The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure

Bianca I. F. Klasens, Maike Thiesen, Anders Virtanen1 and Ben Berkhout*

Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, PO Box 22700, 1100 DE Amsterdam, The Netherlands and 1Department of Medical Genetics, Uppsala University, Biomedical Center, Box 589, SE-751 23 Uppsala, Sweden

Received October 14, 1998; Revised and Accepted November 19, 1998

ABSTRACT

The 5′ and 3′ ends of HIV-1 transcripts are identical in sequence. This repeat region (R) folds a stem–loop structure that is termed the poly(A) hairpin because it contains polyadenylation or poly(A) signals: the AAUAAA hexamer motif, the cleavage site and part of the GU-rich downstream element. Obviously, HIV-1 gene expression necessitates differential regulation of the two poly(A) sites. Previous transfection experiments indicated that the wild-type poly(A) hairpin is slightly inhibitory to the process of polyadenylation, and further stabilization of the hairpin inhibited polyadenylation completely. In this study, we tested wild-type and mutant transcripts with poly(A) hairpin structures of differing thermodynamic stabilities for the in vitro binding of polyadenylation factors. Mutant transcripts with a destabilized hairpin efficiently bound the polyadenylation factors, which were provided either as purified proteins or as nuclear extract. The RNA mutant with a stabilized hairpin did not form this ‘poly(A) complex’. Additional mutations that repair the stability of this hairpin restored the binding capacity. Thus, an inverse correlation was measured between the stability of the poly(A) hairpin and its ability to interact with polyadenylation factors. The wild-type HIV-1 transcript bound the polyadenylation factors suboptimally, but full activity was obtained in the presence of the USE enhancer element that is uniquely present upstream of the 3′ poly(A) site. We also found that sequences of the HIV-1 leader, which are uniquely present downstream of the 5′ poly(A) site, inhibit formation of the poly(A) complex. This inhibition could not be ascribed to a specific leader sequence, as we measured a gradual loss of complex formation with increasing leader length. We will discuss the regulatory role of RNA structure and the repressive effect of leader sequences in the context of differential HIV-1 polyadenylation.

INTRODUCTION

Polyadenylation is one of the mRNA maturation steps in eukaryotic cells. Poly(A) tails can endow critical properties on the mRNA, such as stability, nuclear export and translatability (1–5). Polyadenylation involves cleavage of the primary transcript and subsequent addition of ∼250 adenosine ribonucleotides (6–11). Polyadenylation of mRNA precursors depends on specific RNA sequences that are recognized by nuclear protein factors (6,12–14). At least two sequence elements define a poly(A) site. The AAUAAA hexamer is the almost invariant poly(A) signal that is positioned ∼15 nt upstream of the site of cleavage. A poorly conserved GU-rich or U-rich element is usually positioned 20–30 nt downstream of the hexamer (15–21). These two elements form the core poly(A) site. Additional upstream enhancer (USE) elements have been described, in particular for viral RNAs (22–29). Although generally U-rich, these enhancer elements exhibit little sequence or structural similarity. The efficiency of RNA processing has been shown to correlate with the binding of polyadenylation factors to these signals (30).

The two key components responsible for the specific recognition of poly(A) sites are the cleavage and polyadenylation specificity factor (CPSF) and the cleavage stimulation factor (CstF) (6). CPSF consists of four polypeptide subunits with molecular weights of 160, 100, 73 and 30 kDa (31,32). CstF consists of three polypeptides of 77, 64 and 50 kDa. CPSF binds specifically to the AAUAAA signal and CstF binds to the downstream enhancer element (13,32–35). The presence of CPSF is required for efficient binding of CstF, and the interaction of CPSF with the RNA is stabilized by CstF. The stability of this RNA–CPSF–CstF ternary complex has been shown to correlate with the efficiency of mRNA processing (30). These factors form a 3′-processing complex upon association of the cleavage factors CFI and CFII and the 82 kDa poly(A) polymerase (PAP). It is generally assumed that CstF and the cleavage factors leave the complex after the cleavage reaction. CPSF and PAP remain bound to the RNA precursor and facilitate the addition of the poly(A) tail.

