Efficient cleavage and polyadenylation at the human immunodeficiency virus type-1 (HIV-1) poly(A) site requires an upstream 3'-processing enhancer to overcome the suboptimal sequence context of the AAUAAA hexamer. The HIV-1 3'-processing enhancer functions to stabilize the association of the pre-mRNA with cleavage and polyadenylation specificity factor (CPSF), the factor responsible for recognition of the AAUAAA hexamer. Intriguingly, in the absence of the 3'-processing enhancer, CPSF binding and polyadenylation efficiency could be restored to near wild-type levels upon replacement of the 14-nucleotide region immediately 5' of the HIV-1 AAUAAA hexamer (the B segment) by the analogous sequences from the efficient adenovirus L3 poly(A) site. To further investigate the contributions of RNA sequence and structure to poly(A) site recognition, we have used an in vitro selection system to identify B segment sequences that enhance the polyadenylation efficiency of a pre-cleaved RNA lacking a 3'-processing enhancer. The final RNA selection pool was composed of two predominant classes of RNAs. Nuclease probing revealed that the selected sequences restored an RNA conformation that facilitates recognition of the AAUAAA hexamer by CPSF. These results indicate that both the sequence and structural context of the AAUAAA hexamer contribute to poly(A) site recognition by CPSF.

The 3'-ends of all eukaryotic mRNAs are generated by endonucleolytic cleavage and, with the exception of those encoding histones, subsequently polyadenylated. In mammals, this coupled reaction requires two sequences within the pre-mRNA that together compose a core poly(A) site, a highly conserved AAUAAA hexamer 10–30 nt upstream of the cleavage site and a highly variable U- or GU-rich element downstream of the cleavage site (reviewed in Refs. 1 and 2). The AAUAAA hexamer is recognized by the multisubunit cleavage and polyadenylation specificity factor (CPSF) (3–6). The binding of CPSF to the pre-mRNA is relatively unstable, but upon interaction with cleavage stimulation factor (CstF) (7), which binds to the downstream element (8), a stable ternary complex is formed (9). Endonucleolytic cleavage of the pre-mRNA also requires the cleavage factors Ii (10, 11), and in most cases poly(A) polymerase (PAP) (9, 10, 12). The subsequent poly(A) addition reaction is carried out by CPSF, PAP, and poly(A) binding protein II (PAB II). In vitro, CPSF and PAP are sufficient to reconstitute the poly(A) addition reaction on a “pre-cleaved” RNA substrate (an RNA that extends only to the site of endonucleolytic cleavage) (3, 4). However, PAB II enhances the processivity of PAP and confers poly(A) tail length control to the reaction (13, 14).

Human immunodeficiency virus type-1 (HIV-1) and equine infectious anemia virus contain 3'-processing enhancers upstream of their core poly(A) sites (15–19). These enhancers are functionally equivalent and act to stabilize the binding of CPSF to the pre-mRNA (18, 19). A linker substitution mutation (Δuse) within the HIV-1 3'-processing enhancer has been shown to reduce 3'-processing in vitro (15–17). This mutation dramatically reduced the binding of CPSF and the efficiency of poly(A) addition to a pre-cleaved RNA in vitro (18).

We have recently described an in vitro selection technique that facilitates the isolation of RNA molecules that are efficiently polyadenylated from a large pool of variants (20). By applying this method to study the sequence requirements of the HIV-1 3'-processing enhancer, we demonstrated that the structural context within which the AAUAAA hexamer resides is vital to its recognition by CPSF. In addition, we observed that the Δuse mutation altered the conformation of the pre-mRNA such that the sequences encompassing the AAUAAA hexamer were sequestered.

