JOURNAL OF VIROLOGY, Oct. 1993, p. 6265–6269
0022-538X/93/106265-05$02.00/0

Vol. 67, No. 10

Influence of Sequences in the Long Terminal Repeat and Flanking Cell DNA on Polyadenylation of Retroviral Transcripts

AMANDA SWAIN† AND JOHN M. COFFIN*

Department of Molecular Biology and Microbiology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, Massachusetts 02111

Received 8 March 1993/Accepted 29 June 1993

Readthrough transcripts are formed during retrovirus infection by polyadenylation of viral RNA in cellular sequences adjacent to the provirus. We have studied such transcripts in avian leukosis virus-infected cell clones containing a single provirus, either the wild type or one with an inactivating mutation in the poly(A) addition signal. All individual wild-type proviruses produced readthrough transcripts, implying that this property is not restricted to a few integration sites. The range of sizes of viral RNA in the mutant lacking a correct signal for poly(A) addition reflected both the occurrence of functional polyadenylation sites within flanking cell DNA and increased usage of cryptic sites within viral sequences.

Retroviruses replicate via a DNA intermediate, the provirus, integrated within the cell DNA. Viral transcripts are produced by host cell machinery which recognizes transcription and processing signals present in the long terminal repeats (LTRs) which flank the viral genes. The 5′ LTR signals initiation of transcription, while cleavage and polyadenylation to generate the RNA 3′ end occur in sequences derived from the 3′ LTR. Sequences important for cleavage and polyadenylation of viral transcripts are similar to signals found in eukaryotic mRNA. All retroviruses contain the canonical AAUAAA sequence essential for correct 3′ processing of mRNAs (1, 5). A characteristic of retroviral RNA 3′-end processing is the presence of readthrough transcripts that extend into neighboring cellular sequences and constitute up to 15% of viral RNA in cells infected with avian leukosis virus (3). A high level of readthrough transcripts, caused by a mutation in the hexanucleotide to AAGGAA introduced into Rous-associated virus type 1 (RAV-1), abolishes cleavage and polyadenylation at the normal site within 3′ LTR-related sequences but is not grossly detrimental to virus replication (8), and such transcripts can be intermediates in the process of transduction of cellular genes by retroviruses (9). The experiments presented here were designed to study the process of readthrough transcript formation in individual proviruses.

For this purpose, clones of QT6 cells (4) containing single proviruses of RAV-1 (rescued from plasmid pRAV-1 [7, 8]) with either the wild-type (AAUAAA) or the mutant (AAGGAA [8]) hexanucleotide in the LTRs were isolated. Four clones with the wild-type hexanucleotide (designated WT-) and six with the mutant hexanucleotide (designated M-) were selected for further analysis. The purity of four (WT13, WT33, WT34, and M55) of the nine clones was confirmed by presence of a single provirus at the same site in the DNAs of at least four subclones of each (data not shown).

To determine the levels of readthrough RNA produced by the individual proviruses, cellular RNAs from the four clones containing a wild-type provirus were analyzed by hybridization to an RNA probe complementary to the LTR and leader sequences (3, 8) (Fig. 1). All of the cell clones analyzed generated detectable levels of readthrough transcripts (Fig. 1A). In most clones, the level of readthrough relative to normally processed RNA was less than that in the polyclonal infection. Nevertheless, most or all individual proviruses seem to be capable of giving rise to readthrough transcripts, at levels which probably reflect the inefficiency of viral polyadenylation signals and the occurrence of functional 3′-end processing signals in the neighboring cellular sequences. Such cellular signals must be close enough to be transcribed before cleavage at the viral site can occur. The lack of wild-type proviruses which failed to yield readthrough transcripts implies that such sequences must be quite frequent in cell DNA. Inefficient polyadenylation leading to the formation of readthrough transcripts has also been reported for cellular mRNAs (5).