The HIV-1 retroviral genome contains a 97 nt repeat region (R) that forms both the extreme 5′ and 3′ ends of the mature RNA. Because the poly(A) signal is encoded within the R region, it is present at both ends of the viral transcript. This necessitates differential regulation to either repress recognition of the 5′ poly(A) signal or enhance usage of the 3′ signal. HIV-1 has been reported to have both regulatory features (Fig. 1). Usage of the 3′ poly(A) site is promoted by an enhancer motif that is uniquely present upstream of this poly(A) site (22–25,36). This USE
interacts directly with CPSF and stabilizes its binding at the AAUAAA hexamer motif (27). Repression of the 5′ poly(A) site is mediated by several mechanisms. First, the 5′ poly(A) site becomes active when moved further downstream in the transcript, suggesting that the 5′ HIV-1 poly(A) site is repressed because it is positioned too close to the transcription initiation site (37,38). It is possible that the transcription complex engaged in synthesis of the HIV-1 leader transcript is not yet competent for polyadenylation, which is consistent with the recent evidence that polyadenylation factors gain access to the nascent transcript through the RNA which is consistent with the recent evidence that polyadenylation

MATERIALS AND METHODS

Plasmid constructs

The HIV-1 sequences used in this study are derived from the infectious molecular clone pLAI (54). Several mutations were introduced into the poly(A) hairpin of the pBlue-5LTR subclone as described previously (44). Revertant viruses were obtained in prolonged virus culture experiments, and part of their genome was subcloned in the pBlue-5LTR plasmid (55). pBlue-5LTR contains an XbaI–ClaI fragment of pLAI, which includes the complete 5′ LTR and downstream sequences up to 40 nt downstream of the startcodon of the Gag open reading frame (56). pBlue-3LTR contains the complete 3′ LTR (up to LAI position +183) as an Xhol–BglII fragment (57), with downstream non-HIV sequences derived from the pBR322 plasmid. Control transcriptions were synthesized from plasmids that contain the poly(A) site of the adenovirus-2 L3 mRNA, with either the wild-type poly(A) signal AAUAAA (Adwt) or the mutant AAGAAA motif (Admut) (58).

In vitro transcription reactions

The pBlue-5LTR and pBlue-3LTR plasmids were used as template for PCR amplification and subsequent in vitro transcription. All nucleotide positions relate to the position on the viral RNA genome, with +1 being the transcriptional start site, and nucleotide positions in the 3′R region will be numbered according to the corresponding 5′R position. The sense primers T7-1 and T7-2 contain the T7 RNA polymerase promoter sequence.

Figure 1. Schematic of transcripts that mimic the 5′ or 3′ end of HIV-1 RNA. The HIV-1 DNA provirus is shown at the top, with a close up of the long terminal repeats (LTRs). The LTRs are divided in the U3, R and U5 regions. Nucleotide numbers refer to positions on the genomic RNA, with +1 being the capped G residue. The primary transcript is initiated at the U3R border within the 5′ LTR, and polyadenylation occurs at the R/U5 border within the 3′ LTR. The R region encodes the TAR hairpin and the poly(A) hairpin, and the latter structure contains the AAUAAA poly(A) signal (indicated by an arrow-head). The major SD (closed box) in the untranslated leader region has been reported to inhibit polyadenylation at the 5′ poly(A) site (40,41). Mutational inactivation of the SD context has been demonstrated to negatively influence the inhibitory effects of the SD. The major SD (closed box) in the untranslated leader region has been reported to inhibit polyadenylation at the 5′ poly(A) site (40,41). Mutational inactivation of the SD context has been demonstrated to negatively influence the inhibitory effects of the SD.

in vitro RNA–protein interaction studies and electrophoretic mobility shift assays (EMSA) to test this possibility.
Figure 2. RNA structure of the wild-type poly(A) hairpin and the mutants/revertants used in this study. Nucleotide numbers refer to positions on the wild-type (wt) genomic RNA, with +1 being the capped G residue. The poly(A) signal AAUAAA is marked by shading. Structure prediction and free-energy minimization were performed with the MFOLD program (70) in the GCG package, and the thermodynamic stability calculated for 37°C is presented below the structures (kcal/mol).

