between the early and late poly(A) sites may significantly influence poly(A) site choice in BPV-1-infected cells (8, 27). The BPV-1 late polyadenylation signal (AAUAAA) and downstream GU-rich element function efficiently in nonpermissive cells. However, a negative element in the late 3′ UTR acts at a posttranscriptional level to decrease steady-state levels of cytoplasmic late mRNAs in transfection studies and may contribute to the lack of BPV-1 late gene expression in nonpermissive cells (29). The results of actinomycin D chase experiments suggest that this element does not function by destabilization of cytoplasmic mRNA (29). The BPV-1 negative 3′ UTR element is also able to inhibit gene expression when it is cloned into the 3′ UTR of a chloramphenicol acetyltransferase (CAT) expression vector (29). A negative element located at the 3′ end of the human papillomavirus type 16 (HPV-16) L1 coding region and extending into the 3′ UTR also inhibits CAT expression in transfection assays in nonpermissive cells (40, 41). This suggests that the presence of an inhibitory element in the late 3′ UTR may be a common regulatory feature of papillomaviruses. Although the HPV-16 negative element has been shown to destabilize RNA in an in vitro RNA degradation assay (41), this function has never been confirmed by an in vivo assay and may be an artifact of the in vitro assay.

In a continuation of our previous studies (29), we have now demonstrated by deletion and point mutagenesis that sequences homologous to a 5′ splice site (AAG/GUAAGU) are both necessary and sufficient for the inhibitory activity of the BPV-1 late 3′ UTR element. Furthermore, the function of this element requires the base pairing between the element and the 5′ end of the U1 snRNA. In addition, we further mapped the HPV-16 late 3′ UTR inhibitory element to a 51-nt region which contains four overlapping sequence motifs with partial homology to 5′ splice sites. At least one of these motifs appears to be required for the activity of the HPV-16 element. Since no mRNAs which utilize either the BPV-1 or HPV-16 potential 5′ splice sites present in the viral late 3′ UTRs have been identified, it is likely that the primary function of these sequences is the inhibition of late gene expression in nonpermissive cells. We propose that these sequences reduce the efficiency of polyadenylation at the late poly(A) site by interfering with 3′-terminal exon definition.

MATERIALS AND METHODS

Expression vectors. Nucleic acid manipulations, bacterial transformations, recombinant screening, and plasmid preparations were performed by standard methods (53). High-level-expression SV40-based CAT vectors with a polylinker in the 3′ UTR have been described previously (5, 29). These vectors were used to test the effects of specific nucleotide mutations and deletions on the inhibitory activity of the BPV-1 3′ UTR fragment, to map the HPV-16 late gene inhibitory element, and to test the effects of specific nucleotide mutations in the HPV-16 element. The structure of pOBCAT20 (p2518, CCB186-1) is shown in Fig. 2. Plasmid pOBCAT19 (p2643, CCB185-1) is similar but contains the 3′ UTR polylinker in the opposite orientation. BPV-1 and HPV-16 DNA fragments with *Sal*I- and *Bgl*II-compatible ends were generated by oligonucleotide synthesis and cloned into pOBCAT20 and pOBCAT19 linearized at the *Sal*I and *Bgl*II polylinker cloning sites in the 3′ UTR to generate constructs which contain the fragments in either the sense or antisense orientation as described previously (29). The sequences of the cloned fragments were verified by DNA sequencing. In several cases, spontaneous mutations were also generated. The plasmids containing sense inserts are identified in the legends to Fig. 3, 4, and 7. Antisense inserts had no significant effect on CAT expression; these plasmids have therefore not been included in this report. CAT expression vectors containing the wild-type (wt) BPV-1

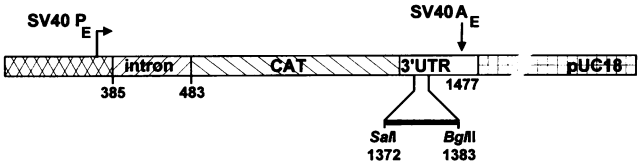

FIG. 2. Map of the pOBCAT20 expression vector. The map shows the parental CAT vector pOBCAT20 (p2518, CCB186-1) (5) used to assay the effects of the BPV-1 late 3′ UTR element on gene expression. The relative positions of the SV40 early promoter (P_E) and SV40 early poly(A) site (A_E) are indicated by arrows. Numbers below the map indicate nucleotide positions of the 5′ and 3′ splice sites, *Sal*I and *Bgl*II cloning sites, and poly(A) site.

53-bp fragment in sense (CCB250) and antisense (CCB249) orientations have been described previously (29).

Plasmids containing wt and mutant U1 snRNA genes were obtained from Alan Weiner (83, 85). A β-galactosidase expression vector (pCH110) was obtained from Pharmacia.

Cells. BPV-1-transformed murine C127 cells (clone H2) were maintained in Dulbecco's modified Eagle medium supplemented with 10% fetal calf serum, antibiotics, and glutamine.

Transfections. BPV-1-transformed C127 cells were transfected with the CAT expression vectors by the calcium phosphate method and harvested at 48 h as described previously (29). These transfections used 5 μg of CAT plasmid, 2.5 μg of pCH110, and 12.5 μg of pUC18 DNA. The medium was supplemented with 5 mM sodium butyrate until harvest.

Transfection mixtures for U1 suppression assays contained 5 μg of CAT plasmid, 5 μg of U1 plasmid, 2.5 μg of pCH110, and 7.5 μg of pUC18 DNA. Cells were harvested at 48 h posttransfection after incubation in medium which did not contain sodium butyrate.

Assays for CAT and β-galactosidase. Cellular extracts were prepared by freeze-thawing. CAT activity was assessed by standard methods; the acetylated and unacetylated forms of chloramphenicol were quantitated by thin-layer chromatography and radioanalytic imaging (AMBIS or Molecular Dynamics PhosphorImager) and expressed as percent acetylation. β-Galactosidase activity was measured and used to correct CAT activity for differences in transfection efficiency and recovery as previously described (5). Fold inhibition for each insert was then calculated by dividing the normalized CAT activity for the parent CAT vector with no insert by the normalized CAT activity for the CAT vector with the insert. Values for fold inhibition were the average of several independent transfections. Standard deviations were calculated using the Lotus 123 @std function.

Computer predictions of RNA secondary structures. RNA secondary structure analysis of the BPV-1 late region 3′ UTR was performed by computer as described by Zuker and colleagues (39, 86), using PCGene (Intelligenetics). The predicted secondary structure of the 53-nt late 3′ UTR fragment was analyzed in the context of the L1 gene as well as in the context of the CAT expression vector. The predicted secondary structures of the mutated fragments were analyzed in the context of the CAT expression vector.

Nuclease S1 protection analysis of bovine fibropapilloma mRNA. Total cellular RNA was isolated from previously quick-frozen bovine fibropapilloma tissue by the method of Chirgwin et al. (22). Polyadenylated RNA was selected by using microcrystalline oligo(dT)-cellulose, using a modification of the method of Aviv and Leder (3). A uniformly labeled single-stranded DNA S1 probe spanning BPV-1 nt 6958 to 7197 was generated by using an antisense oligonucleotide (nt 7197 to 7174) containing 20 nt of extraneous sequence at the 5′ end and a single-stranded M13 DNA containing the BPV-1 long control region following the prime-cut method of Biggins et al. (10). Nuclease S1 protection analysis was carried out on polyadenylated RNA selected from 12.5 μg of fibropapilloma RNA as described previously (48) except that annealing was for 12 h and digestion was with 240 and 800 U of nuclease S1. Nuclease S1-protected fragments were separated on an 8% sequencing gel. Size markers were generated by dideoxy DNA sequencing using a 5′-labeled 20-nt antisense primer with the 5′ end at BPV-1 nt 208 and the single-stranded M13 DNA described above.

RESULTS

We previously used a CAT expression vector system to identify a 53-nt region (nt 7094 to 7146) of the BPV-1 late 3′ UTR which inhibits CAT expression approximately 10-fold when cloned into the 3′ UTR of the CAT vector (29). This inhibition is both orientation and position dependent. We decided to undertake a mutational analysis of the 53-nt BPV-1 3′ UTR fragment to identify nucleotides critical for the inhibitory function of the fragment, using the CAT expression system.

