MICROBIOLOGICAL REVIEWS, Sept. 1993, p. 511–521
0146-0749/93/030511-11$02.00/0

Vol. 57, No. 3

Transcription Termination and Polyadenylation in Retroviruses

RAMAREDDY V. GUNTAKA

Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri-Columbia, Columbia, Missouri 65212

RETROVIRUSES

Introduction

Retroviruses have been found in almost all metazoans that have been examined so far (86, 87). Retrovirus-related proviruses (endogenous retroviruses) have also been detected in all mammalian genomes that have been investigated. It has been estimated that in some species of mice as many as 1,000 to 2,000 copies of endogenous proviruses or provirus-like elements per haploid genome exist, most of them with the characteristic structure long terminal repeat (LTR) genes-LTR, reminiscent of infectious retroviruses. Some of these elements, which are often referred to as retrotransposons, transpose or rearrange via reverse transcription of mRNA. There is substantial evidence that these retrotransposons are a continuous source of gene transposition and insertional mutagenesis and are a reservoir of LTRs, which are potential activators of transcription (45). Retroviruses act as transducing agents, behave as insertional mutagens, induce a variety of cancers, and have been incriminated as causative agents of many diseases such as immunodeficiencies, arthritis and other autoimmune diseases, and anemias (86, 87).

Retroviruses have several unique features: (i) they contain a diploid RNA genome and the enzyme reverse transcriptase (RNA-dependent DNA polymerase); (ii) their life cycle involves conversion of the genomic RNA into a double-stranded DNA intermediate by the virion-associated reverse transcriptase, a feature that distinguishes retroviruses from all other viruses; and (iii) mandatory integration of the viral linear DNA into host chromosomal DNA results in the formation of proviruses, a step carried out by the viral integrase (93).

Retroviruses, like avian leukosis-sarcoma viruses and murine leukemia viruses, are either simple, with only three genes, *gag*, *pol*, and *env*, required for replication or complex, such as human immunodeficiency viruses and human T-lymphotropic viruses, whose genomes encode several regulatory genes in addition to the three structural genes (16, 20, 21). Despite the intricate network of gene regulation in the latter group of viruses, all retroviruses share some fundamental properties which include reverse transcription, generation of LTRs during reverse transcription, integration, gene expression from the provirus by host RNA polymerase II, splicing and polyadenylation, translation, and maturation and release of virus by budding.

Life Cycle

As with other viruses, the life cycle of retroviruses begins with the entry of extracellular virus particles. Following adsorption and penetration into the cytoplasm, reverse transcription begins within the partially uncoated virus particles by the associated reverse transcriptase. These steps are schematically illustrated in Fig. 1. By using both subunits of RNA and the tRNA primer, the reverse transcriptase synthesizes a linear double-stranded DNA within a nucleoprotein complex in the cytoplasm (16, 61, 93). During this process of reverse transcription, LTRs flanking the structural genes are generated (37, 74). Each LTR contains unique sequences derived from both ends of the genome (U3 and U5) sandwiching a repeat element, R, in a structure U3-R-U5 (Fig. 1). Following migration of the complex to the nucleus, the linear DNA is covalently integrated into host chromosomal DNA by the virus-encoded integrase protein (12, 30). The integrated provirus is colinear with the unintegrated linear DNA. During integration, two nucleotides from the outer end of the U3 of the upstream 5′ LTR and two nucleotides from the outer end of the U5 of the downstream 3′ LTR are lost and a 4- to 6-bp sequence of the host DNA is duplicated at the site of integration. These features put retroviruses into the category of transposable elements.

Following integration, host DNA-dependent RNA polymerase II transcribes viral DNA into RNA (55), which is

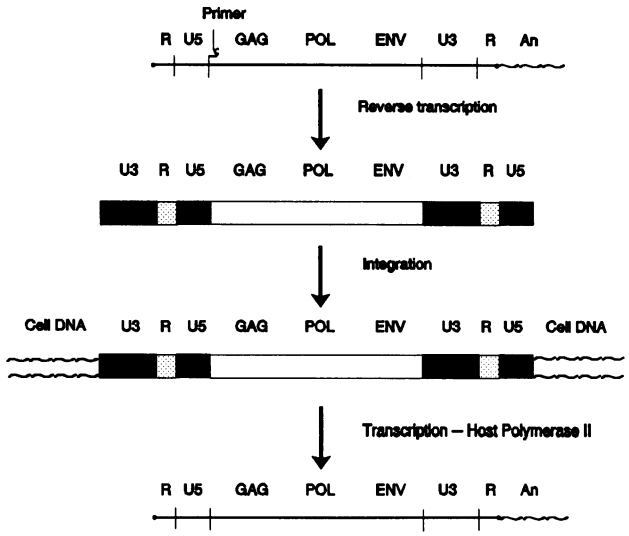

FIG. 1. Replication of retroviruses. Virion RNA is converted to double-stranded DNA by the virion-associated reverse transcriptase, using tRNA primer to initiate DNA synthesis. In this process U5 and U3 sequences, present only once in the RNA, are copied twice in such a way that U3 goes to the 5′ end and U5 goes to the 3′ end. The structure with U3-R-U5 is referred to as the LTR. Both linear and circular intermediates are found in the infected cell. The linear DNA integrates into the host chromosome in a reaction carried out by the viral integrase. Following integration, host RNA polymerase II transcribes viral DNA into RNA. The details of these various steps can be found in several excellent reviews (16, 61, 93).

then processed by splicing into various mRNAs. The *gag* and *pol* genes are translated from the primary transcript as a fusion protein, which is subsequently cleaved to the precursor *gag* and *pol* proteins (16). In all the retroviruses *gag* and *pol* are separated by an overlapping sequence containing a termination codon. In some retroviruses (e.g., murine leukemia virus) the amber codon between *gag* and *pol* is read as *glu* by a rare $tRNA^{Glu}$ (41). In most retroviruses *gag* and *pol* are translated in different reading frames (−1), and in these cases the termination codon is suppressed by ribosomal frameshifting (41). Regulation of suppression ensures synthesis of only smaller amounts of *pol* protein relative to the *gag* precursor. The other protein (*env* glycoprotein) is synthesized from a spliced mRNA (16). In complex retroviruses, such as human immunodeficiency virus (HIV) and human T-cell leukemia virus (HTLV), many other regulatory proteins such as *tat*, *tax*, *rev*, and *rex* proteins are synthesized from different overlapping exons (21, 32).

TRANSCRIPTION TERMINATION AND POLYADENYLATION IN EUKARYOTES

Without exception, the genes of all living species have to be transcribed into RNA. Transcription starts at a specific site and terminates at a region downstream of the gene. Initiation and termination of transcription define the boundaries of the functional unit, the gene. In general, transcriptional units are much longer than the coding requirements. Most primary transcripts are processed by splicing and 3′-end maturation to yield mRNAs. All eukaryotic mRNAs, with the exception of histone mRNAs, have a poly(A) sequence added posttranscriptionally to the 3′ end in a stepwise fashion by a multimeric protein complex (9, 51, 56, 102). Transcription is regulated differentially by a variety of environmental signals as well as by several internal factors interacting in an intricate manner.

An initiated transcript has to be terminated, and both initiation and termination are remarkably regulated. Termination is coupled to 3′-end processing, which includes both cleavage and polyadenylation. These reactions appear to occur almost simultaneously and are tightly regulated by both *cis* elements in the transcript and *trans*-acting protein factors (102). Although the function of poly(A) at the 3′ end of mRNAs is not fully understood, it has been implicated in a variety of important properties of mRNA, such as stability, translatability, and translocation from the nucleus to the cytoplasm (9, 42, 51).

In the genesis of mature mRNAs, the primary transcript (or pre-mRNA) is endonucleolytically cleaved and then a poly(A) tail of 200 to 300 nucleotides (nt) is added by the poly(A) polymerase (51, 102). At least two *cis*-acting recognition signals, a highly conserved consensus AAUAAA sequence approximately 10 to 30 nt upstream of the polyadenylation site and a weakly conserved G+U- or U-rich sequence 1 to 30 nt downstream of the poly(A) site, are required for efficient cleavage and polyadenylation (38, 51, 63, 102). The sequence of AAUAAA or its variant AGUAAA is crucial, whereas the requirement for the GU sequence is somewhat less stringent. At least six to eight factors and a poly(A) polymerase appear to be involved in the processing of 3′ ends of mRNAs. A model describing the interaction of various factors in the 3′-end processing of pre-mRNA has recently been presented (82, 102), and the major steps are shown in Fig. 2. These include the following. (i) RNA polymerase II, aided by several transcription factors including site selection factors βγ (RAP30/74 or TFIIF) and TATA-binding protein, binds to the core promoter element (P in Fig. 2) and initiates transcription of the gene X; the rate of transcription is regulated by several TBP-associated coactivator factors and transcription enhancer factors which interact specifically with the upstream *cis*-regulatory (E in Fig. 2) elements (reviewed in references 17, 31, 72, and 105). (ii) The transcription apparatus continues well past the poly(A) site and dissociates, producing a pre-mRNA (49, 59). (iii) Specific factors such as the 500-kDa cleavage-polyadenylation specificity factor (CPSF) or PF2 and cleavage stimulation factor (CstF) together form a stable complex by interacting with the AAUAAA sequence of the pre-mRNA (14, 28, 29, 44, 81a, 82). (iv) The pre-mRNA is cleaved about 15 to 30 nt downstream of the AAUAAA by an unknown activity of the complex, which is stimulated by the multicomponent CstF (82) producing 3′-OH ends. (v) Finally, polyadenylation is carried out by the poly(A) polymerase in a biphasic manner (102); the details of this step are not yet understood, but with the recent cloning of the gene encoding the poly(A) polymerase (67, 97), significant insights into the mechanism of addition of poly(A) tails to mRNAs should be forthcoming.