Here, we have used in vitro selection to further examine the contributions of RNA sequence and structure to poly(A) site recognition by CPSF. Our previous work suggested that the HIV-1 AAUAAA hexamer resides within a suboptimal sequence context, necessitating its juxtaposition to an upstream 3'-processing enhancer for efficient processing (17, 18). Replacement of the 14-nucleotide region of the HIV-1 poly(A) site between the TAR element and the AAUAAA hexamer (a sequence element we refer to here as the B segment) by the corresponding sequence of the adenovirus L3 poly(A) site was sufficient to restore CPSF binding and efficient poly(A) addition to a pre-cleaved RNA in the absence of a functional 3'-processing enhancer (18). These results suggested a potential contribution of sequence context to poly(A) site recognition by CPSF. We have therefore selected for B segment sequences capable of restoring efficient poly(A) addition to a pre-cleaved RNA containing the Δuse mutation. In stark contrast to the wide array of sequences identified in our previous selection for efficient upstream 3'-
processing enhancers (20), selection of efficient polyadenylation substrates, from a pool of 268,435,456 potential B segment variants, yielded two predominant classes of RNAs. RNase probing data revealed that the selected sequences restored an RNA conformation comparable with that of the wild-type HIV-1 RNA. These results strongly suggest that both the RNA sequence and structural context of the region encompassing the AAUAAA hexamer contribute to poly(A) site recognition by CPSF.

**Experimental Procedures**

**Nuclear Extract and Protein Purification**—HeLa cell nuclear extract was prepared as described by Rüegsegger et al. (11). Purified human CPSF and recombinant bovine PAP were prepared as described previously (18).

**Preparation of RNA Substrates**—The transcription templates encoding the RNAs in which the TAR element and B segments were exchanged were generated using combinations of unique restriction sites and oligonucleotide cartesians. The templates encoding pre-cleaved RNAs containing the Δuse mutation were PCR amplified with an SP6 primer and a primer ending at the cleavage site, and the PCR products were transcribed with SP6 RNA polymerase. Templates encoding the wild-type HIV-1 enhancer, use, were digested with XhoI prior to transcription with SP6 RNA polymerase. The templates encoding RNAs lacking sequences upstream of TAR were generated by either digestion with XhoI or PCR amplification, and the templates were transcribed with T7 RNA polymerase. The templates encoding the selected RNAs containing the HIV-1 downstream element were generated by digesting the respective plasmids with EcoRI prior to transcription with SP6 RNA polymerase.

**In Vitro Selection**—In vitro selection was performed essentially as described previously (20). The initial pool in which the 14 nt region between TAR and the AAUAAA hexamer was randomized in the context of the Δuse mutation was created as follows. Two partially overlapping oligonucleotides were annealed and extended with Klenow. The double-stranded DNA was PCR amplified with the SP6-Δuse primer (5'-ATTTCGTTGACACTTAGAAGATGCATATAAGAAGCTCATATG-CTCGAGGTCGACACT-3') and a 3'-primer 2 (5'-TGAAGACGCTACGAGCTAAGGCAAGCTTTATT-3'). The PCR product was transcribed with SP6 RNA polymerase and the RNA gel purified prior to use. 75 femtomoles of the initial pool (each variant should be represented 168 times) was incubated with 100 femtomoles of purified CPSF and 300 femtomoles of PAP for 10 min at 30 °C. ATP and MgCl₂ were added to 0.7 and 1 mM, respectively, to initiate polyadenylation, and the reaction continued for 2 min at 30 °C. The reaction was stopped by the addition of 175 μl of ETS (10 mM EDTA, 10 mM Tris (pH 7.8), and 0.5% SDS), sequentially extracted with phenol and phenol:chloroform, and ethanol precipitated. The RNA was selected on Dynabeads Oligo-d(T)₅₀ (Dynal) as described previously (20). The selected RNA was annealed to 20 pmol of 3'-primer 1 (5'-TGAAGACGCTACGAGCTAAGGCAAGCTTTATT-3') and reverse transcribed with avian myeloblastosis virus reverse transcriptase at 37 °C for 30 min. The reaction was extracted with phenol:chloroform, ethanol precipitated, and resuspended in H₂O. The cDNA was PCR amplified with the SP6-Δuse primer and 3'-primer 2 for 40 cycles at 94°C for 15 s, 55°C for 1 min, and 72°C for 1 min. The PCR product was transcribed with SP6 RNA polymerase, and the RNA gel purified. The new RNA pool was used for the next round of selection. The stringency of selection was increased by raising the KCl concentration and decreasing the CPSF concentration in the reaction as described (20). During the final round of selection, the selected RNAs were separated on a denaturing 5% gel, and the polyadenylated RNAs were excised and eluted. The eluted RNAs were then reverse transcribed and PCR amplified as described above. The PCR products encoding the initial and final pools were digested with NcoI and HindIII and ligated into pIVS-1 (20). All clones were sequenced with the T7 primer using Sequenase v2.0.