Mutational analysis of a stem-loop structure in the BPV-1 late 3′ UTR. Computer analysis for potential regions of secondary structure in the 3′ UTR predicted a short stem-loop structure for nt 7117 to 7137. Plasmids containing clustered 6-nt mutations within the 53-nt fragment were made to examine the role of this predicted stem-loop structure (Fig. 3). The stem-loop structure was disrupted by mutating either the 5′ stem (Fig. 3, line 4) or the 3′ stem (Fig. 3, line 6). In addition, the stem-loop was reconstructed by exchanging the 5′ and 3′ stems (Fig. 3, line 5). Computer analysis of the secondary structures of the mutated 53-nt fragments within the surrounding vector sequences predicted that the individual 5′ and 3′ stem mutations disrupted the putative secondary structure and that the reconstructed stem-loop construct recreated a secondary structure with the same predicted free energy as that of the wt fragment. In addition, a fragment containing just the stem-loop was analyzed (Fig. 4, line 3). The activities of the different 3′ UTR fragments were assayed in BPV-1-transformed C127 cells by cotransfection of the CAT vectors with a β-galactosidase internal control vector. BPV-1-transformed C127 cells harbor extrachromosomal BPV-1 genomes which express only early mRNAs. CAT expression levels were measured 48 h after transfection, and β-galactosidase activity was used to correct for differences in transfection efficiency.

In agreement with our previous study, the wt BPV-1 53-nt 3′ UTR fragment inhibited CAT expression approximately 18-fold (Fig. 3, line 1) compared with the parent CAT expression vector containing no insert, and this inhibition was orientation dependent (data not shown). The short stem-loop fragment (nt 7116 to 7137) had no inhibitory activity by itself (Fig. 4, line 3). In addition, a fragment with mutation of the 5′ stem of the stem-loop (nt 7117 to 7122) exhibited significantly greater inhibition than the wt 53-nt fragment (Fig. 3, line 4). These results indicate that the computer-predicted stem-loop is neither sufficient nor required for the inhibitory activity of the 3′ UTR element. However, fragments which contained clustered 6-nt mutations of the 3′ stem (nt 7132 to 7137) inhibited CAT expression only about two- to threefold (Fig. 3, lines 5 and 6). These results suggest that an inhibitory element is located in the 3′ part of the 53-nt fragment.

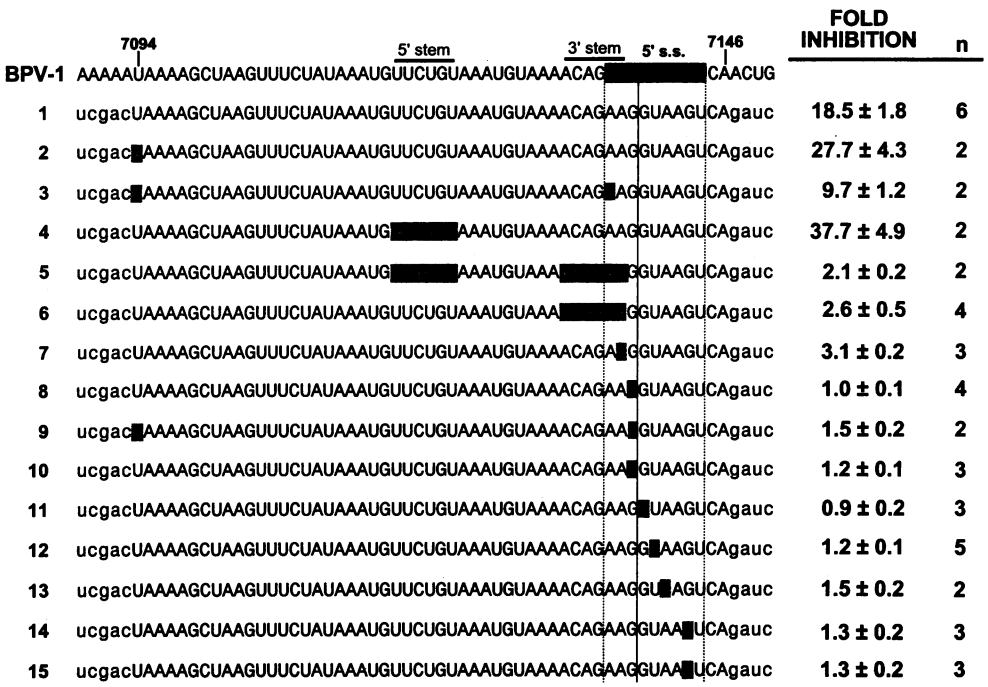

FIG. 3. Analysis of the effects of point mutations on the function of the BPV-1 late 3′ UTR inhibitory element. The top sequence line shows the wt BPV-1 sequence, with the consensus 5′ splice site shaded and labeled 5′ s.s. The 5′ and 3′ stems of a short stem-loop structure are overlined and labeled above the sequence. The nucleotide positions of the ends of the 53-nt fragment are indicated above. The RNA sequences of the BPV-1 inserts in the pOBCAT vectors (Fig. 2) are presented below along with the 5′ and 3′ flanking restriction sites. Lines are numbered at the left. Wild-type BPV-1 sequences are indicated by uppercase letters. Mutations and flanking sequences from the vector are indicated by lowercase letters. Deleted nucleotides are represented as dashes. All point mutations are highlighted with a gray background. The solid vertical line indicates the exon/intron junction within the 5′ splice site. The broken vertical lines indicate the boundaries of the 5′ splice site sequence. The effects of wt and mutant 3′ UTR fragments on CAT expression were assayed by transient transfection of BPV-1-transformed C127 cells. CAT activity was corrected for transfection efficiency by using β-galactosidase activity expressed from a cotransfected β-galactosidase expression vector (pCH110) as described in Materials and Methods. Fold inhibition was calculated by using the CAT activity expressed from the appropriate parent vector with no insert. The average fold inhibition and standard deviation are shown to the right of each sequence. n, number of independent transfections for each mutant. The plasmids used in this analysis were as follows: line 1, p2621; line 2, p2632; line 3, p2076; line 4, p2053; line 5, p2054; line 6, p2052; line 7, p2637; line 8, p2633; line 9, p2634; line 10, p2638; line 11, p2639; line 12, p2521; line 13, p2640; line 14, p2641; and line 15, p2642.

Localization of the BPV-1 late 3′ UTR inhibitory element to a 9-nt sequence homologous to a 5′ splice site. Deletion analysis of the 53-nt BPV-1 fragment was performed to further localize the inhibitory sequences within the fragment. Synthetic fragments containing the 5′ 30 nt of the fragment (nt 7094 to 7123; Fig. 4, line 2), the middle 22 nt of the fragment (nt 7116 to 7137; Fig. 4, line 3), the 3′ 24 nt of the fragment (nt 7123 to 7146; Fig. 4, line 4), and the 3′ 15 nt of the fragment (nt 7132 to 7146; Fig. 4, line 5) were cloned into the 3′ UTR of the CAT expression plasmid and analyzed as before. Fragments from the 5′ and middle portions of the 53-nt 3′ UTR fragment showed no inhibitory activity by themselves (Fig. 4, lines 2 and 3). However, all fragments containing wild-type BPV-1 nt 7132 to 7146 inhibited CAT expression (Fig. 4, lines 1, 4, and 5). In fact, a fragment containing just this region showed more inhibitory activity than the full length 53-nt fragment (Fig. 4, line 5). These results indicate that the inhibitory element is located within BPV-1 nt 7132 to 7146 and suggest that sequences upstream of this element may only moderate the effect of the element.

Inspection of BPV-1 nt 7132 to 7146 revealed the 9-nt sequence AAGGUAAGU, which has perfect homology with the consensus sequence for 5′ splice sites (Fig. 4) (65). This 9-nt sequence was assayed by itself and shown to inhibit CAT expression more than 100-fold (Fig. 4, line 8). In fact, the 9-nt sequence was the most inhibitory fragment tested, indicating that this sequence is responsible for the inhibition of gene expression.

Mutations which are known to block splicing also inactivate the BPV-1 negative element. If the inhibitory element is actually functioning as a 5′ splice site, then it should show the same sensitivity to mutations as a 5′ splice site. Initially, two 5′ splice site mutations were made within the BPV-1 15-nt inhibitory fragment to assess their effects on the fragment's inhibitory activity (Fig. 4). The first mutation, a U-to-A mutation in the conserved GU dinucleotide (Fig. 4, line 6), should allow at least initial splicing complexes to assemble since 5′ splice sites with this mutation still undergo cleavage, although at a reduced efficiency (2). A second, more extensive mutation (Fig. 4, line 7) which would eliminate base pairing with U1 snRNP and prevent recognition of the 5′ splice site (63) was also made. The data show that the two mutations were equally effective at abrogating the ability of the 5′ splice site to inhibit CAT expression (Fig. 4, lines 6 and 7). These results indicate that the integrity of the 5′ splice site is required for the inhibitory activity of the 3′ UTR element.

