MOLECULAR AND CELLULAR BIOLOGY, Mar. 1991, p. 1624–1630
0270-7306/91/031624-07$02.00/0

Vol. 11, No. 3

Involvement of Long Terminal Repeat U3 Sequences Overlapping the Transcription Control Region in Human Immunodeficiency Virus Type 1 mRNA 3′ End Formation

JAMES D. DEZAZZO, JAY E. KILPATRICK, AND MICHAEL J. IMPERIALE*

Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0620

Received 26 September 1990/Accepted 26 December 1990

In retroviral proviruses, the poly(A) site is present in both long terminal repeats (LTRs) but used only in the 3′ position. One mechanism to account for this selective poly(A) site usage is that LTR U3 sequences, transcribed only from the 3′ poly(A) site, are required in the RNA for efficient processing. To test this possibility, mutations were made in the human immunodeficiency virus type 1 (HIV-1) U3 region and the mutated LTRs were inserted into simple and complex transcription units. HIV-1 poly(A) site usage was then quantitated by S1 nuclease analysis following transfection of each construct into human 293 cells. The results showed that U3 sequences confined to the transcription control region were required for efficient usage of the HIV-1 poly(A) site, even when it was placed 1.5 kb from the promoter. Although the roles of U3 in processing and transcription activation were separable, optimal 3′ end formation was partly dependent on HIV-1 enhancer and SP1 binding site sequences.

The biosynthesis of eukaryotic mRNAs requires a number of posttranscriptional processing steps (23). One of these steps, 3′ end formation, is catalyzed by a large, multicomponent complex and is characterized by two tightly coupled reactions: endonucleolytic cleavage of the primary transcript and addition of a polyadenylate tail to the newly formed 3′ end (reviewed in reference 21). The search for *cis*-acting signals which control 3′ end processing has led to the identification of a critical core region in the RNA flanking the poly(A) site. The core consists of the highly conserved AAUAAA signal, which is positioned 10 to 30 nucleotides (nt) upstream of the cleavage site and is absolutely required for both cleavage and polyadenylation (29, 30), and less well conserved sequences located 5 to 50 nt downstream of the cleavage site, which stabilize binding of processing factors to the RNA (11).

Although the majority of processing studies suggest that sequences within the core region are necessary and sufficient to confer optimal poly(A) site usage, this view is currently being challenged by studies of several animal viruses (4, 8, 25). At the simian virus 40 (SV40) late and ground squirrel hepatitis virus (GSHV) poly(A) sites, critical elements found upstream of AAUAAA increase the efficiency of 3′ end processing in simple transcription units (4, 25). Moreover, at the adenovirus L1 poly(A) site, sequences upstream of AAUAAA play a similar yet distinct role, enhancing selection of the L1 poly(A) site in complex transcription units (8). By extending the functional definition of poly(A) site, these observations suggest that in systems in which differential poly(A) site usage occurs, the search for regulatory signals should include sequences outside the core region.

One long-standing question particularly amenable to this expanded view concerns the mechanism of regulated 3′ end formation during the replication of many retroviruses. In human immunodeficiency virus type 1 (HIV-1), for example, a poly(A) site is present at the boundary of the R and U5 regions in each proviral long terminal repeat (LTR; Fig. 1), yet only the site in the 3′ LTR is used. Examination of the LTR sequence does not provide an obvious explanation for this phenomenon, since the AAUAAA signal and downstream sequences are located in R and U5, respectively, and are therefore transcribed from both LTRs (3). In contrast, the U3 region of the LTR, which contains most of the transcriptional control elements of the provirus, is transcribed only from the 3′ LTR. Perhaps, then, HIV-1 U3 sequences are required in the RNA for efficient 3′ end formation. Indeed, retroviral U3 sequences can partly restore processing activity to a defective GSHV poly(A) site (25) and may even activate the spleen necrosis virus poly(A) site (9). Alternatively, the proximity of the poly(A) site in the 5′ LTR to the promoter may somehow preclude 3′ end formation at that site.

To directly test HIV-1 upstream sequences for 3′ end processing function in the absence of spacing constraints, a complete LTR with flanking viral sequences was cloned into an expression vector and the upstream region was subjected to deletion and point mutagenesis. The results of this study are presented here and indicate that U3 sequences confined to the promoter region are required for optimal HIV-1 poly(A) site usage.

MATERIALS AND METHODS

Reagents. Enzymes and molecular linkers were purchased from New England BioLabs, Inc., and Bethesda Research Laboratories, Inc. Radiochemicals were purchased from Amersham Corp.

Transfections and RNA isolation. Maintenance and transfection of human 293 cells and isolation of total, cytoplasmic, and nuclear RNA were done as previously described (8). Amounts of control and assay plasmid used in each transfection are indicated in the figure legends. The pattern of poly(A) site usage for each assay plasmid was independent of the amount of control plasmid used in the transfection.

