Characterization of the Polyomavirus Late Polyadenylation Signal

DAVID B. BATT AND GORDON G. CARMICHAEL*

Department of Microbiology, University of Connecticut Health Center, Farmington, Connecticut 06030

Received 16 March 1995/Returned for modification 24 April 1995/Accepted 3 June 1995

The polyomavirus late polyadenylation signal is used inefficiently during the late phase of a productive viral infection. Inefficient polyadenylation serves an important purpose for viral propagation, as it allows a splicing event that stabilizes late transcripts (G. R. Adami, C. W. Marlor, N. L. Barrett, and G. G. Carmichael, *J. Virol.* 63:85–93, 1989; R. P. Hyde-DeRuyscher and G. G. Carmichael, *J. Virol.* 64:5823–5832, 1990). We have recently shown that late-strand readthrough transcripts serve as natural antisense molecules to downregulate early-strand RNA levels at late times in infection (Z. Liu, D. B. Batt, and G. G. Carmichael, *Proc. Natl. Acad. Sci. USA* 91:4258–4262, 1994). Thus, poor polyadenylation contributes to the early-late switch by allowing the formation of more stable late RNAs and by forming antisense RNA to early RNAs. The importance of late poly(A) site inefficiency in the viral life cycle has prompted us to map the *cis* elements of this site. Since the polyomavirus late site proved a poor substrate for *in vitro* polyadenylation, we used an *in vivo* assay which allowed us to map the *cis* sequences required for its function. In this assay, various fragments containing the AAUAAA and different surrounding sequences were placed 1.4 kb upstream of a second, wild-type signal. The second signal served to stabilize transcripts that are not processed at the upstream site, allowing accurate quantitation of relative poly(A) site use by an RNase protection assay. Processing was primarily at the upstream site when a large fragment surrounding the poly(A) signal (50 nucleotides [nt] upstream and 90 nt downstream) was tested in this assay, demonstrating that this fragment contains the essential *cis* elements. Deletion analysis of this fragment revealed that most but not all upstream sequences can be removed with little effect on polyadenylation efficiency, indicating the absence of a strong stimulatory upstream element. Deletion of all but 25 nt downstream of the AAUAAA reduced polyadenylation activity only by half, demonstrating that processing can occur at this site despite the lack of downstream sequences. Thus, the core *cis* element for polyadenylation is quite small, with most important *cis*-acting elements lying within 19 nt upstream and 25 nt downstream of the AAUAAA sequence. This core contains the AAUAAA hexanucleotide, an upstream A/U-rich element, and three identical repeats of a 6-nt sequence, UAUUCA. Polyadenylation was eliminated or greatly reduced when either the AAUAAA or the three repeats were mutated.

The polyadenylation of pre-mRNAs requires both *cis*-acting elements and *trans*-acting factors. Factors important for 3'-end processing include the cleavage/polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), cleavage factors 1 and 2, and poly(A) polymerase (46). As many of these factors include multiple polypeptides, the addition of adenosines to the 3' end of a pre-mRNA is clearly a complex reaction. Two *cis*-acting elements have been identified as important for polyadenylation in mammalian cells. The first element is the sequence AAUAAA, located 10 to 30 nucleotides (nt) upstream of the cleavage site. This element interacts specifically with CPSF (25). The second important *cis* element is the U/GU-rich stretch, typically located within 50 nt downstream of the cleavage site (49). This element interacts with CstF (44, 45, 51). It is believed that CPSF binds to the AAUAAA sequence first but that this interaction is unstable (20) and is greatly influenced by sequence context of the AAUAAA element (19). The binding of CstF, which depends on the prior binding of CPSF, stabilizes the CPSF interaction (20, 21, 50). Cleavage factors 1 and 2 interact immediately downstream of CPSF to allow cleavage of the nascent RNA (50). Although most processing sites contain both an AAUAAA and a U/GU element, it has become clear that some sites contain other *cis*-acting sequences which modulate the efficiency of the cleavage and/or polyadenylation reaction. Systems in which *cis* elements modulate cleavage/polyadenyl-

ation efficiency include human immunodeficiency virus, ground squirrel hepatitis virus, the simian virus 40 (SV40) late genes, the U1A gene, adenovirus type 2, and cauliflower mosaic virus (8, 10, 13–15, 37–39, 42, 47, 48).

The polyomavirus late poly(A) site is inefficiently processed at late times in the viral life cycle (1, 23, 24). This inefficient polyadenylation of the late-strand RNA may serve an important purpose for viral propagation, as it allows a splicing event that stabilizes late transcripts (4, 24). Further, transcription downstream of the late poly(A) site serves to downregulate early mRNA levels at late times (28). The efficiency of late-strand polyadenylation at early times is difficult to address because the levels of late pre-RNA molecules before viral DNA replication are very low and difficult to detect by RNase protection assays. Indirect evidence that the polyadenylation efficiency may change during the viral life cycle is that transcription termination at the late polyadenylation site changes with time during a productive infection. Nuclear run-on assays indicated that transcription termination is efficient at early times but inefficient at late times (23).

Since the polyomavirus late poly(A) signal is weak *in vivo* and might be regulated, we have been interested in characterizing it further. As this site was used inefficiently *in vitro*, we designed an *in vivo* competition assay similar to the one used by Russnak (40). The assay is based on the observation that if two weak poly(A) sites are spaced far apart, the first site is overwhelmingly preferred as the site of polyadenylation (30). Using this *in vivo* assay, we found that the core *cis* element for late polyomavirus polyadenylation is quite small, with most important *cis*-acting elements lying within 19 nt upstream and

* Corresponding author. Phone: (203) 679-2259. Fax: (203) 679-1239.

