

# Effect of RNA secondary structure on polyadenylation site selection

Pamela H. Brown, Laurence S. Tiley, and Bryan R. Cullen<sup>1</sup>

Howard Hughes Medical Institute and Department of Microbiology and Immunology, Duke University Medical Center, Durham, North Carolina 27710 USA

**Functional polyadenylation [poly(A)] sites consist of two sequence elements, the AAUAAA and G/U box signals, that closely flank the site of mRNA 3'-end formation. In agreement with previous results, random sequence insertions between the AAUAAA and G/U box signals were observed to inhibit poly(A) site function. However, sequence insertions of similar size that were predicted to form RNA stem-loop structures were found to have little effect on the efficiency of polyadenylation and instead induced a 3' shift in the site of polyadenylation that was equal to the length of the inserted stem-loop. The *in vivo* utilization of a poly(A) site bearing an internal RNA stem-loop structure was inhibited by mutations that destabilized the predicted stem but was restored by compensatory mutations. These results strongly support the hypothesis that the appropriate spacing of the AAUAAA and G/U box signals is critical for poly(A) site function. Sequence insertions that are able to form RNA secondary structures that maintain the correct spacing of these two RNA target sequences are well tolerated, whereas sequence insertions that disturb this spacing inhibit poly(A) site recognition. It is proposed that the effect of sequence insertions on poly(A) site function may be sufficiently predictable to allow the development of an assay for *in vivo* RNA secondary structure that uses poly(A) site selection as a readout.**

[Key Words: RNA secondary structure; mRNA polyadenylation; protein-RNA interaction]

Received November 28, 1990; revised version accepted April 12, 1991.

An essential step in the generation of mature eukaryotic mRNAs is the formation of a defined 3' terminus. In the vast majority of mRNAs, this processing step involves a sequence-specific endonucleolytic cleavage followed by the stepwise addition of  $\leq 250$  adenylate residues to the newly formed 3' end (for review, see Birnstiel et al. 1985; Wickens 1990). Although the purpose of this polyadenylation event remains to be fully defined, current data suggest a role for the poly(A) tail in enhancing transcript stability and translation (for review, see Jackson and Standart 1990). Many genes encode multiple distinct polyadenylation [poly(A)] sites, and in some cases, alternative site selection has been shown to regulate the synthesis of distinct protein products from the same initial gene transcript (Birnstiel et al. 1985). The process of poly(A) site selection, therefore, represents an interesting example of gene regulation at the level of RNA sequence recognition.

The primary *cis*-acting target signal for polyadenylation is a highly conserved "AAUAAA" hexanucleotide found 10–30 nucleotides 5' to the poly(A) site (Proudfoot and Brownlee 1976; Fitzgerald and Shenk 1981). Almost all mutations of the AAUAAA signal have been shown to inhibit poly(A) site function (Wickens 1990). Recently, it has become clear that a second signal, termed

the G/U box, is also required for efficient polyadenylation (McDevitt et al. 1986; Gil and Proudfoot 1987; Proudfoot 1991). Although the G/U box has no clearly defined consensus sequence or size, the presence of a G/U-rich sequence 3' to the site of polyadenylation is nevertheless critical for efficient poly(A) site utilization. Of particular importance is the proximity of the AAUAAA signal to the G/U box. If this distance is increased to  $>40$  nucleotides, poly(A) site usage is essentially lost (McDevitt et al. 1986; Gil and Proudfoot 1987).

An interesting exception to this rule occurs in the case of transcripts derived from the human T-cell leukemia virus type I (HTLV-I). In this virus, polyadenylation has been shown to occur  $>250$  nucleotides 3' to the only evident AAUAAA signal (Seiki et al. 1983). The sequences located between the AAUAAA and the poly(A) sites of HTLV-I are predicted to form a highly stable RNA stem-loop structure, and it was therefore suggested that this RNA structure could function to approximate the AAUAAA and G/U box signals in such a way that mRNA polyadenylation can occur (Seiki et al. 1983; Hanly et al. 1989). Recently, the existence of this RNA secondary structure has been substantiated (Ahmed et al. 1990; Toyoshima et al. 1990), and the critical role of this structure in mediating efficient polyadenylation of HTLV-I transcripts has been confirmed (Ahmed et al. 1991). Overall, it then appears that the HTLV-I poly(A)

<sup>1</sup>Corresponding author.

Brown et al.

site is used efficiently despite or, rather, because of the presence of a large RNA secondary structure located between the AAUAAA and G/U box signals.

In this report we have examined whether this unique phenomenon could be generalized to other poly(A) sites and RNA secondary structures. We demonstrate that a consensus synthetic poly(A) site continues to be utilized efficiently after insertion of any one of several RNA stem-loop structures between the AAUAAA and the site of cleavage but was no longer recognized when unstructured sequences of the same length were tested. Mutations designed to destabilize an introduced RNA secondary structure were shown to result in reduced poly(A) site usage, whereas compensatory mutations were observed to restore function. These results shed light on the process of poly(A) site selection and suggest a possible *in vivo* assay for the validation of computer-predicted RNA secondary structures.

## Results

To assay the effect of sequence insertions on polyadenylation efficiency, we constructed an expression vector bearing two functional poly(A) sites (Fig. 1, top). The 3' poly(A) site remained invariant in all expression constructs, whereas the 5' poly(A) site served as the substrate for insertion mutagenesis. Relative *in vivo* utilization of the experimental site, termed PA1, and the invariant site, termed PA2, was assayed by transfection into COS cells followed by quantitative S1 nuclease protection analysis of cytoplasmic RNA.

