Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses

Roland Russnak¹ and Don Ganem¹⁻³

Departments of ¹Microbiology and Immunology and ²Medicine, University of California Medical Center, San Francisco, California 94143 USA

Most genetic elements that employ reverse transcription generate a terminally redundant genomic RNA that serves as the template for this reaction. Because the identical polyadenylation signal is present in each terminally redundant segment, synthesis of this RNA requires that this signal be ignored on the first pass of the transcription machinery, then recognized and used on the second pass. We have studied the mechanism of this differential poly(A) site use in one family of retroviral elements, the hepatitis B viruses (hepadnaviruses). Our results indicate that two features are involved: the presence of a variant poly(A) signal (TATAAA) and the participation of multiple sequences 5' to this signal that act to increase the efficiency of its use. Deletion of these upstream elements abolishes proper poly(A) site use, despite the presence of the poly(A) signal and downstream GT- and T-rich motifs known to be required for polyadenylation. Sequences from the corresponding regions of retroviral genomes can restore proper processing to these hepadnaviral deletion mutants. Thus, functionally analogous upstream elements exist in other classes of retroviral elements, including those employing the canonical AATAAA hexanucleotide signal.

[**Key Words:** Reverse transcription; hepatitis B viruses; polyadenylation]

Received January 8, 1990; revised version accepted February 27, 1990.

Eukaryotic transcripts are subject to a variety of important post-transcriptional processing reactions on the pathway to mature mRNA, including capping, splicing, 3' cleavage, and poly(A) addition. Proper 3'-end formation is particularly important to assure the stability and function of most mRNAs. In recent years, great progress has been made in understanding the mechanism of polyadenylation. For most genes, a relatively simple set of sequences has been implicated as essential *cis*-acting elements (for review, see Friedman et al. 1987). *Cis* requirements were identified first through sequence inspection of a variety of eukaryotic mRNA sequences (Proudfoot and Brownlee 1976) and their importance confirmed subsequently by mutational analyses *in vivo*. These include the hexanucleotide AATAAA, located ~10–20 nucleotides 5' to the poly(A) addition site (Fitzgerald and Shenk 1981; Montell et al. 1983; Wickens and Stephenson 1984), and more degenerate T-rich or GT-rich sequences located 3' to this site (Simonsen and Levinson 1983; Gil and Proudfoot 1984, 1987; McDevitt et al. 1984, 1986; Woychik et al. 1984; Bhat and Wold 1985; Conway and Wickens 1985; Hart et al. 1985; McLauchlan et al. 1985; Sadofsky et al. 1985; Stacy and Cole 1985; Kessler et al. 1986; Mason et al. 1986; Zhang et al. 1986; Bohnlein et al. 1989). In the case of the rabbit

β -globin gene, a synthetic oligonucleotide composed only of these elements has been shown to be sufficient to confer correct 3'-end processing on a heterologous transcript (Levitt et al. 1989). The development of *in vitro* systems that carry out correct polyadenylation of precursor RNAs (Manley 1983; Moore and Sharp 1984, 1985) has also allowed the identification of at least two classes of *trans*-acting factors required for the reaction, including cleavage/specificity factor(s) and poly(A) polymerase activity (Christofori and Keller 1988, 1989; Takagaki et al. 1988; Terns and Jacob 1989).

For most genes, polyadenylation is presumed to be unregulated; that is, the efficiency with which their poly(A) sites are used is not known to be controlled. However, recent findings indicate that for many other genes, 3'-end formation is subject to regulation. The adenovirus major late transcription unit is characterized by differential use of the poly(A) signals L1–L3 during late stage versus early stage of infection (Nevins and Wilson 1981; Falck-Pedersen and Logan 1989). Many important cellular genes [e.g., those for tropomyosin (Helfman et al. 1986) and for immunoglobulin M (IgM)] also contain multiple polyadenylation signals whose differential recognition leads to the production of multiple mRNAs that encode distinct protein isoforms. In the case of IgM, differential use of these signals controls the synthesis of secreted versus transmembrane forms of the

³Corresponding author.

protein, and this decision is strictly regulated during B-cell development (Alt et al. 1980; Early et al. 1980). However, investigation of the mechanism of differential polyadenylation in the IgM gene has been complicated by the concurrent occurrence of differential splicing, raising the issue of whether splice-site selection influences poly(A) site use (Danner and Leder 1985; Galli et al. 1988; Peterson and Perry 1989). In the calcitonin/CGRP gene, splice site commitment clearly plays the dominant role in differential polyadenylation (Leff et al. 1987). The additional variable of splicing has thus made it difficult to define the minimal sequence requirements for regulated poly(A) site use.

A potentially simpler example of differential polyadenylation occurs among the so-called retroid elements, genetic elements that replicate via reverse transcription of an RNA intermediate. These include retroviruses, hepatitis B viruses (hepadnaviruses), caulimoviruses and many families of transposable elements (e.g., yeast Ty and *Drosophila copia*). These elements all synthesize unspliced, terminally redundant RNA molecules of greater than unit length that serve as the template for reverse transcription (the terminal redundancy being required for strand transfer steps during DNA synthesis). To generate the terminal redundancy, identical genomic sequences must be transcribed twice. In all of the above cases, the doubly transcribed sequences include the signals for polyadenylation. Thus, to produce this essential RNA, these signals must be ignored on the first pass of the transcription machinery, then recognized and used on the second pass. We have examined the mechanism of this differential poly(A) site use in one family of retroid elements, the hepatitis B viruses. Our results indicate that in this system, multiple novel accessory elements located 5' to the previously described hexanucleotide are required for proper poly(A) site use, and that functionally analogous upstream signals exist among other retroid elements.