* Corresponding author.

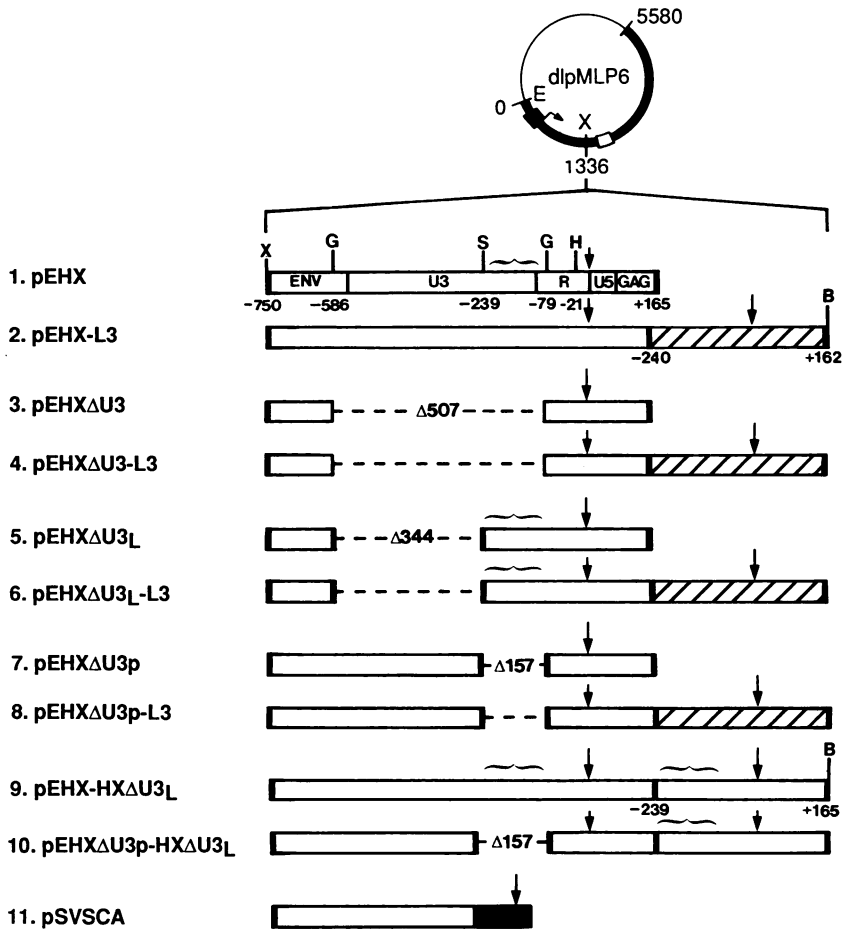

FIG. 1. Single and tandem poly(A) site constructs. dlpMLP6 (10) encodes the first 5,580 bp of the adenovirus type 5 sub360 genome, plus the adenovirus type 2 major late promoter in place of the E1A promoter (solid box with arrow), minus the E1B promoter and E1A poly(A) site (open box). Single and tandem poly(A) site constructs (lines 1 to 11) were made by inserting the appropriate fragments into dlpMLP6. Details of plasmid constructions are described in Materials and Methods. The relative distances (in base pairs) upstream and downstream of the cleavage sites are indicated by negative and positive numbers, respectively. The deleted HIV-1 sequences (△), cleavage sites (▼) HIV-1 transcription control region (brace), L3 poly(A) site (▨), and SV40 early poly(A) site (■ in line 11) are shown. Restriction sites: E, *Eco*RI; B, *Bam*HI; G, *Bgl*II; H, *Hin*dIII; S, *Sca*I; X, *Xba*I.

Plasmids and bacteria. All plasmids were grown, analyzed, and quantitated as previously described (8).

Plasmid constructions. (i) Deletion mutants. The parent plasmid for deletion mutagenesis, pEHX, was constructed stepwise, using a set of HIV-1 strain HXB2 genomic subclones spanning both provirus LTRs. Consequently, a 901-bp *Xho*I-*Bss*HII fragment consisting of the entire LTR, 188 bp of *env* sequence, and 77 bp of *gag* sequence was inserted into the *Xba*I site of dlpMLP6 (10). To facilitate cloning, the *Bss*HII site was filled in and ligated to a 12-bp *Bam*HI-*Xba*I linker. Similarly, the *Xho*I site was filled in and ligated to a 10-bp *Xba*I linker. From pEHX, a set of deletion mutants was then generated. For pEHXΔU3, a *Bgl*II fragment (HXB2 nt 9054 to 9560) was excised. For pEHXΔ$U3_L$, a *Bgl*II-*Sca*I fragment (nt 9054 to 9397) was removed, requiring addition of a *Bgl*II linker to the blunt *Sca*I site to close the vector. Similarly, removal of the *Sca*I-*Bgl*II fragment (nt 9398 to 9560) in pEHXΔU3p was followed by addition of a *Bgl*II linker to the *Sca*I site in the vector.