25 nt downstream of the AAUAAA sequence. This core contains the AAUAAA hexanucleotide, an upstream A/U-rich element, and three identical repeats of a 6-nt sequence, UAUUCA. Polyadenylation was eliminated or greatly reduced when either the AAUAAA or the three repeats were mutated.

MATERIALS AND METHODS

Restriction enzymes, T4 DNA ligase, DNA polymerase I large fragment, T4 DNA polymerase, and T4 polynucleotide kinase were from New England Biolabs and were used as suggested by the manufacturer. RNase T₂ was prepared as previously described (27). [α -³²P]UTP was from New England Nuclear. Polyomavirus strain 59RA was used and has been described elsewhere (16, 17). All recombinant plasmids used here were propagated in *Escherichia coli* JM83.

Cell culture and transfections. Mouse NIH 3T3 cells were obtained from the American Type Culture Collection. The techniques for their propagation and transfection have been described in detail elsewhere (9). Before transfection, all recombinant plasmids were cut with *Eco*RI to release intact viral genomes and ligated under dilute conditions with T4 DNA ligase as described previously (3).

Constructs used. All constructs were created from the parent vector IR10. IR10 was created by removing the *Hinc*II fragment from the wild-type virus (59RA) and replacing it with a *Hinc*II-*Pvu*II fragment from pBR325. This created a construct in which a unique *Hinc*II site was 1.4 kb upstream of the late polyadenylation site. Different-size fragments spanning the wild-type polyomavirus late polyadenylation site were then cloned into the *Hinc*II site, creating constructs which contained two polyadenylation sites. First, a large fragment (50 nt upstream and 90 nt downstream of the AAUAAA) was cloned into the *Hinc*II site. This large fragment was efficiently used in the *in vivo* assay. To define what *cis* elements within this fragment are important, both 5' and 3' deletions of the fragment were made. The fragments used to create the 5' deletion constructs were generated by PCR amplification by using a constant 3' primer and one of several 5' primers. The 5' primers were designed to bind progressively closer to the 3' primer, thereby generating DNA products with constant 3' borders. These fragments were then treated with T4 kinase and ligated into the unique *Hinc*II site of IR10. The fragments used to create the 3' deletion constructs were created either by PCR or by restriction endonuclease digestion followed by ligation into the *Hinc*II site of IR10. The fragments used to generate IR14 and IR11 were generated by *Hinc*II and *Alu*I or *Nde*I, respectively. The fragments for IR40 and IR12 were created by PCR followed by ligation into the *Hinc*II site of IR10. Constructs IR22, IR23, IR24, IR37, and IR38, which contain small fragments of the poly(A) site, were created by hybridizing two oligonucleotides together, treating them with T4 kinase, and cloning them into the *Hinc*II site of IR10. Construct IR100 was made by site-directed PCR mutagenesis (5). Briefly, this procedure involved two independent PCRs, each of which used one mutagenic primer and one normal primer. The amplified DNAs from the two PCRs were mixed, allowing the mutagenized region to hybridize. This product was extended and amplified in a second round of PCR with the two outer primers. Construct IR30 was created by ligating the DNA fragment used to make IR24 upstream of the polyadenylation signal of construct IR12. This placed three repeats of the sequence UAUUCA upstream of the polyadenylation signal in IR12.

Harvesting of cytoplasmic RNA. Cells were harvested 44 to 48 h after initiation of transfection. Each plate was rinsed with 1× phosphate-buffered saline (PBS) without Mg²⁺ and Ca²⁺ (PBS⁻), and RNA was harvested by a protocol described elsewhere (24). Briefly, cells were scraped into PBS⁻ and pelleted by centrifugation at 1,000 × *g* for 3 min. The pellet was resuspended in PBS and centrifuged again at 1,000 × *g* for 3 min. The cell pellet was resuspended in 1.5 ml of Nonidet P-40 lysis buffer and left on ice for 5 min. Nuclei were removed by centrifugation at 12,000 × *g* for 2 min. To the supernatant (the cytoplasmic RNA fraction) was added 1.0 g of guanidine thiocyanate. RNA was separated from DNA by centrifugation through a 5.7 M cesium chloride cushion for 24 h at 110,000 × *g* (11). The pelleted RNAs were drained well and resuspended in 300 μl of RNase-free water.

Riboprobe preparation and RNase T₂ protection assays. All riboprobes were made from clones by using *in vitro* transcription with T7 RNA polymerase and [α -³²P]UTP, creating an internally labeled probe. It was necessary to maintain colinearity between the probe and each IR transcript being studied; therefore, different probes were used to map the processing site in each IR construct. A probe for each IR construct was made by PCR amplification with primers 135 bp upstream of the inserted polyadenylation test fragment and 90 bp downstream of the fragment. One of the PCR primers contained an *Eco*RI restriction site, and the other contained an *Xho*I site which allowed for the insertion of the fragment into the same sites in pBlueScript SK after it was treated with the appropriate restriction endonucleases. The probe construct for IR15 contains an additional 19 nt upstream, which produces protected fragments that are slightly longer. Each probe construct was digested with *Bam*HI, and riboprobes for each of the IR constructs were generated by runoff transcription. Uniformly labeled riboprobes were annealed to 25 to 50 μg of cytoplasmic RNA at 57°C for 15 to 20 h. The resulting hybrids were digested with 20 μl of RNase T₁/T₂ (27) at 37°C for 2 h. After addition of 200 μl of guanidine thiocyanate-glycogen (4 M guanidine thiocyanate, 25 mM sodium citrate [pH 7.0], 0.1 M 2-mercaptoethanol, 0.5%

FIG. 1. *In vitro* polyadenylation reactions. Precleaved substrates were generated by *in vitro* runoff transcription in the presence of [α -³²P]UTP as described in Materials and Methods. The 3' ends of the radiolabeled precleaved substrates were within 20 nt of the AAUAAA element. Cleavage/polyadenylation (C/P) substrates were generated similarly to the precleaved substrates except that the 3' ends were further downstream (about 50 nt). Radiolabeled RNAs were incubated in HeLa nuclear extracts for 1 h at 30°C, and products were resolved on a denaturing 6% polyacrylamide gel. Polyoma, polyomavirus; Ad, adenovirus.