Results

The use of the hepadnaviral poly(A) signal is regulated

Figure 1A shows the genetic and transcriptional map of the 3.3-kb circular genome of the ground squirrel hepatitis virus (GSHV), the mammalian hepadnavirus used in this study. Infected cells contain two major classes of poly(A)⁺ RNA (Enders et al. 1985). Subgenomic transcripts of 1.9 kb encode the production of the major viral envelope glycoprotein (surface antigen, sAg), and are processed at the unique polyadenylation signal located within the gene for the core (or nucleocapsid) protein. Genomic transcripts (3.5 kb) are the terminally redundant RNAs that are the templates for reverse transcription; in addition, they also serve as mRNAs for the core protein and possibly for the viral reverse transcriptase (for review of the hepadnavirus life cycle, see Ganem and Varmus 1987). Because the 3.5-kb RNAs are reverse-transcribed into the DNA genome, they are also referred to as "pregenomic" RNAs.

As depicted in Figure 1A, pregenomic RNA synthesis initiates at several sites ~100 and 130 nucleotides upstream of the poly(A) signal in the core gene. Transcription proceeds through this region and around the genome until this signal is again encountered, at which point efficient 3'-end formation is carried out. This results in production of RNAs with terminal redundancies of 125 or 155 nucleotides. Importantly, inspection of the GSHV DNA sequence reveals that the hexanucleotide presumed to be part of the polyadenylation signal is TATAAA rather than the canonical AATAAA; downstream of the cleavage site are both T-rich and GT-rich stretches. Similar TATAAA and downstream elements exist in the human hepatitis B virus (HBV), and Simonsen and Levinson (1983) have shown previously that, as in other cases, the downstream sequences are required for accurate and efficient polyadenylation in HBV.

Although synthesis of pregenomic GSHV RNA requires that the 5'-proximal processing signals be bypassed but the 3' one recognized, this fact alone does not establish that poly(A) site use is regulated. Given that the viral hexanucleotide signal is a variant of the consensus sequence AATAAA, the possibility exists that the GSHV signals are simply inefficient—that is, that only a fraction of nascent transcripts containing these sequences is polyadenylated, and that this fraction is invariant on each pass of the transcription machinery past the signals. This model makes specific predictions about the structure of viral pregenomic RNAs that will be found in an infected cell. If this processing fraction is large, then cells should contain large quantities of very short transcripts corresponding to transcripts arising from the pregenomic promoter and being polyadenylation on the first pass. If, on the other hand, this fraction is low, very few such short transcripts should be present, but readthrough should be frequent even on the second pass of the site, leading to the accumulation of large quantities of dimer and trimer-length RNAs, as, for example, happens during polyomavirus transcription (Acheson 1984).

To explore this possibility, we examined viral RNA species in infected cells both by Northern blotting and by an RNase mapping procedure that uses uniformly labeled probes spanning the region of the processing signals and the genomic RNA start sites. As diagrammed in Figure 1B, the products that result from first-pass processing, first-pass readthrough/second-pass processing, and second-pass readthrough can all be distinguished by RNase mapping. Figure 2 shows the results of these studies.

In Figure 2A (lane 2) is a Northern blot of viral transcripts present in infected liver, demonstrating the pregenomic and subgenomic transcripts of GSHV and showing that dimer and trimer RNAs resulting from second-pass readthrough are rare (see below). In Figure 2B (lane 4) is the RNase protection analysis of the same infected liver RNA. First-pass readthrough of the two genomic RNA species gave rise to protected bands of 284 and 253 nucleotides that were readily detected. Prege-

Figure 1. Transcripts found in GSHV-infected liver and pGSpA.wt-transfected cells. (A) General organization of the circular GSHV genome showing the four open reading frames and the structures of the two major classes of poly(A)⁺ RNA, including the subgenomic (sAg-encoding, 1.9-kb) mRNAs and the terminally redundant 3.5-kb pregenomic RNAs. (B) Relationship of the multiple pregenomic RNA start sites to the single cleavage/polyadenylation site. First-pass processing, first-pass readthrough, and second-pass processing products can be distinguished by hybridization to the uniformly labeled RNA probe spanning sequences +154 to -172. Also shown are the sizes of the expected RNase-resistant fragments (wavy lines). Second-pass readthrough would result in complete protection of the probe (for transcript mapping results, see Fig. 2B). (C) A system for studying 3'-end processing using an SV40-based expression vector. GSHV sequences from -1800 to +835 are flanked upstream by the SV40 early promoter/enhancer and origin of replication and downstream by the SV40 late polyadenylation signal. Wavy lines depict predicted transcripts. Correct processing at the GSHV poly(A) site yields a 0.7-kb spliced mRNA. Defective processing at the GSHVpA site results in extended 1.6-kb transcripts that are processed at the downstream SV40pA signal. The liver-specific promoter elements responsible for the synthesis of pregenomic mRNA are inactive in COS-7 cells as indicated (arrows).

monic transcripts processed on the first encounter with the poly(A) signal would result in short transcripts of 153 and 122 nucleotides (excluding the added poly(A) tract) which in this assay would generate RNase-resistant fragments of the same sizes. Examination of this region of the gel revealed very little signal at these positions above the background of small RNA degradation products that is inevitable in assays of this type. Given the background, we estimate that no more than 10–20% of the total genomic RNA was processed on the first pass, and very probably much less. As expected from the Northern blot analysis, the major protected species was 195 nucleotides, corresponding to proper processing of subgenomic (sAg) mRNA on the first pass, and of genomic RNA on the second pass. Importantly, the subset of these RNAs that are overlength (e.g., dimers), whose signature fragment is the fully protected 330-nucleotide band, represented less than 5% of the total (as expected from the Northern blot analysis). Thus, transcription

initiating 1.9 kb or farther 5' to the GSHV poly(A) site results in mRNA which is very efficiently processed.

