JOURNAL OF VIROLOGY, June 1991, p. 3340–3343
0022-538X/91/063340-04$02.00/0

Vol. 65, No. 6

Efficient Polyadenylation within the Human Immunodeficiency Virus Type 1 Long Terminal Repeat Requires Flanking U3-Specific Sequences

PAMELA H. BROWN, LAURENCE S. TILEY, AND BRYAN R. CULLEN*

Howard Hughes Medical Institute and Department of Microbiology and Immunology, Duke University Medical Center, Durham, North Carolina 27710

Received 31 August 1990/Accepted 28 February 1991

During transcription of the human immunodeficiency virus type 1 provirus, polyadenylation signals present in the 5′ long terminal repeat (LTR) are disregarded while the identical polyadenylation signals present in the 3′ LTR are utilized efficiently. As both transcribed LTR sequences contain all signals known to be required for efficient polyadenylation, the basis for this differential utilization has been unclear. Here, we describe experiments that suggest that transcribed sequences present within the human immunodeficiency virus type 1 LTR U3 region act in *cis* to enhance polyadenylation within the 3′ LTR.

The retroviral replication cycle is characterized by the conversion of the single-stranded viral RNA genome to a double-stranded DNA provirus (9, 14). A critical step in this process is the jump of the reverse transcriptase molecule from the 5′ end to the 3′ end of the RNA genome, a process which is mediated by a short terminal redundancy that is referred to as the R element (9). In the majority of retroviruses, including human immunodeficiency virus type 1 (HIV-1), the generation of this terminal RNA repeat requires that the polyadenylation (PA) signal present in the 5′ long terminal repeat (LTR) of the integrated provirus must be ignored while the identical PA signal in the 3′ LTR is utilized efficiently (9, 14). This inefficient recognition does not result from any evident defect in the PA signal itself, as both sequences known to be required for efficient polyadenylation, i.e., both the conserved AAUAAA motif and the less clearly defined G-U box (1, 15), are present within the HIV-1 LTR (2). In fact, the PA sequences present in the 3′ LTR of HIV-1 have been shown to be very efficiently utilized when examined in transient expression assays in tissue culture (2). Two distinct hypotheses have therefore been advanced to explain the differential utilization of these apparently identical PA sites. The first suggests that inefficient recognition of the 5′ LTR PA site is due simply to its close proximity to the cap site (8, 12). Alternatively, efficient recognition of the 3′ LTR PA site cold be mediated by the action of transcribed viral sequences located 5′ to the downstream, but not the upstream, LTR PA signal (11, 13). In this report, we present data supporting the latter hypothesis.

We have previously described a vector, pU3R+39, in which the 3′ LTR of HIV-1 is positioned 3′ to the active cytomegalovirus immediate-early (CMV-IE) promoter (2) (Fig. 1). In this construct, the LTR is flanked 5′ by appropriate HIV-1 genomic sequences primarily derived from the viral *nef* gene. In this context, the HIV-1 LTR has been shown to function as a highly efficient PA site for transcripts derived from the CMV-IE promoter (2). The pU3R+39 vector also contains a second complete PA signal, derived from the genomic rat preproinsulin II gene (2, 4) (Fig. 1). In the parental pU3R+39 vector this default PA site, which is located 216 bp 3′ to the HIV-1 PA site, is not significantly utilized. However, the 3′ PA site is activated if deleterious mutations are introduced into either the AAUAAA sequence or the GU box of the HIV-1 LTR PA site (2; data not shown).

To examine whether utilization of the HIV-1 LTR PA site is affected by proximity to the CMV-IE promoter, we used convenient restriction sites to construct a series of deletion mutants of the pU3R+39 vector (Fig. 1). These mutations were designed to delete HIV-1 genomic sequences without directly affecting the LTR PA signal itself. The pU3R+39 vector and its derivatives were then transfected into COS cell cultures (4). At 72 h after transfection, total cytoplasmic RNA was harvested and subjected to quantitative S1 nuclease analysis using an end-labeled probe designed to detect the level of utilization of both the 5′ LTR and the 3′ insulin gene PA sites (2, 4) (Fig. 2). It is predicted that any inhibition of the 5′ LTR PA site induced by deletion of flanking viral sequences would result in activation of the 3′ site.