We note that there is considerable biochemical, phylogenetic and virological evidence for the existence of this stem–loop structure (42–44). In mutant A, the hairpin is stabilized by deletion (black triangle) of two bulged nucleotides and one nucleotide substitution (boxed). In mutants B and C, destabilizing mutations were introduced into the left- and right-hand side of the stem, respectively. A2, A4 and A7 are revertants of mutant A and the mutations that mediate the reversion phenotype are marked by black boxes.

(underlined in Table 1) and will anneal to the HIV-1 sequences starting at positions –54 and +1, respectively. The antisense primers used for amplification of different HIV-1 fragments are TU5, l521, BB3, ADSD and AD-AUG, and their positions on the HIV-1 RNA genome are indicated in Table 1. The AatII primer was used for amplification of HIV-1 sequences fused to downstream non-HIV sequences on plasmid pBlue-3′LTR. The PCR reactions were performed in PCR buffer (20 mM Tris–HCl pH 8.3, 2 mM MgCl₂, 50 mM KCl, 0.1 mg/ml BSA) containing 100 ng of each primer, 10 mM of each dNTP, 1.5 U Taq polymerase and 10 ng plasmid.

Adenoviral transcripts were synthesized with T3 RNA polymerase on DraI-linearized plasmids (58). The HIV-1 PCR fragments and the linearized adenovirus plasmid constructs were phenol-extracted, precipitated and dissolved in water. The in vitro transcription reaction was performed in 10 µl transcription buffer (40 mM Tris pH 7.5, 2 mM spermidine, 10 mM DTT and 12 mM MgCl₂) containing ~0.5 µg DNA template, 20 µCi [α-32P]UTP (3000 Ci/mmol), 0.06 µmol ATP, 0.06 µmol GTP, 0.06 µmol CTP, 0.02 µmol UTP, 10 U T7 or T3 RNA polymerase (Boehringer Mannheim) and 20 U RNase inhibitor (Boehringer Mannheim). RNA synthesis was continued for 2 h at 37°C, and then the template was removed by addition of 1 µl (2 U) DNase I (Boehringer Mannheim) and 10 ng plasmid.

Preparation of nuclear extracts and purified CPSF/CstF

HeLa cells (50 × 10⁶ cells) were grown to 60% confluency in 90 mm dishes in Dulbecco’s modified Eagle’s medium containing 8% fetal calf serum (FCS), 100 U of penicillin and 100 µg/ml of streptomycin at 37°C and 5% CO₂. The cells were washed once with prewarmed (37°C) PBS and put on ice, and then 2 ml of ice-cold PBS were added. The cells were collected by scraping with a rubber policeman, and washed twice with cold PBS. The cells were pelleted for 10 min in a Sorvall centrifuge at 2000 g. All further steps were performed in the cold room. The pellet was dissolved in 0.5 ml nuclear extract buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.5% NP-40), incubated on ice for 1 min, and centrifuged at 120 g for 5 min to pellet the nuclei. The pellet was resuspended in one packed cell volume of lysis buffer I [20 mM HEPES pH 7.9, 600 mM KCl, 0.2 mM EDTA and 0.2 mM EGTA, 2 mM Pefabloc (Boehringer Mannheim), 2 mM DTT]. The extract was incubated on ice for 30 min and subsequently cleared by centrifugation at 10 000 g for 10 min. The supernatant was collected and one packed cell volume lysis buffer II (20 mM HEPES pH 7.9, 20% glycerol, 0.2 mM EDTA and 0.2 mM EGTA) was added. The nuclear extract was frozen in liquid nitrogen and stored at –70°C. Protein concentrations were determined by the Bradford Method (59). For RNA binding
assays, we used 3 µl of the HeLa nuclear extract (4 mg/ml). Purified CPSF and CstF were obtained from HeLa cells as fractions IIIC and IA, respectively, according to the protocol of Åström et al. (60). We used 9 µl CPSF (0.6 mg/ml) and 2 µl CstF (1.1 mg/ml) per RNA binding reaction.