To determine the sizes and structures of readthrough viral transcripts, cellular RNAs from five clones containing the polyadenylation mutation in the provirus were analyzed by electrophoresis through a 1% agarose gel containing 8% formaldehyde, transferred to nitrocellulose (6), and hybridized to a labeled DNA probe containing LTR sequences. These clones contained viral transcripts of well-defined sizes, in contrast to the smeared and indistinct pattern in the nonclonal infection (Fig. 2A; an enlargement of the region containing genome-size viral RNA from wild-type and mutant nonclonal virus infections is shown in Fig. 2B). These results suggest that the readthrough transcripts produced during infection are cleaved and polyadenylated predominantly at a specific site for each provirus.

To determine the characteristics of the endings of readthrough RNAs, cDNAs derived from cell clones containing mutant proviruses were amplified by a modification of the RACE technique (2) using an oligo(dT)-containing primer and an LTR-specific primer. Sequence analysis showed that readthrough transcripts were polyadenylated at sites within neighboring cellular sequences that contained sequences identical (clone 84a) or related to AAUAAA, 14 to 22 bp upstream of the polyadenylation site (Fig. 3). No other similarities either in cell-derived sequences or in distance from the LTR to the polyadenylation site were

* Corresponding author.
† Present address: National Institute for Medical Research, London NW7 1AA, England.