As shown previously (2), RNAs derived from the parental vector pU3R+39 were polyadenylated at high efficiency at the 5′ LTR PA site while very little RNA processed at the 3′ site was detected (Fig. 2, lane 2). Deletion of 169 (pΔ78/247) or 286 (pΔ78/361) bp of the HIV-1 sequences present between the CMV-IE promoter and the HIV-1 LTR had no effect on PA site utilization (Fig. 2, lanes 3 and 4). However, deletion of 301 bp of U3-specific sequences, in pΔ470/771, resulted in a modest but readily detectable activation of the 3′ PA site (Fig. 2, lane 5), while a deletion (pΔ396/873) that removed all of the HIV-1 LTR U3 region gave rise to essentially equivalent levels of RNA processed at the 5′ and 3′ sites (Fig. 2, lane 6). These data therefore demonstrate that the HIV-1 LTR PA site is less effective when moved closer to the 5′ promoter element.

To examine whether this decline was due simply to proximity to the promoter or instead reflected the loss of a specific viral sequence, we next examined whether the HIV-1 PA site could be rescued by insertion of nonrelevant DNA sequences between the cap and PA sites. Because the identity and prevalence of any *cis*-acting sequence required for activation of the HIV-1 PA site is not known, we chose to insert sequences derived from prokaryotic genes, which appeared likely to lack such sequences. The mutant used to

* Corresponding author.

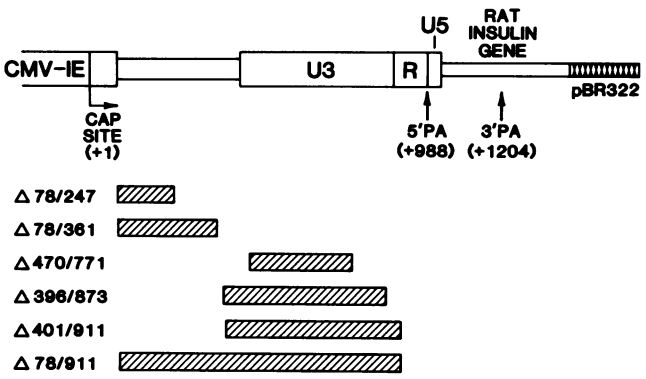

FIG. 1. Structure of the pU3R+39 expression vector. The pU3R+39 vector (2) contains 949 bp of HIV-1 DNA sequences, consisting primarily of the *nef* gene and the 3′ LTR, under the control of the CMV-IE promoter (4). The LTR contains all sequences previously defined as necessary for efficient polyadenylation (5′ PA), including the canonical AAUAAA signal and a GU box (1, 2, 15). The HIV-1 LTR is flanked 3′ by sequences derived from the genomic rat preproinsulin II gene, and these include a second consensus PA site (3′ PA) (2). The first 78 bp of the transcribed region are noncoding sequences derived from CMV. Deletions of the HIV-1-derived sequences were obtained by cleavage with restriction enzymes (see below) followed by blunt ending with Klenow DNA polymerase and religation. The size and location of the introduced deletions are indicated both by the hatched boxes present in the lower part of the figure and, more precisely, by the specific names assigned to each deletion (pΔ) mutant, using coordinates given relative to the start of transcription, which is designated +1: *Hin*dIII, +78; *Xho*I, +247; *Mst*II, +361; *Pvu*II, +396 and +873; *Eco*RV, +470 and +771; *Bgl*II, +401 and +911; start of U3, +434; start of R, +890; start of AAUAAA signal, +963; end of U5 sequences, +1027.