N-lauroylsarcosine, 0.2 μg of glycogen per μl) and ethanol precipitation, the samples were dissolved in denaturing dye, and protected bands were resolved on 6% polyacrylamide-urea sequencing gels. Band intensities were quantified with a Packard Instant Imager, and percentages were assigned after adjustment for uridine content.

***In vitro* cleavage/polyadenylation assays.** *In vitro* cleavage/polyadenylation reactions were performed as described previously (35). RNAs were synthesized *in vitro*, using either T7 or SP6 RNA polymerase in the presence of m⁷G(5')ppp(5')G, ATP, CTP, UTP, and [α -³²P]UTP. The reaction mixture contained rATP (500 μM), rCTP (500 μM), rUTP (50 μM), rGTP (50 μM), m⁷G(5')ppp(5')G (500 μM), dithiothreitol (25 mM), 1× reaction buffer (40 mM Tris-HCl [pH 7.9], 6 mM MgCl₂, 2 mM spermidine, 10 mM NaCl), 20 U of RNasin, 45 μCi of radiolabeled UTP (800 Ci/mmol), 0.5 μg of linearized DNA template, and 20 U of T7 RNA polymerase. The transcription was allowed to proceed for 1 h at 37°C. Transcription was stopped by the addition of RQ DNase and incubation for 15 min at 37°C. Labeled RNAs were then incubated in the presence of HeLa cell nuclear extracts at 30°C for 1 h. Reaction mixtures were extracted with phenol-chloroform and ethanol precipitated. The products were separated by electrophoresis on 6% denaturing polyacrylamide gels.

RESULTS

In initial attempts to define the *cis*-acting elements important for polyomavirus late polyadenylation, we used HeLa cell nuclear extracts competent for *in vitro* polyadenylation (Fig. 1). Two types of substrates were provided. Precleaved substrates were created by *in vitro* runoff transcription producing capped RNAs which terminated shortly downstream of the AAUAAA signal. Such precleaved molecules do not require the cleavage factors and therefore provide a less stringent assay for polyadenylation site strength. *In vitro* reactions in which the 3' ends of the runoff transcripts were near the normal cleavage site (precleaved RNAs) showed that both the L3 site of adenovirus (construct provided by C. Moore) and, less efficiently, the polyomavirus late polyadenylation site can serve as substrates for the addition of adenosines (Fig. 1, lanes 2 and 5). A better measure of polyadenylation strength *in vitro* is the cleavage/polyadenylation reaction. This assay demands that the extract first cleave the nascent RNA and then add the adenosines to the newly formed 3' end. When such substrates are used, a clear difference can be seen between the L3 site and the polyomavirus late site. The L3 site, which is known to be a strong site both *in vitro* and *in vivo*, showed strong cleavage and/or polyadenylation (Fig. 1, lane 4), while the polyomavirus site failed to show any signal in the *in vitro* assay (Fig. 1, lane

FIG. 2. Schematic diagram of the in vivo polyadenylation competition assay. Test fragments surrounding the polyomavirus late polyadenylation site were inserted 1.4 kb upstream of another polyomavirus polyadenylation signal in an otherwise wild-type polyomavirus background. A probe construct specific for each of the IR series plasmids was made as described in Materials and Methods. Uniformly labeled antisense RNA for each IR construct was generated by in vitro runoff transcription from the specific probe clone. Radiolabeled RNA probes specific for the upstream site were hybridized to cytoplasmic RNA, and relative poly(A) site usage was determined by a quantitative RNase protection assay. Full-length protection indicates processing at the downstream site; a shorter protected fragment results if processing occurs at the upstream site. The probe for IR15 is slightly longer than those for the other constructs, resulting in longer protected fragments at both the upstream and downstream sites.

3). The lack of signal in this assay indicated that the polyomavirus late polyadenylation site is extremely inefficient in vitro. Such an assay could therefore not be used to map the *cis* elements important for its function.

We next developed an in vivo competition assay for poly(A) site efficiency. The assay is based on the observation that if a construct contains two weak polyomavirus late polyadenylation sites separated by a large distance (here 1.4 kb), the first polyadenylation site is overwhelmingly preferred as the site of polyadenylation (30). The assay that we used placed different fragments surrounding the polyadenylation site 1.4 kb upstream of the wild-type polyadenylation site. To maintain colinearity between the probe and each transcript of interest, a different uniformly labeled riboprobe was made for each construct and used to measure relative polyadenylation site use on

FIG. 4. RNase protection assays to measure relative polyadenylation site use of 3' deletion constructs. The in vivo assay described in the text was used to test polyadenylation efficiency of the constructs with a full-length insert (IR15) and 3' deletion fragments (IR14, IR11, and IR40). The mock lane represents RNA harvested from untransfected 3T3 cells. The upper band corresponds to use of the downstream site, and the lower band corresponds to use of the upstream (test) site. Band intensities were measured with a Packard Instant Imager and corrected for uridine content, and relative percentages were assigned.

cytoplasmic RNAs harvested from transfected cells (Fig. 2). Preferred use of the upstream site indicates that the inserted fragment can support processing. Preferred use of the downstream site indicates that the inserted fragment is poor at supporting cleavage and/or polyadenylation. When a fragment encompassing 50 nt upstream and 90 nt downstream of the polyadenylation signal was placed 1.4 kb upstream of the wild-type polyadenylation site, it was used 81% of the time (construct IR15; Fig. 3). This finding demonstrates that most elements required for the polyomavirus late poly(A) signal are contained within this fragment. Deleted variants of this fragment were then examined in order to determine which *cis* sequences are required for cleavage and/or polyadenylation.