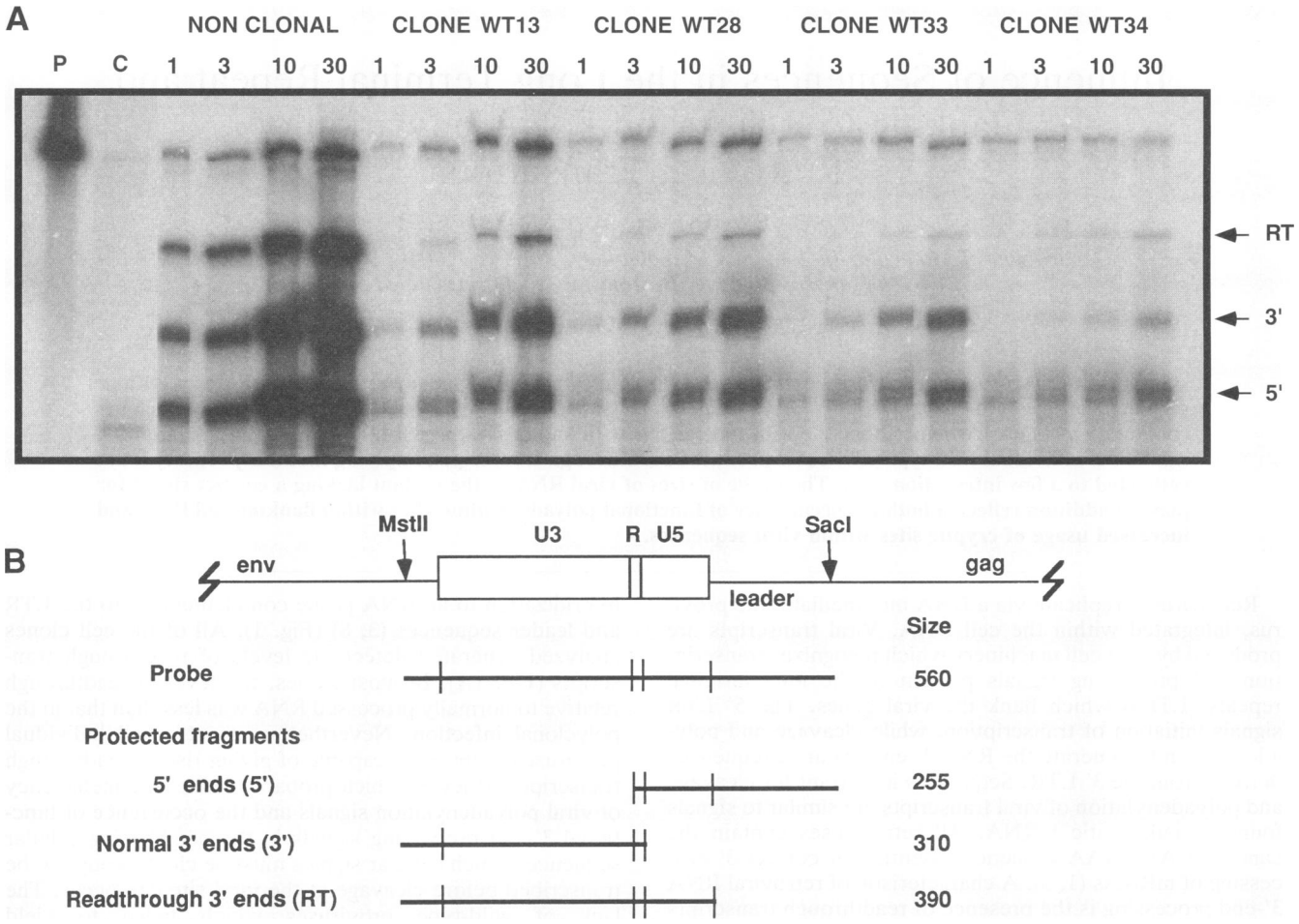

FIG. 1. Readthrough transcripts in clones containing a single wild-type provirus. (A) RNAs isolated from four clones of QT6 cells containing one wild-type (WT) provirus and from turkey embryo fibroblasts infected with wild-type RAV-1 (nonclonal) were subjected to nuclease mapping as described previously (8). Hybridizations were with 0.1 pmol of labeled RNA probe, 0.3, 1, 3, or 9 μg of total cellular RNA from the cell clones, and 0.1, 0.3, 1, or 3 μg of total cellular RNA from the infected turkey embryo fibroblasts (lanes 1, 3, 10, and 30, respectively, corresponding to relative concentrations). The sample in lane C contained no cellular RNA and was treated like the other samples. Lane P contained a 1-to-100 dilution of the untreated probe. The protected fragments derived from different areas of viral transcripts are indicated: RT, readthrough; 3′, normal 3′ ends; 5′, 5′ ends of viral RNA. For presentation of the data, autoradiographs were scanned by using a Hewlett-Packard DeskScan Plus flatbed scanner and a Macintosh computer. The resulting 72-dots-per-inch, 8-bit gray-scale files were adjusted for size, brightness, and contrast levels by using Enhance (version 2.0) and merged with the line drawings by using Canvas (version 3.01). (B) A representation of the structures and sizes (in base pairs) of the antisense RNA probe used and the fragments protected from RNase digestion by viral RNA. The probe was derived from a single-LTR-containing clone of RAV-1 permuted at the *Sac*I site (7, 8). The LTR with its different components is represented by an open box.

found among the clones. Clone 94 showed the use of two polyadenylation sites within the same cell-related sequence. The two unrelated sequences and viral RNA endings in clone 84 were probably due to a mixture of different clones in the sample analyzed. These results imply that the pattern of sizes of the readthrough transcripts produced during infection of the polyadenylation mutant virus (Fig. 2A) reflects the incidence of integration near a potentially functional polyadenylation signal. Variants of the AAUAAA polyadenylation signal of this type have been shown to be relatively efficient in 3′-end processing of different genes (1). It seems that in more that 80% of viral integrations (Fig. 4A) the neighboring cell sequences contain similar functional polyadenylation signals.

Two of the five clones containing a mutant provirus, M82 and M96, generated RNAs shorter than genome size (Fig. 2A). To map the 3′ ends of these smaller transcripts, cellular RNA was hybridized to a labeled RNA probe complementary to the LTR, leader sequences, and the 3′ untranslated region of the same virus (Fig. 4B). Both clones yielded a fragment corresponding to readthrough transcripts although at lower levels relative to clone M94. RNAs from these clones also protected five fragments of different sizes that were present in the RNA from the nonclonal infection with the mutant virus but not present in the RNA from clone M94. Such fragments were not seen in analyses of RNA from cultures or clones of cells infected with wild-type virus (Fig. 1 and data not shown). The presence of the identical set of fragments in samples from both clones as well as following a nonclonal infection (Fig. 4A, lane NC) implies that they