The invariant PA2 site present in the psPA vector is a natural poly(A) site derived from the genomic rat preproinsulin II gene (Lomedico et al. 1979; Böhnlein et al. 1989). In contrast, the experimental PA1 site, is a synthetic poly(A) site closely modeled on the consensus poly(A) site described by Levitt et al. (1989). This synthetic site contains consensus AAUAAA and G/U box motifs spaced at the appropriate distance apart (Fig. 1, bottom). Endonucleolytic cleavage is predicted to occur

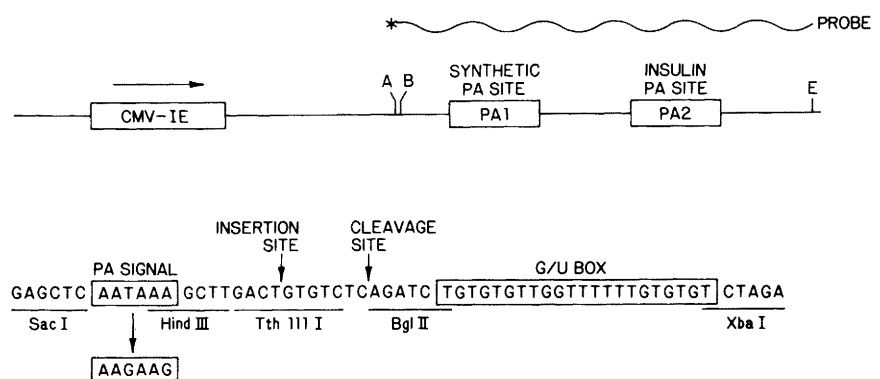
at the indicated CA dinucleotide sequence. The synthetic site also includes several convenient unique restriction sites, including a *Tth*III site suitable for insertion of test sequences. The minimally palindromic *Tth*III site (5'-GACNNNGTC-3'), as well as the overall sequence context of this poly(A) site, was designed to minimize the propagation of introduced RNA secondary structures. A derivative of the psPA vector, termed pΔPA, is identical to psPA except that the AAUAAA signal has been modified to AAGAAG (Fig. 1, bottom). This mutation is predicted to inactivate the synthetic poly(A) site; therefore, pΔPA serves as a negative control.

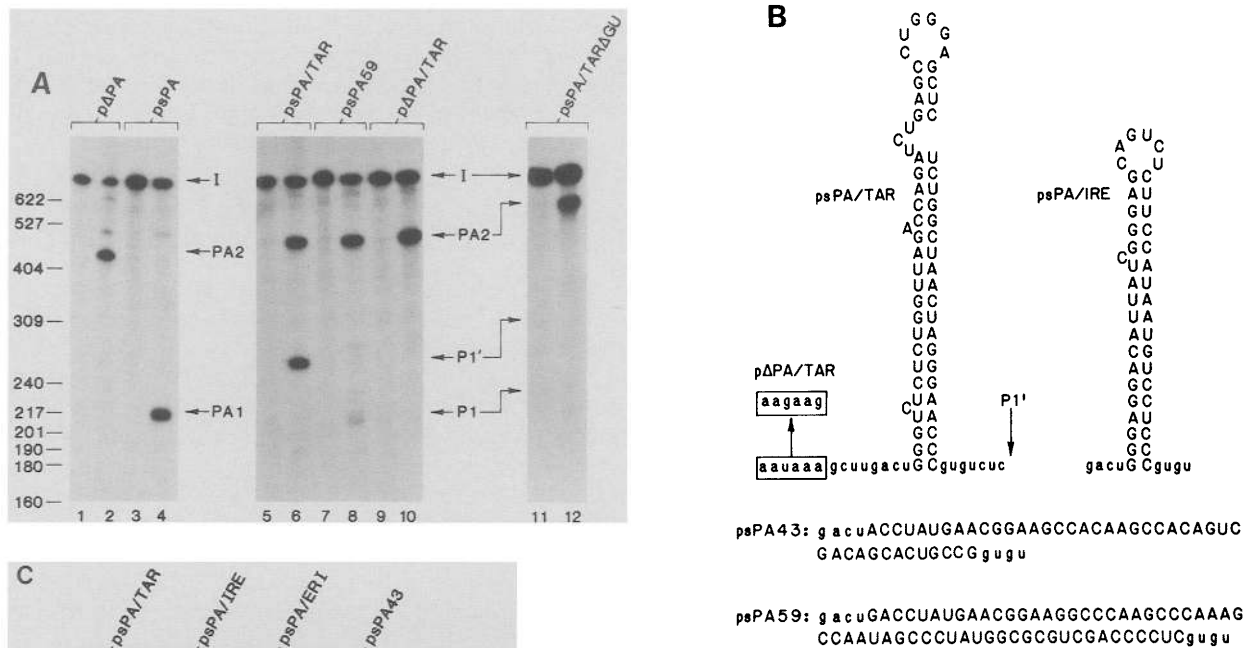
The *in vivo* pattern of RNA 3'-end formation, displayed by mRNAs transcribed from the psPA and pΔPA vectors in transfected COS cell cultures, was quantitated by S1 nuclease analysis (Fig. 2A). RNA transcripts derived from the psPA vector were observed to rescue a single probe fragment of the size predicted to result from utilization of the PA1 poly(A) site (Fig. 2A, lane 4). In contrast, transcripts derived from pΔPA exclusively rescued the larger probe fragment predicted to result from utilization of the PA2 poly(A) site (Fig. 2A, lane 2). Cultures transfected with a negative control vector gave no specific protected probe fragments (Fig. 2A, lanes 1 and 3). Overall, the results presented in Figure 2A indicate that both poly(A) sites present in the psPA vector are fully functional and also demonstrate that the 3' PA2 poly(A) site can be activated by introduction of a deleterious mutation into the 5' PA1 poly(A) site.