Constructs with deletions downstream of the AAUAAA element. Constructs IR40, IR11, and IR14 contain progressively larger 3' deletions leaving 49, 25, and 12 nt, respectively, downstream of the AAUAAA (Fig. 3). RNase protection assays revealed that construct IR40 uses the upstream polyadenylation site 64% of the time, with the remaining polyadenylation taking place at the downstream site (Fig. 4, lane 4). Constructs

FIG. 3. Constructs used and summary of results of in vivo polyadenylation of test fragments. Thick solid lines indicate fragments which were inserted into the upstream position of construct IR10. The sequence of the wild-type polyomavirus late polyadenylation region is shown at the top. The relative use of each of the tested fragments is shown at the left. Relative use was determined in RNase protection assays with uniformly labeled antisense RNA probes spanning the upstream site. After correction for uridine content, percentages were assigned. Percentages shown are the averages of multiple independent experiments. Individual values did not deviate from those shown by more than 5%.

FIG. 5. RNase protection assays to measure relative polyadenylation site use. The *in vivo* assay described in the text was used to test polyadenylation efficiency of the constructs with a full-length insert (IR15), 5' deletion fragments (IR16, IR17, IR18, IR21, and IR34), and constructs with both 3' and 5' deletions (IR12 and IR37). The upper band corresponds to use of the downstream site, and the lower band corresponds to use of the upstream (test) site. Band intensities were measured with a Packard Instant Imager and corrected for uridine content, and relative percentages were assigned. Most bands located between the two expected bands were not reproducible. The extra bands in IR21, IR34, and IR12 were generally lighter than in this experiment and could either be an artifact of the specific riboprobe used or the result of a minor level of cryptic polyadenylation.

IR11 and IR14 use the upstream polyadenylation site 45 to 50% of the time (Fig. 4, lanes 2 and 3). Results with IR14 indicate that a fragment with only 12 nt downstream of the AAUAAA and only 1 nt downstream of the normal cleavage site can support moderate levels of polyadenylation. Inspection of the pBR325 sequence which is just downstream of the cleavage site in constructs with large 3'-end deletions revealed no obvious G+U-rich sequence and contained no more than two consecutive uridines. This finding argues against any large contribution by the downstream pBR325 sequences. If pBR325 sequences had a modest positive effect on processing, the percentages obtained for the 3' deletion constructs would represent an overestimate of the polyadenylation strength and would thus not change our conclusions. Lanoix et al. reported that a fragment similar to that in IR14 ending 12 nt downstream of the AAUAAA could not serve as a processing site (26). In this construct, however, the fragment was placed only about 50 nt upstream of the natural polyadenylation signal sequence. Polyadenylation sites with this close spacing will compete with each other, and polyadenylation will take place at the stronger site (6, 8). As shown here, a construct with only 12 nt of downstream sequence is greatly weakened and would be expected to compete poorly with the downstream poly(A) site in the closely spaced construct. Comparison of constructs IR11 and IR40 indicates that sequences between +25 and +49 contribute about 18% to the processing efficiency. This region contains a U/GU domain (10 of 14 nt) but is interspersed with cytosines. Comparison of constructs IR40 and IR15 indicates that sequences between +49 and +90 contribute about 17% to the polyadenylation activity. This region is not U/GU rich (37% G+U).

Constructs with deletions upstream of the AAUAAA element. Constructs IR16, IR17, IR18, IR21, and IR34 contain progressively larger 5' deletions leaving 44, 30, 15, 3, and 1 nt, respectively, upstream of the polyadenylation site (Fig. 3). Constructs IR16, IR17, and IR18 use the upstream polyadenylation site at 77, 84, and 74%, respectively, and showed little difference from IR15 (Fig. 5, lanes 1 to 4). However, constructs IR21 and IR34, which contain larger 5' deletions, show much

FIG. 6. Analysis of constructs IR22, IR23, and IR24. (A) RNase protection assays on the minimal fragment (IR22), a mutated hexanucleotide box construct (IR23), and a mutated poly(A) site construct (IR24). The top band represents use of the downstream site, and the bottom band represents use of the upstream poly(A) site. Bands located between the two expected bands were generally lighter than in this experiment and could either be an artifact of the specific riboprobe used or the result of a minor level of cryptic polyadenylation. (B) Sequences of the fragments tested in panel A. The poly(A) signal is underlined, and mutations are shown in lowercase boldface letters.

lower activity in the *in vivo* competition assay. Construct IR21 uses the upstream site only 43% of the time, while IR34 uses the upstream site 26% of the time (Fig. 5, lanes 5 and 6). This finding indicates that constructs IR21 and IR34 lack an important *cis* element which is present in IR15 to IR18. This element lies between 4 and 15 nt upstream of the AAUAAA signal and is almost entire (11 of 12 nt) A and U residues. If this element were important for RNA stability rather than for RNA processing, similar results would be expected because the upstream processed RNA would be preferentially degraded. Actinomycin D treatment, however, indicated that the upstream and downstream processed RNAs from construct IR34 were equally stable (data not shown).