EMSA

Complexes between HIV-1 RNA and polyadenylation factors were assembled as described (27) in 25 µl band-shift buffer (16.5 mM HEPES pH 7.9, 83 mM KCl, 8.3% glycerol, 1% polyvinylalcohol (Sigma), 0.016% NP-40, 16 µg/ml tRNA (Escherichia coli, Boehringer Mannheim), 0.17 mM EDTA, 0.41 mM DTT). A standard amount of the 32P-labeled transcript was used per assay (1 nCi). The samples were incubated with the purified polyadenylation factors or nuclear extract for 10 min at 30°C, placed on ice and then heparin (Sigma) was added to 5 mg/ml. The samples were analyzed on a non-denaturing 3% acryl:bisacrylamide (19:1)/0.25× TBE with 5% glycerol. Electrophoresis was performed for 3 h at 200 V at room temperature. The gel was dried and signals were visualized and quantitated by PhosphorImager (Molecular Dynamics). It was difficult to compare the activity in the poly(A) complex with that of free RNA because of the formation of an aspecific RNA–protein complex of low molecular weight that is seen as a smear in most EMSA experiments. Therefore, we used equal counts of the different transcripts with the same specific activity per EMSA sample and thus per lane, and we simply measured the counts of the poly(A) complex. The efficiency of poly(A) complex of the wild-type –54/+134 transcript was set at 100% for each experiment.

RESULTS

Stable RNA structure interferes with the binding of polyadenylation factors

We analyzed the effect of RNA secondary structure of the HIV-1 poly(A) site on the binding of polyadenylation factors. To do this, wild-type and mutant transcripts were synthesized that mimic the 3’-context within the HIV-1 genome (Fig. 1). The –54/+134 RNA fragment contains the USE (marked as open box), the AAUAAA hexamer (marked as a black triangle), the cleavage site and two GU-rich downstream elements (15). Such transcripts were synthesized for the wild-type HIV-1 sequence that encodes the poly(A) hairpin (Fig. 2, ΔG = –15.3 kcal/mol) and mutants thereof stabilized hairpin (lanes 5 and 6, respectively). Several EMSAs were quantitated by PhosphorImager to calculate the efficiency of poly(A) complex formation for the RNA mutants compared with the wild-type –54/+134 transcript, of which the activity was arbitrarily set at 100%. These results are summarized in Table 2.

Table 2. Poly(A) complex formation with different HIV-1 templates

<table>
<thead>
<tr>
<th>HIV-1 transcript</th>
<th>ΔG (kcal/mol)</th>
<th>Efficiency of poly(A) complex formation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>–54/+134 (+USE)</td>
<td>-11.4</td>
<td>5</td>
</tr>
<tr>
<td>–54/+134 (+USE)</td>
<td>-25.7</td>
<td>20</td>
</tr>
<tr>
<td>–54/+134 (+USE)</td>
<td>-15.3</td>
<td>50</td>
</tr>
<tr>
<td>–54/+134 (+USE)</td>
<td>-6.8</td>
<td>100</td>
</tr>
</tbody>
</table>

aΔG was calculated at the temperature of the EMSA (30°C).
bAverage of 3–7 experiments with HeLa nuclear extract, except for the +/1/+368 RNAs that were tested once. The activity measured for the wild-type –54/+134 template (+USE) was arbitrarily set at 100% (value marked in bold).
cnd, not determined.
dTranscript –54/+368 (+USE).

We next tested the wild-type and mutant HIV-1 transcripts in EMSA with the purified CPSF and CstF polyadenylation factors (Fig. 4, lanes 1–4). For comparison, we also performed EMSA with HeLa nuclear extract (lanes 5–8). Consistent with the previous results, binding of the purified factors was completely blocked in the stable A mutant compared with the wild-type control (lanes 2 and 1, respectively). We also observed some differences in the EMSA pattern obtained with purified factors versus nuclear extract. With the purified factors, we measured a significant increase in the efficiency of poly(A) complex formation upon opening of the wild-type hairpin as in mutants B and C (lanes 3 and 4). This effect was not observed in EMSA with nuclear extract (Fig. 4, lanes 5–8; see also Fig. 3). These results suggest that even the wild-type hairpin structure is somewhat inhibitory to poly(A) complex formation.