Figure 2B (lane 3) shows the results of a similar analysis of RNA from cultured permissive human hepatoma cells (HepG2) transfected with cloned GSHV DNA (pBA27.3; Seeger et al. 1984). The similarity in the patterns and intensities of protected species to those observed in infected liver (lane 4) indicates that the pattern of 3' processing in productively infected cells in culture is conserved. (Minor variation is observed in the degree of genomic second-pass and sAg mRNA readthrough, perhaps indicative of slightly more efficient processing in infected liver).

These results indicate that very little processing of RNAs initiated at the genomic promoter occurs on the first pass of the poly(A) signals (<10–20%), but that extremely efficient processing (>95%) occurs on the second pass. This shows that the GSHV poly(A) signal is not simply leaky, but that the efficiency of its use is

Figure 2. Northern (A) and RNase mapping (B) analysis of either total poly(A)⁺ RNA from GSHV-infected liver or cytoplasmic poly(A)⁺ RNA isolated from the indicated transfected cell lines. The probe used in these experiments is the SP6-derived RNA (+154 to –172) shown in Fig. 1B. Positions of the protected species predicted in Fig. 1B are indicated by arrows at right, together with notation of the transcripts from which they are derived. The 330-nucleotide protected species is just shorter than the undigested probe as a result of the removal of flanking polylinker sequences during RNase treatment. pBA27.3 contains three tandem copies of the full-length GSHV genome.

modulated during mRNA biogenesis. Because this conclusion is based in part on the failure to find appreciable quantities of the small transcripts resulting from the first-pass polyadenylation, an alternative interpretation would be that this RNA is simply too unstable to permit detection. As will be shown below, however, a single base change in the hexanucleotide allowed accumulation and ready detection of this RNA, strongly suggesting that instability does not account for these findings.

A system for analyzing GSHV 3'-end processing

The preceding results showed that when transcription is initiated at –1900 relative to the poly(A) signal, first-pass processing is nearly quantitative, whereas when it initiates at –100 or –130, first-pass readthrough is the rule rather than the exception. These observations are consistent with many different models for the mecha-

nism of regulation—for instance, that a minimal distance of greater than 130 nucleotides between the cap site and the poly(A) site is required for efficient processing. Another model would envision that specific sequences located between –100 and –1900 are required for efficient use of the viral poly(A) signals. To test these models, we designed a system to analyze the requirements for proper polyadenylation in GSHV.

Our strategy was to drive GSHV transcription with the SV40 early promoter, cloned at various positions upstream of the poly(A) signal. For example, clone pGSpA.wt (Fig. 1C) has this promoter positioned at –1805 relative to the poly(A) signal, thereby generating an analog of the subgenomic (1.9-kb) mRNA. (In this work, all nucleotide positions will be numbered relative to the first T of the TATAAA sequence as +1.) The presence of the SV40 ori region on this plasmid allows copy number amplification of COS cells, thereby augmenting the yield of RNA and greatly facilitating the analysis. To

allow quantitation of the processing efficiency at the GSHV poly(A) site, an SV40 late poly(A) signal was cloned into the GSHV genome 900 nucleotides 3' to the GSHV poly(A) signal. Thus, transcripts that fail to be processed at the hepadnaviral signal (which would ordinarily be lost to degradation) can be rescued by processing at the downstream SV40 signal. Thus (assuming equal stabilities of the two RNA species; see Discussion), the ratio of the RNAs processed at each site allows estimation of the GSHV processing efficiency.

Because in this system GSHV poly(A) site use is always assayed in the context of a heterologous poly(A) site, it is important to show that the normal pattern of GSHV poly(A) site selection is preserved in this context. The RNAs produced from pGSpA.wt in COS-7 cells are shown in lane 1 of Figure 2A. The 1.8-kb species represents the expected product of first-pass polyadenylation, and no RNA of 2.7 kb (corresponding to first-pass readthrough) was seen. Unexpectedly, however, a highly abundant additional transcript of ~700 nucleotides was observed. This species proved to be due to efficient RNA splicing leading to the removal of a 1.1-kb intron. Amplification of the region straddling the splice junction by PCR and subsequent sequencing (results not shown) placed the donor site at -1697 (between the codons for amino acids 14 and 15 of the major sAg protein) and the acceptor site at -593 (60 nucleotides 5' to the X gene initiation codon). Translation of this spliced message would result in a truncated form of surface antigen, a stop codon being introduced 17 amino acids downstream of the splice junction. [A similar splice from within sAg codon 102 to an acceptor located 70 nucleotides upstream of the X gene ATG has also been described for HBV in COS cells (Simonsen and Levinson 1983).] The removal of this intron in GSHV is not peculiar to COS cells or to transcription from the SV40 promoter, since the majority of subgenomic messages were similarly spliced in HepG2 cells transfected with pBA27.3 (Fig. 2A, lane 3).

The significance of this spliced RNA in the GSHV life cycle (if any) is unknown. However, relevant to the present work is that the 0.7-kb transcript was efficiently and accurately processed at the GSHV poly(A) site. Failure to process this spliced RNA at the GSHV site would result in readthrough transcripts of 1.6 kb. Northern analysis (Fig. 2A, lane 1) revealed only very low levels of these RNAs, and RNase protection experiments (Fig. 2B, lane 5) indicated that these readthrough RNAs represent less than 5% of the total. Thus, in this context, processing occurred on the first pass of the GSHVpA site at the same efficiency (95%) as in the analogous subgenomic RNA in infected liver *in vivo*. In COS cells, no activity of the pregenomic RNA promoter was demonstrable (see Fig. 2B, lane 5), a result that is expected given the known liver-specificity of this element (Ganem and Varmus 1987). However, as will be detailed later, when the SV40 promoter is substituted for the GSHV pregenomic promoters (at nucleotide -107), processing on the first pass by the GSHV poly(A) site drops to <10–20%, exactly as in pregenomic RNA synthesis

in the infected liver (see Fig. 4, clone pGSpA.1, and accompanying text). We conclude that GSHV processing signals are functional in COS cells and that the system faithfully reproduces the differential polyadenylation observed in authentic viral infection.