Constructs with deletions both upstream and downstream of the AAUAAA element. Two additional fragments were next assayed for their abilities to support polyadenylation in the *in vivo* assay. Construct IR12 contains a fragment which has 1 nt upstream of the AAUAAA signal and 49 nt downstream of this element (Fig. 3). This construct consistently failed to show detectable use of the upstream site (Fig. 5, lane 8), indicating that sequences immediately upstream of the AAUAAA element are critical for polyadenylation. The possibility that a strong inhibitory sequence may have been brought into proximity to the AAUAAA in this construct is ruled out by analysis of construct IR34, which can process at the upstream site and has identical upstream elements. Construct IR37 contains a fragment of 18 nt upstream of the AAUAAA signal and 25 nt downstream of it (Fig. 3). This small, 49-nt fragment was capable of supporting about 50% polyadenylation in the *in vivo* competition assay (Fig. 5, lane 9). Construct IR22 is similar to IR37 but lacks the upstream A/T-rich region just upstream of the polyadenylation signal. This construct can support processing (about 33%) but at a level lower than that of IR37 (about 52%) (compare Fig. 5, lane 9, with Fig. 6, lane 1).

Functional importance of a hexanucleotide element repeated three times in the poly(A) signal region. Examination

FIG. 7. Analysis of constructs IR23, IR37, and IR38. (A) RNase protection assays of constructs lacking the repeated hexanucleotide elements (IR23 and IR38) and a construct containing it (IR37). The top band is undigested probe, the second band represents use of the downstream polyadenylation site, and the bottom band represents use of the upstream site. (B) Sequences of the fragments used in panel A. The poly(A) signal is underlined, and mutations are shown in lowercase boldface letters.

of the minimal sequences necessary for polyadenylation at the polyomavirus late site revealed the presence of three repeats of the hexanucleotide sequence UAUUCA within a 34-nt fragment spanning the AAUAAA signal. This hexanucleotide sequence is not found anywhere else on the late strand of the viral RNA. Further, lower activity in the *in vivo* assay correlated with constructs that lacked the 5' most of these hexanucleotide elements. Therefore, several additional experiments were performed to address the role that the hexanucleotide sequences may play in late polyadenylation. First, we tested a set of constructs based on a minimal polyadenylation-proficient fragment (34 nt). This fragment included only the polyadenylation signal and the three repeats of the hexanucleotide (construct IR22; Fig. 6A, lane 1). Two types of mutations were independently introduced into this fragment. First, double point mutants were introduced into each of the three hexanucleotide repeats, changing each of them to UAUGGA (construct IR23; Fig. 6A, lane 2). Second, two base changes were introduced into the conserved polyadenylation signal, changing it from AAUAAA to AAGCAA (construct IR24; Fig. 6A, lane 3). When these constructs were compared in the *in vivo* system, construct IR22 showed detectable activity (33%); however, the hexanucleotide mutant and the polyadenylation signal mutant both failed to exhibit any observable cleavage and/or polyadenylation within the test fragment (Fig. 6A). Therefore, both the hexanucleotide repeats and the polyadenylation signal appear to be critical for polyadenylation. In changing the hexanucleotide sequences in construct IR23, we also converted the CA at the likely cleavage site to a GA. Concern that this may have resulted in the low use of the upstream site in IR23 prompted us to make constructs IR38 and IR100 (see below).

A second set of constructs was tested in order to evaluate the importance of the hexanucleotide repeats. Construct IR37 (Fig. 7B) included the minimal region noted above (IR22) in addition to the A/U-rich region just upstream of the first hexanucleotide sequence which was shown to be important in the deletion study (compare IR18 with IR21 in Fig. 3). To destroy

FIG. 8. RNase protection assays of the full-length fragment (IR15) and a similar construct containing mutated hexanucleotide box elements (IR100). The top bands represent use of the downstream site, and the bottom bands represent use of the upstream site. Mutations in the hexanucleotide repeats were the same as those made in construct IR23.

the hexanucleotide sequences in this fragment, four base changes were introduced into each of the three hexanucleotide repeats, changing each to UGAAUA (construct IR38; Fig. 7B). Although IR38 changes the normal cleavage site from CA to TA, it does restore a CA sequence 3 nt downstream of the normal cleavage position. RNase protection assays revealed that construct IR37 uses the upstream site about 50% of the time whereas IR38 uses the upstream site less than 5% of the time (Fig. 7A, lanes 3 and 4). This result clearly demonstrates the importance of the UAUUCA elements.

Although the hexanucleotide repeats are critical in the short fragments, it remained possible that the repeats are dispensable in the context of a longer fragment. To confirm the role of the hexanucleotide repeats in the full-length IR15 construct, double point mutations were introduced into each of the three hexanucleotide repeats, changing them from UAUUCA to UAUGGA (IR100). Construct IR100 processes at the upstream site only about 37% of the time (Fig. 8, lane 2). This result indicates that even in the full-length background, the repeats are important for processing (Fig. 8; compare lanes 1 and 2).

Finally, to determine whether the hexanucleotides could restore activity to a nonfunctional fragment, we made a construct (IR30; Fig. 9B) in which hexanucleotide elements were placed immediately upstream of the nonfunctional IR12 fragment. The nonfunctional IR12 fragment contains the AAUAAA signal and downstream sequences but exhibits little or no activity, presumably because it lacks essential upstream sequences. The missing upstream sequences include the 5'-most hexanucleotide repeat element and the upstream A/U-rich region. IR30 showed about 25% activity in the *in vivo* assay (Fig. 9A, lane 2), again demonstrating that the repeated hexanucleotide sequences contribute to polyadenylation efficiency.