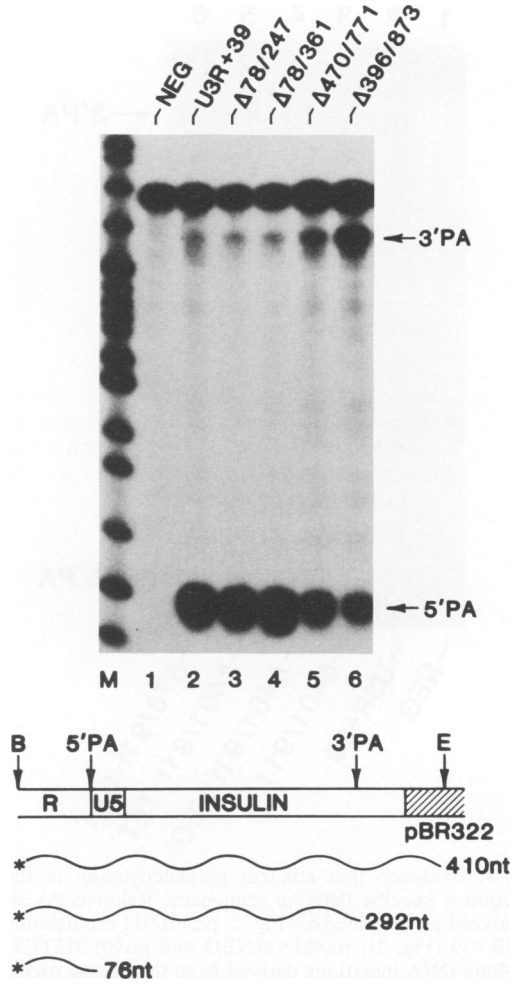

FIG. 2. Effect of 5′ deletions on PA site selection. COS cells were transfected with pU3R+39 and its derivatives, and cytoplasmic RNA was harvested at 72 h posttransfection (2, 4). Samples of RNA (5 μg) were then used in an S1 nuclease protection assay designed to quantitate the relative level of RNA polyadenylated at each PA site (2). Rescued probe fragments were resolved by electrophoresis on a 10% acrylamide-urea gel and visualized by autoradiography. The probe was uniquely end labeled at an HIV-1 R-region-specific *Bgl*II site (B) and extends 410 nucleotides (nt) to an *Eco*RI (E) site present in the flanking pBR322 sequence. RNA polyadenylated at the 5′ LTR PA site is predicted to rescue a probe fragment of 76 nucleotides, while RNA polyadenylated at the 3′ site is predicted to rescue a 292-nucleotide fragment. The DNA size marker lane (M) utilized end-labeled *Msp*II-cleaved pBR322 DNA. The negative control lane (NEG) was derived from a culture transfected with the parental vector pBC12/CMV (4). Other lanes were derived from cultures transfected with the indicated constructions.

test the effect of these substitutions, termed pΔ401/911, is closely similar to pΔ396/873 in both the size and location of the introduced deletion (Fig. 1), and this mutant also displays a high level of polyadenylation at the 3′ insulin PA site (Fig. 3, lane 3). Substituted prokaryotic sequences were derived from the bacterial *neo* (pΔ401/911NEO) or *tet* (pΔ401/911TET) gene (5). In each case, a 510-bp sequence was excised from the coding sequence of these prokaryotic genes by using the polymerase chain reaction (10) and was then substituted for the 510-bp LTR-specific sequence lacking in pΔ401/911. A third construct, pΔ78/911CAT, was derived by removal of a 833-bp fragment containing essentially all flanking HIV-1-specific sequences, other than the PA signals themselves, and substitution of a comparably sized 794-bp fragment bearing the complete prokaryotic *cat* gene (7). These plasmids were then transfected into COS cells, and the relative steady-state levels of mRNA polyadenylated at the 5′ (LTR) or 3′ (default) PA site were assayed by S1 nuclease protection analysis (Fig. 3). Levels of each of these RNA species were then quantitated by scanning the resultant autoradiograph with a laser densitometer (2) (Table 1).

As previously noted, transcripts derived from the parental pU3R+39 plasmid were efficiently polyadenylated at the LTR-specific PA site (Fig. 3, lane 2). In contrast, deletion of the LTR U3 region, pΔ401/911, resulted both in activation of the default 3′ PA site and in reduced polyadenylation at the LTR PA site (Fig. 3, lane 3; Table 1). Although substitution of an identically sized sequence derived from the prokaryotic *tet* gene had relatively little effect on the pattern of PA site utilization, a similarly sized DNA fragment derived from the *neo* gene was observed to result in a clear, if partial, restoration in the effectiveness of the LTR PA site (Fig. 3, lanes 4 and 5; Table 1). This may suggest that the *neo*, but not the *tet*, sequence was able to partially substitute for sequence information present in the HIV-1 LTR U3 sequence. It is important to note, however, that neither insertion restored recognition of the 5′ LTR PA site to the level observed in the parental pU3R+39 construct (Fig. 3, lane 2).