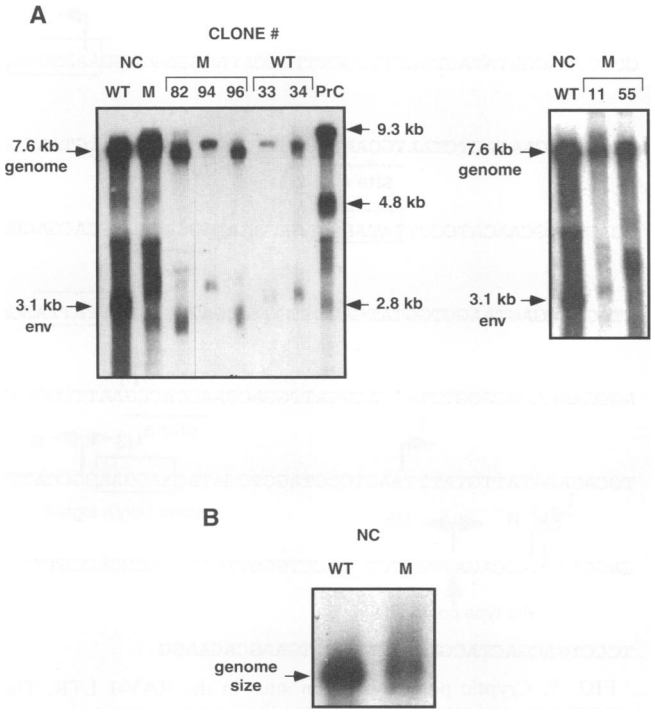

FIG. 2. Size of viral RNA in clones with a single provirus containing the hexanucleotide mutation. (A) About 10 μg of total RNA from clones containing the wild-type (WT) or mutant (M) provirus and about 2 μg of total cellular RNA from the nonclonal (NC) infection of turkey embryo fibroblasts with either wild-type or mutant virus were analyzed by agarose gel electrophoresis and hybridized to the RNA probe shown in Fig. 1 but labeled at a higher specific activity. RNA from turkey embryo fibroblasts infected with Rous sarcoma virus (Prague C) was used as a size standard. (B) Threefold enlargement of the region of the blot around the genome size transcript corresponding to lanes marked NC WT and NC M in panel A. RNAs from clones containing wild-type proviruses WT33 and WT34 were included as size standards.

reflect the use of cryptic polyadenylation sites in the LTR and are not due to specific, rare integration sites or mutant proviruses.

The 3′ ends of the smaller transcripts produced by the mutant virus were mapped within and very close to the U3 region. Other studies done with RNA probes containing different amounts of RAV-1 sequence confirmed the presence of three of the 3′-end sites observed (data not shown). These mapped to about 75, 160, and 250 bp (sites a to c, respectively) upstream of the wild-type cleavage site at the end of R. The regions upstream of these putative cleavage sites in the virus were searched for sequences that may serve as polyadenylation signals (Fig. 5). Several variants of the hexanucleotide were found, including UUAUUA, about 40 bp upstream of site a; UAUAAG, about 15 bp away from site