DISCUSSION

The polyomavirus late polyadenylation site represents one of the weakest naturally occurring sites so far reported, with only 50 to 70% polyadenylation efficiency at late times in productive viral infections (2, 24). Weak polyadenylation and transcription termination at late times in the viral life cycle may benefit the virus by creating long primary transcripts that are more stable than their shorter counterparts and serve as antisense RNA to early pre-mRNAs (24, 28). The antisense RNA serves to lower the accumulation of the early RNAs in the late

FIG. 9. Analysis of constructs IR12 and IR30. (A) RNase protection assays showing the effect of placing repeated hexanucleotide elements upstream of the AAUAAA element. The top band corresponds to use of the downstream site, and the bottom band corresponds to use of the upstream site. (B) Sequences of the fragments used in panel A. The poly(A) signal is underlined, and mutations are shown in lowercase letters. Construct IR12 includes only one nucleotide upstream of the AAUAAA element. IR30 is the same except a 34-nt fragment containing three hexanucleotide elements has been inserted immediately upstream of the AAUAAA signal.

phase of viral infection and thereby contributes to the early-late switch (28).

In addition to the highly conserved AAUAAA element, other sequences both upstream and downstream of the polyadenylation signal have been reported to contribute to polyadenylation efficiency (7, 8, 10–15, 18, 22, 33, 34, 37–39, 41, 42, 47, 48, 52). The presence of downstream elements (DSEs) has been described for many polyadenylation sites (7, 12, 18, 22, 33, 34, 41, 52). Generally, these elements do not contain a defined sequence but do contain a region of high U or G/U content. The U/GU-rich region typically lies 15 to 70 nt downstream of the polyadenylation site and is thought to be the binding site for CstF (44, 45, 51). The SV40 late polyadenylation site may contain two DSEs, one which is a uridine-rich region and the other of which is a guanosine-rich region which specifically interact with a cellular factor (36). The uridine-rich (four consecutive uridines) region of the SV40 late polyadenylation site has been shown to interact with CstF (32).

The presence of upstream elements (USEs) has been reported for the polyadenylation sites of several viruses, including SV40 (late site), human immunodeficiency virus, adenovirus type 2, cauliflower mosaic virus, ground squirrel hepatitis virus, and hepatitis B virus (8, 10, 13–15, 37–39, 42, 47, 48). Several of the USEs show homology and seem to be able to functionally replace each other (47). In the case of SV40, the USE, located 13 to 48 nt upstream of the AAUAAA element, has been shown to interact with the U1A protein (a component of the U1 ribonucleoprotein particle) (31, 43). Recent reports indicate that the sequences immediately surrounding the polyadenylation signal sequence contribute to the efficiency of the site, possibly by affecting the stability of interacting factors (19, 50).

The polyomavirus late polyadenylation site seems to have two distal components and three major proximal components which contribute to its recognition. The first distal component is a U/GU (10 of 14 nt) domain located 24 nt downstream of the cleavage site. Deletion of this region led to a slightly lower processing efficiency (compare results for IR40 and IR11 in

this study). Although this region is 71% G+U, it does not contain the four consecutive uridine residues which are present downstream of most cleavage sites (32). The second distal element is located between +49 and +90 and also contributes to the processing efficiency. This region is not U/GU rich and does not show any obvious motifs.

The first proximal element is the AAUAAA sequence. As expected, point mutations in this sequence showed no activity (Fig. 6, lane 3). The second proximal component important for the polyomavirus late polyadenylation site is an A/U-rich stretch located –16 to –6 nt relative to the AAUAAA signal. The significance of this region can best be seen by comparing constructs IR22 and IR37 (Fig. 6, lane 1; Fig. 7, lane 3). The importance can also be seen by comparing IR18 and IR21; here the removal of the A/U-rich region reduced the activity of the site by about twofold (Fig. 3). It is interesting that the USE in SV40 is also A/U rich but does not contain the same sequence.

The third component which is important for the core polyomavirus late polyadenylation site consists of the repeats of the sequence UAUUCA. These repeats seem to possess properties of both USEs and DSEs. One hexanucleotide is found upstream of the polyadenylation signal as with a USE, and two hexanucleotides are found downstream as with a DSE. Similar to the USE in SV40, the hexanucleotide element in polyomavirus consists of three repeats which are located both close to each other and close to the polyadenylation site (43). The element, however, differs from most DSEs and USEs in its proximity to the AAUAAA sequence. Like most processing sites, polyomavirus contains a U/GU-rich DSE; however, it differs in that its removal only slightly reduces processing efficiency.

The functional importance of the hexanucleotide repeat elements in polyadenylation is apparent from the deletion mapping and point mutants; however, these elements may not contribute equally to polyadenylation. The 3'-most hexanucleotide appears to be dispensable, as measured by our *in vivo* assay, because deletion of this hexanucleotide did not negatively affect the use of the polyadenylation site (compare IR14 and IR11 in Fig. 3). However, it should be pointed out that this conclusion is based on the comparison of two constructs, both of which lack large regions of downstream sequence. Therefore, the deletion of the 3'-most hexanucleotide may not result in a further reduction because the polyadenylation fragment is already crippled. The importance of the upstream hexanucleotide is harder to evaluate because deletions which remove it have also removed the A/U-rich region. However, a construct in which the first two hexanucleotides were mutated so as to leave only the third hexanucleotide intact showed a low level of polyadenylation (data not shown). Taken together, these data suggest that two hexanucleotide elements must be present for efficient polyadenylation and that a third hexanucleotide does not further increase polyadenylation as measured in this assay. To confirm this finding and to understand the relative importance of each repeat, each hexanucleotide must be mutated individually in the full-length construct.