### *The effect of RNA sequence insertions on poly(A) site selection is determined by secondary structure*

The target sequence (TAR) for the HIV-1 Tat *trans*-activator, a 59-nucleotide RNA stem-loop structure located at the 5' end of all HIV-1 transcripts (for review, see Cullen 1990), was chosen as the first sequence to be inserted into the PA1 poly(A) site, to give the psPA/TAR vector (Fig. 2B). A second construct, termed psPA59, was derived by insertion into the PA1 site of a 59-nucleotide

**Figure 1.** Experimental strategy. (Top) Structure of the psPA vector. The psPA vector contains a 5' synthetic poly(A) site separated by 204 nucleotides from a natural 3' poly(A) site derived from the genomic rat preproinsulin II gene. The psPA vector also contains the highly active cytomegalovirus immediate early (CMV-IE) promoter. Transcribed sequences located between the CMV-IE promoter and the synthetic poly(A) site are primarily derived from the U3 region of the HIV-1 LTR. The *Ava*I (A) and *Bsp*MII (B) sites used in the preparation of S1 probes are located 217 and 203 nucleotides 5' to the predicted PA1 poly(A) site, respectively. (Bottom) Sequence of the synthetic poly(A) site introduced into the psPA vector. The AAUAAA and G/U box signals are boxed, and relevant restriction enzyme sites are underlined. Test sequences were inserted at the indicated *Tth*III site. The predicted poly(A) cleavage site is also indicated. The pΔPA vector is identical to psPA except for the indicated mutation of the AAUAAA signal sequence.





**Figure 2.** Effect of introduced RNA secondary structure on poly(A) site selection. (A) The indicated constructs were introduced into COS cells by transfection (Cullen 1987). Total cytoplasmic RNA was harvested at 70 hr post-transfection (Malim et al. 1988), and the utilization of the various available poly(A) sites was quantitated by S1 nuclease protection analysis (Böhnlein et al. 1989). Each of the indicated constructs differed in primary sequence, thus necessitating the preparation of individual matched end-labeled S1 nuclease protection probes. Therefore, a control S1 nuclease analysis (odd lanes) was run for each probe using cytoplasmic RNA derived from cells transfected with the negative control vector pBC12/CMV. All probes used were end labeled at *Ava*I (see Fig. 1B). Predicted sizes of the input (I) and rescued probe

fragments are indicated and are as follows: (Lanes 1–4) Input probe was 732 nucleotides in length; RNA polyadenylated at the PA2 site is predicted to rescue a 422-nucleotide fragment; RNA polyadenylated at PA1 should rescue a 217-nucleotide fragment. (Lanes 5–10) Input probe, 791 nucleotides; PA2 site usage, 481-nucleotide probe fragment; P1' site usage, 276-nucleotide probe fragment; P1 site usage, 217-nucleotide probe fragment. (Lanes 11 and 12) Input probe, 769 nucleotides; PA2 site usage, 459-nucleotide probe fragment; P1' site usage, 276-nucleotide probe fragment; P1 site usage, 217-nucleotide probe fragment. Size markers (left) were obtained by end-labeling *Msp*I-digested pBR322 DNA. (B) Primary RNA sequence and predicted secondary structures of the sequences introduced into the synthetic poly(A) site of the psPA and pΔPA vectors (see Fig. 1B). Inserted sequences are in uppercase letters; vector sequences are in lowercase letters. The predicted site of mRNA 3'-end formation in the psPA/TAR vector (P1') is indicated. This site is identical to that used in the parental psPA vector (Fig. 1) but is shifted 59 nucleotides 3' because of the inserted mRNA stem-loop structure. (C) Analysis of poly(A) site selection in transcripts derived from the indicated vectors was performed as described in Fig. 2A. Even-numbered lanes are experimental RNA samples; odd-numbered lanes represent control RNA samples. The probes used were end-labeled at *Bsp*MII (see Fig. 1B). Predicted sizes of the input (I) and rescued probe fragments are indicated and are as follows: (Lanes 1 and 2) Input probe, 586 nucleotides; PA2 site usage, 467-nucleotide probe fragment; P1' (TAR) site usage, 262-nucleotide probe fragment; P1 site usage, 203-nucleotide probe fragment. (Lanes 3–8) Input probe, 570 nucleotides; PA2 site usage, 451 nucleotides; P1' (IRE) site usage, 246 nucleotides; P1 site usage, 203 nucleotides. Size markers (left) were obtained by end-labeling *Msp*I-digested pBR322 DNA.

sequence predicted to lack secondary structure (Fig. 2B). Three possible effects were predicted to result from the insertion of these sequences into the functional PA1 poly(A) site: (1) The PA1 site could be rendered nonfunctional, thereby activating cleavage at the 3' PA2 poly(A) site. (2) Polyadenylation might continue to occur ~15 nucleotides 3' to the PA1 AAUAAA signal and within the introduced RNA sequence (cleavage site P1). (3) The

site of polyadenylation might be shifted 3' by the length of the inserted RNA sequence, that is, by ~59 nucleotides (cleavage site P1'; Fig. 2B).

Analysis of the pattern of *in vivo* mRNA 3'-end formation in transcripts derived from the psPA/TAR vector revealed that polyadenylation occurred predominantly ~74 nucleotides 3' to the AAUAAA signal at the P1' site, that is, well beyond the normally predicted distance

(Fig. 2A, lane 6). In fact, no polyadenylation at the normal distance from the AAUAAA motif, at the predicted P1 cleavage site, was detected in this assay. However, ~40% of the transcripts derived from the psPA/TAR vector were found to be polyadenylated at the 3' PA2 poly(A) site (Fig. 2A, lane 6). As the PA2 site was not used in the parental psPA vector (Fig. 2A, lane 4), this suggests that insertion of the TAR RNA stem-loop did somewhat inhibit functional recognition of the PA1 poly(A) signal. However, the insertion of a nonstructured sequence of identical size, in psPA59, resulted in a far more drastic effect on PA1 site usage in vivo. The small amount of 3'-end formation detected at the psPA59 PA1 poly(A) signal, ~5% of normal, was exclusively at the predicted P1 cleavage site, ~15 nucleotides 3' to the AAUAAA motif and within the inserted nucleotide sequence (Fig. 2A, lane 8). Mutation of the AAUAAA poly(A) signal of psPA/TAR, in  $\Delta$ psPA/TAR, or deletion of the G/U box, in psPA/TAR $\Delta$ GU, resulted in the complete loss of polyadenylation at the PA1 site of psPA/TAR (Fig. 2A, lanes 10 and 12). It is therefore evident that utilization of the P1' poly(A) site in psPA/TAR is dependent on the functional recognition of both the AAUAAA and G/U box sequences and is not signaled by the TAR sequence itself.