The role of the G+T- or T-rich downstream sequence in transcription termination and polyadenylation is somewhat uncertain. About 60 to 70% of mammalian genes contain a consensus sequence YGTGTTYY (Y = T or C) approximately 5 to 30 bp downstream of the poly(A) site. The rest of the genes contain either a stretch of six or seven T residues or a G+T-rich sequence (54). Mutational analysis revealed the requirement of this sequence in the termination and polyadenylation of some but not other genes (51). In most of the retroviruses, such a consensus sequence is absent (Table 1). The importance of this G+T-rich sequence in retroviral

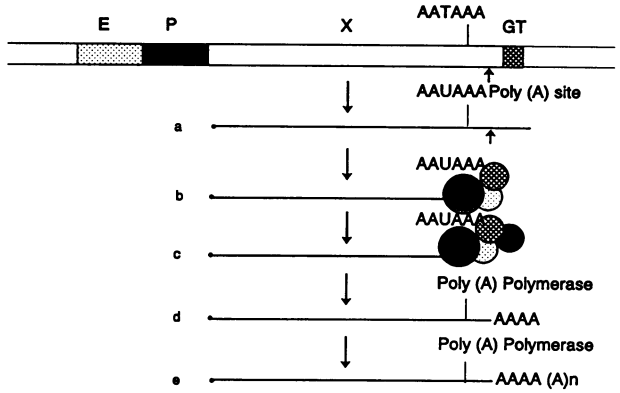

FIG. 2. Model to illustrate transcription of gene X, termination, and 3′ formation of processed mRNAs. Gene X is transcribed by RNA polymerase II by using the promoter (P) and enhancer (E) signals. Initiation takes place about 20 to 30 bp downstream of the TATA box (located in P) and elongates until the transcription complex dissociates at the 3′ end of the G+T-rich sequence (GT). At least five factors are involved in the catalysis of pre-mRNA cleavage (14, 29, 82, 83). These include a CPSF (58; other terms such as cleavage polyadenylation factor [14, 44] or polyadenylation factor 2 [28] were used before the now standard nomenclature [96]) (●) and CstF (82, 83) (or CF1 [29]), and the poly(A) polymerase (3, 83) (●). The CstF appears to directly interact with pre-mRNAs containing AAUAAA via a component of CstF, the 64-kDa protein (58, 82, 104) (●). Other polypeptides of CstF are also indicated (◎). CPSF (28) interacts with the pre-mRNA on the AAUAAA sequence (6) (step b). CstF/CF1 (29) interacts with the pre-mRNA dependent on both the downstream G+U-rich element and the AAUAAA, and a complex between CstF and CPSF is required for the AAUAAA-dependent cross-linking of the 64-kDa polypeptide (29, 104). CPSF complexed with CstF cleaves the pre-mRNA downstream of AAUAAA (29, 105) (step c), and then poly(A) polymerase adds a poly(A) tail in two stages, an AAUAAA-dependent addition of about 10 adenine residues (step d) and an AAUAAA-independent but factor-dependent addition of a poly(A) tail of 200 to 300 nt (75, 102) (step e).

transcription termination/polyadenylation is discussed below in the section on G+T-rich sequences in U5.

LTRs

All retroviruses have evolved a unique mode of replication to ensure genesis of LTRs (U3-R-U5), which contain the controlling elements of transcription. Some acutely transforming retroviruses are defective for replication because they suffered deletions in the structural genes; they nevertheless retain the LTRs derived from their cognate helper viruses, indicating the importance of these repeats. LTRs have a central role in viral replication and integration and in gene expression (24, 50, 93). They are also involved in determining cell and tissue tropism (24, 88). They vary in size from about 300 to 1,200 nt. The 5′ and 3′ LTRs are identical in nucleotide sequence but functionally different. The basis for this functional discrimination is not understood, but recently described experiments may shed some light on this paradox (see the section on differences between 5′ and 3′ LTRs, below).

Some of the important features of a few representative LTRs that are pertinent to this review are schematically depicted in Fig. 3. The U3 region of almost all the retroviruses is relatively longer than the R or U5 region. Except in one group of retroviruses, which include HTLV-I, HTLV-II, and bovine leukemia virus (BLV), the core promoter (TATA and CCAAT boxes) is included within the first 100 bp upstream from the transcription start site, the TATA sequence determines the transcription initiation site, and the CCAAT sequence(s) contributes to the efficiency of transcription initiations. We do not know which sequences substitute for the CCAAT function in retrovirus LTRs that do not contain this core sequence. Other *cis*-acting regulatory sequences, commonly termed enhancers, reside further upstream of the start site and may extend almost to the left boundary of the U3 sequence. In the HTLV group, the lentivirus group (HIV belongs to this group), and the mouse mammary tumor virus (MMTV) group, the U3 region also contains coding sequences (16, 50).

AAUAAA Sequence

The presence of the AAUAAA sequence among animal cell mRNAs was first noticed by Proudfoot and Brownlee (64). Since then it has been shown that this poly(A) signal is present in more than 85 to 88% of all vertebrate mRNAs. In about 12 to 15% mRNAs a variant sequence, AUUAAA, is found (102). Either AAUAAA or AUUAAA is absolutely required for 3′-end maturation of RNA. The presence of U at the third position is essential for endonucleolytic cleavage of pre-mRNA, since substitutions at this position virtually eliminate cleavage (51, 103). Substitution of the A at the second position by G renders the molecules only one-third as efficient for 3′-end processing as AAUAAA is.

In all but one of the retroviruses, AAUAAA is invariant and is located in either the U3 or the R region (Fig. 3). In MMTV, AGUAAA is present instead of AAUAAA (50). In most retroviruses the distance between the poly(A) signal and the poly(A) site is remarkably constant, about 16 to 25 nt (Table 1). However, in the human and bovine lymphotropic virus subfamily (HTLV-I, HTLV-II, and BLV), the distance is quite large, about 260 to 270 nt. The significance of this configuration will be considered later in this review.

In almost all avian leukosis-sarcoma virus, HTLV-BLV, and MMTV groups of retroviruses, the AAUAAA signal is located in the U3 region, only a few nucleotides upstream of the R sequence. Unlike the cellular genes, in which AAUAAA occurs only once at the 3′ end of the mRNA, in retroviruses the AAUAAA sequence occurs twice, once in each LTR. For this group of viruses, the presence of a poly(A) signal in the U3 region, which is outside the transcriptional start site, restricts transcription termination and polyadenylation to the 3′ LTR.

In most of the mammalian retroviruses, in the avian spleen necrosis virus (which resembles more closely the mammalian than the retroviruses), and in all the lentiviruses, AAUAAA is located in the R region (Fig. 3). Transcription starts at the junction between U3 and R (the cap nucleotide is the first nucleotide of R), polyadenylation takes place at the junction of R and U5 (3′ boundary of R), RNA initiation occurs in the 5′ LTR, and termination and polyadenylation are likely to take place in the 3′ LTR. This arrangement raises a question about the mechanism of termination and polyadenylation. Since both LTRs are identical, we expect that the transcript initiated at +1 nucleotide in the 5′ LTR would be polyadenylated at the end of the R sequence in the 5′ LTR itself. This clearly does not happen. For instance, in murine leukemia virus, spleen necrosis virus, and lentiviruses such as HIV, only short polyadenylated transcripts of 65 to 100 nt would be synthesized; these have not been detected in infected cells or in cell-free systems. These

TABLE 1. U5 sequence[a]