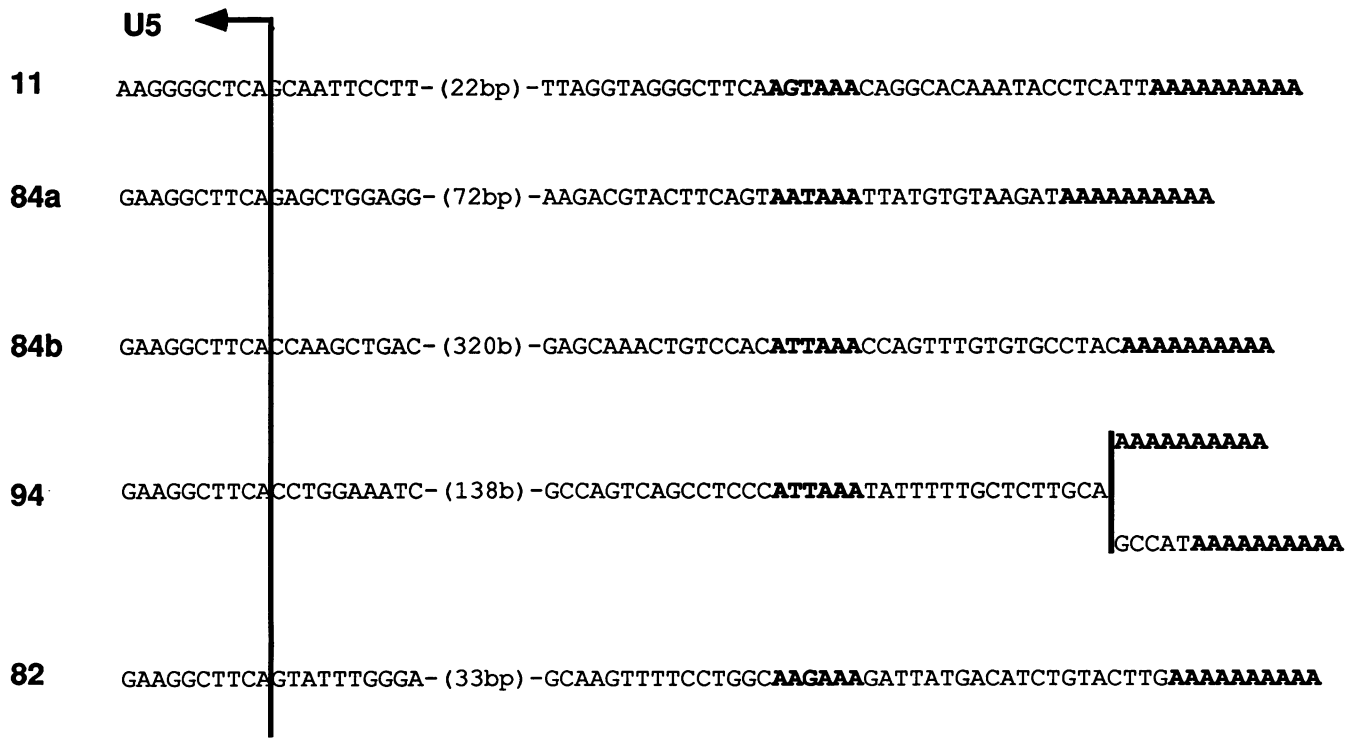

FIG. 3. Sequences of cDNAs derived from cell clones containing mutant proviruses. cDNA corresponding to the region between U3 and the poly(A) sequence was synthesized from virion RNA isolated from 10 ml of infected-cell supernatant from the individual clones containing mutant proviruses by using an oligo(dT) primer with a 5′ adapter of the sequence GACTCGAGTCGACATCGAT$_{17}$, amplified with the virus-specific primer GTCGAGCTCTTGCAACATGCTTAT (bases 33 to 50 of U3 [Fig. 6]) and the adapter primer (Fig. 3 and data not shown), cloned into M13MP18, and sequenced (6). The terminal portions of each sequence are aligned to the position of the LTR and the putative polyadenylation signal sequence (boldfaced).