Basal processing elements are insufficient for proper poly(A) site use

We next asked whether the core or basal elements (i.e., hexanucleotide plus downstream GT- and/or T-rich sequences) implicated in polyadenylation by previous studies (Simonsen and Levinson 1983) are sufficient to mediate GSHV 3'-end formation. In pGSpA.21, GSHV sequences from -1800 to -43 were replaced with a heterologous coding region, in this case the cDNA encoding chicken c-src (Fig. 3A). This chimeric gene retained the GSHV polyadenylation signal, processing site, and 800 bp of downstream sequences; as before, an SV40 poly(A) site is cloned downstream. Northern analysis of RNA from COS cells transfected with this construct showed it to be severely defective in GSHV processing (Fig. 3B, lane 2), resulting in the accumulation of 2.6-kb readthrough transcripts processed at the downstream SV40 poly(A) site. To determine if this inefficient processing was due to the presence of the variant hexanucleotide, we constructed plasmid pGSpA.21/A, in which the variant TATAAA signal was mutated to the consensus AATAAA. Surprisingly, this mutation only modestly improved usage of the GSHVpA site (Fig. 3B, lane 3). This experiment indicates that the core GSHV processing signals are extremely inefficient, and that the variant hexanucleotide is only partially responsible for this defect. The poor level of GSHV processing observed in this experiment, in which the cap site is 1.7 kb from the processing site, renders unlikely the simple model that inefficient first-pass processing of genomic RNA is due solely to the short spacing between the 5' end and the processing signals. Taken together with the results using pGSpA.wt (Fig. 2), it suggests that additional virus-specific sequences 5' to -43 may be required for efficient use of these signals.

Upstream sequences are necessary for efficient GSHV poly(A) site use

To explore the potential role of upstream GSHV sequences in polyadenylation, we constructed a series of plasmids in which the SV40 promoter was positioned at varying distances upstream from the GSHV polyadenylation signal (Fig. 4A). As always, a downstream SV40 poly(A) signal was present in all constructs. Our goal was to determine the minimal amount of sequence required to allow use of the GSHV processing signals with maximum efficiency. In pGSpA.1, the promoter is situated at -107 and initiates RNA synthesis 180 bp 5' to the polyadenylation signal, the first 73 bp being SV40 sequences. The RNA products of this construct are analogous to the pregenomic RNA transcripts of an authentic GSHV infection. RNase protection experiments

Figure 3. Determination of the efficiency of processing by GSHV core elements located downstream of a heterologous gene. (A) Illustration of the chimeric *c-src*-GSHV genes. In this plasmid, 1.6 kb of *src* coding region is followed by 1.0 kb of GSHV DNA extending 3' from -43 relative to the TATAAA sequence. Processing at either the GSHVpA site or the downstream SV40pA site results in polyadenylated transcripts of 1.7 kb or 2.6 kb, respectively. (B) Northern analysis of cytoplasmic poly(A)⁺ RNA purified from COS-7 cells transfected with pGSpA.21 (lane 2) and 21/A (lane 3). Lane 1 shows pGSpA.wt RNA as a control.

(Fig. 4B, lane 4) revealed that (maximally) only 10–20% of these transcripts were processed at the proximal site. As discussed above, this accurately reflects the pattern of poly(A) site use in pregenomic RNAs in GSHV-infected liver.

Movement of the promoter in a 5' direction resulted in increasing levels of 3'-end formation at the GSHVpA site. The relevant constructs are diagrammed in Figure 4A, along with a tabular summary of the fraction of transcripts that escaped processing at the GSHV site; Figure 4B shows several representative RNase protection experiments that form the basis for this tabulation. As summarized in Figure 4A, in plasmids pGSpA.2 and pGSpA.3, in which the SV40 promoter was positioned at -172 and -247, respectively, intermediate levels of processing were observed. Not until the promoter was positioned at -397 (pGSpA.4) was wild-type processing efficiency at the GSHV site restored. Thus, in GSHV, from 250 to 400 nucleotides of upstream sequences are required for the efficient recognition of the basal processing signals.

Additional protected bands (indicated by dots) were observed in the RNase mapping experiment of Figure 4B. Their sizes are consistent with a splice occurring in the readthrough transcripts, the donor junction being located at +114/+115. Although direct evidence for a splice in this position has not been obtained, the size of the protected fragments indicates that they cannot result from processing at the GSHVpA site.

Multiple cis-acting elements affect poly(A) site use

To more precisely localize the sequences required for poly(A) site use in GSHV, deletion mapping was performed on pGSpA.wt (Fig. 5). In this plasmid, the SV40 promoter was positioned at -1800 and processing at the GSHVpA site occurred with >95% efficiency. Deletion

of ~60 bp 5' and 40 bp 3' to the *Cla*I site at -172 (pGSpA.5 and 6, respectively) had no detectable effect on processing. Unexpectedly, the combined deletion from -229 to -135 (pGSpA.11) resulted in 10–20% readthrough (Fig. 5A, lane 3). Perhaps the two halves of this region contain functionally redundant information, both components of which must be deleted to manifest a phenotype. We refer to this region as processing signal 1 (PS1) and note that it is characterized by extensive AT stretches (64% overall).

Further unidirectional (5' to 3') deletions from -172 to -8 have defined a second region (PS2) that affects polyadenylation. While deletions from -172 to -96 had no effect on processing, further deletions to -73 showed readthrough levels of 10–20%; deletions to -43 resulted in substantial (60–80%) readthrough. Further deletions to -8 did not show additional decrements in the use of the GSHV poly(A) site. This defines a second region (termed PS2) from -96 to -43 that is involved in processing. It is noteworthy that PS2 is contained within the 5' portion of the pregenomic RNA, yet this RNA is not efficiently processed on the first pass (see Discussion).