Virus[b]	Sequence	% GT in nt:			Distance from AAUAAA (nt) to poly(A) site	% G+T
		1–40	1–20	21–40		
	1 40					
RSV PrC	TTGGTGTGCACCTGGGTTGATGGCCGGACCTGCGATTCCC	62.5	75	50	22	33
SNV	GATTGGCAGTGAGAGGAGATTTTGTTCGTGGTGTTGGCTG	77.5	65	90	21	52
MoMULV	TCCGACTTGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTC	65.0	65	65	16	50
AKV	TCCGAATCGTGGTCTCGCTGATCCTTGGGAGGGTCTCCTC	60.0	60	60	16	69
MMTV	TTCACCTCTTGTGTGTTTGTGTCTGTTCGCCATCCCGTCT	65.0	75	55	18	33
GA. FeSV	TCTGACTCGTGGTCTCGGTGTTCCGTGGGTACGGGGTCTC	70.0	70	70	16	63
SSV	TCCGAAGCCGTGGTCTCGTTGTTCCTTGGGAGGGTCTCTC	65.0	60	70	17	53
HTLV-I	CGTCTTTGTTTCGTTTTCTGTTCTGCGCCGTTACAGATCG	67.5	80	55	269	
HTLV-II	CTTCCGCGTTCTTGTCTCGTTCTTTCCTCTTCGCCGTCAC	57.5	65	50	269	
BLV	GTTTGTTTCCTGTCTTACTTTCTGTTTCTCGCGGCCCGCG	67.5	75	60	263	
HIV-1	AGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAG	67.5	70	65	19	63
HIV-2	AGTTAAGTGTGTGCTCCCATCTCTCCTAGTCGCCGCCTGG	55.0	60	50	17	47
Visna virus	GCTTGCTGGTTATTATCGGGATTGGTTACTAATTCCGTGC	67.5	75	60	19	68
HSRV	ACTCATATATTATTGTCTCTTTTATACTTTATTAAGTAAA	55.0	55	55	17	40
FIV	TGTCGAGTATCTGTGTAATCTTTTTTACCTGTGAGGTCTC	67.5	65	70	17	53
SIV_{AGM}	AGCCAGTGTGTCTTCCCATCTCTCCTAGTCGCCGCCTGGT	52.5	50	55	17	47
Consensus "GT Box"	YGTGTTYY (Y = T or C)					

[a] The G+T composition of the first 40 bp in the U5 region of various retroviruses is compared with that of the R region of the LTR between the poly(A) signal and the poly(A) site. The most noteworthy feature of this comparison is the absence of a consensus sequence YGTGTTYY, which was deduced from a collection of more than 100 cellular genes (54). In only HIV and simian immunodeficiency virus (SIV_{AGM}) is there a sequence that closely matches with the consensus sequence. In most of the retroviruses, the first 20 bp of the U5 region is relatively enriched for the G and T residues. Except for a few retroviruses, the R region has a relatively lower G+T content. Since there is no consensus sequence in retroviruses and as few as 1 nt in a *cis*-element can relatively affect the interaction of proteins with their cognate sequence, I suggest that formation of a secondary structure or distortion of the DNA in nucleosome in the 3′ LTR is an important signal for the dissociation of the complex and/or cleavage-polyadenylation in these viruses. In a few retroviruses reported, the importance of the GT sequence in the U5 has been directly demonstrated by constructing deletion mutants which drastically reduced 3′-end formation (see the section on G+T-rich sequences in U5). Most of the sequences were taken from reference 101, except those for HIV-2 (47), visna virus (78), feline immunodeficiency virus (FIV) (84), and HSRV (52).

[b] Abbreviations: RSV, Rous sarcoma virus; SNV, spleen necrosis virus; MoMuLV, Moloney murine leukemia virus; AKV, AKR murine leukemia virus; GA. FeSV, feline sarcoma virus; SSV, simian sarcoma virus; HSRV, human spuma retrovirus.

results suggest occlusion of the poly(A) signals in the 5′ LTR. In HIV, a nonpolyadenylated RNA of 60 nt was found (20). The accumulation of this short transcript diminished greatly following expression of the Tat protein, a virus-encoded regulatory protein which acts as an antiterminator by binding to the short *tat*-binding RNA, TAR (20).

If termination of transcription does not occur in the 5′ LTR but only in the 3′ LTR, what is the mechanism of this differential utilization of poly(A) signals? To address this question, several investigators have devised ingenious experiments (13, 23, 39, 99). The conclusion from these experiments was that the distance from the cap site to the poly(A) signal is critical and increasing this spacing allows efficient 3′-end processing. One study (13) revealed an optimal distance of about 300 nt between the cap site and poly(A) signal, whereas the experiments of Iwasaki and Temin (39) indicated an optimal distance of about 1 to 1.5 kb; if the distance was shorter than 500 nt, only 3 to 9% of the RNA was polyadenylated. Iwasaki and Temin (39) further showed that this spacing effect of 3′-end processing occurs only with spleen necrosis virus and probably with related retroviruses but not with thymidine kinase or simian virus 40 late genes. They have further shown that multiple elements located within the U3 and U5 regions are required for efficient processing of 3′ ends (40). Spacing between the cap site and poly(A) site also appears to be critical with certain pararetroviruses such as hepatitis B virus and cauliflower mosaic virus (70, 71).

It has been shown that in some retroviruses including HIV-1, an element in the U3 sequence, upstream of the AAUAAA, contributes to the 3′-end processing (10, 23, 90,

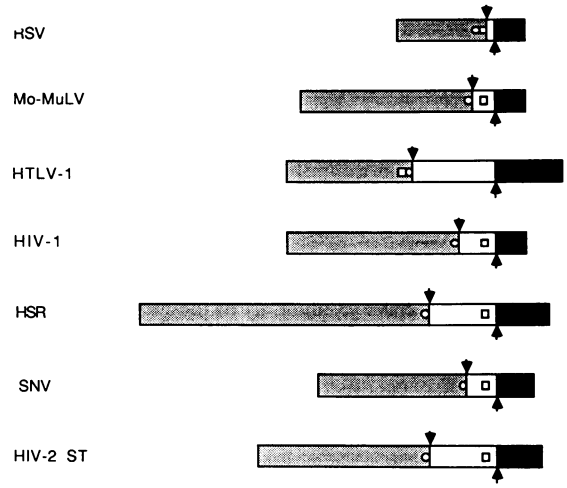

FIG. 3. Structural features of retroviral LTRs. Retroviral LTRs contain three different segments, U3 (▨), R (□), and U5 (■), and vary in size from about 300 to 1,200 bp. The U3 region contains *cis* elements to which cellular factors and/or virus-encoded factors bind and promote transcription. The TATA box (○), which determines the transcription start site (↓), is located about 20 to 30 bp upstream of the transcription start site. The polyadenylation signal, AATAAA (□), is located either in the U3 or R region. It should be pointed out that 5′ and 3′ LTRs (Fig. 1) are identical in sequence and arrangement. Transcription begins in the 5′ LTR (↓) and ends in the 3′ LTR (↑). The distance between the poly(A) signal and polyadenylation site is fairly constant (16 to 25 bp), with the exception of that in the HTLV subfamily of retroviruses. Another feature that is peculiar to HTLV-I is that the AATAAA sequence is 5′ to the TATA box. The significance of this arrangement is not known.

91). Valsamakis et al. (91) showed that mutations in the sequence from 77 to 94 nt upstream of the AAUAAA signal motif severely reduced the efficiency of polyadenylation. This sequence (77 to 94 nt upstream of AAUAAA) is different from the sequence determined by Brown et al. (10). The detailed deletion analysis with spleen necrosis virus has indicated that the U3 deletion of the U3 sequence also severely curtailed 3'-end formation (40). Other in vitro studies, however, have suggested that there are no upstream elements in the HIV polyadenylation signal (99). These conflicting results can be resolved only by further experiments. If the requirement for the U3 region is true for other lentiviruses, it would suggest that the presence of this sequence distinguishes the 3' LTR from the 5' LTR, since the 5' RNA transcript does not include this sequence.

It is also appropriate to mention another important aspect of the regulation of transcription of retroviruses. Although the mechanism is not known, transcription initiation in the 3' LTR is blocked when the 5' LTR is active (22, 81). However, when the upstream LTR is deleted, the 3' LTR can be activated as a promoter. This phenomenon, originally described with avian leukosis virus, has been observed frequently (reviewed in reference 43). It is also known that the U3 region of all retroviruses contains multiple *cis* elements to which various cellular as well as virus-encoded factors bind and promote transcription (24, 34, 50). It is logical, therefore, to assume that transcription factors bind to the core promoter TATA and CCAAT elements in the U3 region and form initiation complexes with RNA polymerase II, with this multicomponent transcription apparatus advancing down to the R and U5 sequences. In addition, transcription factors such as LBP (leader binding protein) bind to specific sequences in the R region, and this interaction, along with transcription complexes TFIID, RNA polymerase, and other components that destabilize and unwind the DNA between the TATA box and the R sequence, might preclude recognition of AAUAAA sequences by the CPSF-CstF complex.

In many elegant studies it has been shown that the 5' LTR is accessible to S1 and DNase I, implying active chromatin (18, 100) which is free of nucleosomes or deficient in histone H1, a protein shown to act as a repressor of transcription (48, 100). Since active 5' LTR renders the 3' LTR inactive, it is very likely that the configuration of the 3' LTR is in a "closed" state, i.e., present in a highly condensed structure. These structural differences may determine the fate of the transcription complex. Some of these points are discussed in more detail below (see the section on chromatin structure).