*The size of the introduced RNA stem-loop determines the site of polyadenylation*

The results presented in Figure 2A confirm the prediction that insertion of a random, 59-nucleotide sequence between the AAUAAA and G/U box signals would markedly inhibit the in vivo utilization of the PA1 poly(A) site. However, a 59-nucleotide sequence predicted to form an RNA stem-loop only modestly reduced the in vivo recognition of the PA1 poly(A) site and instead shifted the site of polyadenylation ~59 nucleotides 3' of the site predicted by the location of the AAUAAA signal. Therefore, we then asked whether other structured RNA sequence elements could also shift the site of polyadenylation by a distance equal to the size of the inserted RNA stem-loop. The second sequence chosen for insertion, the iron responsive element (IRE) derived from the transferrin receptor gene (Rouault et al. 1989), is predicted to form a 43-nucleotide RNA stem-loop structure (Fig. 2B). The IRE sequence was inserted into the PA1 poly(A) site of psPA in both the sense orientation (psPA/IRE) and in an inverted orientation (psPA/ERI). A 43-nucleotide sequence predicted to lack structure was again inserted into the psPA vector as a control (psPA43) (Fig. 2B). Both psPA/IRE and psPA/ERI gave rise primarily to transcripts polyadenylated at the predicted P1' (IRE) site, that is, 43 nucleotides 3' to the site predicted by the PA1 AAUAAA signal and ~16 nucleotides 5' to the poly(A) site [P1' (TAR)] used by transcripts derived from the psPA/TAR vector (Fig. 2C, lanes 2, 4, and 6). Although both psPA/IRE and psPA/ERI also gave rise to readily detectable levels of transcripts polyadenylated at the PA2 site (~30% of total), the PA2 site was used exclusively by transcripts derived from the

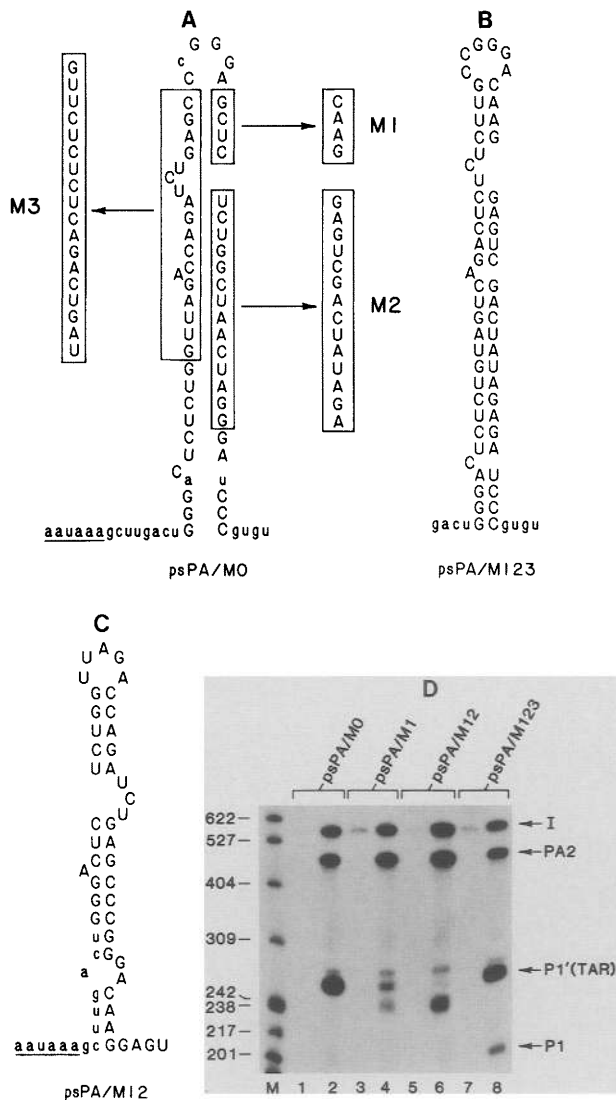
control psPA43 vector. We therefore conclude that insertion of the IRE RNA stem-loop resulted in a shift in the site of polyadenylation within the PA1 sequence and note that the size of this shift was predicted by the size of the inserted IRE RNA stem-loop.

*Mutations that destabilize an introduced RNA stem-loop structure inhibit poly(A) site recognition*

The results presented in Figure 2 demonstrate that the utilization of a poly(A) site that bears a sequence insertion between the AAUAAA and G/U box signals is dependent on the formation of an RNA stem-loop structure that correctly approximates these two RNA sequence elements. It is therefore predicted that polyadenylation would be reduced by mutations that destabilize an introduced RNA stem-loop structure and could then be reactivated by compensatory mutations. The parental plasmid used to test this hypothesis, termed psPA/MO, is essentially identical to psPA/TAR (Fig. 3A). The psPA/M1 construct was derived from psPA/MO by the introduction of a 4-nucleotide mutation near the apex of the predicted 59-nucleotide RNA stem-loop. The psPA/M12 vector bears an additional 14-nucleotide mutation predicted to essentially fully destabilize this stem-loop structure (Fig. 3A). Finally, psPA/M123 contains a compensatory 18-nucleotide mutation that is designed to again permit the formation of an alternate 59-nucleotide RNA stem-loop structure (Fig. 3B). All four vectors were transfected into COS cells, and the pattern of RNA 3'-end formation was analyzed by S1 nuclease protection analysis (Fig. 3D).