Removal of 185 bp containing both PS1 and PS2 in the context of subgenomic RNA (pGSpA.12) nearly abolished processing at the GSHVpA site (Fig. 5B, lane 6). A similar phenotype occurred when this deletion was crossed into pGSpA.4, the plasmid that contains the minimal GSHV information required for efficient processing (Fig. 4A; plasmid pGSpA.4/12). Although the phenotype of this deletion indicates that the involved sequences (PS1 plus PS2) are necessary for poly(A) site use, the data of Figure 4 indicate that they are not sufficient. Additional upstream information (to -397) is required, indicating that still more functional elements are located in this region and participate in regulation. For ease of reference, we provisionally designate this re-

Figure 4. Determination of the minimal upstream sequence requirements for efficient 3'-end formation in GSHV. (A) In these plasmids, the SV40 early promoter (shaded box) has been cloned 5' to varying amounts of GSHV sequences (open boxes) upstream of the GSHV poly(A) signal. Also indicated are the extent of the labeled RNA probe used for RNase mapping (dark lines) and the lengths of the expected RNase-resistant products (wavy lines). The fraction of RNA molecules that escape processing at the GSHV site and are processed at the downstream SV40pA site (referred to as readthrough) is indicated for each construct at right. pGSpA.4/12 contains the 185-bp deletion described in the mutant pGSpA.12 (Fig. 5A). (B) RNase mapping analysis of several of the constructions shown above. The protected bands corresponding to those shown in A are indicated by arrows. Note that the 107-nucleotide RNA from pGSpA.1/A represents processing at the GSHVpA site but is 23 nucleotides shorter than predicted because a 1-bp mismatch between the mutated nucleotide at +1 and the wild-type probe is sensitive to cleavage by RNase. As a result, readthrough transcripts would generate a 3'-protected species of 154 nucleotides. The origin of the protected products indicated by dots is discussed in the text.

gion PS3, while recognizing that it may well contain multiple active elements.

Some insight into the functional properties of the different signals can be derived by examining their abilities to act on poly(A) signals that contain the consensus hexanucleotide AATAAA versus the variant TATAAA. For example, in the context of genomic RNA, PS2 alone is insufficient to activate use of the TATAAA-containing signal (Fig. 4A, plasmid pGSpA.1). However, mutation of the hexanucleotide to AATAAA in this same plasmid (pGSpA.1/A, Fig. 4A) restores efficient processing. (The ability to readily detect the short transcript resulting from first-pass processing in this mutant also argues that instability of this RNA does not explain its absence from the wild-type parent.) The same is true of the elements in PS3: when the canonical hexanucleotide is restored to the PS1–PS2 deletion plasmid

pGSpA.12, GSHV processing returns to levels that are nearly normal (Fig. 5, plasmid pGSpA.12/A). Thus, GSHV subgenomic RNA contains multiple upstream elements that must be present together to activate its variant TATAAA signal efficiently; at least two of these, however, can activate a canonical AATAAA signal when present individually.

Retroviral U3 regions contain analogous functions

With pGSpA.12 we now possessed a processing-defective mutant that retained the core processing signals known to be required for 3'-end formation. It was then possible to test foreign DNA for activating *cis*-acting elements that are functionally analogous to those identified in GSHV. As a control, replacement of the deleted GSHV sequences with the same region of GSHV DNA in

Figure 5. Deletion analysis of the region upstream of the polyadenylation signal in pGSpA.wt. (A) Deleted sequences, originating at the *ClaI* site (–172), are represented by the black bars; end points are numbered. Open bars indicate the identified processing signals PS1 and PS2. The degree of defective processing at the GSHVpA site (readthrough) for each of the mutants is shown in the column on the right and is based on Northern analysis of cytoplasmic poly(A)⁺ RNA (B). (B) Northern analysis of cytoplasmic poly(A)⁺ RNA from COS-7 cells transfected with the indicated plasmids. Transcripts resulting from processing at either the GSHVpA site or the downstream SV40pA site are indicated.

the correct orientation (pGSpA.13) restored wild-type levels of processing as shown in lane 9 of Figure 6B. Interestingly, the opposite orientation of the same fragment re-established low but significant levels of proper site usage (see lane 10). Insertion of 194 bp of bacteriophage DNA in either orientation did not reconstitute processing at the GSHV site (lanes 3 and 4, Fig. 6B). This provides strong evidence for specific sequence requirements and rules out the existence of a signal upstream of –229 which is critically dependent on its distance from the polyadenylation site for function.

As noted earlier, retroviruses face the same dilemma as hepadnaviruses in generating their terminally redundant viral RNAs. In the case of retroviral proviral DNA, the U3 region would be positioned similarly to the PS1 and PS3 regions of GSHV on the transcriptional map. That is, these sequences are 5' to the cap site and thus are absent from the RNA at the time of the first pass of the poly(A) site, but are included in the transcript at the time of the second pass. Interestingly, Dougherty and Temin (1987) have noted previously that retroviral vectors lacking portions of U3 are replication defective. The fact that these viruses could be rescued by provision of an SV40 poly(A) signal led these workers to suggest that sequences within U3 may influence polyadenylation. For all of these reasons, the entire U3 regions of both spleen necrosis virus (SNV) and human immunodeficiency virus (HIV) were also inserted into pGSpA.12. In both cases, 40–50% processing at the GSHV site was restored when the DNA was placed in the correct orientation (pGSpA.17 and 19, lanes 5 and 7) but no reconstitution of function was observed when these sequences were in the opposite orientation (pGSpA.18 and 19, lanes 6 and 8).