Except for one study, a thorough mutational analysis of the requirement of retrovirus AAUAAA sequence has not been done. In a rather more rigorous study, Herman and Coffin (35, 36) showed that (i) readthrough transcripts, which are transcripts initiated in the 5' LTR but continued into the flanking cellular sequences, constitute about 15% of the total RNA, and (ii) transcription initiation did not occur in the downstream LTR, which is in agreement with the original observation by Cullen et al. (22). By comparison, in spleen necrosis virus, more than 96% of transcripts that originated in the upstream LTR are correctly processed in the 3' LTR (39). We have also observed that in a situation as close to the natural provirus, i.e., LTR-gene-LTR, less than 1 to 2% of the total transcripts are readthrough RNAs (15). This quantitation was done by S1 mapping with a vector Neo^r gene driven to expression by the Rous sarcoma virus LTR (pATV-6D3A-Neo). This vector DNA was introduced into quail tumor cells, and G418-resistant transformants were selected. Southern analysis revealed that integration occurred in the plasmid pBR322 sequences and therefore the Neo^r gene sequence was flanked by LTRs. This arrangement was very similar to that of an integrated viral genome, except that in this construct only 23 bp of U5 was present instead of the 81 bp in the provirus. S1 mapping analysis with RNA isolated from stable transformants showed only one band, of 72 nt (15). This S1-protected fragment could come only from the correctly processed *neo* mRNA. Exactly the same band was also present in virus-infected chicken cellular RNA. If 15% of the RNA were due to unprocessed readthrough transcripts, we should have detected a band migrating at 95 nt in the provirus. Neither in the pATV-6D3A-Neo vector nor in the virus-infected cell clones (Pr.C cl.2, Pr.C cl.P) could this band be detected. Further experiments with many individual clones of the deletion mutant pATV-6D3A-319 and by polymerase chain reaction amplification (73) supported the above conclusions (14a).

Coffin's group (35, 36, 81) also demonstrated that polyadenylation at the correct site is not required for retrovirus replication, since mutation of AAUAAA to AAUGAA or AAGGAA increased the frequency of readthrough to about 70 and 100%, respectively. This is rather surprising, because it was thought that longer viral transcripts may not be encapsidated very efficiently. Nevertheless, these experiments conclusively establish (i) the requirement of AAUAAA for cleavage and polyadenylation at the correct site (if this is impaired, no processed transcripts are produced) and (ii) the absence of a rigorous size requirement for RNA packaging, although transcripts of less than 9 to 10 kb, which is the maximum size of all retrovirus genomes, are preferred.

However, closer examination of the results did reveal some preference in the size of the viral RNA that is encapsidated. For example, in the mutant (AAUAAA to AAGGAA) virus, production decreased to about 20 to 25% of the wild-type level, indicating an anomalous behavior of the mutant virus. Also, at least 60 to 70% of the RNA packaged into the virions was only 0.2 to 0.5 kb longer than in the wild type, suggesting preferential termination very close to the U5 sequence of the virus. It is also possible that the virus produced in the first round after transfection of DNA came from proviruses that are integrated at a site closer to the poly(A) signal. Comparison of the amounts of virus-specific genomic RNA inside the cell with those of packaged viral RNA indicated that although more than 60% of the mutant virus RNA is of a single size of about 7.8 to 8.0 kb (the size of the wild type is 7.6 kb), the mutant virus RNA certainly displayed a greater extent of heterogeneity than did the RNA in the virion. These results suggest a restriction in the size of the RNA packaged in virus particles. It is important to understand the nature of the readthrough cellular sequences to reveal where termination and polyadenylation occur. Since most of the readthrough RNA is only slightly larger than the genomic RNA, it is possible that the transcription complex dissociates at the end of U5 in response to the G+U-rich sequences, which have been shown to play a significant role in transcription termination (see below).

G+T-Rich Sequences in U5

In most animal cell mRNAs, a less well conserved G+T- or T-rich sequence is located downstream of the AATAAA signal sequence. Compilation of a vast number of cellular and viral termination and polyadenylation signal sequences revealed the presence of a G+T-rich sequence. Mutational analysis indicated that this sequence is also required, al-

though less stringently, for 3'-end processing (53, 54). Moreover, it appears that the distance between the AATAAA and the GT sequences is also critical (27).

In retroviruses, this G+T-rich sequence appears to be important. Examination of several retroviral LTRs indicates that the first 40 bp in the U5 region is rich in GT sequences (Table 1). This sequence is about 15 to 60 bp downstream of the poly(A) signal, placing it 1 to 40 bp from the poly(A) site. Apart from a few viruses (e.g., AKR murine leukemia virus, feline sarcoma virus, HIV, and visna virus [Table 1]) in which the interval sequence between the AATAAA and poly(A) site is also relatively high in G+T, this region is relatively low in G+T compared with the U5 sequence. In the U5 of many retroviruses, the first 20 bp (1–20, Table 1) contains more G and T residues than the next 20 bp does (21–40, Table 1).

Avian retroviruses. Our laboratory has shown that deletion of the sequence 1 to 23 in U5 (or +21 to 44 from the cap site) severely diminishes the amount of correctly processed viral or reporter gene transcripts (15). In these deletion mutants (pATV-6D3A-319), we kept the AATAAA sequence at −2 to 7 intact but deleted the G+T-rich sequence. Wild-type and deletion-bearing plasmid DNAs were introduced into QT6 cells, and RNAs from individual G418-resistant transformants were analyzed by S1 mapping. The results indicated that the U5 region of Rous sarcoma virus is important for transcription termination and/or polyadenylation. These experiments also suggested that the AAUAAA alone is not sufficient for 3' processing. Since similar results were obtained with RNA from virus-infected cells, it would argue against selection of a dominant marker (Neo^r) being responsible for this effect of the GT region (14a). The results with spleen necrosis virus also support these conclusions in that deletion of the GT sequence severely impaired 3'-end processing (40).

The results obtained by Coffin and his colleagues (36) demonstrated the requirement of AAUAAA for the correct 3' maturation of mRNA, although a significant proportion of transcripts (15%) failed to be processed. Our results, although not in total disagreement with those of Coffin and his colleagues, nevertheless put more weight on the role of the U5 sequence.

HIV. Working with HIV, Bohnlein et al. (8) showed that when an intact HIV LTR was placed downstream of a reporter insulin gene, more than 98% of the transcripts were correctly processed and polyadenylated at the authentic poly(A) site. These results were in excellent agreement with those of Iwasaki and Temin (39) as well as ours, in that no more than 1 to 2% readthrough transcripts could be detected as opposed to 15 to 20% reported by others (36, 81). Bohnlein et al. (8) also demonstrated that deletion of the first few nucleotides of the U5 region in the 3' LTR (nt 5 to 12) virtually eliminated 3'-end processing. In a follow-up study Brown et al. (11) showed that both the AAUAAA and G+U-rich sequences in the HIV RNA were required for efficient and correct formation of 3'-polyadenylated RNAs. Further, they demonstrated that random sequence insertions between the AAUAAA and the GU box inhibited poly(A) site function. However, they observed that if the inserted sequence could form stem-loop structures that maintain the correct spacing between the poly(A) signal and the GU box signal, poly(A) addition would proceed normally (12). From these results they concluded that the appropriate distance between the poly(A) signal and the poly(A) site determines where polyadenylation will occur (12). Similar studies have not been reported with any other member of the lentivirus subfamily.

HTLV subfamily. The HTLV group of retroviruses, which includes HTLV-I, HTLV-II, and BLV, is quite distinct from other retroviruses in that the distance between the AATAAA and the G+T-rich sequence in U5 is 265 nt, quite long compared with those of other retroviruses (Fig. 3; Table 1). These viruses also code for two important proteins, Tax and Rex. Tax is a 40-kDa *trans*-activator protein which potentiates transcription by binding to the 21-bp enhancer element located in the U3 region of the HTLV-I LTR (2, 25, 77, 85). The *Rex* gene product is a 27-kDa protein that binds to the *Rex*-responsive element and increases the export of unspliced genomic RNA and singly spliced *env* mRNA into the cytoplasm from which viral structural proteins are synthesized. The *Rex*-responsive element is predicted to be a highly stable RNA stem-loop structure of 255 nt (2, 89).

The distance between AATAAA and GT elements in this subfamily of retroviruses is about 255 to 263 bp. It might be suspected that this would pose a serious problem, since separation of these sequences by more than 30 to 40 nt would be expected to lead to a total loss of poly(A) site usage in 3'-end processing (11, 27, 53). Recent demonstration of the formation of a stable stem-loop structure between the AAUAAA and GU elements would seem to alleviate the paradox (2, 4). Both groups have convincingly demonstrated that the formation of RNA secondary structure juxtaposes the AAUAAA hexamer motif and GU element (2, 4).