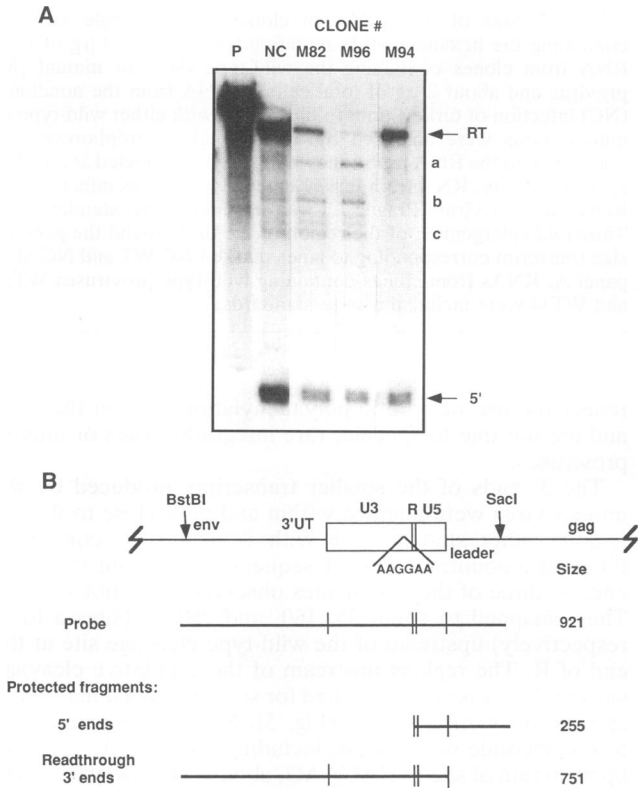

FIG. 4. 3′ ends of shorter than genome-size viral transcripts encoded by mutant proviruses. (A) Hybridizations were done as described in the legend to Fig. 1 with 3 μg of total cellular RNA from the individual clones or from nonclonal (NC) infection of turkey embryo fibroblasts with mutant virus. The protected fragments are as indicated in Fig. 1. P, probe; RT, readthrough. (B) Representation of the structures and sizes (in base pairs) of the antisense RNA probe used and the fragments protected from RNase digestion by viral RNA. The probe was derived from the polyadenylation mutant clone of RAV-1. The specifications are as described in the legend to Fig. 1. 3′UT, 3′ untranslated region of the virus.

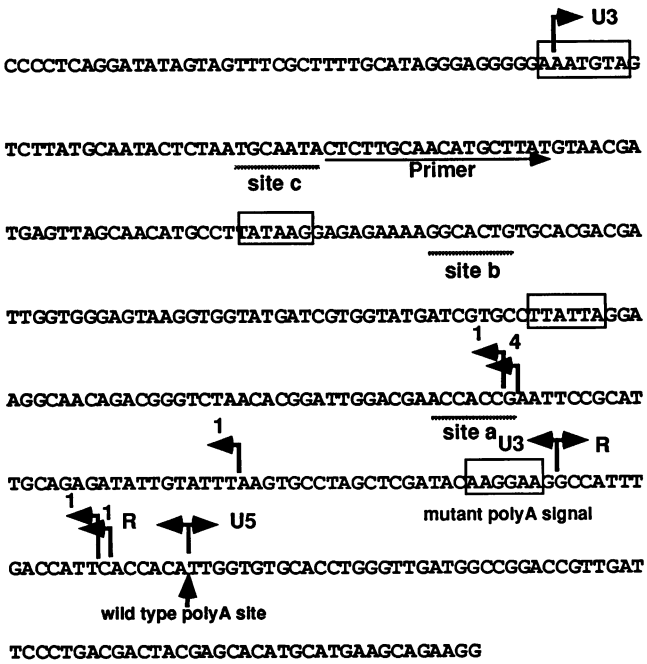

FIG. 5. Cryptic polyadenylation sites in the RAV-1 LTR. The approximate end sites of the shorter than genome size transcripts are underlined and are labeled a, b, and c as in Fig. 4. The boxed sequences contain the putative polyadenylation signals upstream of these sites. The positions of the polyadenylation sites determined by polymerase chain reaction amplification of cDNAs derived from clones 82 and 96 are marked with left-pointing arrows along with the number of times each was found. The U3 primer used for amplification is underlined with an arrow. The wild-type polyadenylation signal and site are marked, as are the boundaries of the U3, R, and U5 regions of the LTR.

b; and AAUGUA, 20 bp upstream of site c. The variant related to site b has been found at the 3′ ends of some eukaryotic genes (1), and AAUGAA is known to be a weak signal (1), suggesting that the variant related to site c (AAUGUA) might have a low polyadenylation efficiency. Amplification and sequencing of cDNAs derived from clones M82 and M96 showed polyadenylation of viral transcripts at a variety of sites within the LTR (Fig. 5), some of which coincided with the major sites derived from mapping analysis. We conclude that integration of the mutant provirus at some sites leads to use of cryptic polyadenylation signals within LTR sequences.