(ii) Point mutants. Plasmids containing mutations in the transcription control region were made by replacing the *Sca*I-*Hin*dIII fragment containing this region in pEHX (HXB2 nt 9398 to 9613) with the corresponding wild-type and mutated fragments from HIV strain H9. The wild type (pEH9), enhancer mutant (pEH9ENH), SP1 mutant (pEH9SP1), and *tar* mutant (pEH9tar) were derived from the WT, ΔENH, ΔSP1, and Δtar(+31/+34) plasmids, respectively, described by Kliewer et al. (18). The point mutations in each construct were confirmed by DNA sequencing using a Sequenase kit (U.S. Biochemicals) according to the manufacturer's protocol.

(iii) Tandem site constructs. To construct plasmids encoding two poly(A) sites, either the adenovirus L3 poly(A) site on a 402-bp *Bgl*II-*Bam*HI fragment or the HIV-1 poly(A) site on a 404-bp *Sca*I-*Bam*HI fragment from pEHX was inserted into the single *Bam*HI site of the appropriate single poly(A) site construct. Insertion of the HIV poly(A) site required the addition of a *Bcl*I linker at the *Sca*I site.

(iv) pSVSCA. The control plasmid was made by removing the 404-bp *Sca*I-*Bam*HI fragment encoding the HIV poly(A) site from pEHX and replacing it with the SV40 early poly(A)

site on a 131-bp *Hpa*I-*Bam*HI fragment obtained from pEC113 (15).

Virus constructions. vEHX-L3 and vEHXΔU3p-L3 were generated by the overlap recombination method (5). Recombinant viral fragments beginning with the left end of the genome were made by linearizing pEHX-L3 or pEHXΔU3p-L3 at the *Eco*RI site (indicated by an E in Fig. 1). Recombinant virus strains were plaque purified and propagated in 293 cells.

Quantitative S1 nuclease mapping. RNA from transfection of each single and corresponding tandem poly(A) site construct was analyzed with a DNA probe made from the tandem-site construct (e.g., pEHX and pEHX-L3 with a probe made from pEHX-L3). Each probe was specifically 3′ end labeled at the *Xba*I site upstream of the HIV poly(A) site (indicated by an X in Fig. 1). Hybridizations, S1 digestion, and densitometric scanning were performed as previously described (8) except that the temperature for hybridization was optimized at 52°C.

RESULTS

Experimental strategy. To study the role of U3 sequences in HIV-1 mRNA 3′ end formation, we adapted the plasmid vector dlpMLP6 (Fig. 1), which has been used in the study of adenovirus mRNA processing (10). This vector contains the adenovirus E1A gene minus its poly(A) site, and a unique *Xba*I site into which the poly(A) site (or sites) of interest can be inserted. Owing to the presence of adenovirus sequences including the left end of the viral genome, dlpMLP6 can also be used as a vector for producing recombinant adenoviruses. We began this study by inserting into dlpMLP6 the entire HIV-1 LTR, as well as 188 bp 5′ of U3 and 77 bp 3′ of U5, to make pEHX (Fig. 1). From pEHX we then constructed a set of mutants with the following deletions in U3 (Fig. 1): the entire U3 region (pEHXΔU3); the 5′ two-thirds (pEHXΔ$U3_L$); and the 3′ one-third, which includes the HIV-1 transcription control region (TCR) (pEHXΔU3p).

Since some upstream sequences influence poly(A) site usage only in complex transcription units (8), we wanted to assay the HIV upstream sequences in this genetic context as well. Therefore, complex transcription units were constructed from pEHX and the mutants by adding either the adenovirus L3 poly(A) site or a version of the HIV-1 poly(A) site encoding just 239 bp of 5′ sequence. Because the distance between the two poly(A) sites in all constructs is constant, any changes in their relative usage cannot be attributed to spacing effects.

Each assay plasmid was cotransfected into the 293 cell line, which efficiently drives transcription from the adenovirus promoter (1), along with the control plasmid pSVSCA, which contains the SV40 early poly(A) site in place of the HIV-1 site (Fig. 1). Total cellular poly(A)$^+$ RNA was isolated 48 h posttransfection and analyzed by S1 protection, using a 3′-end-labeled probe made from the tandem-site version of each construct. Using tandem-site probes to analyze RNA from single-site constructs allowed resolution of poly(A)$^+$ readthrough RNA, which therefore is included in the quantitative S1 analysis. The tandem-site probes also recognized HIV-1 sequences present upstream of the SV40 sequences in control RNA. S1 mapping of control and assay RNAs from each transfection sample could therefore be carried out together, allowing quantitation of 3′ end processing from each assay construct. This strategy was used before in the identification of upstream sequences required for adenovirus L1 3′ end processing (8).