Discussion

We examined the requirements for a seemingly simple

form of differential poly(A) site selection among the mammalian hepatitis B viruses. Our results indicate that the mechanism underlying this process is complex and demonstrates several novel features. First, the virus employs a variant hexanucleotide (TATAAA) in its poly(A) signal, and this signal and its downstream elements are insufficient to allow use of the GSHV poly(A) site in a heterologous transcript. Second, correction of the hexanucleotide back to consensus (AATAAA) does not completely reconstitute the function of these core sequences. Rather, multiple elements upstream of the core elements are required to do so. We have defined three principal regions (designated PS1, 2, and 3) located in the 397 nucleotides 5' to the hexanucleotide that together allow efficient recognition of the downstream poly(A) site. Deletion of any one of these elements appreciably reduces the efficiency of poly(A) site use; deletion of both PS1 and PS2 nearly abolishes polyadenylation at this site. The fact that pregenomic RNA contains only PS2 between its cap site and the processing site explains why the processing signal is used so inefficiently on the first pass (see below).

Before considering the implications of these findings, it is important to discuss several features of the assay system we have employed. First, in our experimental design, the efficiency of processing at the GSHV site is assessed by examining the ratio of the RNAs processed at this site to those polyadenylated at a downstream SV40 poly(A) site. Such an assessment is valid only if the two product RNAs are equally stable; this is particularly important in the case of mutant genomes, where the possibility exists that the deletions may selectively destabilize one of the RNA products, leading to a false impression of the processing frequencies. We have explicitly validated this for the wild-type plasmid (pGSpA.wt) and several key mutants (pGSpA.11 and 12) that alter the ratio of processed and readthrough transcripts. This was done by examining the half-life of the two RNAs over a

Figure 6. Use of the processing-defective mutant pGSpA.12 to screen for functionally analogous sequences. (A) The series of constructions shown were created by inserting the indicated DNA fragments into the *Bgl*II site of pGSpA.12 43 bp upstream of the polyadenylation signals (see Methods). The fraction of the poly(A)⁺ transcripts that were processed at the GSHV poly(A) site is indicated for each construct at right. (B) Northern analysis of cytoplasmic poly(A)⁺ RNA from COS-7 cells transfected with the above plasmids. Transcripts resulting from processing at either the GSHVpA site or the downstream SV40pA site are indicated.

12-hr period in the presence of actinomycin D. In all cases, the transcripts had comparable stabilities (data not shown).

Second, it is possible that some RNAs that escape processing at the upstream poly(A) site may terminate prior to reaching the downstream poly(A) signal. If so, then the assay may tend to underestimate the degree of readthrough at the upstream site. However, since the region between the poly(A) signals is identical in all constructs, this effect is unlikely to alter our assessment of the relative efficiencies of the various mutant GSHV poly(A) signals we have studied.

Third, the presence of an abundant spliced RNA in transfected COS (and HepG2) cells initially raised concerns that splice-site selection may complicate the interpretation of poly(A) site use in our system, as it has in others. Fortunately, this has proven not to be the case. Plasmids (e.g., pGSpA.4) in which both the splice donor and acceptor have been deleted nonetheless display proper regulation of polyadenylation (Fig. 4A).

The nature of the upstream signals

Although the present studies clearly implicate upstream sequences in the regulation of polyadenylation, they do not directly address the question of their mechanism of action. For instance, we have no direct evidence as to whether these signals function as DNA or as RNA. It is perhaps simplest to imagine that the signals act at the level of RNA—for example, by binding additional factors that would augment processing at the next downstream signal. Another model may be that nascent

RNA transcripts from PS1, PS2, and PS3 regions modify the elongating transcription complex so as to favor processing at the next poly(A) site. Our data indicate that if these elements act as RNA, all must be present contiguously in the transcript for efficient processing to occur. This would be the case on the first pass during subgenomic RNA synthesis, but only on the second pass in pregenomic RNA transcription.

If the processing signals are encoded in specific RNA sequences, one might expect them to display strong orientation dependence. This is certainly the case for the retroviral elements (Fig. 6). Interestingly, only partial orientation dependence was noted for the GSHV fragment (Fig. 6). However, such behavior is not incompatible with the signal functioning as RNA: if the relevant signals are the result not of a particular sequence but of a characteristic structure (e.g., a stem-loop), such a structure could be assumed by complementary sequences (see below).

Orientation dependence is also compatible with DNA signals, and it is equally possible that the processing sequences we have identified act at the DNA level—for instance, by modifying the transcription complex with factors involved in 3' processing during assembly of the initiation complex or during elongation. There is precedent for such a mechanism: others have shown that promoter sequences strongly influence 3'-end processing in small nuclear RNA gene transcription (Hernandez and Weiner 1986; Neuman de Vegvar et al. 1986). If the elements defined here are acting at the DNA level, it is simplest to imagine them acting to modify the transcription complex during elongation. In this model, the

elongating transcription complex must traverse all three regions to assemble the requisite polyadenylation machinery. Given the locations of these signals relative to the pregenomic RNA cap site, this would only be possible on the second pass during pregenomic transcription (during subgenomic RNA transcription, of course, all three elements would be encountered sequentially on the first pass). Finally, it is even possible that some of the elements act as DNA and others as RNA. In the last analysis, it will be difficult to decide between action as DNA versus action as RNA by *in vivo* studies alone. We are currently setting up an *in vitro* processing system to examine whether correct regulation can be observed when the signals are presented only as RNA species generated by *in vitro* transcription.

Structural considerations

Computer analysis has revealed no conservation of primary sequence or secondary structure between the GSHV upstream elements themselves or between the GSHV elements and the U3 regions of SNV and HIV. This could indicate either that a common mechanism can be mediated by highly degenerate sequence requirements, or that multiple different mechanisms can promote the same outcome in processing.