**Polyadenylation and CPSF-RNA Complex Assays—Polyadenylation and CPSF-RNA complex formation assays were performed as described previously (20). Data quantitation was performed on a Bio-Rad phosphorimager.**

**RNase Probing**—RNAs were ³²P 3'-end-labeled and probed with 10 picograms of RNase A or 0.1 units of RNase T1 as described previously (20).

**Cleavage Assays**—Cleavage assays were performed in HeLa cell nuclear extract for 10 min at 30 °C as described previously (19).
these pools was a reflection of CPSF binding, each RNA pool was incubated with purified human CPSF for 10 min at 30 °C, treated with heparin at 0 °C, and subjected to electrophoresis on a non-denaturing 3% polyacrylamide gel at 4 °C. The CPSF-RNA complexes were visualized by autoradiography. CPSF bound to the final RNA pool to a much greater extent than the initial pool (Fig. 1D, lanes 1 and 2). The extent of CPSF binding to the final pool was approximately five times greater than to the wild-type HIV-1 poly(A) site, use (Fig. 1D, lanes 2 and 3). This difference was also reflected in the efficiency of poly(A) addition (see below and Fig. 2B). The comparable amount of CPSF-RNA complexes formed with the initial RNA pool and the wild-type HIV-1 poly(A) site suggested that there were a relatively large number of molecules in the initial pool capable of binding CPSF.

**Functional Analysis of Selected RNAs**—The sequences of 22 individual clones from the final pool revealed that 50% of the population could be grouped into a single, highly related class in which all members contained an identical B segment (Fig. 2A, class I). Furthermore, an additional 18% of the clones could be grouped into a second class containing identical B segments distinct from class I (Fig. 2A, class II). The remaining seven clones were apparently unrelated to the class I or II RNAs or to one another (Fig. 2A, class III). All of the clones from the final pool contained mutations within the TAR element, whereas none of the 18 clones sequenced from the initial pool contained mutations within TAR (data not shown). These mutations were most likely acquired during repeated reverse transcription and PCR amplification, possibly conferring a selective advantage on the RNA molecules containing them (see below).

To determine if the selected RNAs were efficient poly(A) addition substrates, each individual clone from the final pool was transcribed and tested for poly(A) addition and CPSF binding. All of the RNAs from class I behaved identically in these assays, with the exception of 8.23, which differs from the other class I RNAs at a single position within TAR (data not shown). Therefore, only RNAs 8.1, which was selected as the representative from this class, and 8.23 are shown here. All members of class II performed identically (data not shown), and only RNA 8.18 is shown here. Each of the unique clones

---

**Fig. 1. In vitro selection of B segment sequences that promote efficient poly(A) addition in the absence of a functional 3′-processing enhancer.** A, nucleotide sequence of the pre-cleaved RNA used for in vitro selection. The nucleotides between G_{42} and C_{100} constitute the RNA stem loop structure of TAR. The **boldfaced** nucleotides upstream of TAR, C_{44} to C_{17}, constitute the linker substitution mutation Δuse that reduced 3′-processing in vivo and in vitro (15–17). The AAUAAA hexamer is **boldfaced** and resides between nucleotides 115 and 120. The 14-nt region between the TAR element and the AAUAAA hexamer, the B segment, was randomized. B, schematic diagram of the in vitro selection procedure. The RNA pool depicted in A was incubated with purified CPSF and recombinant PAP to allow for the formation of polyadenylation complexes. Polyadenylation was initiated upon the addition of MgCl\_2 and ATP and continued for 2 min. The poly(A)-RNA was separated from the poly(A)-RNA with Dynabeads Oligo-(dT)\_25, and the selected RNA was reverse transcribed, PCR amplified, and transcribed with SP6 RNA polymerase to generate a new RNA pool enriched for efficient polyadenylation substrates. The 5′-primer used for PCR amplification contained an SP6 RNA polymerase promoter and extended to the 5′-base of TAR. This was necessary to maintain the sequence of the Δuse mutation. Eight rounds of selection were performed. The cDNAs encoding the final pool were cloned, and individual clones were sequenced in their entirety. C, the final RNA pool from in vitro selection was enriched for RNAs that were efficiently polyadenylated. 32P-Labeled RNAs from selection rounds 0 and 8 were incubated with purified CPSF and rPAP under polyadenylation conditions. Aliquots were removed at the times indicated, and the reaction products were separated on a denaturing 5% polyacrylamide gel. D, the poly(A) addition efficiency of the RNA pools is reflected in CPSF binding. 32P-Labeled RNAs from selection rounds 0 and 8, as well as the wild-type HIV-1 RNA, use, and the HIV-1 RNA containing the Δuse mutation were incubated with purified CPSF for 10 min at 30 °C. The reactions were then treated with heparin to 5 mg/ml at 0 °C and separated on a nondenaturing 3% polyacrylamide gel.
**Fig. 2. Characterization of the selected RNAs.** A, nucleotide sequence of individual RNAs from the eighth round of selection. The sequence between the primer binding sites is depicted and aligned with the sequence of the starting RNA (— = same nucleotide; ● = deleted nucleotide). The clones could be grouped into three classes based on the sequences of the B segments. Classes I and II represent RNAs containing identical B segments. Class III represents RNA that were represented only once. The clones are named by the selection round from which they were derived (8) and the clone number from that pool. B, the selected RNAs were efficiently polyadenylated. 32P-Labeled RNAs transcribed from each of the clones depicted in A were incubated with purified CPSF and rPAP under polyadenylation conditions for the times indicated. Only RNAs 8.1 and 8.23 are shown from class I; all other class I members behaved in a manner identical to that of 8.1. Only RNA 8.18 from class II is shown; all other class II RNAs behaved in a manner identical to that of 8.18. All of the class III RNAs are shown. C, the poly(A) addition efficiency of each selected RNA is reflected in CPSF binding. The RNAs shown in B were incubated with purified CPSF, and the reactions were separated on a nondenaturing 3% polyacrylamide gel.