To demonstrate the requirement of both AATAAA and GT motifs, Ahmed et al. (2) constructed vectors containing the HTLV-I *Rex*-responsive element with its poly(A) site as well as an insulin poly(A) site about 195 nt downstream of the HTLV-I poly(A) site. Transfection into COS cells and analysis of the transcripts by S1 protection assays revealed efficient and correct processing at the HTLV poly(A) site. However, alteration of the AAUAAA to AAUAUU or introduction of deletions which disrupt the formation of the stem-loop structure or deletion of the GU sequence eliminated 3'-end processing (2). This group further showed that deletions preventing the folding of the secondary structure and disrupting the spacing between the two motifs abolished authentic 3'-end processing, suggesting the importance of this arrangement. These results were confirmed by Bar-Shira et al. (4). It is important to note that in these experiments, base changes in the GT sequence in U5 did not drastically affect poly(A) addition. For instance, alteration of the sequence CGTCTTTGTTTCGTTTTCTG (1–20 in U5 [Table 1]) to either CGTCACTTCGTTTTCTG or TCTACGTCTT TGTGTCGACATCTG had only a slight effect on the efficiency of 3'-end processing but deletion of the sequence TTTCTG (15 to 20 bp) severely curtailed the formation of correctly processed mRNA (4). Some additional important points have emerged from these studies: (i) the first 20 nt of U5 is sufficient for pre-mRNA processing; (ii) the sequence per se may not be important, because base changes in this region did not affect 3'-end maturation; and (iii) the bases from 15 to 20 of the U5 region are essential for the addition of the poly(A) at the HTLV-I poly(A) site. It should be noted that the first 20 nt of HTLV-I is extremely G+T rich (80%; Table 1) and that the base changes introduced in this sequence did not affect the G+T content significantly (4).

It appears that some factors specifically recognize the G+U-rich element in the pre-mRNA. One factor, CF1, which is composed of three polypeptides of molecular sizes 76, 64, and 48 kDa, provides specificity for the GU sequence and stabilizes the complexes between the PF2 (CPSF) and

pre-mRNA (29, 96). CF1 appears to be identical to CstF and provides specificity for the downstream GU element via the 64-kDa protein component of the complex by cross-linking in an AAUAAA-dependent manner (29, 82). The results unambiguously establish the interaction of multiple protein factors with both AAUAAA and GU elements of pre-mRNA (14, 29, 66, 103, 104).

In this group of retroviruses, as in avian retroviruses, the AAUAAA motif is located in the U3 region of the LTR; this arrangement eliminates premature termination of the transcript in the 5′ LTR since U3 sequences are lacking in the transcript. Thus, each group of retroviruses has evolved a different strategy for differential utilization of 5′ and 3′ LTRs for transcription initiation and termination-polyadenylation.

Differences between 5′ and 3′ LTRs

One of the most important and strategic features of retrovirus replication is the unique mode of generation of LTRs in which U3 and U5 sequences are duplicated. This arrangement automatically suggests that these viruses do not depend on host cell regulatory *cis* elements for transcription initiation and termination. Such a dependence would be disadvantageous for retroviruses because they cannot depend on integration into a host site that ensures efficient expression, since any of several thousand potential chromosomal loci could be occupied. Had the viruses depended on the host termination polyadenylation sequences, they would have to find a locus immediately downstream of the provirus, otherwise transcripts of various sizes would be produced. Most of these would probably not be encapsidated because of the limited capacity of the virus particles. Although some preference for integration at DNase I-hypersensitive sites (actively transcribing regions) has been reported, this may not be a general event (69, 76, 95). Elegant experiments by Pryciak et al. (65) with in vitro minichromosomes as targets for retroviral integration have demonstrated some preference for integration into the exposed face of the nucleosomal DNA compared with naked DNA. These data, however, do not show whether the preferred sites of integration are indeed DNase I-hypersensitive sites, i.e., sites where the nucleosomal structure is probably accessible to transcription machinery. In retrovirus-infected cells, as much as 1% of the total poly(A) RNA is virus specific, suggesting that the proviruses efficiently use host RNA polymerase II for their own transcription, and the *cis* elements in the LTR must determine the specificity as well as the efficiency of transcription; therefore, this site of integration is critical for viral gene expression.

Several reports have already shown that in a normal infection only the 5′ LTR is active whereas the 3′ LTR is silent for transcription initiation (43, 81). The following questions are then pertinent. What are the structural differences between the 5′ and 3′ LTRs that render them functionally different? How do the two LTRs cross-talk? Although it is difficult to arrive at answers to these important questions, some provocative information has been emerging recently (94).

It has been known for some time that the 5′ LTR contains DNase I-hypersensitive sites, implying that it is in a transcriptionally active state (18, 100). It has been rigorously established that the structure of chromatin and supercoiling of DNA are important parameters governing expression of eukaryotic genes in a variety of systems (57, 100). Therefore, it is likely that the 5′ LTR is in a configuration that is accessible to transcription factors, whereas the 3′ LTR must be inaccessible to the same factors. Recent evidence suggests that there are indeed differences between the 5′ and 3′ LTRs, as revealed by probing with DNase I and restriction enzymes.

The detailed analysis of the 5′ and 3′ LTRs for structural changes under induced and constitutive levels of HIV provirus expression has been carried out by DNase I digestion of isolated nuclei followed by restriction enzyme digestion and Southern hybridization (94). These studies have yielded some interesting results. The U3 region of the 5′ LTR, which supports transcription initiation, contains three major DNase I-hypersensitive sites in constitutively expressing cell lines; these sites are confined to the promoter and enhancer sequences to which specific transcription factors bind. Two more sites, which extend into the R, U5, and leader sequences, were found in cells treated with tumor promoter (tetradecanoyl phorbol acetate) or tumor necrosis factor which induce HIV transcription. In contrast to the 5′ LTR, in the 3′ LTR only one major site appeared on DNase I digestion under both uninduced and induced states. Although some of these sites bind to the overlapping sequences in the 3′ LTR, the pattern of hypersensitive sites in Southern blots indicated that the proteins that interacted in this area in the 5′ LTR did not bind to the homologous region in the 3′ LTR. In other words, the chromatin structure of the 5′ LTR is different from that of the 3′ LTR, with the 5′ LTR DNA more accessible to the transcription and enhancer factors (94). These provocative experiments must be extended to other retroviral systems before a consensus theme can be proposed. Nevertheless, these results are exciting because they demonstrate the structural differences in the chromatin in the vicinity of 5′ and 3′ LTRs and show that the 3′ LTR is in a closed state which may prevent transcription apparatus from moving further down into adjacent cellular sequences. In other words, the transcripts probably terminate downstream of this site and become polyadenylated after correctly removing the additional nucleotides.

The situation with *myc*-activated chicken syncytial proviruses is somewhat different in that a predominance of transcription from the 3′ LTR and little transcription from the 5′ LTR have been reported (7). These authors have shown that the presence of a 0.3-kb leader sequence downstream of the primer-binding sequence suppresses transcription from the 3′ LTR whereas its absence, which is the case in most proviruses in B-cell lymphomas induced by chicken syncyial proviruses, promotes transcription from the 3′ LTR.

Chromatin Structure

The bulk of eukaryotic DNA is in nucleosomes, which consist of DNA of 165 bp, wrapped twice around an octamer of core histones of two H2A-H2B dimers and one $H3_2$-$H4_2$ tetramer (1, 26). The DNA in this configuration is inaccessible to transcription factors. In the active state, some of these nucleosomes, especially those in specific regions containing promoter and enhancer elements, are displaced, leaving stretches of DNA exposed and readily accessible to transcription factors (26, 33). The presence of nucleosomes is sufficient to inhibit interaction of transcription factors (57, 106). In particular, histone H1, which is associated with nucleosomes, is a potent inhibitor of transcription (48). Activation of genes from the inactive state is often accompanied by changes in chromatin structure, particularly by the displacement of nucleosomes and H1 (26, 48, 100, 105, 107).

During transcription it is necessary either that the RNA

polymerase II rotate along the DNA template or that another protein act to relieve the torsional stress of supercoiling. Evidence that topoisomerase I is involved in transcription elongation has been obtained by antibody and inhibitor studies (79, 80, 98). Stewart et al. (79) used camptothecin, an inhibitor of topoisomerase I (80, 108), to study the fate of c-*fos* transcripts when the gene is induced to an active state by the addition of the calcium ionophore A23187. They induced the cells with A23187, and added camptothecin at various times after induction. Under these conditions, camptothecin caused a retardation of transcriptional elongation by trapping the topoisomerase I molecules. These experiments demonstrated the functional involvement of supercoiling during transcription.