A potential stem-loop structure of 63 nucleotides ($\Delta G > -20$ kcal) has been identified between -71 and -8 (i.e., overlapping PS2). In pGSpA.8, a deletion to -73 has a small but measurable effect on 3'-end formation (Fig. 5). An extended deletion to -43 (pGSpA.9), which would effectively destroy this secondary structure, dramatically reduces processing. Further deletions in a 3' direction, such as pGSpA.10, do not alter the defective phenotype. On the basis of the correlation between deletions defining PS2 and the removal of a significant RNA structure, it is tempting to assign a role in 3'-end processing to this stem-loop. Supporting this notion is the fact that similar structures are conserved in all three mammalian hepadnaviruses (R. Colgrove, unpubl.). To define better the role of this structure in processing we are currently conducting a detailed mutational analysis of this region. Here we note only that this stem-loop does not account for the partial orientation independence observed for the GSHV signals in Figure 6: In the reverse orientation construct in which partial restoration of processing occurred, this stem-loop is disrupted at position -43 .

TATAAA and AATAAA respond differently to upstream signals

We have demonstrated that, for GSHV processing, both TATAAA and AATAAA signals are inefficient in the absence of upstream information (Fig. 3), though the canonical hexanucleotide functions somewhat better than the variant. Interestingly, the two hexanucleotides respond differently to the influence of the upstream elements. Full processing is restored in the context of TATAAA only when all three upstream elements are

present (Fig. 4). However, the fact that efficient first-pass processing of pregenomic RNA is restored by the AA-TAAA signal indicates that PS2 alone is capable of fully activating this signal (plasmid pGSpA.1/A, Fig. 4). Similarly, in the PS1/PS2 deletion plasmid pGSpA.12, efficient processing is restored by the AATAAA hexanucleotide, suggesting that sequences within PS3 alone are able to activate this signal.

A fuller understanding of these phenomena must await the development of *in vitro* processing assays that respond to the upstream signals. These observations, however, together with the existence of functionally analogous upstream elements in retroviruses (Fig. 6; Dougherty and Temin 1987), indicate that the potential of upstream elements to modulate processing decisions need not be confined to signals with variant hexanucleotides. Most retroviruses use the canonical hexanucleotide AATAAA, and many (e.g., HIV, RSV, MMTV, and SNV) contain GT-rich clusters in U5, the genomic region immediately 3' to the poly(A) cleavage site. In the one case that has been closely examined (HIV), mutational alterations in this GT-cluster diminished proper processing (Bohnelein et al. 1989). Thus, polyadenylation signals that cannot be distinguished from unregulated signals by standard criteria can nonetheless be subject to this form of modulation. The influence of sequences upstream of AATAAA has now been demonstrated for the SV40 late (Carswell and Alwine 1989) and adenovirus L1 (DeZazzo and Imperiale 1989) polyadenylation sites. This raises the possibility that the processing of some cellular transcripts whose poly(A) signals appear conventional may likewise be influenced by functionally analogous upstream elements.

Methods

Plasmid constructions

For the construction of pGSpA.wt, the large 2.6-kb *Pst*I fragment of GSHV was inserted into pSP65 (Promega Biotec). The SV40 late polyadenylation signal (*Bam*HI–*Bgl*II; 2517–2753) was cloned into the *Bam*HI site of the same polylinker downstream of the GSHV sequences. The *Hind*III (pSP65)–*Kpn*I (GSHV) fragment of the resulting plasmid was replaced with a *Hind*III (converted *Bal*I site; 7503)–*Kpn*I (8296) fragment of mouse leukemia virus (MLV). This created a *Sma*I site immediately upstream of the *Kpn*I site of GSHV. The 2.8-kb *Sma*I (MLV; 8292)–*Sma*I (pSP65 polylinker) fragment was then excised and inserted into the *Hinc*II site of a pSP65-derived plasmid in which the bacterial sequences from *Pvu*II to *Hind*III had been replaced with the SV40 early promoter and origin of replication (*Pvu*II–*Hind*II; 342 bp). In the resulting plasmid, sequences upstream of the GSHV *Kpn*I site (-1805) have been replaced with the SV40 promoter while the SV40 polyadenylation signal is located ~ 100 bp downstream of the GSHV *Pst*I site ($+835$) and ~ 900 bp 3' to the polyadenylation signal of GSHV, the first nucleotide of which has been designated $+1$. Finally, a *Bgl*II linker was inserted into the *Cla*I site (-172) of GSHV.

Deletion mutations (see Fig. 5A) were generated by exonuclease III digestion of pGSpA.wt after cleavage with *Bgl*II. Single-stranded DNA was removed with mung bean nuclease, and the two newly created ends were joined with a *Bgl*II linker.

pGSpA.6, 7, 8, and 9 were constructed by replacing the upstream deleted *HindIII*–*BglII* fragment with the corresponding fragment from pGSpA.wt, whereas for pGSpA.5, substitution was carried out with the downstream deleted *BglII*–*EcoRI* region. In the case of pGSpA.11 and 12, the wild-type *HindIII*–*BglII* fragments from pGSpA.6 and 10, respectively, were substituted with the deletion fragment purified from pGSpA.5. The deletion in pGSpA.10 is simply the removal of a *BglII*–*XhoII* fragment followed by religation of the compatible ends.

To create pGSpA.1, complete removal of upstream GSHV sequences was carried out on an exonuclease III mutant (3' limit of –107). After cleavage with *HindIII* and *BglII* endonucleases, the ends were repaired with T4 DNA polymerase and religated with a *BglII* linker. Plasmids pGSpA.2, 3, and 4 were constructed by carrying out *HindIII*/*BglII*, *HindIII*/*PvuII*, and *HindIII*/*BstEII* digestions, respectively, on pGSpA.wt and processing as above. Removal of the *HindIII*–*BstEII* fragment from pGSpA.12 resulted in pGSpA.4/12.

pGSpA.12 was used to generate a series of replacement plasmids in which four distinct DNA fragments were introduced (in either orientation) into the *BglII* site. These include a 208-bp *PvuII*–*AluI* fragment from GSHV (pGSpA.13 and 14), a 194-bp *HaeIII* fragment from Φ X174 (pGSpA.15 and 16), a 375-bp *SstI*–*NciI* fragment from pB101 (Panganiban and Temin 1984), which encompasses the U3 region of SNV (pGSpA.17 and 18), and a 450-bp *PvuII* fragment from HIV (U3 region of subtype HXB2; pGSpA.19 and 20).