### Table: Characterization of the selected RNAs

| Class I | TAR |
|---------|-----|
| 8.1     | C   |
| 8.3     | C   |
| 8.7     | C   |
| 8.8     | C   |
| 8.10    | C   |
| 8.11    | C   |
| 8.13    | C   |
| 8.19    | C   |
| 8.20    | C   |
| 8.21    | C   |
| 8.23    | C   |

Class II

| Class III |
|-----------|
| 8.6       | C   |
| 8.16      | C   |
| 8.17      | C   |
| 8.18      | C   |

### Diagram: Poly(A) Site Recognition

![Diagram showing poly(A) site recognition](image-url)
from class III is shown here as well.

Polyadenylation of the HIV-1 RNA containing the Δuse mutation was barely detectable (Fig. 2B, lanes 5–8). As expected, each of the selected RNAs was efficiently polyadenylated (Fig. 2B, lanes 9–48). In fact, the efficiency of poly(A) addition of all of the selected RNAs was greater than that of the wild-type HIV-1 RNA, use (Fig. 2B, lanes 1–4). The single nucleotide difference between RNAs 8.1 and 8.23 from class I resulted in a significantly lower polyadenylation efficiency of the 8.23 RNA (Fig. 2B, lanes 9–16). Most likely, the single nucleotide difference between RNA 8.23 and the remainder of the class I RNAs occurred during the final rounds of reverse transcription and PCR amplification.

The efficiency of poly(A) addition was directly reflected in the binding of CPSF to each substrate (Fig. 2C). No detectable complex was formed between CPSF and the HIV-1 RNA containing the Δuse mutation, whereas CPSF bound to the RNA containing the wild-type HIV-1 sequence, use (Fig. 2C, lanes 1 and 2). For each of the selected RNAs, with the exception of RNA 8.23 (lane 4), CPSF bound a greater fraction of the RNA than the wild-type use RNA. The single base difference between RNAs 8.1 and 8.23 resulted in a reduction in the amount of 8.23 RNA bound by CPSF (lanes 3 and 4).

Selected B Segment Sequences, as Well as TAR Mutations, Contribute to Poly(A) Addition Efficiency—All of the selected RNAs contained mutations within the TAR RNA stem loop structure. To determine if these mutations contributed to poly(A) addition efficiency, we focused our attention on the class I RNAs. The TAR mutations were separated from the class I B segment, and the resulting RNAs containing each element alone were assayed for poly(A) addition and CPSF binding (Fig. 3). The polyadenylation data shown in Fig. 3B demonstrates that the class I TAR element and B segment were each capable of enhancing poly(A) addition efficiency in the context of the Δuse mutation (cf. Δuse, lanes 5–8; 8.1a, lanes 13–16; and 8.1b, lanes 17–20). However, the poly(A) addition efficiency of the class I RNA containing both selected elements was greater than the efficiency of either RNA containing a single selected element alone (Fig. 3B, 8.1, lanes 9–12). Quantitation of the data in Fig. 3B by phosphorimager revealed that, when combined, the effects of the selected TAR and B segment sequences on poly(A) addition were additive (Fig. 3A).