Another difference in EMSA with the two sources of polyadenylation factors is the gel mobility of the poly(A) complexes, suggesting a difference in the composition of the assembled RNA–protein complexes. The complex made with CPSF and CstF migrated faster than the complex made with total nuclear extract (compare for instance lanes 4 and 5 of Fig. 4). The fast-migrating band is likely to represent the ternary RNA–CPSF–CstF complex because we detected no such RNA–protein complex in assays with the individual factors (not shown). Indeed, binding of CstF to RNA has been reported to depend on...
Figure 3. Stabilization of the poly(A) hairpin occludes binding of polyadenylation factors. EMSA was performed with adenoviral transcripts and 3′-like HIV-1 RNAs (position –54/+134) in combination with HeLa nuclear extract (lanes 1–6). The mock-incubated RNA samples are included for comparison (lanes 7–12). The transcript that was used is indicated above the lanes. The adenoviral transcripts have either the wild-type AAUAAA poly(A) signal (Adwt) or the mutant AAGAAA signal (Admut). The wild-type and mutant poly(A) hairpin in the HIV-1 RNAs are shown in detail in Figure 2. The position of the free RNA and the RNA–protein complex is indicated on the left. A representative experiment is shown, but similar results were obtained in at least six independent assays.

CPSF-binding, and the interaction of CstF subsequently stabilizes the CPSF–RNA complex (6,12,32,61). Because of the slower migration of the RNA–protein complex made with nuclear extract, it is likely that other factors have joined the RNA–CPSF–CstF complex. Likely candidates are PAP and/or CFI, which bind RNA substrates to further stabilize the RNA–CPSF–CstF complex (12,62). Binding of other polyadenylation factors may also add to the molecular weight of the poly(A) complexes assembled in HeLa nuclear extract (14).

The combined EMSA results with the control adenoviral transcripts and the purified factors indicate that we are able to assemble poly(A) complexes on HIV-1 RNA, and this reaction is exquisitely sensitive to stable RNA structure. RNA structure seems to inhibit the initial step of polyadenylation, that is formation of the RNA–CPSF–CstF complex. The negative effect of RNA structure was dominant in case of the A mutant template with a stabilized hairpin. The wild-type HIV-1 template was suboptimal compared with the mutants that contain a destabilized hairpin in EMSA with purified factors, but full activity was measured in a HeLa cell extract that allows for the formation of larger, presumably more mature poly(A) complexes. Thus, whereas binding of the individual CPSF or CstF factors to the HIV-1 AAUAAA signal is inefficient, the increased affinity of the more mature poly(A) complex for the RNA is able to effectively compete with the repressive RNA structure.

Revertants of the stabilized mutant hairpin rescue binding of polyadenylation factors

The A mutant template with the stabilized poly(A) hairpin was a poor template for assembly of the poly(A) complex in all experimental settings. Because this mutant contains three nucleotide changes compared with the wild-type (Fig. 2), it cannot be excluded that the inhibition is caused by mutation of sequence elements that are recognized by the polyadenylation factors. For instance, the two bulging nucleotides of the wild-type hairpin were removed in mutant A to increase the thermodynamic stability, but the C-bulge is in fact part of the CA-cleavage site. To rule out putative sequence effects, we tested variants of the A mutant that were identified in virus reversion experiments (55). We tested three revertant sequences with additional mutations (Fig. 2, marked by a black box) that restore the thermodynamic stability of the hairpin.