Enhancement of HIV-1 poly(A) site usage by sequences within the transcription control region. S1 analysis of RNA from the constructs in Fig. 1 is shown in Fig. 2A, with the densitometric scanning data presented in Table 1. On the basis of at least three criteria, sequences spanning the TCR (indicated by a brace in Fig. 1) are required for efficient HIV-1 poly(A) site usage.

First, the amount of readthrough poly(A)$^+$ RNA (indicated by an R in Fig. 2A) from single-site constructs retaining the TCR is consistently much less than from those lacking the TCR. Whereas greater than 90% of transcripts are processed at the HIV-1 poly(A) site when the TCR is present, less than half are processed at that site when the TCR is absent (Fig. 2A; compare lanes 1 and 5 with lanes 3 and 7). Because the tandem-site probes are protected by the poly(A)$^+$ readthrough RNAs only up to the HIV-1/L3 junction, the actual sites at which these RNAs are processed is not known. The processing sites must, however, map to distal adenovirus and vector sequences, one possible site being the adenovirus E1B poly(A) site 2.3 kb downstream of the HIV-1 site.

Second, when directly compared with the internal control band (indicated by a C in Fig. 2A), processing at the HIV-1 poly(A) site is reproducibly two- to threefold greater from constructs retaining the TCR than from those lacking it. One possible explanation for this result is that the HIV-1 enhancer in the TCR may function in these plasmids, increasing the total amount of transcription and, therefore, 3′ end processing. As indicated by the processing efficiency data, however (Table 1), the total amount of poly(A)$^+$ RNA formed at the HIV-1 and downstream processing sites is essentially invariant for all single-site constructs. Therefore, the TCR appears to function in these constructs as a 3′ end processing element, not as a transcriptional enhancer. This conclusion is supported by the analysis of TCR mutants described below.

Third, in the complex transcription units encoding the HIV-1 and L3 poly(A) sites, preferential 3′ end formation clearly occurs at the HIV-1 site when the TCR is present but at the L3 site when the TCR is absent (Table 1; Fig. 2A; compare lanes 2 and 6 with lanes 4 and 8). Because the total amount of processed RNA from the different constructs remains constant, decreased use of the HIV-1 poly(A) site in RNA from pEHXΔU3-L3 or pEHXΔU3p-L3 is offset by increased usage of the L3 poly(A) site. A role for the TCR in transcript stability would therefore not be predicted from these findings. Furthermore, when transcription from pEHX-L3 or pEHXΔU3p-L3 is blocked with dactinomycin, the relative amount of RNA processed at the HIV-1 and L3 poly(A) sites remains constant, even after 36 h (data not shown). Thus, the sequences in U3 appear to directly affect 3′ end processing, not transcript stability.

Enhancement of HIV-1 poly(A) site usage by U3 sequences in miniproviruses or on the adenovirus chromosome. Comparison of data from the single- and tandem-site constructs (Table 1) shows that the L3 poly(A) site competes more effectively than the readthrough poly(A) sites for HIV-1 processing. This result suggests that competition between the poly(A) signals in the HIV-1 provirus might also contribute to the control of 3′ end processing. To explore this possibility, we constructed miniproviruses by substituting the HIV-1 poly(A) site for the L3 site in selected constructs (Fig. 1, lines 9 and 10). Although total poly(A)$^+$ RNA production from each construct is constant, the distribution of HIV-1 poly(A) site usage changes dramatically (Fig. 2A; compare lanes 9 and 10). When the TCR is transcribed