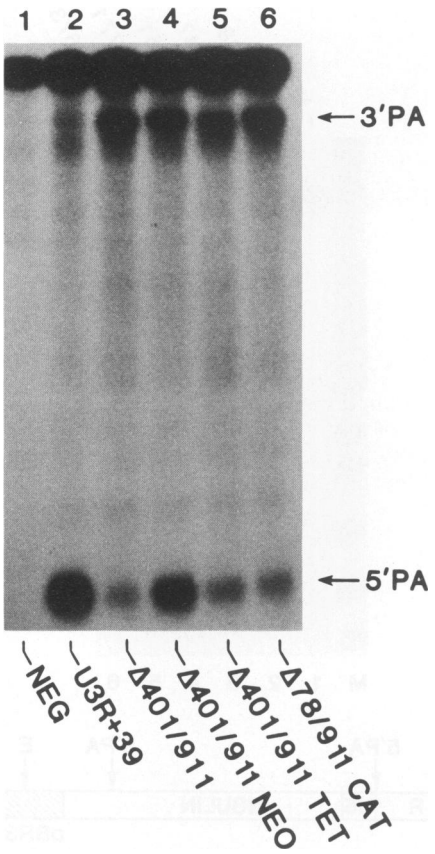

FIG. 3. Evidence that efficient polyadenylation in the HIV-1 LTR requires specific flanking sequences. Relative PA site usage was analyzed as described for Fig. 2. pΔ401/911 is a deletion mutant of pU3R+39 (Fig. 1); pΔ401/911NEO and pΔ401/911TET contain 510-bp-long DNA insertions derived from the coding regions of the bacterial *neo* and *tet* genes, respectively (5). pΔ78/911CAT contains the complete prokaryotic *cat* gene (7) in place of the indicated HIV-1 sequences. Quantitation of the data is presented in Table 1. NEG, Negative control.

Although plasmids Δ396/873 and Δ401/911 display relatively high levels of utilization (50 to 75%) of the default 3′ PA site, the 5′ PA site remains significantly active. Unfortunately, further deletion of sequences located between the cap site and the 5′ PA site resulted in the expression of increasingly unstable mRNA species (data not shown). We therefore addressed the effect of the deletion of all viral sequences located 5′ to the LTR PA site indirectly, by the complete replacement of these viral sequences with the prokaryotic *cat* gene (pΔ78/911CAT). Analysis of this construct revealed a level of 5′ PA site usage approximately similar to that observed in the pΔ401/911 construct (Fig. 3, lanes 3 and 6; Table 1). The residual, ~25% utilization of the HIV-1 LTR PA signal noted in these two constructs therefore appears to be independent of specific viral 5′-flanking sequences.

TABLE 1. Effect of 5′-flanking sequences on utilization of the HIV-1 LTR PA site

Transfected clone	% Efficiency of PA site[a]		Total RNA signal[b]
	HIV-1 LTR (5′)	Default (3′)	
U3R+39	>95	<5	1.36
Δ401/911	26	74	0.89
Δ401/911NEO	54	46	1.26
Δ401/911TET	32	68	0.63
Δ78/911CAT	25	75	0.86

[a] The relative steady-state level of mRNA species polyadenylated at either the 5′ or 3′ PA site for the indicated expression plasmids was derived by densitometric scanning of Fig. 3.

[b] Relative to the average signal, which was set at 1.00.

The data presented in Fig. 2 and 3 show the relative steady-state levels of mRNA species polyadenylated at two distinct PA sites. In addition, these mRNA species differ due to deletions or insertions. Although quantitation of the total RNA signal present in each lane in Fig. 3 (Table 1) suggested a relatively modest level of variability (setting 1.0 as the mean, the observed standard deviation was ±0.30) that could easily be explained by variable transfection efficiency or RNA recovery, we nevertheless wished to confirm that differential mRNA stability was not a major factor in our observations. For this purpose, we transfected several COS cell cultures with either plasmid pU3R+39, which gives mRNA polyadenylated exclusively at the 5′ PA site, or plasmid pΔ78/911CAT, which gives high levels of mRNA polyadenylated at the 3′ PA site. (The pΔ78/911CAT construct was chosen both because it is the most dissimilar to the parental pU3R+39 plasmid and because *cat* mRNA has been suggested to be relatively unstable in eukaryotic cells.) Dactinomycin (5 μg/ml) was added at 48 h after transfection, and the cultures then were sequentially harvested over a 24-h period. The level of mRNA polyadenylated at the 5′ and 3′ PA site was determined by S1 nuclease protection analysis (Fig. 4A) and quantitated by densitometric scanning (Fig. 4B). These results demonstrated that the mRNAs derived from pU3R+39 and pΔ78/911CAT polyadenylated at the 5′ PA site had an essentially identical half-life of ~12 h. The pΔ78/911CAT-specific mRNA polyadenylated at the 3′ site displayed a slightly longer half-life of ~16 h. These data therefore demonstrate that all of the mRNAs derived from the pU3R+39 and Δ78/911CAT vectors are comparatively stable and exclude variation in mRNA stability as an explanation for the apparent differential utilization of the two available PA sites present in these vectors.