EMSA was performed with the wild-type, mutant and revertant transcripts (Fig. 5). Compared with the A mutant, all revertant RNAs demonstrated increased binding of polyadenylation factors. The results were quantitated by PhosphorImager and we calculated.
the relative binding efficiencies of the different templates, with
the wild-type set at 100% (Table 2). Partial recovery of protein
binding was measured for the revertants with an intermediate
stability (A7 and A2), and full binding activity was measured for
revertant A4 with a thermodynamic stability that is comparable
with that of the wild-type poly(A) hairpin. These results indicate
that structural changes, and not sequence changes, are responsible
for the inaccessibility of the poly(A) signal in mutant A.

The USE enhancer specifically overcomes the
structure-mediated repression of the HIV-1 poly(A) site

So far we have tested HIV-1 transcripts that start at position –54
in the U3 region and end at position +134 (Fig. 1). These RNA
substrates mimic the 3′ poly(A) site in that they contain the
upstream USE enhancer. We also wanted to perform EMSA with
HIV-1 templates that mimic the 5′ poly(A) site. Two such
templates were synthesized, both starting at position +1 and
ending either at position +134 or +368. By comparing the
–54/+134 and +1/+134 templates in EMSA, it is clear that the
USE element stimulated binding of polyadenylation factors to the
wild-type HIV-1 RNA ∼2-fold (Fig. 6, compare lanes 1 and 5).
Interestingly, this USE-enhancement is specific for the wild-type
poly(A) site. We also analyzed the extended 5′-like HIV-1 transcripts, but the +1/+368 RNAs were
profoundly defective in binding of the polyadenylation factors, even
when the AAUAAA signal was exposed by destabilization of the
poly(A) hairpin (Fig. 6, lanes 9–12).

Leader-specific inhibition of the HIV-1 poly(A) site

A surprising result is that the +1/+368 HIV-1 transcript is an
inactive template for binding of polyadenylation factors (Fig. 6).
This may indicate that the 5′-specific leader sequences, and in
particular the +138/+368 region, affects the accessibility of the 5′
poly(A) signal, even in mutants with an opened poly(A) hairpin.
To test which part of the leader RNA is responsible for this effect,
we generated a nested set of +1 transcripts, ending at position
+134, +202, +245, +290 or +368 (Fig. 1). These transcripts were
tested in EMSA with HeLa nuclear extract (Fig. 7, lanes 7–11).
We observed a gradual loss of poly(A) complex formation with
increasing leader length, but most binding activity was lost upon
inclusion of the +245/+290 region. This part of the leader RNA encodes the DIS hairpin element that is important for dimerization of the viral genome (63), and indeed dimers were spontaneously formed by the +1/+290 transcript (Fig. 7, lane 16). Nevertheless, only a small percentage of this transcript was in the dimeric form, suggesting that RNA dimerization is not the direct cause of the reduction of poly(A) complex formation. To further exclude this possibility, we performed EMSA with mutant HIV-1 transcripts containing a deletion within the palindromic sequence of the DIS hairpin loop (64). These mutant RNAs did not form dimers in this in vitro system, but were indistinguishable from the wild-type control RNA in poly(A) complex formation (results not shown).

To prove that the leader-mediated inhibition is sequence-specific and not due to 3′-extension of the transcript, we synthesized a 389 nt control transcript in which the 5′ poly(A) site was extended by vector-derived sequences that are unrelated to HIV-1 (+1/+180R, see Fig. 1). Poly(A) complex formation was indeed restored for this control RNA (Fig. 7, lane 12), suggesting that the observed suppression of the binding of polyadenylation factors is specific for the sequence of the HIV-1 leader RNA. We also synthesized a similar 3′-nested set of HIV-1 transcripts starting at position –54, thus including the USE enhancer. Again, a gradual loss of complex formation was observed as a function of the length of the downstream HIV-1 sequences (Fig. 7, lanes 1–5), and binding activity was partially restored in the control transcript in which the HIV-1 poly(A) site was fused to vector-derived sequences (lane 6).