Additional mutations were made at positions −3 to +5 within the 5′ splice site homology to determine if the inhibitory element has the same sensitivity to mutations as a 5′ splice site. These mutations were analyzed in the context of both the 9- and 53-nt fragments. In general, mutations which deviated from the consensus 5′ splice site sequence significantly reduced

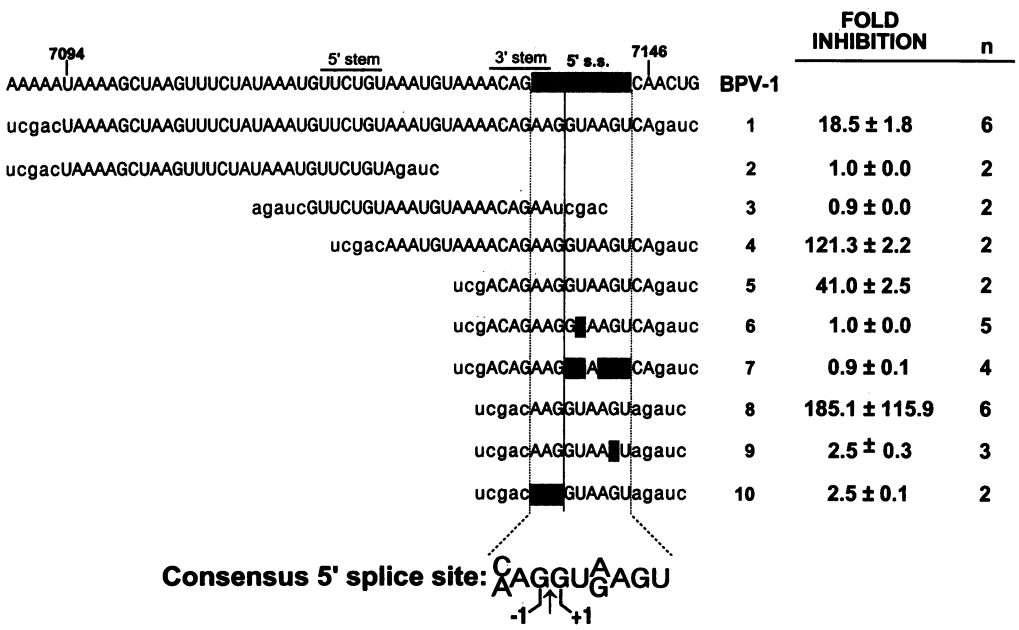

FIG. 4. Deletion analysis of the BPV-1 late 3′ UTR inhibitory element. Wild-type and mutant subfragments of the BPV-1 late 53-nt 3′ UTR fragment are shown. Sequence annotations are as described in the legend to Fig. 3. Lines are numbered in the center. The effects of wt and mutant 3′ UTR fragments on CAT expression were assayed by transient transfection of BPV-1-transformed C127 cells as described in the legend to Fig. 3. Fold inhibition is shown to the right of each insert. n, number of independent transfections for each mutant. The consensus 5′ splice site sequence is shown at the bottom. The arrow indicates the exon/intron junction in a functional 5′ splice site. Positions labeled −1 and +1 are the last nucleotide of the exon and first nucleotide of the intron, respectively. The plasmids used in this analysis were as follows: line 1, p2621; line 2, p2051; line 3, p2622; line 4, p2078; line 5, p2079; line 6, p2623, p2624, and p2625; line 7, p2626 and p2627; line 8, p2628; line 9, p2636; and line 10, p2538 and p2539.

the inhibitory function of the element. The most highly conserved nucleotides in 5′ splice sites are at positions −1 to +5. Mutations in these positions essentially abolished the function of the inhibitory element in the context of the 53-nt fragment (Fig. 3, lines 8 to 15) and reduced inhibition by the highly potent 9-nt fragment to only 2.5-fold (Fig. 4, lines 9 and 10). Mutations at these positions have been shown to strongly inhibit splicing both in vitro (1) and in vivo (17, 72). The 5′ splice site positions −2 and −3 show considerably less sequence conservation, with position −3 being the least conserved. In addition, mutation of these positions has only a modest effect on splicing in vitro (1). Mutations at −2 and −3 can affect splicing in vivo, but the degree of effect may depend on the assay and sequence context (46, 73). Mutation of position −2 in the context of the 53-nt fragment had significantly less effect on the inhibitory activity of the fragment than did mutations at −1 to +5; the mutant fragment still inhibited CAT expression threefold (Fig. 3, line 7). Mutation of position −3 had even less of an effect on the inhibitory activity of the fragment (Fig. 3, line 3). However, the effect of this mutation may be somewhat moderated by a second upstream mutation in the fragment (see below). These data indicate that the mutational sensitivity of the inhibitory element is similar to that expected for a 5′ splice site and strongly suggest that the inhibitory element is a 5′ splice site.

It should be noted that mutations upstream of the 5′ splice site-like sequence in the 53-nt fragment made the fragment more inhibitory (Fig. 3, lines 2 and 4). This is consistent with the observation that fragments with partial or complete removal of the upstream sequences inhibited CAT activity significantly more that the 53-nt fragment (Fig. 4, lines 1, 4, and 5). Recent studies of splicing in the dihydrofolate reductase gene demonstrated that exon mutations away from the splice sites can have both positive and negative effects on splicing, perhaps through changes in RNA secondary structure (13, 17). In addition, Goguel et al. (32) have shown that hairpins as short as 6 nt which involve the 5′ splice site significantly inhibit splicing. In this regard, disruption of the 5′ stem of the stem-loop structure in the BPV-1 late 3′ UTR increased inhibition of CAT activity by the 53-nt fragment twofold (Fig. 3, line 4). This suggests that secondary structure may influence the activity of the BPV-1 late 3′ UTR inhibitory element.

Base pairing between the 5′ end of the U1 snRNA and the 5′ splice site-like sequence is required for function of the 3′ UTR element. Sequences at a 5′ splice site are recognized by base pairing with the 5′ end of the U1 snRNA (Fig. 5). In addition, the strength of a 5′ splice site is influenced by the strength of this interaction (42). Base pairing between U1 snRNA and the 5′ splice site has been shown genetically by suppressing 5′ splice site mutations with U1 snRNAs containing compensatory mutations which restore base pairing (66, 69, 85). A similar approach was used to determine if base pairing between the 5′ end of the U1 snRNA and the 5′ splice site-like sequence in the BPV-1 late 3′ UTR was required for inhibition of gene expression by the 3′ UTR element. We initially chose to assay the CAT vector containing the 9-nt fragment with the G → C mutation at position +5 of the 5′ splice site-like sequence because the 9-nt fragment showed the greatest difference between wt and mutant fragments (Fig. 4, lines 8 and 9). A set of cloned human U1 genes containing all possible mutations at nt 4 and 6 was obtained from Alan Weiner (85). The names of the mutant U1 snRNAs indicate the position and nucleotide substitution within the U1 molecule (Fig. 5). Only

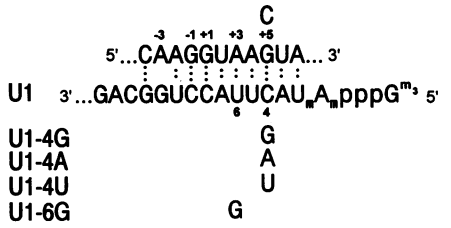

FIG. 5. Base pairing between the 5′ splice site and the 5′ end of the U1 snRNA. The upper sequence is the wt 9-nt insert shown in Fig. 4, line 8. The lower sequence is that of the 5′ end of the U1 snRNA. The standard Watson-Crick base pairing is indicated between the two sequences. Note that the sequences flanking the 9-nt insert can form two additional base pairings compared with the larger BPV-1 inserts. Mutations in the 5′ splice site and U1 snRNA are shown above and below the respective sequences. Only the U1-4G mutant can complement the C_{+5} splice site mutation.

the U1-4G suppressor snRNA would be expected to restore function to the C_{+5} mutant (Fig. 5). The effects of the U1 suppressors on CAT expression were tested in a transient cotransfection assay in BPV-1-transformed C127 cells. Human U1 genes have previously been shown to express functional U1 snRNAs in BPV-1-transformed C127 cells (51). The results of two experiments are shown in Table 1. In these experiments, the CAT vectors containing either the wt or C_{+5} mutant 9-nt fragment were cotransfected with either a wt U1 clone or clones expressing either U1-4G, U1-4A, or U1-4U snRNA. Only cotransfection with the U1-4G suppressor, which contains a compensatory mutation restoring base pairing, significantly increased inhibition of CAT expression compared with the wt U1 control (Table 1). Although the activity of the mutant fragment was not restored to wt levels, the 3.6- to 5.5-fold increase in activity in the presence of the suppressor U1 snRNA was highly significant. It should be noted that this experiment was highly controlled for effects of the suppressor U1 snRNAs not related to the 3′ UTR element. First, the data were corrected for transfection efficiency as well as nonspecific effects of the suppressors by normalizing the data to β-galactosidase activity expressed from the cotransfected vector pCH110. The β-galactosidase mRNA expressed from this vector is unspliced. Second, fold inhibition was calculated by comparison with the CAT activity expressed from pOBCAT20 cotransfected with the same suppressor. This corrects for any specific effects of the suppressor on splicing (and therefore expression) of the CAT pre-mRNA. Thus, the differences seen in Table 1 should reflect only changes in expression levels due to the 3′ UTR element. The magnitude of the effect of the suppressor U1 snRNA also correlated with the relative transfection efficiencies of the two experiments (Table 1). This is not surprising since the level of suppressor U1 snRNA which accumulates in transfected cells should be proportional to the amount of U1 genomic DNA which gets into the cell and is competent for expression. It is also not surprising that the suppressor U1 does not restore activity of the mutant fragment to wt levels. At early times after transfection, there will be only very low levels of suppressor U1 snRNA, the mutant 5′ splice site will not inhibit CAT expression, and CAT mRNA and protein will accumulate. Since CAT protein is very stable, this protein will contribute significantly to total CAT protein levels in the cells at the time of harvest. Only at later times will levels of suppressor U1 snRNA sufficient to inhibit CAT expression be present.