An analysis of the ability of each RNA in Fig. 3A to form a complex with CPSF revealed that, again, each selected RNA element alone could overcome the effect of the Δuse mutation to promote the binding of CPSF (Fig. 3C, cf. Δuse, lane 2; 8.1a, lane 4; and 8.1b, lane 5). Consistent with the poly(A) addition data, the binding of CPSF to the class I RNA containing both RNA elements was greater than to the RNAs containing either element alone (Fig. 3C, 8.1, lane 3). Phosphorimager quantitation of the data in Fig. 3C revealed that the two RNA elements combine synergistically to promote CPSF binding (Fig. 3A). 26% of the RNA containing both elements was bound by CPSF in contrast to 5 and 6% of the RNA containing only the selected B segment or selected TAR element, respectively. These results demonstrate that the mutations within the TAR element of the class I RNAs enhance the efficiency of poly(A) addition and CPSF binding above the level of the selected B segment alone. The mutations within the class I TAR element therefore contributed to the selective advantage of the class I RNAs.

The Selected RNAs Exhibit a Structural Conformation Similar to That of the Wild-type HIV-1 RNA—We previously found that RNAs selected for efficient poly(A) addition share a common structural conformation that appears to place the sequences encompassing the AAUAAA hexamer in open structure accessible to CPSF (20). To determine if the selected RNAs adopted a similar structural conformation, we probed the structures of the class I and class II RNAs with the RNases A and T1. Equal moles of 32P 3′-end-labeled RNA were incubated in the absence of protein at 30 °C for 5 min under poly(A) addition
conditions (except that ATP was omitted; see “Experimental Procedures”) followed by the addition of RNase A or T1 to the reaction mixture, and the incubation was continued for 10 min. The RNase sensitivity profiles were analyzed by electrophoresis of the samples on denaturing 10% polyacrylamide gels and autoradiography. We have previously demonstrated that the accessibility of G121 (which resides immediately 3' to the AAUAAA hexamer; see Fig. 1A) to RNase T1 directly correlated with the ability of that RNA to be utilized as a poly(A) addition substrate (20). The class I and class II RNAs contained this nucleotide in a conformation that was accessible to RNase T1 (Fig. 4A, lanes 2, 4, 5, 6). In contrast, G121 was inaccessible to RNase T1 in the RNA containing the Δuse mutation and the wild-type HIV-1 B segment (Fig. 4A, lane 3). The doublets observed were due to 3'-end heterogeneity generated by SP6 RNA polymerase.

We have previously demonstrated that the accessibility of nucleotide C40 (see Fig. 1A) to RNase A also correlated with poly(A) addition efficiency and CPSF binding (20). The class I RNAs isolated in this selection, as well as the wild-type HIV-1 RNA, used, contained C40 in a conformation that was accessible to RNase A (Fig. 4A, lanes 8, 10, 11). This nucleotide was not accessible to RNase A in the Δuse RNA or, surprisingly, in the class II RNAs (Fig. 4A, lane 9, 12). However, the TAR element in the class II RNAs clearly adopted a very different structure from that of the wild-type RNA. Both RNase T1 and RNase A cleaved within the class II TAR element in regions that are normally inaccessible to these single strand-specific nucleases (Fig. 4A, lanes 6, 12). Therefore, the inaccessibility of C40 to RNase A in the class II RNAs is likely due to large changes that disrupt the integrity of the TAR RNA stem loop and therefore the overall RNA structure.

To determine if the enhancement of poly(A) addition observed with the class I TAR and B segments alone was also due to a restoration of RNA structure, we subjected the RNAs 8.1a and 8.1b to RNase probing. G121 and C40 were accessible to RNase T1 and RNase A, respectively, in each of these RNAs (Fig. 4B, lanes 13–18).