In this report, we present an initial examination of the factors that govern the usage of the HIV-1 LTR PA site. When present distal to a promoter element, and flanked 5′ by appropriate viral genomic sequences, the HIV-1 LTR PA site was found to be highly effective. However, deletion of flanking LTR U3 region-specific sequences resulted in a significant reduction in the utilization of the LTR PA site (Fig. 2). Substitution of the deleted LTR sequences with prokaryotic DNA sequences of identical or similar size resulted in, at best, a partial restoration of PA site function (Fig. 3; Table 1). This loss of polyadenylation efficiency at the LTR PA site was shown not to result from differential RNA stability (Fig. 4) but instead reflected inefficient recognition of the HIV-1 LTR PA site. We therefore conclude that transcribed sequences present in the HIV-1 LTR U3 region act in *cis* to enhance the in vivo utilization of this retroviral PA site.

Although our results demonstrate that the inefficient utilization of the HIV-1 proviral 5′ LTR PA site must result, at least in part, from the different RNA sequence context of this site, we cannot exclude the possibility that proximity to the cap site may also inhibit PA site utilization (8, 12). However, our results suggest that this proximity effect, if it exists, is likely to be a less significant determination of HIV-1 LTR PA site utilization efficiency.

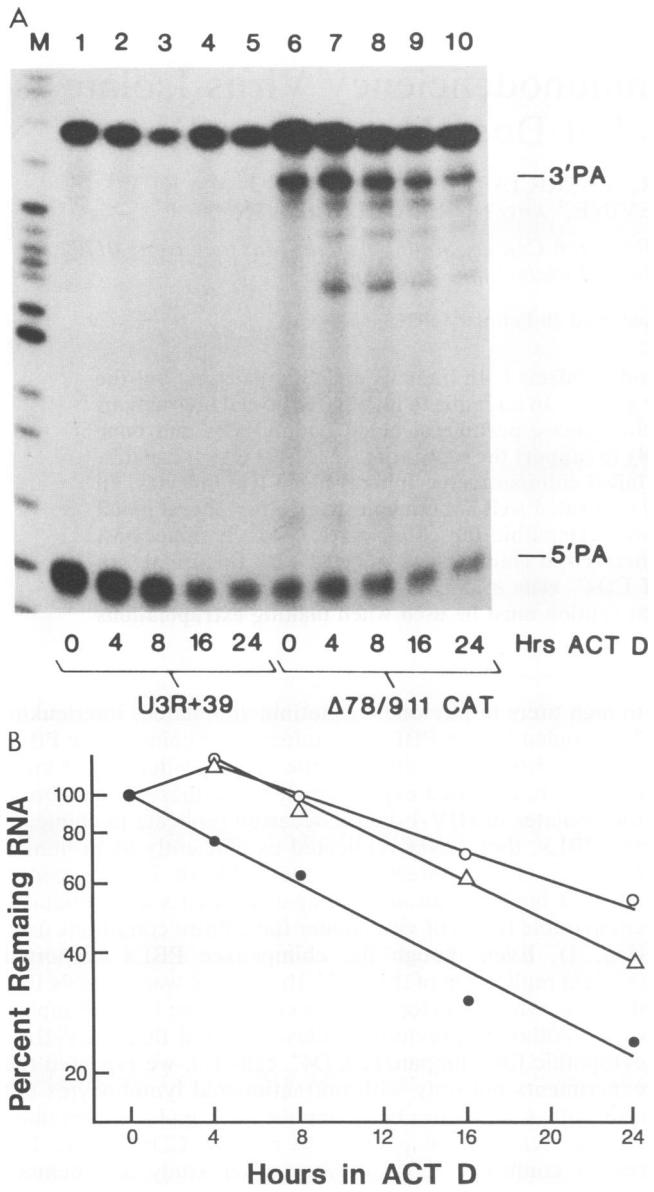

FIG. 4. Analysis of the stability of different mRNA species. (A) S1 analysis. Cells transfected with pU3R+39 (lanes 1 to 5) or pΔ78/911CAT (lanes 6 to 10) were treated with 5 μg of dactinomycin (ACT D) per ml for the indicated time periods prior to RNA harvest. Cell equivalent RNA samples were then utilized for S1 nuclease protection analysis. Additional specific bands noted in lanes 7 to 10 may be degradation products of a *cat* mRNA polyadenylated at the 3′ PA site. These bands were detected only after treatment with dactinomycin. (B) Plot of the decay rate of the different mRNA species in the presence of dactinomycin, derived by densitometric scanning of Fig. 4A. The 0-h signal was set at 100. The 0-h sample of the Δ78/911CAT-transfected culture displays a low RNA signal that is believed to reflect variable transfection efficiency or variable RNA recovery. All other samples appear to fall on a linear plot. Symbols: ●, U3R+39, 5′ PA; △, Δ78/911CAT, 5′ PA; ○, Δ78/911, 3′ PA.

The basis for the differential utilization of promoter-proximal and promoter-distal PA sites has been previously examined both in retroviruses and in the somewhat similar pararetroviruses. In agreement with our results, both Russnak and Ganem (11), working with hepatitis B virus, and Valsamakis et al. (13), working with HIV-1, have presented data arguing that specific RNA sequences located 5′ to these viral PA sites activate their utilization. In contrast, Iwasaki and Temin (8), working with spleen necrosis virus, and Sanfacon