TABLE 1. Restoration of function with mutant U1 snRNAs

Insert[a]	U1	Fold inhibition[b]	
		Expt 1[c]	Expt 2[d]
wt	wt	61.30	86.1 ± 10.9
C_{+5}	wt	2.2 ± 0.1	1.5 ± 0.1
C_{+5}	U1-4G	12.0 ± 1.3	5.4 ± 0.5
C_{+5}	U1-4A	ND	2.2 ± 0.5
C_{+5}	U1-4U	ND	1.8 ± 0.1

[a] The 9-nt 5′ splice site sequence was cloned into the CAT vector shown in Fig. 1. wt, the wt insert shown in Fig. 4 line 8; C_{+5}, the insert with a G-to-C mutation at position +5 (see Fig. 4, line 9).

[b] CAT data were first normalized to β-galactosidase expression to correct for differences in transfection efficiency. Fold inhibition was then calculated by using the normalized CAT data for the parent CAT vector with no insert but cotransfected with the same U1 clone. This corrects for specific effects of the mutant U1 snRNAs on expression of the CAT vector which are not related to the insert. Values are averages of two samples where standard deviations are given. ND, not determined.

[c] Relative transfection efficiency = 12.6.

[d] Relative transfection efficiency = 1.

In an attempt to delay expression of the CAT vectors to allow time for levels of suppressor U1 snRNAs to build up, cotransfection experiments were carried out in COS-1 cells. COS-1 cells make SV40 T antigen, which allows the SV40-based expression vectors to replicate. Under these conditions, CAT activity is much higher after 3 days compared with 2 days, suggesting that these vectors are transcriptionally active for at least 3 days. To our surprise, however, both the 9- and 53-nt wt inhibitory fragments inhibited CAT expression 5- to 10-fold less in COS-1 cells compared with BPV-1-transformed C127 cells (see Discussion). However, both the 9- and 53-nt C_{+5} mutant fragments inhibited CAT expression approximately twofold more when cotransfected with the U1-4G suppressor clone compared with the wt U1 clone. Cotransfection with the U1-4A, U1-4U, and U1-6G clones had no significant effect (data not shown). This finding confirms the results presented in Table 1.

These experiments strongly suggest that the 5′ end of the U1 snRNA interacts with the BPV-1 late 3′ UTR element in vivo by base pairing and that this interaction is essential for inhibition of gene expression by the element. This conclusion is supported by the observation that the most inhibitory fragment which we have tested is the 9-nt consensus 5′ splice site (Fig. 4, line 8). The increased activity seen for this fragment may be due to the potential base pairing between the two flanking nucleotides and the 5′ end of U1 snRNA (Fig. 5), which would considerably increase the strength of the interaction.

The consensus 5′ splice site in the BPV-1 late 3′ UTR is not used for splicing in fibropapillomas. The binding of splicing factors to the BPV-1 late 3′ UTR element may be sufficient for the function of the element, perhaps through interference with polyadenylation or nucleocytoplasmic transport. Alternatively, the mechanism of action could be as simple as competition between splicing and polyadenylation reactions. However, the structures of BPV-1 mRNAs in both productively infected fibropapillomas and BPV-1-transformed C127 cells have been extensively studied (see reference 6 and references therein). No mRNAs which are spliced by using the 5′ splice site in the late 3′ UTR have ever been identified. A nuclease S1 protection assay was used to confirm that the 5′ splice site in the late 3′ UTR is not used for splicing. A single-stranded antisense DNA fragment spanning BPV-1 nt 6958 to 7197 was used as the probe. Several different protection fragments are possible, depending on the RNA processing site used in the pre-mRNA (Fig. 6B). RNAs cleaved and polyadenylated at the late poly(A) site (nt 7175) would generate a fragment 218 nt long.

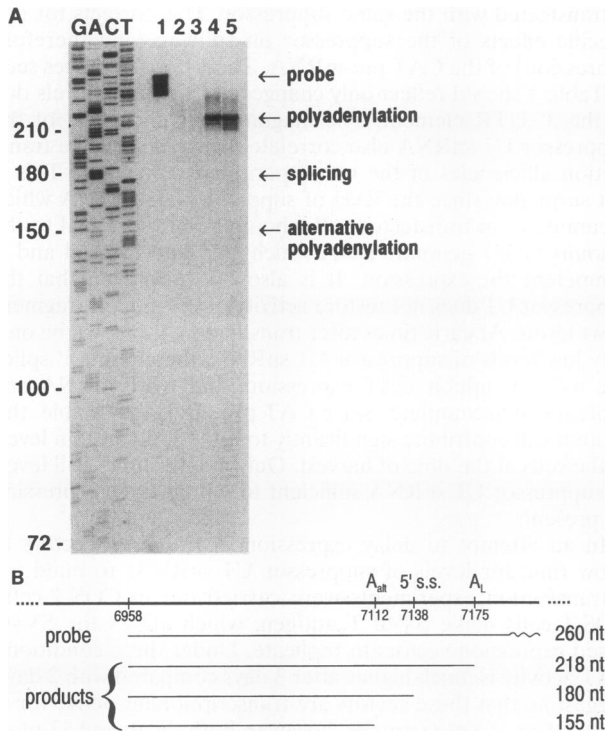

FIG. 6. S1 analysis of mRNA isolated from a BPV-1-infected bovine fibropapilloma. (A) Polyadenylated RNA isolated from a BPV-1-infected bovine fibropapilloma was analyzed by nuclease S1 protection, using a uniformly labeled single-stranded antisense DNA probe which spans BPV-1 nt 6958 to 7197 as described in Materials and Methods. Samples were as follows: lane 1, undigested probe; lanes 2 and 3, tRNA; and lanes 4 and 5, fibropapilloma RNA. Hybridizations in lanes 2 and 4 were digested with 240 U of S1; hybridizations in lanes 3 and 5 were digested with 800 U of S1. Lanes labeled G, A, C, and T are the respective dideoxy sequencing reactions generated by using a 5'-labeled 20-nt antisense primer with 5' nt at BPV-1 nt 208 and a single-stranded M13 DNA containing BPV-1 nt 6958 to 471. Products were separated on an 8% sequencing gel. Exposure time was 5 h with a Dupont Cronex Lightning-Plus intensifying screen. Approximate sizes in nucleotides are labeled on the left side. Bands corresponding to full-length probe (probe) and RNA cleaved at the late poly(A) site (polyadenylation) are labeled at the right. The expected positions of bands corresponding to RNAs cleaved at the 5' splice site in the late 3' UTR (splicing) and cleavage at a potential alternative polyadenylation site downstream of an AAUAAA sequence (nt 7092 to 7097) (alternative polyadenylation) are also indicated. (B) Schematic showing nuclease S1 probe and expected products. Sizes of the probe and expected products are shown at the right. The primer used to generate the probe contained an extra 20 nt of extraneous sequence (indicated by wavy line) at its 5' end. 5' s.s., 5' splice site.

Use of the 3' UTR element as a 5' splice site (nt 7138) would yield a protected fragment of 180 nt. In addition, a second potential polyadenylation signal (AAUAAA) is located at nt 7092 to 7097. If this signal were used, then cleavage and polyadenylation would occur at approximately nt 7112 and an S1 protection fragment of 155 nt would be generated. This latter point is important because use of this alternative poly(A) site would prevent the 3' UTR element from functioning.

Total polyadenylated RNA from BPV-1-infected fibropapillomas was chosen for the analysis because all stages of the viral life cycle should be represented in this RNA population. A prominent product consistent in size with that expected for cleavage at the late poly(A) site was generated with fibropapilloma RNA but not tRNA (Fig. 6A, lanes 4 and 5 compared with lanes 2 and 3). No protection products corresponding to splicing at the 3' UTR element or polyadenylation at the alternative poly(A) site were detected. We cannot rule out, however, that these reactions occur, but the resulting mRNAs are very unstable. These results suggest that the late 3' UTR element functions by a mechanism other than a direct competition between splicing and polyadenylation reactions. In addition, the alternative poly(A) site does not appear to function at any stage of the viral life cycle.