From these results I propose that the provirus genome is incorporated into nucleosomes with the exception of the enhancer region of the 5′ LTR region. I further propose that transcription factors interact with the 5′ LTR and displace the inhibitory protein histone H1 or H5 that bound to the core histone-DNA complex and to linker DNA (1, 26) and that this interaction induces higher-order supercoiling, resulting in a higher level of compaction and hence in the 3′ LTR becoming inaccessible to the same proteins. It has been shown that at least 6 to 10 proteins interact with different motifs in the LTR. For example, at least 10 different factors (not counting the transcription factors which recognize the core promoter) bind to the LTR of HIV-1, with at least 2 or 3 of them binding to the R region and 6 or 7 recognizing the U3 region (34). These "activator" proteins not only disrupt the nucleosome structure but also interact with transcription complexes (46). This means that these proteins keep the 5′ LTR in an "open" state at all times, and this configuration may block the interaction between the multicomponent transcription complex with AAUAAA and GU elements, preventing any premature termination of the transcripts. The open state of chromatin would then induce supercoiling further down the DNA, creating a torsional strain that could be relieved only when the transcription apparatus including topoisomerase I advanced past the GT sequence in the 3′ LTR.

The observations made with MMTV promoter are consistent with some of these ideas. Transcription from MMTV promoter is regulated by steroid hormones, which induce a receptor that binds to the hormone-responsive element, HRE (5). Prior to hormone induction, the chromatin region in the U3 was insensitive to DNase I, implying that the LTR is organized into a nucleosome structure (68). After treatment with hormone, the region from −200 to −60 became hypersensitive to DNase I, suggesting that a nucleosome has been removed or altered (19). Subsequently, it has been shown that hormone receptors bind to both naked and reconstituted nucleosomes equally well but that the transcription factor NF1, which is required for MMTV transcription, binds to the naked DNA very tightly but not to the promoter in the nucleosomes (62). These authors have further demonstrated that precise positioning of one nucleosome could repress transcription. The results discussed above (94) also indicated structural differences between the 5′ and 3′ LTR and lend further credence to the notion that the 3′ LTR usually is present in an inactive state.

Sequences other than the U3 and U5 may also contribute toward differential utilization of the 5′ and 3′ LTRs for transcription initiation and termination. As shown by Boerkoel and Kung (7), the sequence downstream of the primer-binding site (which is outside the LTR) may influence which LTR is to be utilized as transcription activator. In addition to the transcription factors that recognize the U3, R, and U5 sequences, it is probable that sequences downstream of the 5′ LTR and upstream of the 3′ LTR contribute to the functional differences between the two LTRs.

INHERENT PROBLEMS IN TRANSCRIPTION TERMINATION ANALYSIS

It has been practically impossible to distinguish between actual termination and efficient cleavage of the pre-mRNA to produce 3′ ends. In retroviruses, except in some abnormal situations, LTRs usually flank structural genes and transcription initiation occurs almost exclusively in the 5′ LTR. This preferential initiation of transcription in the 5′ LTR immediately suggests that the 5′ and 3′ LTRs, although identical in sequence and arrangement, are functionally different. To understand the respective roles of these LTRs in initiation and termination, experiments should be done with proviruses in their natural state. Unfortunately, most of the studies on transcription termination and polyadenylation, with the exception of one study (39), have involved artificial vector constructs in which termination signals of the LTR were placed under heterologous promoters including the thymidine kinase and the simian virus 40 late genes. In addition, most of the results on termination and polyadenylation were obtained under transient-expression conditions. We do not know whether the plasmid vector DNAs, containing the deletions or point mutations of sequences under investigation, assume nucleosome structures that resemble chromosomal arrangement of the provirus. Therefore, I believe that studies on transcription initiation and termination should be carried out with clones of host cells carrying single proviruses. To facilitate these analyses when possible, only small deletions or base substitutions in the appropriate region under investigation should be introduced into the viral genome itself. In our studies, we have taken these issues into account and constructed vectors with a selectable marker flanked by the LTRs (14a). We have already shown that deletion of the GT sequence in the U5 of only the downstream LTR obviously reduces the efficiency of termination and/or cleavage-polyadenylation as observed with natural proviruses. In another study, in which substitutions in the AATAAA sequence were made in the provirus, the results clearly indicated the requirement of AATAAA for 3′-end processing (81). Most of the transcripts, however, are only slightly longer than the genome, suggesting that termination does occur immediately downstream of the GT sequence (81). However, until rigorous analysis is done, most of the results obtained so far must be viewed with caution and only as guides for future experiments.

Of all the viruses known, only the retrovirus family requires integration as an essential step in its life cycle. Retroviruses have also evolved a unique mechanism of replication to ensure genesis of two LTRs flanking the structural genes. This arrangement instructs the host cell machinery to initiate and terminate viral transcripts within its LTRs. In cells infected with an exogenous retrovirus, upto 1% of total RNA or as much as 5 to 10% of poly(A) RNA is virus specific. This suggests that viral LTRs carry the control elements not only for efficient transcription initiation but also for an equally efficient termination. Therefore, it can be speculated that the U5 DNA in the 3′ LTR, which has a relatively higher G+T content, may be present in tightly coiled or highly compact nucleosomes, making it impossible for the transcription apparatus to advance further

as it is inaccessible for transcription factors. This impasse may not arise with the 5′ LTR because of the binding of transcription factors. Recent analysis of about 110 unidentified *Drosophila pol*II genes indicated that about 60% exhibited no evidence of cleavage at the 3′ end, suggesting that transcription termination rather than RNA processing to generate mature mRNAs (60). The results with retroviruses are also consistent with some of these observations. The most important topic in future work will certainly be the dissection of the 5′ and 3′ LTRs for their functional differences. We predict that the chromatin structure, the selection and binding of transcription initiation and termination factors, and the supercoiling of the provirus all will be important in the control of transcription termination.

If the transcripts fail to terminate, they will read through into the adjacent genes. A consequence of this readthrough is the activation of otherwise silent genes, some of which may result in tumorigenesis. These have been amply reviewed (43, 50, 92) and therefore will not be discussed here.

CONCLUSIONS

On the basis of a limited number of studies, it appears clear that in retroviruses, transcription termination and cleavage-polyadenylation follow the same principles that are involved in the processing of animal cell pre-mRNAs. The polyadenylation signal, AAUAAA, or a variant, AGUAAA, is present in all retroviruses; its location, however, varies among different groups of retroviruses. The G+T-rich sequence is present in the first 20 to 30 bp of the U5 region, immediately downstream of the poly(A) site. As in animal cell RNAs, both AAUAAA and G+U-rich sequences apparently contribute to the 3′-end processing of retroviral RNAs. At least in some studies it has been demonstrated that additional sequences in the U3 region determine the efficiency of 3′-end processing. If these studies are extended to other retroviruses, it would resolve a dilemma about the absence of prematurely terminated RNAs in the 5′ LTR. Some studies have also indicated that the distance between the 5′ cap site and AAUAAA distinguishes and/or determines 3′-end processing, which occurs only in the 3′ LTR.

The 5′ and 3′ LTRs are identical in sequence and arrangement, but they function differently. It is speculated that the formation of chromatin structure and possibly supercoiling cause functional heterogeneity between the two LTRs. It is also possible that flanking cellular DNA sequences, sequences downstream of the 5′ LTR and upstream of the 3′ LTR, also influence transcription initiation and termination and 3′-end processing. The reason that retroviruses have developed a unique strategy to generate an extra U3 sequence at the 5′ end and an extra U5 sequence at the 3′ end may be to ensure initiation and termination of its genome within its limits while keeping the minimum AAUAAA and G+U-rich elements required for proper processing.

ACKNOWLEDGMENTS

I thank B. R. Cullen, J. Alwine, J. Cherrington, D. Ganem, and several others for sending me reprints or preprints and H. M. Temin, J. Svoboda, D. Pintel, K. Wise, K. Bennett, J. C. Kandala, and P. Cleavinger for suggestions on the review.

Work in my laboratory over the years was supported by grants from the National Institutes of Health, American Cancer Society, and U.S. Environmental Protection Agency.

REFERENCES

1. **Adams, C. C., and J. L. Workman.** 1993. Nucleosome displacement in transcription. Cell **72**:305–306.
2. **Ahmed, Y. F., G. M. Gilmartin, S. M. Hanly, J. R. Nevins, and W. C. Greene.** 1991. The HTLV-1 Rex response element mediates a novel form of mRNA polyadenylation. Cell **64**:727–737.
3. **Bardwell, V. J., D. Zarkower, M. Edmonds, and M. Wickens.** 1990. The enzyme that adds poly(A) to mRNAs is a classical poly(A) polymerase. Mol. Cell. Biol. **10**:846–849.
4. **Bar-Shira, A., A. Panet, and A. Honigman.** 1991. An RNA secondary structure juxtaposes two remote genetic signals for human T-cell leukemia virus type 1 RNA 3′ end processing. J. Virol. **65**:5165–5173.
5. **Beato, M.** 1989. Gene regulation by steroid hormones. Cell **56**:335–344.
6. **Bienroth, S., E. Wahle, C. Suter-Crazzolera, and W. Keller.** 1991. Purification of the cleavage and polyadenylation factor involved in the 3′-processing of messenger RNA precursors. J. Biol. Chem. **266**:19768–19776.
7. **Boerkoel, C. F., and H.-J. Kung.** 1992. Transcriptional interaction between retroviral long terminal repeats (LTRs): mechanism of 5′ LTR suppression and 3′ LTR promoter activation of c-*myc* in avian B-cell lymphomas. J. Virol. **66**:4814–4823.
8. **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.
9. **Brawerman, G.** 1981. The role of the poly(A) sequence in mammalian messenger RNA. Crit. Rev. Biochem. **10**:1–38.
10. **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.
11. **Brown, P. H., L. S. Tiley, and B. R. Cullen.** 1991. Effect of RNA secondary structure on polyadenylation site selection. Genes Dev. **5**:1277–1284.
12. **Brown, P. O.** 1990. Integration of retroviral DNA. Curr. Top. Microbiol. Immunol. **157**:19–48.
13. **Cherrington, J., and D. Ganem.** 1992. Regulation of polyadenylation in human immunodeficiency virus (HIV); contribution of promoter proximity and upstream sequences. EMBO J. **11**:1513–1524.
14. **Christofori, G., and W. Keller.** 1988. 3′ cleavage and polyadenylation of mRNA precursors *in vitro* requires a poly(A) polymerase. Cell **54**:875–889.