The mechanism of how the hexanucleotide repeats function to stimulate cleavage and/or polyadenylation is unknown. At least two possibilities exist. First, the sequence of the hexanucleotides and their close proximity to the AAUAAA may provide a favorable context for the interaction of CPSF. Since CPSF interaction is the first step in the formation of the polyadenylation complex, increasing the stability of the CPSF interaction would be expected to result in higher processing efficiency. Second, the hexanucleotide repeats may serve as the binding site for a *trans*-acting factor. The *trans*-acting factor

would serve to recruit and or stabilize the interaction of CPSF or other components of the polyadenylation complex. CstF is thought to bind to consecutive uridine residues located downstream of the cleavage site and to stabilize the CPSF interaction. The polyomavirus late polyadenylation site lacks consecutive uridine residues, and this may contribute to its inefficiency. The stability of CPSF interaction with the polyomavirus late polyadenylation site, therefore, may rely more on context and less on the stability provided by CstF.

Several lines of evidence point to a possible role of the polyadenylation site in the early-late switch in the viral life cycle. The early-late switch cannot be regulated at the level of transcription because run-on assays indicate only small changes in the transcription rates of the early and late promoters during the life cycle (23). The early-late switch seems dependent on the amount of late RNA. A marked virus unable to leave the early phase of the life cycle can be induced to enter late phase by supplying late RNA *in trans*, indicating that the late RNA may participate in a titration event necessary for the early-late switch (29). As most of the late region has been excluded as the putative titrating agent by deletion analysis (unpublished results), it is tempting to speculate that part of the titration event necessary for the early-late switch might be mediated by the hexanucleotide sequences found at the polyadenylation site. Ongoing work in the laboratory is directed toward further elucidating the role of the hexanucleotide repeats in late polyadenylation and in regulation of the viral life cycle.

ACKNOWLEDGMENTS

This work was supported by grant CA45382 from the National Cancer Institute.

We thank Claire Moore for providing the adenovirus L3 polyadenylation constructs and Jeff Wilusz for advice on preparing nuclear extracts. We thank other members of the laboratory for helpful comments on the manuscript.