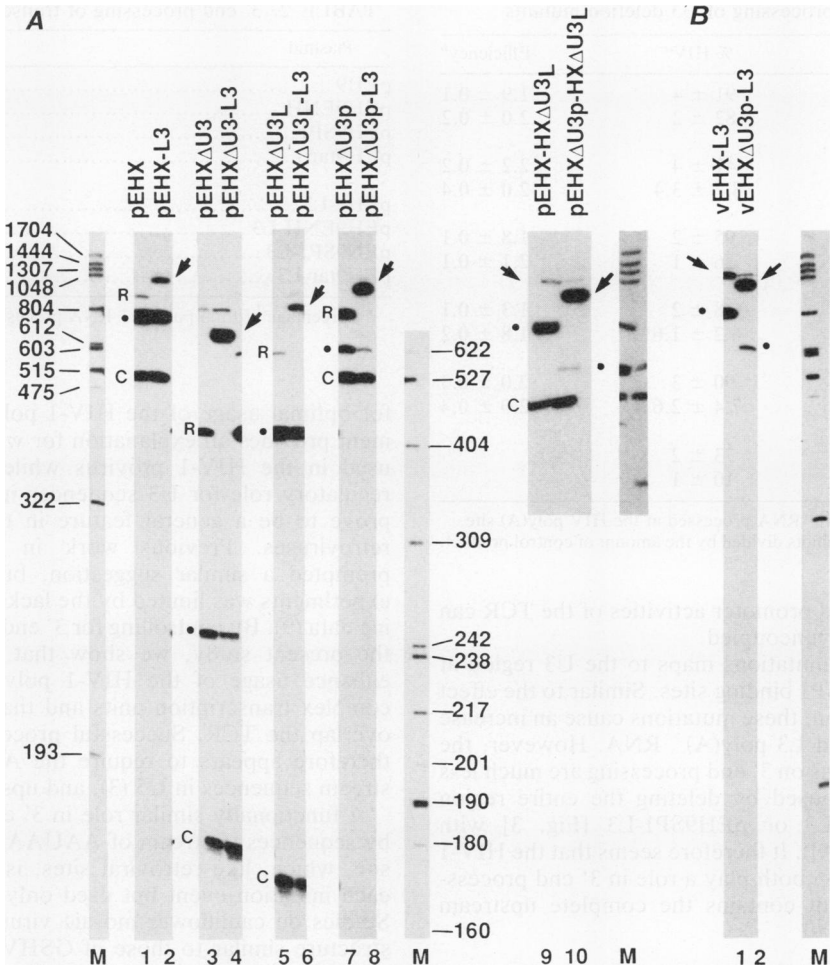

FIG. 2. Effects of U3 deletions on HIV-1 poly(A) site usage. Total, poly(A)$^+$ RNA isolated from transfections of 10 μg of assay and 3 μg of control plasmid (A) or from late infections with recombinant adenoviruses (B) was assayed as described in Materials and Methods. S1 products representing control RNA processed at the SV40 poly(A) site (C) and assay RNA cleaved at the HIV-1 (●), readthrough (R), and L3 (arrows) poly(A) sites are indicated. Sizes (in nucleotides) of DNA markers (M) are shown.

before both poly(A) sites, processing occurs predominantly at the 5′ poly(A) site. However, when the TCR is transcribed only before the 3′ poly(A) site, mimicking the situation on the actual provirus, processing shifts almost completely to that poly(A) site. Thus, even at a distance of about 1.5 kb from the promoter, the TCR directs predominant usage of the 3′ HIV-1 poly(A) site.

Another facet of retroviral infection is that the poly(A) sites are transcribed from a linear chromosome. To test whether HIV-1 U3 sequences would enhance poly(A) site usage on a linear molecule, the HIV-1 transcription units encoded on pEHX-L3 and pEHXΔU3p-L3 were crossed onto the adenovirus chromosome. When total poly(A)$^+$ RNA from a late infection of human 293 cells is assayed by S1 mapping, the pattern of poly(A) site usage from the recombinant adenoviruses is very similar to that seen from the corresponding plasmids (Table 1; Fig. 2B). The same results for the two viruses are observed when nuclear and cytoplasmic poly(A)$^+$ RNAs are assayed independently (data not shown). It appears, then, that U3 sequences function as 3′ end processing signals on both circular plasmids and linear chromosomes.

Role of HIV-1 transcription control signals in 3′ end processing. Having established that sequences overlapping the TCR are essential for optimal usage of the HIV-1 poly(A) site, we began a more detailed dissection of this region. We replaced the TCR in pEHX and in pEHX-L3 with corresponding fragments from HIV-1 strain H9. These inserts encode either the wild-type TCR or defective TCRs with clusters of point mutations in the enhancer, SP1 binding sites, or *tar* region. Previous work has shown that these mutations repress the activity of the HIV promoter in 293 cells to 34.5% (enhancer), 2.8% (SP1), or 9.1% (*tar*) of wild-type levels (18). In this study, we asked whether any of the TCR mutations also affected the efficiency of HIV-1 3′ end processing in simple and complex transcription units.

Each plasmid was therefore assayed by the same method used for the previous constructs; the resulting S1 analysis is shown in Fig. 3, and densitometric scanning values are given in Table 2. Two classes of mutants emerge from these experiments. First, in the R region of the LTR, a 4-bp change in the *tar* loop, which depresses E1A/E1B-induced transcription from the HIV promoter 10-fold (18), still permits wild-type levels of HIV-1 processing in simple and complex transcription units (Fig. 3; compare lanes 1 and 4 and lanes

TABLE 1. 3′ end processing of U3 deletion mutants

Construct	% HIV[a]	Efficiency[b]
pEHX	91 ± 4	1.9 ± 0.1
pEHX-L3	82 ± 2	2.0 ± 0.2
pEHXΔU3	41 ± 4	2.2 ± 0.2
pEHXΔU3-L3	8.7 ± 3.4	2.0 ± 0.4
pEHXΔ$U3_L$	95 ± 2	1.8 ± 0.1
pEHXΔ$U3_L$-L3	76 ± 1	2.1 ± 0.1
pEHXΔU3p	38 ± 2	1.3 ± 0.1
pEHXΔU3p-L3	6.2 ± 1.6	1.8 ± 0.2
pEHX-HXΔ$U3_L$	90 ± 3	2.0 ± 0.3
pEHXΔU3p-HXΔ$U3_L$	7.4 ± 2.6	2.0 ± 0.4
vEHX-L3	73 ± 2	
vEHXΔU3p-L3	10 ± 1	

[a] Percentage of total poly(A)$^+$ RNA processed at the HIV poly(A) site.
[b] Total amount of assay products divided by the amount of control product.