14a. **Cleavinger, P., and R. V. Guntaka.** Unpublished observations.

15. **Cleavinger, P., J. C. Kandala, and R. V. Guntaka.** 1991. Sequences in the U5 region of Rous sarcoma virus LTR are required for transcription termination and polyadenylation. J. Cell. Biochem. Suppl. **15G**:246.
16. **Coffin, J. M.** 1990. Retroviridae and their replication, p. 1437–1500. B. N. Fields and D. M. Knipe (ed.), Virology. Raven Press, New York.
17. **Conaway, J. W., and R. C. Conaway.** 1991. Initiation of eukaryotic messenger RNA synthesis. J. Biol. Chem. **266**: 17721–17724.
18. **Conklin, K. F., and M. Groudine.** 1984. Chromatin structure and gene expression, p. 293–351. *In* A. Razin, H. Cedar, and A. D. Riggs (ed.), DNA methylation. Springer-Verlag, New York.
19. **Cordingley, M. G., A. T. Riegel, and G. L. Hager.** 1987. Steroid-dependent interaction of transcription factors with the inducible promoter of mouse mammary tumor virus *in vivo*. Cell **48**:261–270.
20. **Cullen, B. R.** 1991. Regulation of HIV-1 gene expression. FASEB J. **5**:2361–2368.
21. **Cullen, B. R.** 1992. Mechanism of action of regulatory proteins encoded by complex retroviruses. Microbiol. Rev. **56**:375–394.
22. **Cullen, B. R., P. T. Lomedico, and G. Ju.** 1984. Transcriptional intereference in avian retroviruses-implications for the promoter insertion model of leukaemogenesis. Nature (London) **307**:241–245.
23. **DeZazzo, J. D., J. M. Scott, and M. J. Imperiale.** 1992. Relative roles of signals upstream of AAUAAA and promoter proximity in regulation of human immunodeficiency virus type 1 mRNA

3′-end formation. Mol. Cell. Biol. **12**:5555–5562.
24. **Fan, H.** 1990. Influences of the long terminal repeats on retrovirus pathogenicity. Semin. Virol. **1**:165–174.
25. **Felber, B. K., H. Paskalis, C. Kleiman-Ewing, F. Wong-Staal, and G. N. Pavlakis.** 1985. pX protein of HTLV-1 is a transcriptional activator of its long terminal repeats. Science **229**:675–679.
26. **Felsenfeld, G.** 1992. Chromatin as an essential part of the transcriptional mechanism. Nature (London) **355**:219–224.
27. **Gill, A., and N. J. Proudfoot.** 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β-globin mRNA 3′ end formation. Cell **49**:399–406.
28. **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.
29. **Gilmartin, G. M., and J. R. Nevins.** 1991. Molecular analysis of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. Mol. Cell. Biol. **11**:2432–2438.
30. **Grandgenett, D. P., and S. R. Mumm.** 1990. Unraveling retrovirus integration. Cell **60**:3–4.
31. **Greenblatt, J.** 1991. RNA polymerase-associated transcription factors. Trends Biochem. Sci. **16**:408–411.
32. **Greene, W. C., and B. R. Cullen.** 1990. The Rev-Rex connection: convergent strategies for the post-transcriptional regulation of HIV-1 and HTLV-I gene expression. Semin. Virol. **1**:195–204.
33. **Gross, D. S., and W. T. Garrard.** 1988. Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. **57**:159–197.
34. **Haseltine, W. A.** 1991. Molecular biology of the human immunodeficiency virus type 1. FASEB J. **5**:2349–2360.
35. **Herman, S. A., and J. M. Coffin.** 1986. Differential transcription from the long terminal repeats of avian leukosis virus DNA. J. Virol. **60**:497–505.
36. **Herman, S. A., and J. M. Coffin.** 1987. Efficient packaging of read-through RNA in ALV: implications for oncogene transduction. Science **236**:845–848.
37. **Hsu, T. W., J. L. Sabran, G. E. Mark, R. V. Guntaka, and T. M. Taylor.** 1978. Analysis of unintegrated avian RNA tumor virus double-stranded DNA intermediates. J. Virol. **28**:810–818.
38. **Humphrey, T., and N. J. Proudfoot.** 1988. A beginning to the biochemistry of polyadenylation. Trends Genet. **4**:243–245.
39. **Iwasaki, K., and H. M. Temin.** 1990. The efficiency of RNA 3′-end formation is determined by the distance the cap site and the poly(A) site in spleen necrosis virus. Genes Dev. **4**:2299–2307.
40. **Iwasaki, K., and H. M. Temin.** 1992. Multiple sequence elements are involved in RNA 3′ end formation in spleen necrosis virus. Gene Expression **2**:7–18.
41. **Jacks, T.** 1991. Translational suppression in gene expression in retroviruses and retrotransposons. Curr. Top. Microbiol. Immunol. **157**:93–124.
42. **Jackson, R. J., and N. Standart.** 1990. Do the poly(A) tail and 3′ untranslated region control mRNA translation? Cell **62**:15–24.
43. **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**:179–219.
44. **Keller, W., S. Bienroth, K. M. Lang, and G. Christofori.** 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. EMBO J. **10**:4241–4249.
45. **Keshet, E., R. Schiff, and A. Itin.** 1991. Mouse retrotransposons: a cellular reservoir of long terminal repeat (LTR) elements with diverse transcriptional specificities. Adv. Cancer Res. **56**:215–251.
46. **Kornberg, R. D., and Y. Lorch.** 1991. Irresistable force meets immovable object transcription and the nucleosome. Cell **67**:833–836.
47. **Kumar, P., H. Hui, J. C. Kappes, B. S. Haggarty, J. A. Hoxie, S. K. Arya, G. M. Shaw, and B. H. Hahn.** 1990. Molecular characterization of an attenuated human immunodeficiency virus type 2 isolate. J. Virol. **64**:890–901.
48. **Laybourn, P. J., and J. T. Kadonaga.** 1991. Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. Science **254**:238–245.
49. **Logan, J., E. Falck-Pederson, J. E. Darnell, Jr., and T. Shenk.** 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse β^{mas} globin gene. Proc. Natl. Acad. Sci. USA **84**:8306–8310.
50. **Majors, J.** 1990. The structure and function of retroviral long terminal repeats. Curr. Top. Microbiol. Immunol. **157**:49–92.
51. **Manley, J.** 1988. Polyadenylation of mRNA precursors. Biochim. Biophys. Acta **950**:1–12.
52. **Maurer, B., H. Bannert, G. Darai, and R. M. Flugel.** 1988. Analysis of the primary structure of the long terminal repeat and the *gag* and *pol* genes of the human spumaretrovirus. J. Virol. **62**:1590–1597.
53. **McDevitt, M. A., R. P. Hart, W. W. Wong, and J. R. Nevins.** 1986. Sequences capable of restoring poly(A) site function define distinct downstream elements. EMBO J. **5**:2907–2913.
54. **McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AAUAAA signal is required for efficient formation of mRNA 3′ termini. Nucleic Acids Res. **13**:1347–1368.
55. **Monroy, G., M. Jacquet, Y. Groner, and J. Hurwitz.** 1974. AMV RNA transcription in cell-free systems and properties of *in vitro* chromatin-directed RNA synthesis. Cold Spring Harbor Symp. Quant. Biol. **39**:1033–1041.
56. **Moore, C. L., H. Skolnik-David, and P. A. Sharp.** 1986. Analysis of RNA cleavage at the adenovirus-2 L3 polyadenylation site. EMBO J. **5**:1929–1939.
57. **Morse, R. H., and R. T. Simpson.** 1988. DNA in the nucleosome. Cell **54**:285–287.
58. **Murthy, K. G. K., and J. L. Manley.** 1992. Characterization of the multi-subunit cleavage-polyadenylation specificity factor from calf thymus. J. Biol. Chem. **267**:14804–14811.
59. **Nevins, J. R.** 1983. The pathway of eukaryotic mRNA formation. Annu. Rev. Biochem. **52**:441–466.
60. **Osheim, Y., and A. Beyer.** 1991. Transcript release from the DNA template appears frequently determined by a DNA-associated phenomenon, p. 101. *In* RNA processing. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
61. **Panganiban, A. T.** 1990. Retroviral reverse transcription and DNA integration. Semin. Virol. **1**:187–194.
62. **Pina, B., U. Bruggemeier, and M. Beato.** 1990. Nucleosome positioning modulates accessibility of regulatory proteins to the mouse mammary tumor virus promoter. Cell **60**:719–731.
63. **Proudfoot, N. J.** 1991. Poly(A) signals. Cell **64**:671–674.
64. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature **263**:211–214.
65. **Pryciak, P. M., A. Sil, and H. E. Varmus.** 1992. Retroviral integration into minichromosomes *in vitro*. EMBO J. **11**:291–303.
66. **Qian, Z., and J. Wilusz.** 1991. An RNA-binding protein specifically interacts with a functionally important domain of the simian virus 40 late polyadenylation signal. Mol. Cell. Biol. **11**:5312–5320.
67. **Raabe, T., F. J. Bollum, and J. L. Manley.** 1991. Primary structure and expression of bovine poly(A) polymerase. Nature (London) **359**:229–234.
68. **Richard-Foy, H., and G. L. Hager.** 1987. Sequence-specific positioning of nucleosomes over the steroid-inducible MMTV promoter. EMBO J. **6**:2321–2328.
69. **Rohdewohld, H., H. Weiher, W. Reik, R. Jaenish, and M. Breindl.** 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. **61**:336–343.
70. **Russnak, R. 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.