**DISCUSSION**

Several of the sequences that comprise the HIV-1 poly(A) site are positioned within an RNA stem–loop structure that we termed the poly(A) hairpin. Occluded by basepairing are part of the poly(A) hairpin, and binding activity was partially restored in the control transcript in which the 5′ poly(A) site was extended by vector-derived sequences that are unrelated to HIV-1 (+1/+180R, see Fig. 1). Poly(A) complex formation was indeed restored for this control RNA (Fig. 7, lane 12), suggesting that the observed suppression of the binding of polyadenylation factors is specific for the sequence of the HIV-1 leader RNA. We also synthesized a similar 3′-nested set of HIV-1 transcripts starting at position –54, thus including the USE enhancer. Again, a gradual loss of complex formation was observed as a function of the length of the downstream HIV-1 sequences (Fig. 7, lanes 1–5), and binding activity was partially restored in the control transcript in which the HIV-1 poly(A) site was fused to vector-derived sequences (lane 6).

We analyzed the 3′-specific leader region, and thus 

\[ G \text{ value below } -19.9 \text{ kcal/mol} \]

(A2 revertant, free energy was calculated at 30°C, which is the temperature of the EMSA assay). Even the wild-type hairpin (\( G = -18.1 \text{ kcal/mol at 30°C} \)) was repressive under certain experimental conditions. For instance, inefficient binding of polyadenylation factors was observed in HeLa nuclear extract with the wild-type HIV-1 transcript that mimics the 5′ poly(A) site (+1/+134), which lacks the USE enhancer. The ability of this wild-type transcript to form poly(A) complexes could be improved in two ways; by inclusion of the USE enhancer (transcript –54/+134) or by destabilization of the hairpin structure (mutants B and C). In other words, it appears that local RNA structure partially suppresses the activity of the HIV-1 poly(A) signal, which is the ideal situation to enable further up-regulation by the USE in the 3′-context and further down-regulation by SD/leader sequences in the 5′ context. Thus, a complex interplay of polyadenylation and splicing signals, repressive RNA structure and enhancer/silencer elements are involved in regulated HIV-1 polyadenylation.

The USE element is necessary for efficient processing at the HIV-1 poly(A) site (22–25), and was shown to act through binding of CPSF (27). We propose that the wild-type HIV-1 RNA template, which cannot optimally interact with CPSF because of local RNA structure, uses the upstream USE motif to overcome this deficiency in CPSF-binding. A striking similar observation has been made by Gilmartin et al. (27). These authors showed that the USE is necessitated by the suboptimal sequence context of the HIV-1 AUAAGG hexamer. We think that this suboptimal context
represents the local RNA structure. Indeed, RNA mutants with a destabilized hairpin are not responsive to the USE enhancer, presumably because their AAUAAA motif is not occluded. It has been suggested that the presence of the USE may enable CPSF to identify the correct AAUAAA hexamer (27). We now propose a new mechanistic model for recognition of the 3′ HIV-1 poly(A) site that incorporates all of the above findings (Fig. 8). It is suggested that the USE acts as the entry site for CPSF, which may subsequently either slide along the RNA in the direction of the poly(A) site or bind directly to the AAUAAA signal upon transient opening or ‘breathing’ of the poly(A) hairpin. Although not included in the model, the flanking TAR hairpin may be important to appropriately space the USE and poly(A) site, and we cannot exclude the possibility that structural interactions exist between the USE enhancer and the poly(A) site. Two types of RNA–CPSF complexes are depicted. CPSF may be able to interact with both the USE and AAUAAA sequences (complex shown on the right). Consistent with this idea is the fact that USE does not contain an AAUAAA-like sequence, suggesting that it may occupy another RNA-binding domain within CPSF. The second possibility is that AAUAAA-binding leads to displacement of the USE from CPSF (complex shown on the left). According to this scenario, the binding affinity for the AAUAAA hexamer should be greater than that for the USE sequence. We note that an entry site will be beneficial only if the system is not allowed to reach the equilibrium state. In other words, it will work only if the polyadenylation steps subsequent to recognition of the USE are irreversible or very fast compared with the reverse reaction, which is a likely scenario (6).