5 and 8). Processing and promoter activities of the TCR can therefore be genetically uncoupled.

The second class of mutations maps to the U3 region at either the enhancer or SP1 binding sites. Similar to the effect of deletions in this region, these mutations cause an increase in both readthrough and L3 poly(A)$^+$ RNA. However, the effects of these mutations on 3′ end processing are much less dramatic than those caused by deleting the entire region (compare pEH9ENH-L3 or pEH9SP1-L3 [Fig. 3] with pEHXΔU3p-L3 [Fig. 2A]). It therefore seems that the HIV-1 enhancer and SP1 region both play a role in 3′ end processing, but neither element contains the complete upstream processing signal.

DISCUSSION

This work demonstrates that LTR U3 sequences, located upstream of the AAUAAA signal, are required in the RNA

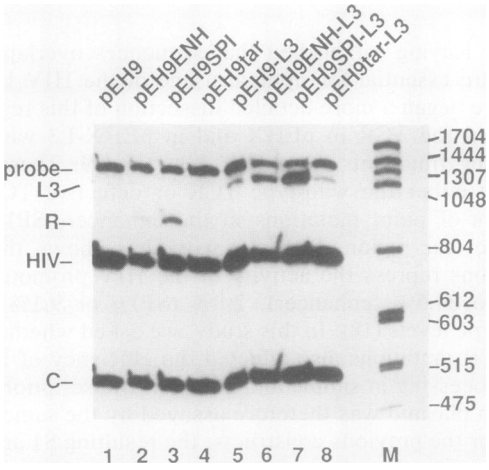

FIG. 3. Effects of TCR point mutations on HIV-1 poly(A) site usage. Total, poly(A)$^+$ RNA isolated from transfections of 5 μg each of assay and control plasmid was assayed as described in Materials and Methods. S1 products representing control RNA processed at the SV40 poly(A) site (C) and assay RNA processed at the HIV-1, readthrough (R), and L3 poly(A) sites are indicated. Sizes (in nucleotides) of DNA markers (M) are shown.

TABLE 2. 3′ end processing of transcriptional control mutants

Plasmid	% HIV[a]
pEH9	99 ± 1
pEH9ENH	87 ± 6
pEH9SP1	83 ± 7
pEH9tar	95 ± 4
pEH9-L3	89 ± 5
pEH9ENH-L3	67 ± 4
pEN9SP1-L3	63 ± 1
pEH9tar-L3	89 ± 5

[a] Percentage of total poly(A)$^+$ RNA processed at the HIV poly(A) site.

for optimal usage of the HIV-1 poly(A) site. This requirement provides an explanation for why the 3′ poly(A) site is used in the HIV-1 provirus while the 5′ site is not. A regulatory role for U3 sequences in 3′ end formation may prove to be a general feature in the replication of many retroviruses. Previous work in spleen necrosis virus prompted a similar suggestion, but interpretation of the experiments was limited by the lack of quantitative processing data (9). By controlling for 3′ end processing efficiency in the present study, we show that U3 sequences directly enhance usage of the HIV-1 poly(A) site in simple and complex transcription units and that the critical sequences overlap the TCR. Successful processing of HIV-1 mRNA therefore appears to require the AAUAAA signal, downstream sequences in U5 (3), and upstream sequences in U3.

A functionally similar role in 3′ end processing is played by sequences upstream of AAUAAA at the GSHV poly(A) site, which, like retroviral sites, is transcribed twice with each initiation event but used only in the 3′ position (25). Studies on cauliflower mosaic virus, which has a genomic structure similar to those of GSHV and retroviruses, indicate that both sequences upstream of AAUAAA (14a) and proximity of the poly(A) site to the promoter (26) influence the efficiency of 3′ end processing. The constructs used in the latter study did not contain U3 sequences, however, so the relative contributions of U3 and spacing cannot be inferred. We eliminated potential negative effects of spacing on processing by positioning the HIV-1 poly(A) site 1.5 kb from the adenovirus promoter. Even at this distance, the poly(A) site is used optimally only when U3 sequences are present. Therefore, poly(A) site proximity to the promoter appears to play, at best, only a minor role in the control of HIV-1 3′ end processing. TCR sequences, on the other hand, appear to play a major role.