Nucleotides homologous to a 5' splice site are also required for the inhibitory activity of the HPV-16 3' late-region element. An inhibitory element is also present in the HPV-16 late coding region and 3' UTR (40, 41). The CAT expression system used to map the BPV-1 inhibitory element was used to further map the HPV-16 element and to determine if the two elements function through the same mechanism. A 51-nt fragment (nt 7130 to 7180) located at the junction of the HPV-16 L1 coding sequences and 3' UTR was cloned into the 3' UTR of the CAT expression vector (Fig. 7). This 51-nt fragment was chosen on the basis of direct sequence comparison with the BPV-1 fragment and published data which defined the 5' extent of the HPV-16 fragment (41). CAT expression levels were measured 48 h after transfection into BPV-1-transformed C127 cells and are presented in Fig. 7. The 51-nt HPV-16 fragment inhibited CAT expression approximately sixfold compared with the parental CAT vector (Fig. 7, line 1). This degree of inhibition is approximately threefold less than that seen with the 53-nt BPV-1 late 3' UTR fragment (Fig. 3, line 1).

Analysis of the sequence of the 51-nt HPV-16 fragment revealed four tandem overlapping sequence motifs (underlined in Fig. 7) which partially match the 5' splice site consensus sequence: AAC/GUAAGC, GCU/GUAAGU, AUU/GUAUGU, and UAU/GUAUGU. All sequence motifs deviate significantly from the consensus sequence; however, the second motif is a perfect match in the intron portion. To test if the integrity of any of these sequences is required for the inhibitory activity, each of the motifs was mutated at position +2 (U → A) in the context of the 51-nt HPV-16 fragment. This mutation had been shown to abolish the activity of the 53-nt BPV-1 fragment (Fig. 3, line 12). Only mutation of the second 5' splice site-like sequence abolished the inhibitory activity of the fragment (Fig. 7, line 5), suggesting that only this copy is essential for the activity of the fragment. However, the second 5' splice site-like sequence inhibited CAT expression only 2.5-fold when assayed by itself (Fig. 7, line 8), suggesting that additional upstream or downstream sequences are required for the full activity of the HPV-16 3' UTR element. Possible candidates are the other 5' splice site-like sequences. The mutation at +2 in the third motif is simultaneously a −3 mutation in the overlapping fourth motif which brings this sequence closer to the consensus sequence. This mutation caused a 50% increase in the activity of the 51-nt fragment (Fig. 7, line 6), supporting this hypothesis. A spontaneous mutation which lies upstream of the cluster of 5' splice site-like sequences had no effect on the activity of the fragment (Fig. 7, line 2). These results suggest that the HPV-16 3' UTR element consists of a cluster of overlapping 5' splice sites and therefore functions through the same mechanism as the BPV-1 late 3' UTR element.

DISCUSSION

These experiments have mapped the inhibitory element in the BPV-1 late 3' UTR (29) to a 9-nt sequence (AAG

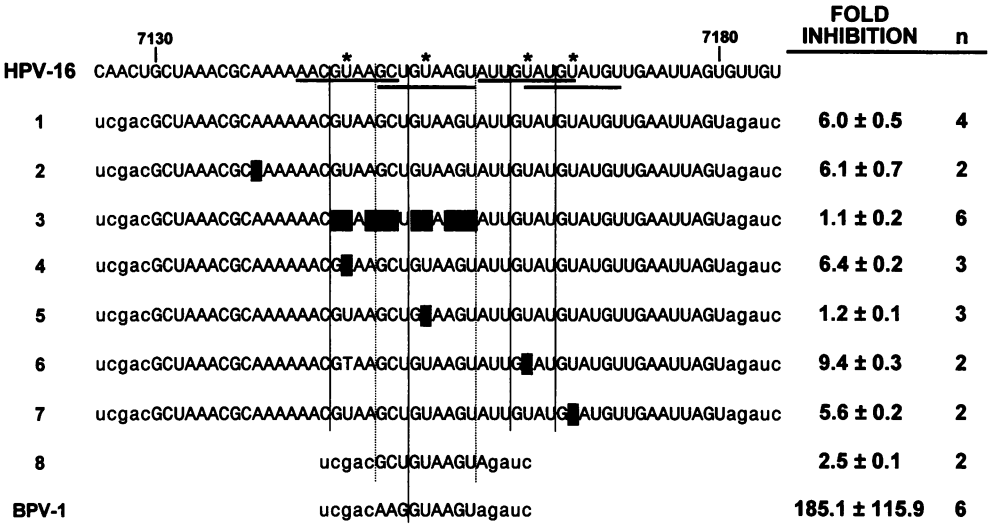

FIG. 7. Mutational analysis of the HPV-16 3′ late-region inhibitory element. The top sequence line shows the wt HPV-16 sequence, with the four 5′ splice site-like sequences underlined and the nucleotide positions of the ends of the 51-nt fragment indicated above. The termination codon for the L1 ORF is at nt 7154 to 7156. The RNA sequences of the HPV-16 inserts are presented below along with the 5′ and 3′ flanking restriction sites. Lines are numbered at the left. Sequence annotations are as described in the legend to Fig. 3. Solid vertical lines indicate the exon/intron junctions for each 5′ splice site motif. Dashed vertical lines delimit the boundaries of the second motif. Asterisks indicate the U residues at position +2 in each motif. The line labeled BPV-1 presents data from Fig. 4, line 8. The effects of wt and mutant 3′ UTR fragments on CAT expression were assayed by transient transfection of BPV-1-transformed C127 cells as described in the legend to Fig. 3. Fold inhibition is shown to the right of each insert. n, number of independent transfections for each mutant. The plasmids used in this analysis were as follows: line 1, p2635; line 2, p2048; line 3, p2629, p2630, and p2631; line 4, p2522, p2523, and p2524; line 5, p2525, p2526, and p2527; line 6, p2536; line 7, p2537; and line 8, p2538 and p2539.

GUAAGU) from nt 7136 to 7144. This sequence has perfect homology with the 5′ splice site consensus sequence ([C/A]AG/GU[A/G]AGU) (65). All fragments which contained these 9 nt intact significantly inhibited CAT expression. In addition, the 9-nt fragment actually inhibited CAT expression to a greater degree than the 53-nt 3′ UTR fragment. Furthermore, this 9-nt sequence showed the same sensitivity to mutations as would be expected for a 5′ splice site. Positions −1 to +5 are the most highly conserved in 5′ splice sites, and mutations in these positions have been shown to block splicing both in vivo and in vitro (1, 17, 72). Mutational analysis of the BPV-1 inhibitory element showed that similar mutations essentially abolished the function of the inhibitory element in the context of the 53-nt fragment (Fig. 3, lines 8 to 15) and reduced inhibition by the 9-nt fragment to only 2.5-fold (Fig. 4, lines 9 and 10). In contrast, positions −2 and −3 are considerably less conserved, and mutations in these positions have only a modest effect on splicing in vitro (1). Similarly, the 53-nt fragments containing mutations in positions −2 and −3 still retained partial inhibitory activity (Fig. 3, lines 7 and 3, respectively). Further evidence that the inhibitory element is a 5′ splice site was obtained through genetic suppression studies. This approach has been used previously to demonstrate that the 5′ end of the U1 snRNA base pairs with the 5′ splice site (66, 69, 85). We used a similar approach to demonstrate that the 5′ end of the U1 snRNA also base pairs with the 5′ splice site in the BPV-1 inhibitory element. Furthermore, this interaction was required for inhibition of gene expression by the element (Table 1). This conclusion is further supported by the observation that the most inhibitory fragment tested was the 9-nt consensus 5′ splice site fragment (Fig. 4, line 8). The nucleotides flanking the 9-nt 5′ splice site sequence are different from those in the other fragments and permit two additional base pairs between the 5′ end of U1 snRNA and the 5′ splice site (Fig. 5). Taken together, these data strongly suggest that the functional component of the BPV-1 inhibitory element is a 5′ splice site.

We have also localized an HPV-16 inhibitory element (40) to a 51-nt fragment at the junction of the L1 open reading frame (ORF) and the late 3′ UTR (nt 7130 to 7180). This fragment contains four motifs (AAC/GUAAGC, GCU/GUAAGU, AUU/GUAUGU, and UAU/GUAUGU) that show partial homology with the 5′ splice site consensus sequence. Mutational inactivation of each of these motifs demonstrated that only the integrity of the second of these motifs is required for the inhibitory activity of the HPV-16 fragment (Fig. 7). Although all four motifs deviate significantly from the consensus sequence, only the second motif is a perfect match in the intron portion. However, the second motif was not sufficient for the full inhibitory activity exhibited by the 51-nt fragment. It is likely that multiple weak interactions between U1 snRNA and the HPV-16 element enhance the inhibitory effect of the element. In support of this is the observation that two tandem consensus 5′ splice sites are considerably more inhibitory than a single site (8a). The deviation of the HPV-16 5′ splice sites from consensus most likely explains the threefold lower inhibitory activity of the 51-nt HPV-16 fragment compared with the 53-nt BPV-1 element.