and Hohn (12), working with cauliflower mosaic virus, have concluded that the efficiency of viral PA site utilization is predominantly determined by the distance between the cap site and the PA site. It remains uncertain whether these different observations reflect real mechanistic differences in these distinct viral systems. However, it is noteworthy that *cis*-acting sequences located 5′ to the simian virus 40 late and adenovirus L1 PA signals have also been reported to enhance the utilization of these sites (3, 6). It will therefore be important to determine whether these disparate viruses do in fact regulate PA site selection via similar mechanisms. Finally, it is important to note that this regulation occurs in the absence of any virus-encoded gene products, thus suggesting the involvement of one or more cellular factors. The possibility that PA sites found in some cellular genes may also be activated by *cis*-acting sequence elements located 5′ to the known PA site consensus must therefore be seriously considered.

We thank Sharon Goodwin for preparation of the manuscript.

REFERENCES

1. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41**:349–359.
2. **Böhnlein, S., J. Hauber, and B. R. Cullen.** 1989. Identification of a U5-specific sequence required for efficient polyadenylation within the human immunodeficiency virus long terminal repeat. J. Virol. **63**:421–424.
3. **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.
4. **Cullen, B. R.** 1986. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. Cell **46**:973–982.
5. **Davies, J., and D. I. Smith.** 1978. Plasmid-determined resistance to antimicrobial agents. Annu. Rev. Microbiol. **32**:469–518.
6. **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.
7. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
8. **Iwasaki, K., and H. M. Temin.** 1990. The efficiency of RNA 3′-end formation is determined by the distance between the cap site and the poly(A) site in spleen necrosis virus. Genes Dev. **4**:2299–2307.
9. **Ju, G., and B. R. Cullen.** 1985. The role of avian retroviral LTRs in the regulation of gene expression and viral replication. Adv. Virus Res. **30**:179–223.
10. **Mullis, K. B., and F. A. Faloona.** 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. **155**:335–350.
11. **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.
12. **Sanfacon, H., and T. Hohn.** 1990. Proximity to the promoter inhibits recognition of cauliflower mosaic virus polyadenylation signal. Nature (London) **346**:81–84.
13. **Valsamakis, A., S. Zeichner, S. Carswell, and J. C. Alwine.** 1991. The human immunodeficiency virus type I polyadenylation signal: a 3′-LTR element upstream of the AAUAAA necessary for efficient polyadenylation. Proc. Natl. Acad. Sci. USA **88**: 2108–2112.
14. **Varmus, H.** 1988. Retroviruses. Science **240**:1427–1435.
15. **Wickens, M.** 1990. How the messenger got its tail: addition of poly(A) in the nucleus. Trends Biol. Sci. **15**:977–981.