Insight into the processing function of the HIV-1 TCR has been gained through analysis of mutations in well-characterized transcriptional control elements. The *tar* mutant provides evidence that promoter activity of the HIV-1 LTR can be genetically uncoupled from processing activity: although the cluster of four point mutations lowers the strength of the promoter at least 10-fold in 293 cells, these mutations have no effect on 3′ end processing. Whether any transcriptional control elements are also required for processing is not clear. Point mutations which disrupt either the enhancer or SP1 binding sites have only small, negative effects on 3′ end processing. Furthermore, linker scanning mutations in the U3 region, with the exception of replacements between the TATA box and cap site, have minor effects on processing (28). Therefore, we feel that stimulation of the HIV poly(A) site is mediated through a large structured region, which is unlikely to be perturbed by most point or linker scanning

mutations. A complete description of the TCR signals and associated factors will obviously bear upon an understanding of the coevolution of DNA signals controlling transcription and RNA signals specifying 3′ end processing.

A comparison of upstream and downstream elements controlling poly(A) site function reveals a number of common properties. Like downstream elements (reviewed in reference 21), upstream elements appear to lack any obvious conservation in primary sequence or secondary structure. Nevertheless, genetically dissimilar pairs of upstream or downstream signals can be functionally exchanged. For example, upstream sequences from the HIV-1 poly(A) site can replace those from the GSHV and SV40 late poly(A) sites (25, 28), and downstream sequences from the adenovirus E2 poly(A) site can substitute for those from the SV40 early poly(A) site (22). Whether an upstream element can be functionally exchanged with a downstream element is still an open question. Point mutagenesis of upstream and downstream elements suggests that the processing information in these signals is similarly organized. For example, clustered point mutations in critical sequences either upstream of the HIV-1 or downstream of the E2 poly(A) site (22) have only slight negative effects on the efficiency of 3′ end processing. Similarly, linker scanning mutations upstream of the HIV-1 site have, for the most part, little or no effect (28). Given their common attributes, upstream and downstream signals may play a similar role in poly(A) site complex formation: like downstream signals, the HIV-1 upstream processing signal may not bind processing factors directly, but instead may influence their association with the poly(A) site (11).

Future in vitro studies of 3′ end processing should elucidate the biochemical mechanisms underlying the action of upstream elements and indicate whether these elements fall into different functional classes, as has been proposed for downstream elements (22). Indeed, there is already evidence supporting two distinct roles for upstream sequences in mRNA 3′ end processing. In some systems, upstream sequences seem to act as household processing elements, required for efficient or exclusive poly(A) site usage (4, 25; this work). In other cases, however, they appear to play a fine-tuning role, controlling poly(A) site selection only in complex transcription units (8). In the adenovirus major late transcription unit, for example, sequences upstream of the L1 poly(A) site are necessary for temporal regulation of 3′ end processing during infection (7a). Interestingly, all reports of upstream sequences so far derive from studies of animal or plant viruses. Whether this reflects the importance of RNA processing as a means of genetic control in viruses or simply the popularity of viruses as experimental systems remains undetermined.

Upstream processing sequences at poly(A) sites represent yet another example of the positional diversity of regulatory signals controlling mRNA biogenesis and function. For example, sequences mediating the binding of transcription factors were once thought to lie only upstream of the transcription start site but are now known to occur downstream as well (e.g., references 17 and 20). Similarly, elements which regulate splicing choices in mRNA precursors have been found in both introns (12, 14) and exons (13, 27). In addition, destabilizing elements governing mRNA turnover have been discovered in both 3′ noncoding and 5′ coding regions (reviewed in reference 7). Finally, elements influencing the efficiency of mRNA translation have been located in both 5′ and 3′ noncoding regions (16, 19). It is therefore perhaps not too surprising that 3′ end formation, once thought to depend on just the AAUAAA signal, cleavage site, and sequences downstream of the cleavage site, can sometimes require additional sequences upstream of the AAUAAA. Given that the common cellular processes of RNA splicing and 3′ end cleavage were first discovered in viruses (2, 6, 24), we conjecture that sequences upstream of AAUAAA controlling poly(A) site strength will soon be found in cellular genes.

ACKNOWLEDGMENTS

We thank Joe Garcia and Richard Gaynor for HIV-1 H9 wild-type and mutant clones, Don Court for HIV-1 HXB2 plasmids, and the members of our laboratory for helpful discussions. We also thank Jim Alwine and Arthur Hunt for communicating results before publication.

This work was supported by PHS grant GM34902 from the NIH. M.J.I. is the recipient of Faculty Research Award FRA-338 from the American Cancer Society.