A review of the genomic sequences of other papillomaviruses reveals that sequences homologous to 5′ splice sites can also be found in the late coding regions and late 3′ UTRs of BPV-2, HPV-6B, HPV-1A, and HPV-18, although most deviate somewhat from the consensus sequence. BPV-2 contains the same 9-nt consensus sequence present in BPV-1. This is not unexpected since BPV-2 is highly homologous to BPV-1. HPV-6B contains the first nonconsensus 9-nt sequence present in the HPV-16 fragment. HPV-1A contains the sequence GGC/GUAAGG, and HPV-18 contains the sequence CGU/

GUACGU. It is not known whether these sequences inhibit expression of the late genes of these viruses. However, terminal exons rarely contain sequences resembling 5′ splice sites (12). The presence of 5′ splice site-like sequences in the 3′ late regions of multiple papillomaviruses supports the hypothesis that these homologies are of functional significance.

The structures of BPV-1 and HPV-16 mRNAs have been extensively studied by cDNA cloning and electron microscopy (7, 25, 67, 71, 82) (for a review of BPV-1 and HPV-16 mRNA structures, see reference 6). Surprisingly, no mRNAs which utilize the 5′ splice sites present in the late 3′ UTR of either BPV-1 or HPV-16 have ever been identified. Furthermore, nuclease S1 protection analysis of bovine fibropapilloma mRNA failed to detect any BPV-1 RNA spliced at the 5′ splice site in the late 3′ UTR (Fig. 6). This suggests that these sites do not function through the direct competition between splicing and polyadenylation reactions. Rather, the primary function of these 5′ splice site-like sequences appears to be the inhibition of late gene expression, most likely by reducing the efficiency of polyadenylation. This could involve either direct interactions between the U1 snRNP or other factors bound at the 5′ splice site-like sequence and the polyadenylation machinery or interference with 3′-terminal exon definition. The exon definition model of splicing suggests that internal exons are defined by an interaction or communication between factors bound at 3′ and 5′ splice sites (62). Hoffman and Grabowski (35) have shown that binding of the U1 snRNP to the 5′ splice site of an internal exon facilitates binding of U2AF65 to the 3′ splice site. Similarly, terminal exons are defined by an interaction between factors bound at a 3′ splice site and at a polyadenylation site (58). Recognition of a terminal exon as a unit appears to enhance the efficiency of polyadenylation (21, 37, 49, 55, 58). The observation that inclusion of introns in histone genes activates cryptic polyadenylation sites and interferes with normal histone 3′-end processing provides further support for the concept that factors bound at 3′ splice sites can interact with factors bound at polyadenylation sites (59). Wassarman and Steitz (77) showed that the coupling between splicing and polyadenylation is mediated by weak interactions between the U1 snRNP and regions of the pre-mRNA upstream of the polyadenylation site which have limited complementarity with the 5′ end of the U1 snRNA. However, when the region of limited complementarity was replaced with a consensus 5′ splice site, the coupling between splicing and polyadenylation was lost. Thus strong interactions between the U1 snRNP and a 5′ splice site located in a 3′ UTR appear to interfere with terminal exon definition and can therefore alter gene expression. This was also demonstrated by Niwa et al. (57), who showed that insertion of a 5′ splice site within a 3′-terminal exon depresses polyadenylation both in vivo and in vitro. They showed that depression of polyadenylation was accompanied by depression of binding of the polyadenylation factor cleavage stimulation factor (CstF). Inhibition of polyadenylation is independent of the distance between the 5′ splice site and the poly(A) site. This suggests that inhibition does not involve steric hindrance between factors bound at the two sites (57). Similarly, the BPV-1 3′ UTR element inhibits both BPV-1 L1 expression when located only 11 nt upstream of the BPV-1 late AAUAAA recognition signal and CAT expression when cloned more than 80 nt upstream of the SV40 early polyadenylation signal (29). In addition, the lack of inhibition of CAT expression when the 53-nt fragment was cloned an equivalent distance downstream of the poly(A) cleavage site also argues against steric interference with polyadenylation (29).

Additional evidence for the involvement of the U1 snRNP in polyadenylation was obtained through studies of the SV40 late poly(A) site. This poly(A) site contains a USE which stimulates polyadenylation (14, 64). Lutz and Alwine (50) have recently demonstrated that the U1 snRNP is required for the activity of this USE. The U1 snRNP-A protein can interact simultaneously with both the USE and the U1 snRNA through two separate RNA recognition motifs and this interaction stimulates polyadenylation. Thus the involvement of the U1 snRNP in polyadenylation can be mediated through either RNA-RNA or RNA-protein interactions with the pre-mRNA.

Another way to look at the effects of a 5′ splice site on polyadenylation is that insertion of a 5′ splice site upstream of a poly(A) site essentially converts a 3′-terminal exon into an internal exon. Thus, the BPV-1 3′ UTR element inhibits CAT expression when located in the 3′ UTR but not when cloned upstream of the 3′ splice site or downstream of the poly(A) site, locations which would not be expected to influence 3′-terminal exon definition (29). Vertebrate internal exons are small, with an average size of 137 nt (34). The exon definition model suggests that internal exons are short because there is a limit to the distance over which factors bound at 3′ and 5′ splice sites can interact, and this has been confirmed experimentally for splicing in vitro (62). Similarly, Niwa et al. (57) showed that a 5′ splice site inserted in a 3′-terminal exon depresses binding of CstF in vitro only if it is inserted less than 300 nt away from the 3′ splice site. However, restrictions on internal exon size may not be as rigid as previously proposed. For example, the effects of exon size on splicing can be modulated by cell-type-specific factors (60). In addition, Chen and Chasin (18) have recently demonstrated that internal exons as large as 1,400 nt can be spliced into mRNAs. The present study shows that both the 9-nt consensus 5′ splice site and the BPV-1 53-nt late 3′ UTR fragment strongly inhibited CAT expression when located 936 nt downstream of the 3′ splice site in the CAT expression vector (Fig. 2). Furthermore, the 3′ UTR inhibitory element lies more than 1,500 nt downstream of the 3′ splice site in its normal location within the BPV-1 genome. Deletion of the 53-bp late 3′ UTR region from a two-intron late minigene vector increased expression of the BPV-1 L1 mRNA approximately sixfold, indicating that the late 3′ UTR element is also functional in its normal context (data not shown).

Previously we showed that the BPV-1 late 3′ UTR element inhibits gene expression when located 1,820 nt downstream of the cap site in a BPV-1 L1 cDNA expression vector with no introns (29). The observation that the 3′ UTR element inhibits expression in the absence of an intron suggests that the BPV-1 3′ UTR element does not function solely by preventing interactions between factors bound at 3′ splice sites and poly(A) sites. However, the observations that polyadenylation and splicing are stimulated in vitro by the presence of a cap on the pre-mRNA suggest that there may be interactions between cap-binding factors and splicing and polyadenylation factors which are similar to the interactions between splicing and polyadenylation factors (26, 56).

Although we favor a mechanism whereby the papillomavirus 3′ UTR elements inhibit polyadenylation through interference with terminal exon definition, we cannot rule out that an unpaired 5′ splice site blocks nucleocytoplasmic transport of the mRNA. Chang and Sharp (15, 16) showed that mutation of a 5′ or 3′ splice site in the β-globin pre-mRNA leads to retention of the RNA in the nucleus. This effect could be overcome by the human immunodeficiency virus Rev-responsive element in *cis* and the Rev protein in *trans*, suggesting that unpaired splice sites might be used to regulate gene expression.

This study did not directly address whether the papillomavirus late 3′ UTR inhibitory elements are constitutive or

regulated. In the absence of a cell culture system permissive for late gene expression, it is difficult to ascertain whether the inhibitory mechanism is inactivated in permissive cells. However, expression of the papillomavirus late genes follows vegetative viral DNA replication and activation of a strong late promoter (9), so it is possible that inhibitory factors are titrated out by abundant late pre-mRNAs. Since the CAT vectors used in the present study contain the SV40 origin of replication, the effects of replication and overexpression could be tested by transfection into COS-1 cells. Surprisingly, both the 9-nt and 53-nt wt inhibitory fragments inhibited CAT expression 5- to 10-fold less in COS-1 cells than in BPV-1-transformed C127 cells (data not shown). This difference could not be attributed to cell type or BPV-1-encoded proteins since there was no difference in inhibition in uninfected CV-1 cells (the parent cell type for COS-1 cells) compared with the BPV-1-transformed C127 cells. In contrast, the fragments were fully inhibitory in COS-1 cells treated with sodium butyrate, which inhibits initiation of cellular and viral DNA replication (78, 81). These data suggest that factors required for the activity of the papillomavirus late 3′ UTR elements may be in limiting supply and therefore could be titrated out at late stages of the viral life cycle.