These results demonstrate that the selected RNAs adopt a structural conformation surrounding the AAUAAA hexamer, as measured by G121 accessibility to RNase T1, that correlates with their ability to serve as poly(A) addition substrates. In addition, the distinct structure of the class II RNAs suggests that the overall structure of poly(A) sites need not be conserved but only the local conformation encompassing the AAUAAA hexamer.

**The Selected B Segment Sequences Enhance CPSF Binding and Poly(A) Addition in the Absence of Upstream Δuse Sequences**—The B segments selected for efficient poly(A) addition place the AAUAAA hexamer in a favorable context that allows for the negative impact of the Δuse mutation to be overcome. We have therefore examined the impact of the selected B segment upon poly(A) addition in the complete absence of upstream sequences. To this end, we constructed a set of RNAs containing the class I TAR element and B segment, either alone or together in the absence of sequences upstream of TAR (Fig. 5A). These RNAs were assayed for poly(A) addition and CPSF binding.

The RNAs containing the class I B segment were polyadenyl-
the context did not result in a further increase in the enhancement of poly(A) addition. However, combining the class I TA element and B segment in binding (Fig. 6).

The class I B segment enhanced the poly(A) addition efficiency of RNAs lacking upstream sequences. A, RNAs lacking sequences upstream of TAR were synthesized. These RNAs contained the class I B segment and TAR element, either alone or in combination. These RNAs were then assayed for polyadenylation efficiency (B) and CPSF binding (C). The data in parts (B) and (C) were quantitated and compared with RNA CPS 8.1, which was arbitrarily set at 1.00 (A). The calculations were performed as described in Fig. 3.

Impact of B Segment Sequences Selected for Efficient Poly(A) Addition upon the Efficiency of CPSF Binding and Endonucleolytic Cleavage When Placed within the Context of the HIV-1 Downstream Element—Although endonucleolytic cleavage requires a considerably more complex set of RNA-protein and protein-protein interactions, we were interested in determining whether the B segment sequences selected for efficient poly(A) addition in the context of the use/dse mutation would also serve to enhance the endonucleolytic cleavage reaction. Each of the selected RNAs illustrated in Fig. 2, panels B and C, was joined with the HIV-1 downstream element, and the resulting pre-mRNAs were examined for endonucleolytic cleavage in HeLa cell nuclear extract. The cleavage efficiency of each RNA was assessed by denaturing polyacrylamide gel electrophoresis. Although in at least one case (Fig. 7A, 8.1/dse, lane 3) a selected B segment enhanced the efficiency of cleavage above the level of use/dse, most of the selected B segments had little impact on the efficiency of endonucleolytic cleavage (Fig. 7A). Selection for efficient poly(A) addition substrates therefore did not yield RNAs that, when placed in the context of the downstream element, necessarily exhibited an equivalent enhancement in cleavage efficiency.

We also examined the impact of the addition of the downstream element on the ability of the selected RNAs to form a complex with CPSF. As illustrated in Fig. 7B, nearly all of the selected RNAs tested formed a CPSF-RNA complex more efficiently than the use/dse RNA. In fact, several of the CPSF-RNA complexes approached wild-type (use/dse) levels. A comparison of the efficiencies of cleavage and CPSF-RNA complex
formation in Fig. 7B indicates that, in contrast to poly(A) addition on pre-cleaved RNAs, CPSF-RNA complex formation did not directly correlate with the efficiency of the endonucleolytic cleavage reaction. As discussed below, the complexity of the RNA-protein and protein-protein interactions required for endonucleolytic cleavage may account for the differential impact of changes within the B segment.

**DISCUSSION**

The only conserved element shared by mammalian poly(A) sites is an AAUAAA hexamer, which adjoins a highly variable U- or GU-rich downstream element. The 3′-processing machinery must be capable of discriminating between authentic poly(A) sites and other sequences within the message sharing similarity to the conserved sequence. Recently, it has become clear that sequences outside of the core elements influence poly(A) site efficiency. We have previously shown that the HIV-1 3′-processing enhancer functions to place the AAUAAA hexamer in an RNA conformation that facilitates CPSF binding (20). Replacement of the HIV-1 3′-processing enhancer with the D use linker substitution mutation significantly decreased the efficiency of endonucleolytic cleavage, poly(A) addition, and CPSF binding (18), apparently by sequestering the sequences encompassing the AAUAAA hexamer (20). In this report, we have used in vitro selection to identify compensatory sequences adjacent to the AAUAAA hexamer that restore efficient poly(A) addition, as illustrated by the large overall structural differences between the class I and class II RNAs. In each case, however, G 121 was found to be accessible to RNase T1. A variety of RNA structures has been shown to influence 3′-processing in vivo as well as in vitro. The extensive RNA structure of the HTLV-I RexRE has been shown to be required for the spatial juxtaposition of the AAUAAA hexamer to the cleavage site located 274 nt downstream (21). Conversely, an RNA structure that placed the hexamer in a double-stranded region has been shown to have a detrimental impact on 3′-processing in vivo (22). Taken together, these data suggest that while the overall structure of a poly(A) site can vary greatly, the local structural conformation of the RNA encompassing the AAUAAA hexamer is of primary importance for the recognition of a poly(A) site by CPSF.