According to the proposed mechanism, HIV-1 RNA has a bipartite CPSF binding site in the 3′ poly(A) site (USE and AAUAAA). In contrast, only the AAUAAA motif is available in the 5′ poly(A) site, which will be occluded by the repressive hairpin structure that is expected to fold rapidly. For instance, early investigations into the folding of the tRNA molecule established an approximate time scale for the formation of RNA secondary structure in the 10^{-4} to 10^{-5} s range (reviewed in 65). However, the RNA structure-mediated inhibition of polyadenylation may not be sufficient for complete suppression of 5′ polyadenylation because the hairpin structure will be in equilibrium with the open form, and this ‘breathing’ will eventually expose the AAUAAA signal. Rapid folding of the poly(A) hairpin on the nascent transcript will delay recognition by polyadenylation factors, such that sufficient time is available for the additional repressive mechanisms to become effective. For instance, the growing RNA chain will at a certain point be recognized by the splicing machinery, and assembly of spliceosomes may hinder the interaction with polyadenylation factors. It has been reported that binding of U1 snRNP to the major SD site in the HIV-1 leader RNA leads to suppression of the 5′ poly(A) site (41). Molecular details of this mechanism are currently unknown, but this example adds to the growing list of cases in which the splicing machinery influences the process of polyadenylation (14). Our RNA binding studies suggest that the leader region can also inhibit the binding of polyadenylation factors to the 5′ poly(A) site in a more direct manner. This negative effect is specific for the HIV-1 leader sequences, although we were unable to identify a distinct repressive element. We measured that the binding of polyadenylation factors is gradually decreased upon inclusion of additional HIV-1 sequences, and it is possible that the leader RNA will bit-by-bit adopt a higher order structure that step-by-step restricts the accessibility of the 5′ poly(A) site. This direct repressive effect of the leader region may add to the SD-mechanism (40,41) and suppression by promoter proximity (37,38) to strongly inhibit the 5′ poly(A) site. Despite these multiple repressive mechanisms, we measured that ∼5–10% of the HIV-1 transcripts are prematurely polyadenylated at the 5′ poly(A) site in virus-infected cells (66).

In conclusion, we propose that the regulated step of HIV-1 mRNA polyadenylation is the binding of CPSF to the AAUAAA sequence motif. This is mediated by sequestering of the polyadenylation signal within the poly(A) hairpin structure. In the context of the 5′ poly(A) site, additional repressive effects lead to nearly complete inhibition of polyadenylation. In the context of the 3′ poly(A) site, the USE element acts as an entry site for the CPSF factor, and thereby facilitates poly(A) complex formation. Some of the predictions of this model were tested recently in full-length HIV-1 constructs with specific mutations introduced in either the 5′ or 3′ poly(A) hairpin (66). Opening of the 5′ hairpin activated premature polyadenylation up to 30–40% efficiency, causing a concomitant decline in viral RNA synthesis. This result demonstrates that this RNA structure is pivotal for repression of the 5′ poly(A) site. Whereas the wild-type hairpin obviously does not interfere with efficient polyadenylation at the
3′ poly(A) site, this mRNA processing step is blocked by the stabilized hairpin mutant. Thus, the thermodynamic stability of this RNA structure is delicately balanced to allow efficient repression of the 5′ poly(A) site, yet full activity of the 3′ poly(A) site. This may explain the apparent conservation of this structured RNA motif and its thermodynamic stability among all human and simian immunodeficiency viruses (42). It is likely that the proposed mechanism of regulated polyadenylation represents a more general gene expression strategy of the complex retroviruses. A common mechanism has been proposed for the equine infectious anemia virus (EIAV), which is a related lentivirus (26). Furthermore, we reported that both the lentivirus and spumavirus groups can fold a similar RNA structure that occludes part of the viral genome that juxtaposes the AAUAAA hexamer and the cleavage site, which are separated by 274 nucleotides in the linear sequence. These examples underscore the versatile usage by retroviruses of RNA structure as a key component of regulatory circuits.

ACKNOWLEDGEMENTS

We thank Maarten de Smit for advice and critical reading of the manuscript and Wim van Est for artwork. This work was supported in part by the Dutch AIDS Fund (AIDS Fonds) and the EC (grant 950675).

REFERENCES