The psPA/MO construct, as with the essentially identical psPA/TAR vector, gave rise to transcripts able to rescue probe fragments consistent with predominant polyadenylation (~70%) at the predicted P1' (TAR) site and significant polyadenylation (~30%) at the PA2 site (Fig. 3D, lane 2). Although no cleavage was detected at the predicted P1 site in RNA derived from this vector, we did observe a minor probe fragment that was ~10 nucleotides larger than the fragment rescued by RNA polyadenylated at the P1' (TAR) site. Similar levels of this minor band were detected after S1 nuclease analysis of RNA samples derived from cultures transfected with any vector based on the psPA/MO construct (Fig. 3D). We therefore believe that this band is likely to be non-specific.

Introduction of the M1 mutation into the 59-nucleotide TAR-like structure present in psPA/MO had multiple effects. The first was a marked reduction in the level of polyadenylation detected at the predicted P1' (TAR) site (Fig. 3D, lane 4); the second was the appearance of a novel, minor rescued probe band ~13 nucleotides smaller than the fragment rescued by the P1' (TAR) cleavage; the third was a slight increase in the relative utilization of the PA2 poly(A) site (Fig. 3D, lane 4). Introduction of the extensive M12 mutation into the predicted TAR-like RNA stem-loop present in psPA/MO enhanced all three of these effects. S1 nuclease analysis of transcripts derived from cells transfected with



**Figure 3.** Mutations that destabilize introduced RNA structure affect poly(A) site selection. (A) Primary sequence of the 59-nucleotide stem-loop structure introduced into the synthetic poly(A) site present in the psPA/MO vector. The sequences of the cumulatively introduced M1, M12, and M123 mutations are also indicated. The psPA/MO construct is identical to the psPA/TAR vector except for three nucleotide changes within the TAR-like RNA stem-loop, indicated by lowercase letters. Flanking sequences derived from the psPA vector, including the AAUAAA signal, are also in lowercase letters. (B and C) Predicted secondary structures formed at the PA1 poly(A) site after insertion of the M123 or M12 mutations into psPA/MO. The predicted RNA secondary structure formed in the psPA/M12 construct not only involves inserted sequences (uppercase letters) but also flanking sequences drawn from the vector (lowercase letters). (D) The sites of 3'-end formation in transcripts derived from the psPA/MO, M1, M12, and M123 vectors were determined using S1 nuclease protection analysis, as described in Fig. 2A. Even-numbered lanes represent experimental RNA samples; odd-numbered lanes represent negative control RNA samples. The probes used were end-labeled at *Bsp*MII (see Fig. 1B). Predicted sizes of the input (I) and rescued probe are indicated and are as follows: Input probe, 586 nucleotides; PA2 site usage, 467-nucleotide probe fragment; P1' (TAR) site usage, 262-nucleotide probe fragment; P1 site usage, 203-nucleotide probe fragment. Size markers (lane M) were obtained by end-labeling *Msp*I-digested pBR322 DNA.

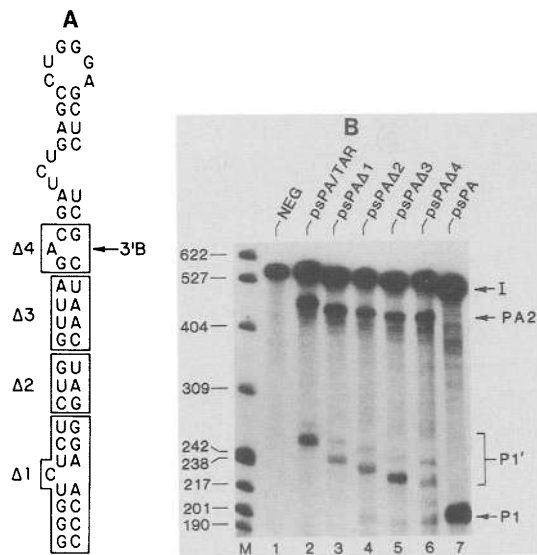
psPA/M12 revealed that utilization of the P1' (TAR) site was essentially lost (Fig. 3D, lane 6). In contrast, the level of a smaller rescued probe fragment, apparently similar to the fragment rescued at low level by transcripts derived from the psPA/M1 vector, increased significantly (Fig. 3D, cf. lanes 4 and 6). Finally, relative utilization of the invariant PA2 poly(A) site was again enhanced so that ~75% of the transcripts derived from the psPA/M12 vector appeared to be polyadenylated at this site.

The final construct, psPA/M123, bears an additional 18-nucleotide mutation, designed to compensate for the introduced M12 mutation and thereby restore the formation of a 59-nucleotide RNA stem-loop structure within the PA1 poly(A) site. As predicted, transcripts derived from this latter construct gave a pattern of probe rescue that was very similar to that observed with the initial psPA/MO vector (Fig. 3D, cf. lanes 2 and 8). In particular, we observed restoration of efficient polyadenylation at the site, here termed P1' (TAR), predicted by the inser-

tion of a 59-nucleotide RNA stem-loop structure in the PA1 poly(A) site. Relative utilization of the PA2 poly(A) site was also reduced to the level observed with psPA/MO. However, one difference between transcripts derived from the psPA/MO and psPA/M123 vectors was the detection, in the latter case, of a low level of RNA polyadenylated ~15 nucleotides 3' to the PA1 AAUAAA signal at the predicted P1 cleavage site. The significance of this latter observation, as well as the possible origin of the novel probe fragment rescued by transcripts derived from the psPA/M1 and psPA/M12 vectors, will be discussed below.