What causes the use of these cryptic sites? Since the best available matches to consensus poly(A) signals in the vicinity of the processing sites are quite poor matches to the canonical AAUAAA, the use of these sites must reflect the absence of stronger signals nearby. This conclusion is supported by the absence of such transcripts in clones or cultures infected with wild-type virus as well as in clones containing the mutant site and an efficient sequence (ANUAAA) near the integration site. Furthermore, one clone (clone 82), which yielded significant levels of both readthrough and truncated transcripts, contained the sequence AAGAAA, known to be quite inefficient (1), near the readthrough processing site.

The presence of truncated transcripts in clones M82 and M96 may imply the existence of some sequence or structure in RNA transcripts containing the 3′ LTR sequence that forces 3′ processing in the vicinity of the correct end, even at very poor matches to the consensus. This putative signal must extend its influence over a range of at least a few hundred bases. Thus, somewhat contrary to our expectation, most aberrant transcripts do not vary greatly in length from normal ones. At first, this effect might seem to preclude the use of such transcripts as intermediates in transduction (9). However, we have found that when a mutant provirus is within a normal transcription unit, no such transcripts are seen. Instead, all RNAs are processed at the correct site about 3 kb downstream of the LTR (9). Thus, the presence of an intact processing unit can override whatever signals remain in the mutant LTR, even at some distance. Alternatively, the LTR-terminated transcripts may be due to signals in the flanking sequence, such as transcriptional termination sites, which could prevent the use of downstream polyadenylation sites.

Our studies have shown that retroviruses have adapted and exploited the properties of 3′-end processing in eukaryotic cells. The properties of viral polyadenylation signals mimic those in host mRNAs, ensuring a relatively efficient production of viral RNA of the correct size. The inefficiency (about 15% for avian retroviruses) of these signals does not

seem to affect viral replication, because readthrough transcripts can be used as genomes (8). One of the consequences of this inefficiency is the ability of retroviruses to activate and transduce cellular sequences. These studies have shown that most proviruses can produce hybrid transcripts of a specific size for every integration event and suggest that a wide variety of cellular sequences can be incorporated into the retroviral genome by the same transduction mechanism.

We thank Stephen Dobson and Swee-Kee Wong for valuable assistance and Claire Moore and Naomi Rosenberg for advice.

This work was supported by grant R35 CA44385 to J.M.C. from the National Cancer Institute.

REFERENCES

1. **Birnsteil, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41**:349–359.
2. **Frohman, M.** 1990. RACE: rapid amplification of cDNA ends, p. 28–38. *In* M. Innis, D. Gelfand, J. Sninsky, and T. White (ed.), PCR protocols. A guide to methods and applications. Academic Press, Inc., New York.
3. **Herman, S. A., and J. M. Coffin.** 1986. Differential transcription from the long terminal repeats of integrated avian leukosis virus DNA. J. Virol. **60**:497–505.
4. **Moscovici, C., M. G. Moscovici, H. Jimenez, M. M. C. Lai, M. J. Hayman, and P. K. Vogt.** 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. Cell **11**:95–103.
5. **Proudfoot, N. J.** 1991. Poly(A) signals. Cell **64**:671–674.
6. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. **Sealy, L., M. L. Privalsky, G. Moscovici, C. Moscovici, and J. M. Bishop.** 1983. Site-specific mutagenesis of avian erythroblastosis virus: erb-B is required for oncogenicity. Virology **130**:155–178.
8. **Swain, A., and J. M. Coffin.** 1989. Polyadenylation at correct sites in genome RNA is not required for retrovirus replication or genome encapsidation. J. Virol. **63**:3301–3306.
9. **Swain, A., and J. M. Coffin.** 1992. Mechanism of transduction by retroviruses. Science **255**:841–845.