Alternatively, it is possible that factors or mechanisms which control splicing regulate the activity of the BPV-1 and HPV-16 inhibitory elements (52). Modulation of the concentration or activity of general splicing factors and heterogeneous nuclear ribonucleoproteins influences 5′ splice site use (see references 28, 43, 54, and 84 and references therein). In addition, specific splicing factors regulate 5′ splice sites through binding to sequences within introns or exons (36, 68). The formation of RNA secondary structures can also affect the use of a specific splice site (see reference 32 and references therein). For example, masking of a splice site by a stem-loop structure in the β-tropomyosin gene regulates its use in specific tissues (23, 47), and artificial hairpins which sequester a 5′ splice site inhibit yeast pre-mRNA splicing (32). The formation of the stem-loop structure in the BPV-1 late 3′ UTR could mask the consensus 5′ splice site sequence in a manner analogous to the artificial hairpins in yeast cells. Evidence for the influence of exon sequences on the activity of the BPV-1 inhibitory element is the observation that sequences immediately upstream of the 5′ splice site moderate the activity of the element (Fig. 3, lines 2 and 4; Fig. 4, lines 1 to 5). Sequences surrounding a splice site can have significant effects on the efficiency of splicing at a designated 5′ splice site (13, 17).

In summary, both the BPV-1 late 3′ UTR and the HPV-16 late gene inhibitory fragments described in this report required the presence of sequences homologous to 5′ splice sites for their activities. Since positioning a 5′ splice site upstream of a polyadenylation site in a 3′-terminal exon impairs binding of CstF and inhibits polyadenylation, it is likely that the papillomavirus inhibitory elements affect gene expression by interfering with the efficiency of polyadenylation.

ACKNOWLEDGMENTS

Won-Taek Choe was supported by a summer fellowship from the Howard Hughes Foundation.

We thank Alan Weiner for providing the wt and mutant U1 clones.

REFERENCES

1. **Aebi, M., H. Hornig, R. A. Padgett, J. Reiser, and C. Weissmann.** 1986. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. Cell **47**:555–565.
2. **Aebi, M., H. Hornig, and C. Weissmann.** 1987. 5′ cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5′ splice region, not by the conserved 5′ GU. Cell **50**:237–246.
3. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
4. **Baker, C. C.** 1990. Bovine papillomavirus type 1 transcription, p. 91–112. *In* H. Pfister (ed.), Papillomavirus and human cancer. CRC Press, Inc., Boca Raton, Fla.
5. **Baker, C. C.** 1990. An improved chloramphenicol acetyltransferase expression vector system for mapping transcriptional and post-transcriptional regulatory elements in animal cells, p. 75–86. *In* M.-L. Huhtala and A. Vaheri (ed.), Recombinant systems in protein expression. Elsevier Science Publishers B.V., Amsterdam.
6. **Baker, C. C.** 1993. The genomes of the papillomaviruses, p. 1.134–1.146. *In* S. O'Brien (ed.), Genetic maps: locus maps of complex genomes. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
7. **Baker, C. C., and P. M. Howley.** 1987. Differential promoter utilization by the bovine papillomavirus in transformed cells and productively infected wart tissues. EMBO J. **6**:1027–1035.
8. **Baker, C. C., and J. S. Noe.** 1989. Transcriptional termination between bovine papillomavirus type 1 (BPV-1) early and late polyadenylation sites blocks late transcription in BPV-1-transformed cells. J. Virol. **63**:3529–3534.

8a. **Barksdale, S., and C. Baker.** Unpublished data.

9. **Barksdale, S. K., and C. C. Baker.** 1993. Differentiation-specific expression from the bovine papillomavirus type 1 P_{2443} and late promoters. J. Virol. **67**:5605–5616.
10. **Biggins, M., P. Farrell, and B. Barrell.** 1984. Transcription and DNA sequence of the Bam HI L fragment of B95-8 Epstein-Barr virus. EMBO J. **3**:1083–1090.
11. **Boelens, W. C., E. J. R. Jansen, W. J. Van Venrooij, R. Stripecke, I. W. Mattaj, and S. I. Gunderson.** 1993. The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. Cell **72**:881–892.
12. **Brunak, S., J. Engelbrecht, and S. Knudsen.** 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J. Mol. Biol. **220**:49–65.
13. **Carothers, A. M., G. Urlaub, D. Grunberger, and L. A. Chasin.** 1993. Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. Mol. Cell. Biol. **13**:5085–5098.
14. **Carswell, S., and J. C. Alwine.** 1989. Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. Mol. Cell. Biol. **9**:4248–4258.
15. **Chang, D. D., and P. A. Sharp.** 1989. Regulation by HIV Rev depends upon recognition of splice sites. Cell **59**:789–795.
16. **Chang, D. D., and P. A. Sharp.** 1990. Messenger RNA transport and HIV *rev* regulation. Science **249**:614–615.
17. **Chen, I.-T., and L. A. Chasin.** 1993. Direct selection for mutations affecting specific splice sites in a hamster dihydrofolate reductase minigene. Mol. Cell. Biol. **13**:289–300.
18. **Chen, I.-T., and L. A. Chasin.** 1994. Large exon size does not limit splicing in vivo. Mol. Cell. Biol. **14**:2140–2146.
19. **Cherrington, J., and D. Ganem.** 1992. Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences. EMBO J. **11**:1513–1524.
20. **Cherrington, J., R. Russnak, and D. Ganem.** 1992. Upstream sequences and cap proximity in the regulation of polyadenylation in ground squirrel hepatitis virus. J. Virol. **66**:7589–7596.
21. **Chiou, H. C., C. Dabrowski, and J. C. Alwine.** 1991. Simian virus 40 late mRNA leader sequences involved in augmenting mRNA accumulation via multiple mechanisms, including increased polyadenylation efficiency. J. Virol. **65**:6677–6685.
22. **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.
23. **Clouet, O., Y. D'Aubenton Carafa, P. Sirand-Pugnet, M. Gallego, E. Brody, and J. Marie.** 1991. RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. Science **252**:1823–1828.