In a final experiment, we examined the effect that increasingly large stem truncations would have on the ability of an inserted stem-loop to influence poly(A) site selection. These nested deletions, shown schematically in Figure 4A, were designed to shorten the TAR RNA stem-loop past the point (indicated by 3'B) at which significant loss of TAR structural integrity is believed to occur based on phenotypic analysis *in vivo* (Hauber and Cullen 1988; Marciniak et al. 1990). These results, presented in Figure 4B, demonstrate that none of the introduced truncations are sufficient to fully prevent TAR stem-loop formation in this RNA context. Notably, each predicted TAR stem-loop derivative was observed to give rise to a readily detectable level of polyadenylation at the P1' site predicted by the increasingly shorter introduced RNA stem (Fig. 4B). However, the  $\Delta 4$  mutation did result in reduced utilization (about threefold) of the P1' site and in a modest activation of the P1 site present in the psPA $\Delta 4$  vector (Fig. 4B, cf. lanes 5 and 6). It therefore appears likely that the residual TAR RNA stem-loop present in transcripts derived from the

Brown et al.



**Figure 4.** Effect of stem truncations on the ability of an inserted TAR RNA stem-loop to affect poly(A) site selection. (A) The parental construct, psPA/TAR, contains the full-length 59-nucleotide TAR element (Fig. 2B). The psPAΔ1–psPAΔ4 plasmids contain a nested set of TAR deletions in which this RNA stem-loop is progressively reduced in size because of the loss of the indicated, boxed nucleotides, that is, psPAΔ2 lacks the nucleotides indicated in both Δ1 and Δ2 boxes. (3'B) The 3' border of the TAR element, as determined by mutational analysis in vivo (Hauber and Cullen 1988; Marciniak et al. 1990). (B) Poly(A) site utilization in transcripts derived from the indicated plasmids was quantitated by S1 nuclease analysis, as described in Fig. 2A. The probes used were end-labeled at *Bsp*MII [Fig. 1B]. Predicted sizes of the input (I) and rescued probe fragments are indicated and are as follows: (Lanes 1 and 2) I, 586 nucleotides; PA2, 467 nucleotides; P1', 262 nucleotides. (Lane 3) I, 571 nucleotides; PA2, 452 nucleotides; P1', 247 nucleotides. (Lane 4) I, 565 nucleotides; PA2, 446 nucleotides; P1', 241 nucleotides. (Lane 5) I, 557 nucleotides; PA2, 438 nucleotides; P1', 233 nucleotides. (Lane 6) I, 552 nucleotides; PA2, 433 nucleotides; P1', 228 nucleotides. (Lane 7) I, 531 nucleotides; PA2, 412 nucleotides. RNAs polyadenylated at the P1 site are predicted to rescue a 203-nucleotide probe fragment in every lane. (Neg) Negative control culture.

psPAΔ4 plasmid forms with significantly reduced efficiency.

## Discussion

A functional poly(A) site normally consists of two RNA sequence elements, the AAUAAA and G/U box signals that closely flank the site of mRNA 3'-end formation (Birnstiel et al. 1985; Proudfoot 1991). Formation of a polyadenylation complex is believed to involve the cooperative binding of cellular factors to both these RNA target sequences (Gilmartin and Nevins 1989; Takagaki et al. 1989). Random sequence insertions that increase the distance between the AAUAAA and G/U box signals to >40 nucleotides have been reported to inhibit effectively the functional recognition of poly(A) sites, pre-

sumably by inhibiting this cooperative binding event (McDevitt et al. 1986; Gil and Proudfoot 1987). However, in the case of HTLV-I, polyadenylation occurs ~250 nucleotides 3' to the only available AAUAAA signal (Seiki et al. 1983; Hanly et al. 1989). To explain this unique phenomenon, it was proposed that these intervening RNA sequences might fold into an RNA stem-loop structure that positions the AAUAAA and G/U box signals appropriately for polyadenylation factor binding (Ahmed et al. 1990; Toyoshima et al. 1990), and this hypothesis has recently been confirmed (Ahmed et al. 1991). In this discussion we have examined whether the apparent tolerance of this viral poly(A) site for large inserted RNA secondary structures represents a highly selected peculiarity of this particular virus or whether this tolerance could instead be generalized to other, unrelated poly(A) sites and RNA stem-loop structures.

Although the insertion of nonstructured sequences between the AAUAAA and G/U box signals was observed to inhibit dramatically poly(A) site utilization, sequence insertions of the same size that were predicted to form stable RNA stem-loop structures were found to have little effect on the efficiency of polyadenylation. Instead, and as predicted from the precedent set by polyadenylation in HTLV-I, such structured sequence insertions were observed to shift the site of polyadenylation 3' by a distance equal to the length of the inserted RNA stem-loop structure (Figs. 2C and 4B). The biological recognition of a poly(A) site bearing an inserted RNA stem-loop structure was reduced by the introduction of mutations that destabilized the inserted RNA secondary structure but was restored by mutations that were compensatory (Fig. 3D). These results, in total, argue that poly(A) sites are capable of tolerating sequence insertions if the inserted sequences are able to form RNA secondary structures that correctly approximate the AAUAAA and G/U box elements. These results therefore support the hypothesis (Gilmartin and Nevins 1989) that formation of a polyadenylation complex involves the spatially constrained cooperative interaction of protein factors with the AAUAAA and G/U box signals.