REFERENCES

1. **Aiello, L., R. Guilfoyle, K. Huebner, and R. Weinmann.** 1979. Adenovirus 5 DNA sequences present and RNA sequences transcribed in transformed human embryo kidney cells (HEK-Ad-5 or 293). Virology **94**:460–469.
2. **Berget, S. M., C. Moore, and P. A. Sharp.** 1977. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. USA **74**:3171–3175.
3. **Bohnlein, 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.
4. **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.
5. **Chinnadurai, G., S. Chinnadurai, and J. Brusca.** 1979. Physical mapping of a large-plaque mutation of adenovirus type 2. J. Virol. **32**:623–628.
6. **Chow, L. T., R. E. Gelinas, T. R. Broker, and R. J. Roberts.** 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell **12**:1–8.
7. **Cleveland, D. W., and T. J. Yen.** 1989. Multiple determinants of eukaryotic mRNA stability. New Biol. **1**:121–126.

7a. **DeZazzo, J.** 1990. Ph.D. thesis. University of Michigan, Ann Arbor.

8. **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.
9. **Dougherty, J. P., and H. M. Temin.** 1987. A promoterless retroviral vector indicates that there are sequences in U3 required for 3′ RNA processing. Proc. Natl. Acad. Sci. USA **84**:1197–1201.
10. **Falck-Pedersen, E., and J. Logan.** 1989. Regulation of poly(A) site selection in adenovirus. J. Virol. **63**:532–541.
11. **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**:2180–2190.
12. **Goux-Pelletan, M., D. Libri, Y. d'Aubenton-Carafa, M. Fiszman, E. Brody, and J. Marie.** 1990. In vitro splicing of mutually exclusive exons from the chicken beta-tropomyosin gene: role of the branch point location and very long pyrimidine stretch. EMBO J. **9**:241–249.
13. **Hampson, R. K., L. La Follette, and F. M. Rottman.** 1989. Alternative processing of bovine growth hormone mRNA is influenced by downstream exon sequences. Mol. Cell. Biol. **9**:1604–1610.
14. **Helfman, D. M., R. F. Roscigno, G. J. Mulligan, L. A. Finn, and K. S. Weber.** 1990. Identification of two distinct intron elements involved in alternative splicing of beta-tropomyosin pre-mRNA. Genes Dev. **4**:98–110.

14a. **Hunt, A.** Personal communication.

15. **Imperiale, M. J., R. P. Hart, and J. R. Nevins.** 1985. An enhancer-like element in the adenovirus E2 promoter contains

sequences essential for uninduced and E1A-induced transcription. Proc. Natl. Acad. Sci. USA **82**:381–385.
16. **Jackson, R. J., and N. Standart.** 1990. Do the poly(A) tail and 3′ untranslated region control mRNA translation. Cell **62**:15–24.
17. **Jansen-Durr, P., G. Mondesert, and C. Kedinger.** 1989. Replication-dependent activation of the adenovirus major late promoter is mediated by the increased binding of a transcription factor to sequences in the first intron. J. Virol. **63**:5124–5132.
18. **Kliewer, S., J. Garcia, L. Pearson, E. Soultanakis, A. Dasgupta, and R. Gaynor.** 1989. Multiple transcriptional regulatory domains in the human immunodeficiency virus type 1 long terminal repeat are involved in basal and E1A/E1B-induced promoter activity. J. Virol. **63**:4616–4625.
19. **Kozak, M.** 1989. The scanning model for translation: an update. J. Cell Biol. **108**:229–241.
20. **Leong, K., W. Lee, and A. J. Berk.** 1990. High-level transcription from the adenovirus major late promoter requires downstream binding sites for late-phase-specific factors. J. Virol. **64**:51–60.
21. **Manley, J. L.** 1988. Polyadenylation of mRNA precursors. Biochim. Biophys. Acta **950**:1–12.
22. **McDevitt, M. A., R. P. Hart, W. W. Wong, and J. R. Nevins.** 1986. Sequences capable of restoring poly(A) site function define two distinct downstream elements. EMBO J. **5**:2907–2913.
23. **Nevins, J. R.** 1983. The pathway of eukaryotic mRNA formation. Annu. Rev. Biochem. **52**:441–466.
24. **Nevins, J. R., and J. E. Darnell, Jr.** 1978. Steps in the processing of Ad2 mRNA: poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell **15**:1477–1493.
25. **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.
26. **Sanfacon, H., and T. Hohn.** 1990. Proximity to the promoter inhibits recognition of cauliflower mosaic virus polyadenylation signal. Nature (London) **346**:81–84.
27. **Streuli, M., and H. Saito.** 1989. Regulation of tissue-specific alternative splicing: exon-specific cis-elements govern the splicing of leukocyte common antigen pre-mRNA. EMBO J. **8**:787–796.
28. **Valsamakis, A., S. Zeichner, S. Carswell, and J. C. Alwine.** Proc. Natl. Acad. Sci. USA, in press.
29. **Wickens, M., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3′ end formation. Science **226**:1045–1051.
30. **Zarkower, D., P. Stephenson, M. Sheets, and M. Wickens.** 1986. The AAUAAA sequence is required both for cleavage and for polyadenylation of simian virus 40 pre-mRNA in vitro. Mol. Cell. Biol. **6**:2317–2323.