It should be noted that the accessibility of G 121 to RNase T1 alone cannot be interpreted as a quantitative indication of the efficiency of poly(A) addition and CPSF binding. For instance, while G 121 is equally accessible to RNase T1 in the RNAs use and 8.1, RNA 8.1 is a much better substrate for CPSF binding and poly(A) addition than use. Therefore, while in all cases examined, the accessibility of G 121 correlated with poly(A) addition, additional parameters must influence efficiency.

The apparent lack of conserved RNA structures shared by mammalian poly(A) sites suggests that multiple structures, or higher order non-Watson-Crick structures, may serve to present the AAUAAA hexamer for CPSF binding. Multiple RNA structures are clearly compatible with efficient poly(A) addition, as illustrated by the large overall structural differences between the class I and class II RNAs. In each case, however, G 121 was found to be accessible to RNase T1. A variety of RNA structures has been shown to influence 3′-processing in vivo as well as in vitro. The extensive RNA structure of the HTLV-I RexRE has been shown to be required for the spatial juxtaposition of the AAUAAA hexamer to the cleavage site located 274 nt downstream (21). Conversely, an RNA structure that placed the hexamer in a double-stranded region has been shown to have a detrimental impact on 3′-processing in vivo (22). Taken together, these data suggest that while the overall structure of a poly(A) site can vary greatly, the local structural conformation of the RNA encompassing the AAUAAA hexamer is of primary importance for the recognition of a poly(A) site by CPSF.
CPSF binding and poly(A) addition and promoted an “open” RNA conformation. It therefore appears unlikely that RNA structure alone can account for the high efficiency of the selected RNAs. Taken together, our results suggest that sequence preferences within the B segment may facilitate optimal CPSF binding and poly(A) addition.

Considerable work will be necessary to fully understand how flanking sequences influence poly(A) site recognition. The scope of this problem is best illustrated by comparing the RNAs Δuse and 8.1b (Fig. 3), which differ at only three positions within TAR (a single nucleotide deletion and two point mutations). These three changes restored the efficiency of poly(A) addition and CPSF binding to near wild-type levels in the context of Δuse. Furthermore, these changes restored the nuclease accessibility of G121 and C40. These data illuminate the complex interplay among the sequences flanking the AAUAAA hexamer. Moreover, subtle changes in flanking sequences may have profound effects on the ability of CPSF to recognize the AAUAAA hexamer.

The impact of the selected B segment sequences upon cleavage at the HIV-1 poly(A) site suggests that a complex set of RNA-protein interactions, in addition to the binding of CPSF at the AAUAAA hexamer, influences the recognition of authentic 3'-processing sites. We observed that in the context of the downstream element, the majority of sequences selected for their ability to enhance poly(A) addition to a pre-cleaved RNA had little impact upon the endonucleolytic cleavage reaction. In each case, however, the binding of CPSF to the RNAs containing the selected B segments in the presence of the HIV-1 downstream element exceeded that of the Δuse RNA. These results are consistent with our earlier observation that replacement of the HIV-1 B segment with the analogous sequences had little impact upon the endonucleolytic cleavage reaction. In each case, however, the binding of CPSF to the RNAs containing the selected B segments in the presence of the HIV-1 downstream element exceeded that of the Δuse RNA. These results are consistent with our earlier observation that replacement of the HIV-1 B segment with the analogous sequences had little impact upon the endonucleolytic cleavage reaction (18) (it is interesting to note that the L3 sequence and the class I B segment are identical in 7 of 14 positions). The cleavage results should hardly be surprising, however, given the relative complexity of the two reactions. In contrast to poly(A) addition, which was accomplished here by two purified factors (CPSF and PAP), the cleavage reaction was carried out in a crude HeLa cell nuclear extract. Endonucleolytic cleavage requires the assembly of an RNA-protein complex composed of at least 12 polypeptides (constituents of the processing factors CPSF, CstF, CFIm, and PAP). Seven UV cross-links between the pre-mRNA and six of the protein components (CPSF 160- and 30-kDa subunits, CstF 64-kDa subunit, and CFIm80, 68-, 59-, and 25-kDa subunits) have been reported (6, 8, 11, 18, 23), although only three of these cross-links have been localized on the pre-mRNA (6, 8, 18). Changes within the B segment may therefore be expected to have multiple ramifications for the assembly of a 3'-processing complex in addition to an impact on CPSF binding.