### *Polyadenylation as an assay for RNA secondary structure*

The results presented in this discussion demonstrate that poly(A) sites are tolerant of RNA sequence insertions between the AAUAAA and G/U box signals, but only if these sequences are able to form an RNA stem-loop that correctly restores the spacing of these two signals. Therefore, it appears possible that polyadenylation could be used as a novel way of assaying the ability of a particular sequence to form an RNA stem-loop structure in vivo. In addition, the polyadenylation assay described here might also be used to confirm the approximate size of an RNA stem-loop structure. This conclusion derives from the observation that the inserted RNA stem-loops shift the site of polyadenylation 3' of the normal site by a distance equal to the size of these predicted RNA structures (Figs. 2C and 4B).

In all of the cases examined in this discussion, insertion of a structured RNA sequence into the experimental PA1 poly(A) site resulted in at least some activation of the 3' PA2 site. Two explanations exist for this observation. The first is that the predicted RNA stem-loops form efficiently but that these structures then reduce the ability of cellular factors to recognize the PA1 poly(A) site. The second explanation is that some transcripts fail to form the predicted RNA stem-loop and are thus structurally equivalent to RNAs bearing random sequence insertions into the PA1 poly(A) site. In the latter case, it is predicted that polyadenylation would occur at the 3' PA2 site or, perhaps, if the sequence insertion was G/U rich, at the appropriate distance 3' to the PA1 AAUAAA signal within the inserted RNA sequence (e.g., psPA59; Fig. 2A, lane 8). In the case of the psPA/M123 vector, we detected a modest level of polyadenylation within the introduced sequence at a site that is predicted to form part of an RNA stem (Fig. 3D, lane 8). This latter result therefore suggests that a significant percentage of the transcripts derived from the psPA/M123 vector was probably polyadenylated before the formation of the predicted RNA stem-loop structure (Fig. 3B).

If polyadenylation is to have general utility as an *in vivo* assay for RNA secondary structure, it must not only be able to confirm but to predict the existence of RNA stem-loop structures. We believe the results presented in Figure 3D may have at least partly achieved this aim. The purpose of this experiment was to show that utilization of the site of polyadenylation predicted by the inserted 59-nucleotide TAR-like stem-loop structure present in the psPA/MO vector would be inhibited by a mutation (M12) that destabilizes this structure and restored by a compensatory mutation (M123). This was precisely the observed result. However, the M12 mutation also led to the utilization of a novel site of mRNA 3'-end formation located ~13 nucleotides 5' to the P1' (TAR) site (Fig. 3D, lanes 4 and 6). As this site is still ~60 nucleotides 3' to the AAUAAA signal, that is, beyond the normally permitted distance, we concluded that this was likely to represent a novel poly(A) site activated by an RNA stem-loop structure of ~46 nucleotides in size. Computer analysis of the psPA/M12 poly(A) site predicted the existence of a stable 46-nucleotide RNA secondary structure involving the inserted M12 sequences, as well as flanking nucleotides derived from the PA1 poly(A) site itself (Fig. 3C). The results presented in Figure 3D, therefore, were able to predict the existence of an RNA stem-loop structure that was not intended by the experimental design.

A number of questions will need to be addressed before a polyadenylation-based assay for RNA secondary structure can be considered of general value. In particular, these must include the consideration of sequence context effects and the question of whether RNA structures more complex than simple RNA stem-loops can be accommodated within a functional poly(A) site. However, the results presented in this discussion do suggest that this assay approach may eventually have significant use-

fulness in testing the *in vivo* existence of computer-predicted RNA stem-loop structures.

## Materials and methods

### *Construction of molecular clones*

The psPA and pΔPA vectors were both derived from the previously described pU3R+1 construct (Böhnlein et al. 1989). pU3R+1 contains the CMV-IE promoter 5' to sequences derived from the 3' end of the HIV-1 genome that primarily consists of the HIV-1 LTR U3 and R regions. The pU3R+1 vector also contains additional 3' sequences derived from the rat preproinsulin II gene that include the genomic rat insulin gene poly(A) site (Lomedico et al. 1979). The HIV-1 LTR poly(A) signal was removed from the pU3R+1 vector by excision of a 61-bp *SacI*-*XbaI* fragment. The synthetic poly(A) site, described in Figure 1, or a negative mutant thereof, was then inserted between the *SacI* and *XbaI* sites to give the psPA and pΔPA vectors, respectively. Synthetic oligonucleotides, as described in the text, were introduced into the *HindIII*, *TthIII*, and/or *XbaI* sites of the psPA or pΔPA vector using standard techniques. The psPA/TARΔGU plasmid is identical to psPA/TAR except that the G/U sequence, boxed in Figure 1, has been excised.

The CMV-IE promoter cap site in the psPA vector is located 951 nucleotides 5' to the predicted cleavage site present within the synthetic PA1 poly(A) site. These transcribed sequences are derived from the CMV-IE leader (79 nucleotides), the 3' end of the HIV-1 genome (845 nucleotides), and the synthetic poly(A) site itself (27 nucleotides).

### *Cell culture and transfection*

COS cells were maintained as described previously (Cullen 1986). COS cell cultures (100-mm dishes) were transfected using DEAE-dextran and chloroquin (Cullen 1987).

### *S1 nuclease analysis*

Total cytoplasmic RNA was harvested (Malim et al. 1988) at ~70 hr after transfection, and equal aliquots were used for quantitation of S1 nuclease-protected probe fragments (Cullen 1986; Böhnlein et al. 1989). The probe strategy used to quantitate the relative level of 3'-end formation at each of the two poly(A) sites present in the psPA vector series has been described previously (Böhnlein et al. 1989). Double-stranded DNA probes were end labeled, using Klenow DNA polymerase, at either an *AvrI* site or at an adjacent *BspMII* site located within the HIV-1 LTR U3 region 5' to the PA1 poly(A) site (Fig. 1B). The unlabeled end of the probe was generated by subsequent cleavage with either *SspI* (Fig. 2A) or *EcoRI* (Figs. 2C and 3D).