24. **DeZazzo, J. D., E. Falck-Pedersen, and M. J. Imperiale.** 1991. Sequences regulating temporal poly(A) site switching in the adenovirus major late transcription unit. Mol. Cell. Biol. **11**:5977–5984.
25. **Doorbar, J., A. Parton, K. Hartley, L. Banks, T. Crook, M. Stanley, and L. Crawford.** 1990. Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. Virology **178**:254–262.
26. **Edery, I., and N. Sonenberg.** 1985. Cap-dependent RNA splicing in a HeLa nuclear extract. Proc. Natl. Acad. Sci. USA **82**:7590–7594.
27. **Enriquez-Harris, P., N. Levitt, D. Briggs, and N. J. Proudfoot.** 1991. A pause site for RNA polymerase II is associated with termination of transcription. EMBO J. **10**:1833–1842.
28. **Fu, X.-D., A. Mayeda, T. Maniatis, and A. R. Krainer.** 1992. General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5′ and 3′ splice site selection. Proc. Natl. Acad. Sci. USA **89**:11224–11228.
29. **Furth, P. A., and C. C. Baker.** 1991. An element in the bovine papillomavirus late 3′ untranslated region reduces polyadenylated cytoplasmic RNA levels. J. Virol. **65**:5806–5812.
30. **Gilmartin, G. M., E. S. Fleming, and J. Oetjen.** 1992. Activation of HIV-1 pre-mRNA 3′ processing *in vitro* requires both an upstream element and TAR. EMBO J. **11**:4419–4428.
31. **Gilmartin, G. M., and J. R. Nevins.** 1989. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. Genes Dev. **3**:2180–2189.
32. **Goguel, V., Y. Wang, and M. Rosbash.** 1993. Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. Mol. Cell. Biol. **13**:6841–6848.
33. **Gunderson, S. I., K. Beyer, G. Martin, W. Keller, W. C. Boelens, and I. W. Mattaj.** 1994. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. Cell **76**:531–541.
34. **Hawkins, J. D.** 1988. A survey on intron and exon lengths. Nucleic Acids Res. **16**:9893–9908.
35. **Hoffman, B. E., and P. J. Grabowski.** 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3′ splice site by a network of interactions spanning the exon. Genes Dev. **6**:2554–2568.
36. **Horabin, J. I., and P. Schedl.** 1993. *Sex-lethal* autoregulation requires multiple *cis*-acting elements upstream and downstream of the male exon and appears to depend largely on controlling the use of the male exon 5′ splice site. Mol. Cell. Biol. **13**:7734–7746.
37. **Huang, M. T., and C. M. Gorman.** 1990. Intervening sequences increase efficiency of RNA 3′ processing and accumulation of cytoplasmic RNA. Nucleic Acids Res. **18**:937–947.
38. **Imperiale, M. J., and J. D. DeZazzo.** 1991. Poly(A) site choice in retroelements: deja vu all over again? New Biol. **3**:531–537.
39. **Jacobson, A. B., L. Good, J. Simonetti, and M. Zuker.** 1984. Some simple computational methods to improve the folding of large RNAs. Nucleic Acids Res. **12**:45–52.
40. **Kennedy, I. M., J. K. Haddow, and J. B. Clements.** 1990. Analysis of human papillomavirus type 16 late mRNA 3′ processing signals in vitro and in vivo. J. Virol. **64**:1825–1829.
41. **Kennedy, I. M., J. K. Haddow, and J. B. Clements.** 1991. A negative regulatory element in the human papillomavirus type 16 genome acts at the level of late mRNA stability. J. Virol. **65**:2093–2097.
42. **Kuo, H., F. H. Nasim, and P. J. Grabowski.** 1991. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. Science **251**:1045–1050.
43. **Lamond, A. I.** 1991. ASF/SF2: a splice site selector. Trends Biochem. Sci. **16**:452–453.
44. **Lancaster, W. D.** 1981. Apparent lack of integration of bovine papillomavirus DNA in virus-induced equine and bovine tumor cells and virus-transformed mouse cells. Virology **108**:251–255.
45. **Law, M. F., D. R. Lowy, I. Dvoretzky, and P. M. Howley.** 1981. Mouse cells transformed by bovine papillomavirus contain only extrachromosomal viral DNA sequences. Proc. Natl. Acad. Sci. USA **78**:2727–2731.
46. **Lear, A. L., L. P. Eperon, I. M. Wheatley, and I. C. Eperon.** 1990. Hierarchy for 5′ splice site preference determined in vivo. J. Mol. Biol. **211**:103–115.
47. **Libri, D., A. Piseri, and M. Y. Fiszman.** 1991. Tissue-specific splicing in vivo of the β-tropomyosin gene: dependence on an RNA secondary structure. Science **252**:1842–1845.
48. **Linz, U., and C. C. Baker.** 1988. Promoters of bovine papillomavirus type 1: in vitro activity and utilization. J. Virol. **62**:2537–2543.
49. **Luo, Y., and G. G. Carmichael.** 1991. Splice site choice in a complex transcription unit containing multiple inefficient polyadenylation signals. Mol. Cell. Biol. **11**:5291–5300.
50. **Lutz, C. S., and J. C. Alwine.** 1994. Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. Genes Dev. **8**:576–586.
51. **Mangin, M., M. Ares, Jr., and A. M. Weiner.** 1985. U1 small nuclear RNA genes are subject to dosage compensation in mouse cells. Science **229**:272–275.
52. **Maniatis, T.** 1991. Mechanisms of alternative pre-mRNA splicing. Science **251**:33–34.
53. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
54. **Mayeda, A., D. M. Helfman, and A. R. Krainer.** 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. Mol. Cell. Biol. **13**:2993–3001.
55. **Miller, J. T., and C. M. Stoltzfus.** 1992. Two distant upstream regions containing *cis*-acting signals regulating splicing facilitate 3′-end processing of avian sarcoma virus RNA. J. Virol. **66**:4242–4251.
56. **Moore, C. L., and P. A. Sharp.** 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell **41**:845–855.
57. **Niwa, M., C. C. MacDonald, and S. M. Berget.** 1992. Are vertebrate exons scanned during splice-site selection. Nature (London) **360**:277–280.
58. **Niwa, M., S. D. Rose, and S. M. Berget.** 1990. In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes Dev. **4**:1552–1559.
59. **Pandey, N. B., N. Chodchoy, T. J. Liu, and W. F. Marzluff.** 1990. Introns in histone genes alter the distribution of 3′ ends. Nucleic Acids Res. **18**:3161–3170.
60. **Peterson, M. L., M. B. Bryman, M. Peiter, and C. Cowan.** 1994. Exon size affects competition between splicing and cleavage-polyadenylation in the immunoglobulin μ gene. Mol. Cell. Biol. **14**:77–86.
61. **Proudfoot, N.** 1991. Poly(A) signals. Cell **64**:671–674.
62. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10**:84–94.
63. **Rosbash, M., and B. Séraphin.** 1991. Who's on first? The U1 snRNP-5′ splice site interaction and splicing. Trends Biochem. Sci. **16**:187–190.
64. **Schek, N., C. Cooke, and J. C. Alwine.** 1992. Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. Mol. Cell. Biol. **12**:5386–5393.
65. **Senapathy, P., M. B. Shapiro, and N. L. Harris.** 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol. **183**:252–278.
66. **Seraphin, B., L. Kretzner, and M. Rosbash.** 1988. A U1 snRNA: pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5′ cleavage site. EMBO J. **7**:2533–2538.
67. **Sherman, L., N. Alloul, I. Golan, M. Durst, and A. Baram.** 1992. Expression and splicing patterns of human papillomavirus type-16 mRNAs in pre-cancerous lesions and carcinomas of the cervix, in human keratinocytes immortalized by HPV 16, and in cell lines established from cervical cancers. Int. J. Cancer **50**:356–364.
68. **Siebel, C. W., L. D. Fresco, and D. C. Rio.** 1992. The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5′ splice site control U1 snRNP binding. Genes Dev. **6**:1386–1401.
69. **Siliciano, P. G., and C. Guthrie.** 1988. 5′ splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes Dev. **2**:1258–1267.
70. **Spalholz, B. A., and P. M. Howley.** 1989. Papillomavirus-host cell

interactions, p. 27–53. *In* G. Klein (ed.), Advances in viral oncology, vol. 8. Raven Press, New York.

71. **Stenlund, A., J. Zabielski, H. Ahola, J. Moreno-Lopez, and U. Pettersson.** 1985. Messenger RNAs from the transforming region of bovine papilloma virus type I. J. Mol. Biol. **182**:541–554.
72. **Talerico, M., and S. M. Berget.** 1990. Effect of 5′ splice site mutations on splicing of the preceding intron. Mol. Cell. Biol. **10**:6299–6305.
73. **Tantravahi, J., M. Alvira, and E. Falck-Pedersen.** 1993. Characterization of the mouse β^{maj} globin transcription termination region: a spacing sequence is required between the poly(A) signal sequence and multiple downstream termination elements. Mol. Cell. Biol. **13**:578–587.
74. **Valsamakis, A., N. Schek, and J. C. Alwine.** 1992. Elements upstream of the AAUAAA within the human immunodeficiency virus polyadenylation signal are required for efficient polyadenylation in vitro. Mol. Cell. Biol. **12**:3699–3705.
75. **Valsamakis, A., S. Zeichner, S. Carswell, and J. C. Alwine.** 1991. The human immunodeficiency virus type 1 polyadenylation signal: a 3′ long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation. Proc. Natl. Acad. Sci. USA **88**:2108–2112.
76. **Van Gelder, C. W. G., S. I. Gunderson, E. J. R. Jansen, W. C. Boelens, M. Polycarpou-Schwarz, I. W. Mattaj, and W. J. Van Venrooij.** 1993. A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. EMBO J. **12**:5191–5200.
77. **Wassarman, K. M., and J. A. Steitz.** 1993. Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP. Genes Dev. **7**:647–659.
78. **Wawra, E., E. Pockl, E. Mullner, and E. Wintersberger.** 1981. Effect of sodium butyrate on induction of cellular and viral DNA syntheses in polyoma virus-infected mouse kidney cells. J. Virol. **38**:973–981.
79. **Weiss, E. A., G. M. Gilmartin, and J. R. Nevins.** 1991. Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. EMBO J. **10**:215–219.
80. **Wickens, M.** 1990. How the messenger got its tail: addition of poly(A) in the nucleus. Trends Biochem. Sci. **15**:277–281.
81. **Wintersberger, E., I. Mudrak, and U. Wintersberger.** 1983. Butyrate inhibits mouse fibroblasts at a control point in the G1 phase. J. Cell Biochem. **21**:239–247.
82. **Yang, Y. C., H. Okayama, and P. M. Howley.** 1985. Bovine papillomavirus contains multiple transforming genes. Proc. Natl. Acad. Sci. USA **82**:1030–1034.
83. **Yuo, C.-Y., and A. M. Weiner.** 1989. A U1 small nuclear ribonucleoprotein particle with altered specificity induces alternative splicing of an adenovirus E1A mRNA precursor. Mol. Cell. Biol. **9**:3429–3437.
84. **Zahler, A. M., K. M. Neugebauer, W. S. Lane, and M. B. Roth.** 1993. Distinct functions of SR proteins in alternative pre-mRNA splicing. Science **260**:219–222.
85. **Zhuang, Y., and A. M. Weiner.** 1986. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. Cell **46**:827–835.
86. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.