The addition of RNA structure to the repertoire of information utilized by the 3'-processing machinery in poly(A) site recognition may help to explain how poly(A) sites are defined. The sequence AAUAAA occurs within the coding region of many mammalian genes, often in association with U-rich downstream sequences (20, 24). Our work suggests that these potential poly(A) sites may reside within an RNA conformation that sequesters them from the 3'-processing machinery. This model is currently under investigation.

Acknowledgments—We thank Sean Flaherty, Keith Cauley, and Jon Cherry for stimulating discussions and critical reading of the manuscript.
REFERENCES

1. Wahle, E. (1995) Biochim. Biophys. Acta 1261, 183–194
2. Keller, W. (1995) Cell 81, 829–832
3. Bienroth, S., Wahle, E., Suter-Crazzolara, C., and Keller, W. (1991) J. Biol. Chem. 266, 19768–19776
4. Murthy, K. G. K., and Manley, J. L. (1992) J. Biol. Chem. 267, 14804–14811
5. Bardwell, V. J., Wickens, M., Bienroth, S., Keller, W., Sprout, B. S., and Lamond, A. I. (1991) Cell 65, 125–133
6. Keller, W., Bienroth, S., Lang, K. M., and Christofori, G. (1991) EMBO J. 10, 4241–4249
7. Takagaki, Y., Manley, J. L., MacDonald, C. C., Wilusz, J., and Shenk, T. (1990) Genes & Dev. 4, 2112–2120
8. MacDonald, C. C., Wilusz, J., and Shenk, T. (1994) Mol. Cell. Biol. 14, 6647–6654
9. Gilmartin, G. M., and Nevins, J. R. (1989) Genes & Dev. 3, 2180–2189
10. Takagaki, Y., Ryner, L. C., and Manley, J. L. (1989) Genes & Dev. 3, 1711–1724
11. Riegegger, U., Beyer, K., and Keller, W. (1996) J. Biol. Chem. 271, 6107–6113
12. Christofori, G., and Keller, W. (1989) Mol. Cell. Biol. 9, 193–203
13. Bienroth, S., Keller, W., and Wahle, E. (1993) EMBO J. 12, 585–594
14. Wahle, E. (1995) J. Biol. Chem. 270, 2800–2808
15. Valsamakis, A., Zeichner, S., Carswell, S., and Alwine, J. C. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 2108–2112
16. Valsamakis, A., Schek, N., and Alwine, J. C. (1992) Mol. Cell. Biol. 12, 3699–3705
17. Gilmartin, G. M., Fleming, E. S., and Oetjen, J. (1992) EMBO J. 11, 4419–4425
18. Gilmartin, G. M., Fleming, E. S., Oetjen, J., and Graveley, B. R. (1995) Genes & Dev. 9, 72–83
19. Graveley, B. R., and Gilmartin, G. M. (1996) J. Virol. 70, 1612–1617
20. Graveley, B. R., Fleming, E. S., and Gilmartin, G. M. (1996) Mol. Cell. Biol. 16, 4942–4951
21. Ahmed, Y. F., Gilmartin, G. M., Hanly, S. M., Nevins, J. R., and Greene, W. C. (1991) Cell 64, 727–737
22. Gimmi, E. R., Reff, M. E., and Deckman, E. C. (1989) Nucleic Acids Res. 17, 6983–6998
23. Jenny, A., Hauri, H.-P., and Keller, W. (1994) Mol. Cell. Biol. 14, 8183–8190
24. Day, I. N. M. (1992) Gene 110, 245–249