71. **Sanfacon, H., and T. Hohn.** 1990. Proximity of the promoter inhibits recognition of cauliflower mosaic virus polyadenylation signal. Nature (London) **346**:81–84.
72. **Sawadogo, M., and A. Sentenac.** 1990. RNA polymerase B(II) and general transcription factors. Annu. Rev. Biochem. **59**: 711–754.
73. **Seiki, M., J.-I. Inoue, M. Hidaka, and M. Yoshida.** 1988. Two *cis*-acting elements responsible for post-translational trans regulation of gene expression of human T-cell leukemia virus type 1. Proc. Natl. Acad. Sci. USA **81**:1079–1083.
74. **Shank, P. R., S. Hughes, H. J. Kung, J. Majors, N. Quintrell, R. V. Guntaka, J. M. Bishop, and H. E. Varmus.** 1978. Mapping unintegrated avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. Cell **15**:1383–1395.
75. **Sheets, M. D., and M. Wickens.** 1989. Two phases in the addition of a poly(A) tail. Genes Dev. **3**:1401–1412.
76. **Shih, C.-C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53**:531–537.
77. **Shimotohno, K., M. Takanao, T. Teruuchi, and M. Miwa.** 1986. Requirement of multiple copies of a 21-nucleotide sequence in the U3 regions of human T-cell leukemia virus type I and II long terminal repeats for *trans*-acting activation of transcription. Proc. Natl. Acad. Sci. USA **83**:8112–8116.
78. **Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Haase, and S. Wain-Hobson.** 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell **42**:369–382.
79. **Stewart, A. F., R. E. Herrera, and A. Nordheim.** 1990. Rapid induction of *c-fos* transcription reveals quantitative linkage of RNA polymerase II and DNA topoisomerase I enzyme activities. Cell **60**:141–149.
80. **Stewart, A. F., and G. Schutz.** 1987. Camptothecin-induced *in vivo* topoisomerase I cleavages in the transcriptionally active tyrosine amino-transferase gene. Cell **50**:1109–1117.
81. **Swain, A., and J. M. Coffin.** 1989. Polyadenylation at correct sites in genome RNA is not required for retrovirus replication or genomic encapsidation. J. Virol. **63**:3301–3306.
81a. **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.
82. **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-mRNAs. Genes Dev. **4**:2112–2120.
83. **Takagaki, Y., L. C. Ryner, and J. L. Manley.** 1988. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell **52**:731–742.
84. **Talbot, R. L., E. E. Sparger, K. Lovelace, W. M. Fitch, N. C. Pedersen, P. A. Luciw, and J. H. Elder.** 1989. Nucleotide sequence and genomic organization of feline immunodeficiency virus. Proc. Natl. Acad. Sci. USA **86**:5743–5747.
85. **Tan, T.-H., M. Horikoshi, and R. G. Roeder.** 1989. Purification and characterization of multiple nuclear factors that bind to the *tax*-inducible enhancer within the human T-cell leukemia virus type long terminal repeat. Mol. Cell. Biol. **9**:1733–1745.
86. **Teich, N., J. Wyke, and P. Kaplan.** 1985. Pathogenesis of retrovirus-induced disease, p. 187–248. *In* R. Weiss et al. (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
87. **Teich, N., J. Wyke, T. Mak, A. Bernstein, and W. Hardy.** 1982. Pathogenesis of retrovirus-induced disease, p. 785–998. *In* R. Weiss et al. (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
88. **Tong-Starksen, S., and B. M. Peterlin.** 1990. Mechanism of retroviral transcriptional activation. Semin. Virol. **1**:215–227.
89. **Toyoshima, H., M. Itoh, J.-I. Inoue, M. Seiki, F. Takaku, and M. Yoshida.** 1990. Secondary structure of the human T-cell leukemia virus type 1 *rex*-responsive element is essential for *rex* regulation of RNA processing and transport of unspliced RNAs. J. Virol. **64**:2825–2832.
90. **Valsamakis, A., N. Schek, and J. C. Alwine.** 1992. Elements upstream of the AAUAAA within the human immunodeficiency virus polyadenylation signal are required for efficient polyadenylation in vitro. Mol. Cell Biol. **12**:3699–3705.
91. **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.
92. **Varmus, H. E.** 1987. Cellular and viral oncogenes, p. 271–346. *In* G. Stamatoyannopoulos, A. W. Nienhuis, P. Leder, and P. W. Majerus (ed.), Molecular basis of blood diseases. The W. B. Saunders Co., Philadelphia.
93. **Varmus, H. E., and P. Brown.** 1989. Retroviruses, p. 53–108. *In* D. Berg, and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
94. **Verdin, E.** 1991. DNase I-hypersensitive sites are associated with both long terminal repeats and the intragenic enhancer of integrated human immunodeficiency virus type 1. J. Virol. **65**:6790–6799.
95. **Vijaya, S., D. L. Steffen, and H. L. Robinson.** 1986. Acceptor sites for retroviral integrations map near DNaseI-hypersensitive sites in chromatin. J. Virol. **60**:683–692.
96. **Wahle, E., and W. Keller.** 1992. The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. Annu. Rev. Biochem. **61**:419–440.
97. **Wahle, E., G. Martin, E. Schiltz, and W. Keller.** 1991. Isolation and expression of cDNA clones encoding mammalian poly(A) polymerase. EMBO J. **10**:4251–4257.
98. **Wang, S. C.** 1985. DNA topoisomerases. Annu. Rev. Biochem. **54**:665–697.
99. **Weichs an der Glon, C., J. Monks, and N. J. Proudfoot.** 1991. Occlusion of the HIV poly(A) site. Genes Dev. **5**:244–253.
100. **Weintraub, H.** 1985. Assembly and propagation of repressed and derepressed chromosomal states. Cell **42**:705–711.
101. **Weiss, R., N. Teich, H. E. Varmus, and J. Coffin.** 1985. RNA tumor viruses, p. 580–1123. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
102. **Wickens, M.** 1990. How the messenger got its tail: addition of poly(A) in the nucleus. Trends Biochem. Sci. **15**:277–281.
103. **Wilusz, J., and T. Shenk.** 1990. A uridylate tract mediates efficient heterogeneous nuclear ribonucleoprotein C protein-RNA cross-linking and functionally substituted for the downstream element of the polyadenylation signal. Mol. Cell. Biol. **10**:6397–6407.
104. **Wilusz, J., T. Shenk, Y. Takagaki, and J. L. Manley.** 1990. A multi-component complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. Mol. Cell. Biol. **10**:1244–1248.
105. **Wolffe, A. P.** 1992. New insights into chromatin function in transcriptional control. FASEB J. **6**:3354–3361.
106. **Workman, J. L., and R. G. Roeder.** 1987. Binding of transcription factor TFIID to the major late promoter during *in vitro* nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell **51**:613–622.
107. **Zawel, L., and D. Reinberg.** 1992. Advances in RNA polymerase II transcription. Curr. Opin. Cell Biol. **4**:488–495.
108. **Zhang, H., J. C. Wang, and L. F. Liu.** 1988. Involvement of DNA topoisomerase I in the transcription of human ribosomal RNA genes. Proc. Natl. Acad. Sci. USA **85**:1060–1064.