REFERENCES

- Acheson, N. 1981. Efficiency of processing of viral RNA during the early and late phases of productive infection by polyomavirus. *J. Virol.* **37**:628–635.
- Acheson, N. 1984. Kinetics and efficiency of polyadenylation of late polyomavirus nuclear RNA: generation of oligomeric polyadenylated RNAs and their processing into mRNA. *Mol. Cell. Biol.* **4**:722–729.
- Adami, G. R., and G. G. Carmichael. 1986. Polyomavirus late leader region serves an essential spacer function necessary for viability and late gene expression. *J. Virol.* **58**:417–425.
- Adami, G. R., C. W. Marlor, N. L. Barrett, and G. G. Carmichael. 1989. Leader-to-leader splicing is required for the efficient production and accumulation of polyomavirus late mRNAs. *J. Virol.* **63**:85–93.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.). 1993. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York.
- Batt, D. B., Y. Luo, and G. G. Carmichael. 1994. Polyadenylation and transcription termination in gene constructs containing multiple tandem polyadenylation signals. *Nucleic Acids Res.* **22**:2811–2816.
- Bhat, B. M., and W. S. M. Wold. 1985. ATTAAA as well as downstream sequences are required for RNA 3'-end formation in the E3 complex transcription unit of adenovirus. *Mol. Cell. Biol.* **5**:3183–3193.
- Brown, P. H., L. S. Tiley, and B. R. Cullen. 1991. Efficient polyadenylation within the human immunodeficiency virus type 1 long terminal repeat requires flanking U3-specific sequences. *J. Virol.* **65**:3340–3343.
- Cahill, K. B., and G. G. Carmichael. 1989. Deletion analysis of the polyomavirus late promoter: evidence for both positive and negative elements in the absence of early proteins. *J. Virol.* **63**:3634–3642.
- Carswell, S., and J. C. Alwine. 1989. Efficiency of utilization of the simian virus 40 late polyadenylation signal: effects of upstream sequences. *Mol. Cell. Biol.* **9**:4248–4258.
- 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.
- Conway, L., and M. Wickens. 1985. A sequence downstream of AAUAAA is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes. *Proc. Natl. Acad. Sci. USA* **82**:3949–3953.
- DeZazzo, J. D., and M. J. Imperiale. 1989. Sequences upstream of AAUAAA influence poly(A) site selection in a complex transcription unit. *Mol. Cell. Biol.* **9**:4951–4961.
- DeZazzo, J. D., J. E. Kilpatrick, and M. J. Imperiale. 1991. Involvement of long terminal repeat U3 sequences overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3' end formation. *Mol. Cell. Biol.* **11**:1624–1630.
- Falck-Pedersen, E., and J. Logan. 1989. Regulation of poly(A) site selection in adenovirus. *J. Virol.* **63**:532–541.
- Feunteun, J., L. Sompayrac, M. Fluck, and T. L. Benjamin. 1976. Localization of gene functions in polyoma virus DNA. *Proc. Natl. Acad. Sci. USA* **68**:283–288.
- Freund, R., G. Mandel, G. G. Carmichael, J. P. Barncastle, C. J. Dawe, and T. L. Benjamin. 1987. Polyomavirus tumor induction in mice: influences of viral coding and noncoding sequences on tumor profiles. *J. Virol.* **61**:2232–2239.
- Gil, A., and N. J. Proudfoot. 1984. A sequence downstream of AAUAAA is required for rabbit β -globin mRNA 3'-end formation. *Nature (London)* **312**:473–474.
- Gilmartin, G. M., E. S. Fleming, J. Oetjen, and B. R. Graveley. 1995. CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev.* **9**:72–83.
- 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.
- Gilmartin, G. M., and J. R. Nevins. 1991. Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-messenger RNA. *Mol. Cell. Biol.* **11**:2432–2438.
- Hart, R. P., M. A. McDevitt, H. Ali, and J. R. Nevins. 1985. Definition of essential sequences and functional equivalence of elements downstream of the adenovirus E2A and the early simian virus 40 polyadenylation sites. *Mol. Cell. Biol.* **5**:2975–2983.
- Hyde-DeRuyscher, R., and G. G. Carmichael. 1988. Polyomavirus early-late switch is not regulated at the level of transcription initiation and is associated with changes in RNA processing. *Proc. Natl. Acad. Sci. USA* **85**:8993–8997.
- Hyde-DeRuyscher, R. P., and G. G. Carmichael. 1990. Polyomavirus late pre-mRNA processing: DNA-replication-associated changes in leader exon multiplicity suggest a role for leader-to-leader splicing in the early-late switch. *J. Virol.* **64**:5823–5832.
- Keller, W., S. Bienroth, K. M. Lang, and G. Christofori. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-messenger-RNA 3' processing signal AAUAAA. *EMBO J.* **10**:4241–4249.
- Lanoix, J., R. W. Tseng, and N. H. Acheson. 1986. Duplication of functional polyadenylation signals in polyomavirus DNA does not alter efficiency of polyadenylation or transcription termination. *J. Virol.* **58**:733–742.
- Lichtler, A., N. L. Barrett, and G. G. Carmichael. 1992. Simple, inexpensive preparation of T1/T2 ribonuclease suitable for use in RNase protection experiments. *BioTechniques* **12**:231–232.
- Liu, Z., D. B. Batt, and G. G. Carmichael. 1994. Targeted nuclear antisense RNA mimics natural antisense-induced degradation of polyoma virus early RNA. *Proc. Natl. Acad. Sci. USA* **91**:4258–4262.
- Liu, Z., and G. G. Carmichael. 1993. Polyoma virus early-late switch: regulation of late RNA accumulation by DNA replication. *Proc. Natl. Acad. Sci. USA* **90**:8494–8498.
- 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.
- 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.
- MacDonald, C. C., J. Wilusz, and T. Shenk. 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol. Cell. Biol.* **14**:6647–6654.
- McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins. 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. *Cell* **37**:993–999.
- McLaughlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements. 1985. The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res.* **13**:1347–1368.
- Moore, C. L. 1990. Preparation of mammalian extracts active in polyadenylation. *Methods Enzymol.* **181**:49–75.
- Qian, Z. W., and J. Wilusz. 1991. An RNA-binding protein specifically interacts with a functionally important domain of the downstream element of the simian virus-40 late polyadenylation signal. *Mol. Cell. Biol.* **11**:5312–5320.
- Raabe, T., F. J. Bollum, and J. L. Manley. 1991. Primary structure and expression of bovine poly(A) polymerase. *Nature (London)* **353**:229–234.
- Russell, J. E., and S. A. Liehaber. 1993. Double-stranded RNA triggers generalized translational arrest in xenopus oocytes. *Biochem. Biophys. Res. Commun.* **194**:892–900.

39. **Russnak, R., and D. Ganem.** 1990. Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. *Genes Dev.* **4**:764–776.
40. **Russnak, R. H.** 1991. Regulation of polyadenylation in hepatitis-B viruses—stimulation by the upstream activating signal PS1 is orientation-dependent, distance-independent, and additive. *Nucleic Acids Res.* **19**:6449–6456.
41. **Sadofsky, M., and J. Alwine.** 1984. Sequences on the 3' side of hexanucleotide AAUAAA affect efficiency of cleavage at the polyadenylation site. *Mol. Cell. Biol.* **4**:1460–1468.
42. **Sanfacon, H., P. Brodmann, and T. Hohn.** 1991. A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Dev.* **5**:141–149.
43. **Schek, N., C. Cooke, and J. 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.
44. **Takagaki, Y., C. C. MacDonald, T. Shenk, and J. L. Manley.** 1992. The human 64-kDa polyadenylation factor contains a ribonucleoprotein-type RNA binding domain and unusual auxiliary motifs. *Proc. Natl. Acad. Sci. USA* **89**:1403–1407.
45. **Takagaki, Y., J. L. Manley, C. C. Macdonald, J. Wilusz, and T. Shenk.** 1990. A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-messenger RNAs. *Genes Dev.* **4**:2112–2120.
46. **Takagaki, Y., L. C. Ryner, and J. L. Manley.** 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev.* **3**:1711–1724.
47. **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.
48. **van Gelder, C. W., S. I. Gunderson, E. J. 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.
49. **Wahle, E., and W. Keller.** 1992. The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.* **61**:419–440.
50. **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.
51. **Wilusz, J., T. Shenk, Y. Takagaki, and J. L. Manley.** 1990. A multicomponent complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. *Mol. Cell. Biol.* **10**:1244–1248.
52. **Zhang, F., R. M. Denome, and C. N. Cole.** 1986. Fine-structure analysis of the processing and polyadenylation region of the herpes simplex virus type 1 thymidine kinase gene by using linker scanning, internal deletion, and insertion mutations. *Mol. Cell. Biol.* **6**:4611–4623.