### *RNA secondary structure analysis*

The predicted RNA secondary structures of the HIV-1 TAR and transferrin receptor IRE sequences are taken from the literature and have been validated experimentally (Rouault et al. 1989; Marciniak et al. 1990). The RNA secondary structures formed by the sequence insertions present in the psPA/M12 and psPA/M123 constructs were predicted, using the University of Wisconsin Fold program (Devereux et al. 1984) run on a Vax computer.

Brown et al.

## Acknowledgments

The authors thank Sharon Goodwin for excellent secretarial skills.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## References

- Ahmed, Y.F., S.M. Hanly, M.H. Malim, B.R. Cullen, and W.C. Greene. 1990. Structure-function analyses of the HTLV-I Rex and HIV-1 Rev RNA response elements: Insights into the mechanism of Rex and Rev action. *Genes & Dev.* **4**:1014–1022.
- Ahmed, Y.F., G.M. Gilmartin, S.M. Hanly, J.R. Nevins, and W.C. Greene. 1991. The HTLV-I Rex response element mediates a novel form of mRNA polyadenylation. *Cell* **64**: 727–737.
- Birnstiel, M.L., M. Busslinger, and K. Strub. 1985. Transcription termination and 3' processing: The end is in site. *Cell* **4**: 349–359.
- Böhnlein, 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.
- Cullen, B.R. 1986. *Trans*-activation of human immunodeficiency virus occurs via a bimodal mechanism. *Cell* **46**: 973–982.
- . 1987. Use of eukaryotic expression technology in the functional analysis of cloned genes. *Methods Enzymol.* **152**: 684–704.
- . 1990. The HIV-1 Tat protein: An RNA sequence-specific processivity factor? *Cell* **63**: 655–657.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387–395.
- Fitzgerald, M. and T. Shenk. 1981. The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* **24**: 251–260.
- Gil, A. and N.J. Proudfoot. 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit  $\beta$ -globin mRNA 3' end formation. *Cell* **49**: 399–406.
- 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.
- Hanly, S.M., L.T. Rimsky, M.H. Malim, J.H. Kim, J. Hauber, M. Duc Dodon, S.-Y. Le, J.V. Maizel, B.R. Cullen, and W.C. Greene. 1989. Comparative analysis of the HTLV-I Rex and HIV-1 Rev trans-regulatory proteins and their RNA response elements. *Genes & Dev.* **3**: 1534–1544.
- Hauber, J. and B.R. Cullen. 1988. Mutational analysis of the *trans*-activation-responsive region of the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.* **62**: 673–679.
- Jackson, R.J. and N. Standart. 1990. Do the poly(A) tail and 3' untranslated region control mRNA translation? *Cell* **62**: 15–24.
- Levitt, N., D. Briggs, A. Gil, and N.J. Proudfoot. 1989. Definition of an efficient synthetic poly(A) site. *Genes & Dev.* **3**: 1019–1025.
- Lomedico, P., N. Rosenthal, A. Efstratiadis, W. Gilbert, R. Kolodner, and R. Tizard. 1979. The structure and evolution of the two nonallelic rat preproinsulin genes. *Cell* **8**: 545–558.
- Malim, M.H., J. Hauber, R. Fenrick, and B.R. Cullen. 1988. Immunodeficiency virus *rev* *trans*-activator modulates the expression of the viral regulatory genes. *Nature* **335**: 181–183.
- Marciniak, R.A., M.A. Garcia-Blanco, and P.A. Sharp. 1990. Identification and characterization of a HeLa nuclear protein that specifically binds to the *trans*-activation-response (TAR) element of human immunodeficiency virus. *Proc. Natl. Acad. Sci.* **87**: 3624–3628.
- McDevitt, M.A., R.P. Hart, W.W. Wong, and J.R. Nevins. 1986. Sequences capable of restoring poly(A) site function define two distinct downstream elements. *EMBO J.* **5**:2907–2913.
- Proudfoot, N. 1991. Poly(A) signals. *Cell* **64**: 671–674.
- Proudfoot, N.J. and G.G. Brownlee. 1976. The 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**: 211–214.
- Rouault, T.A., M.W. Hentze, D.J. Haile, J.B. Harford, and R.D. Klausner. 1989. The iron-responsive element binding protein: A method for the affinity purification of a regulatory RNA-binding protein. *Proc. Natl. Acad. Sci.* **86**: 5768–5772.
- Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida. 1983. Human adult T-cell leukemia virus: Complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc. Natl. Acad. Sci.* **80**: 3618–3622.
- Takagaki, Y., L.C. Ryner, and J.L. Manley. 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes & Dev.* **3**: 1711–1724.
- 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.
- Wickens, M. 1990. How the messenger got its tail: Addition of poly(A) in the nucleus. *TIBS* **15**: 277–281.





## Effect of RNA secondary structure on polyadenylation site selection.

P H Brown, L S Tiley and B R Cullen

*Genes Dev.* 1991, **5**:

Access the most recent version at doi:[10.1101/gad.5.7.1277](https://doi.org/10.1101/gad.5.7.1277)

---

### References

This article cites 26 articles, 11 of which can be accessed free at:  
<http://genesdev.cshlp.org/content/5/7/1277.full.html#ref-list-1>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

An advertisement banner for Dharmacon Reagents and Horizon. On the left, it says 'Dharmacon Reagents' with the tagline 'Custom synthesis, RNAi, and CRISPR solutions'. In the center, the text 'Infinite Reliability' is displayed in a large, white, sans-serif font. To the right, the 'horizon' logo is shown in a white, lowercase, sans-serif font, with 'a PerkinElmer company' written in smaller text below it. A 'More' button is visible in the bottom right corner of the banner. The background features a colorful, abstract image of what appears to be a DNA double helix or a similar molecular structure.