For the construction of pGSpA.21, a 1.6-kb *NcoI*–*BglII* fragment bearing the avian cSRC coding region was purified from pM5HHB5 [Kmieciak and Shalloway 1987]. Staggered ends were made flush with Klenow fragment before insertion into *HindIII*/*BglII*-digested pGSpA.12, the ends of which had been repaired in a similar manner.

Plasmids pGSpA.1/A, 12/A, and 21/A harbor an A at position +1 within the GSHV polyadenylation signal instead of the wild-type T residue. Mutagenesis was carried out using uracil-containing single-stranded DNA as described by Kunkel et al. (1987) and a 20-nucleotide synthetic oligonucleotide bearing the appropriate base change (CCAAATTCTTTATGGGATC).

Cell culture and transfections

COS-7 and HepG2 cells were maintained in Dulbecco's modified Eagle medium (DME) supplemented with 10% fetal calf serum. Transfection of COS-7 cells was carried out with 5 μ g of plasmid per 100-mm dish in the presence of 500 μ g/ml DEAE-dextran [Sompayrac and Danna 1981]. For HepG2 cells, calcium phosphate-mediated DNA transfections were performed [Graham and van der Eb 1973] using similar quantities of plasmid DNA without carrier.

RNA analysis

Cytoplasmic RNA was isolated from tissue culture cells 48 hr after transfection according to the procedure of Kaufman and Sharp (1982). Total RNA from excised ground squirrel liver was extracted using guanidine thiocyanate [Chirgwin et al. 1979] and purified through CsCl gradients. In both cases, the final RNA pellets were resuspended in 1.0 ml of 0.5 M NaCl, 10 mM Tris (pH 7.5), 1.0 mM EDTA, and 1.0% SDS. Fractionation of poly(A)⁺ RNA using oligo(dT)-cellulose was carried out as described by Hirsch et al. (1988).

Separation of poly(A)⁺ RNA by formaldehyde gel electrophoresis and transfer to nitrocellulose was done using standard methodologies [Maniatis et al. 1982]. Prehybridization and hybridization with RNA probes was performed at 60°C in 50%

formamide, 6 $\times$ SSPE, 1.0% SDS, 0.1% Tween-20, and 100 μ g/ml yeast tRNA. The filter was washed twice in 1 $\times$ SSPE, 0.1% SDS for 30 min per wash at room temperature. The final two washes were in 0.1 $\times$ SSPE, 0.1% SDS for 30 min per wash at 65°C. Alternatively, cytoplasmic poly(A)⁺ RNA was analyzed using RNase mapping techniques as described by Melton et al. (1984). Briefly, 1.0–5.0 μ g of test poly(A)⁺ RNA was hybridized to 100,000–200,000 cpm of probe RNA (see below) in 40 mM HEPES (pH 6.7), 0.4 M NaCl, and 1.0 mM EDTA using a volume of 30 μ l. The hybridization mix was incubated for 10 min at 85°C and then transferred to 65°C for 4 hr. After hybridization, 300 μ l of RNase digestion buffer containing 10 mM Tris-HCl (pH 7.5), 5.0 mM EDTA, 300 mM NaCl, 40 μ g/ml RNase A, and 2.0 μ g/ml RNase T1 was added. RNase digestion was allowed to proceed at 30°C for 60 min and was terminated by the addition of 20 μ l of 10% SDS and 50 μ g of proteinase K followed by an additional incubation for 15 min at 37°C. The reaction was extracted and the resistant RNA hybrids were precipitated with 10 μ g of yeast tRNA as carrier.

The 330-nucleotide labeled RNA probe used throughout this study (corresponding to sequences from *StuI* to *Clal*/*BglII*) was generated from a pGSpA.wt variant harboring a deletion of the sequences between *EcoRI* and *StuI*. In vitro transcription reactions were carried out on the above plasmid (linearized with *BglII*) using SP6 polymerase in the presence of [³²P]UTP (25 μ M cold UTP) according to the protocol of Melton et al. (1984).

Acknowledgments

Special thanks to Tyra Wolfsberg for her participation in several aspects of this project. We also thank Dave Kaplan for the SNV clone pB101, Ned Landau for cloned HIV DNA, Hisamaru Hirai for the chicken cSRC cDNA clone, Robin Colgrove for computer analysis, and Christine Guthrie and Keith Yamamoto for helpful comments on the manuscript. This work was supported by grants from the National Institutes of Health.

References

- Acheson, N. 1984. Kinetics and efficiency of polyadenylation of late polyomavirus nuclear RNA: Generation of oligomeric polyadenylated RNAs and their processing into mRNA. *Mol. Cell. Biol.* **4**: 722–729.
- Alt, F.W., A.L.M. Bothwell, M. Knapp, E. Siden, E. Mather, M. Koshland, and D. Baltimore. 1980. Synthesis of secreted and membrane-bound immunoglobulin μ heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* **20**: 293–301.
- Bhat, B.M. and W.S.M. Wold. 1985. AATAAA as well as downstream sequences are required for RNA 3'-end formation in the E3 complex transcription unit of adenovirus. *Mol. Cell. Biol.* **5**: 3183–3193.
- Bohnlein, S., J. Hauber, and B.R. Cullen. 1989. Identification of a U5-specific sequence required for efficient polyadenylation within the human immunodeficiency virus long terminal repeat. *J. Virol.* **63**: 421–424.
- Carswell, S. and J.C. Alwine. 1989. Efficiency of utilization of the simian virus 40 late polyadenylation site: Effects of upstream sequences. *Mol. Cell. Biol.* **9**: 4248–4258.
- Chirgwin, J.M., A.E. Przybyla, R.J. MacDonald, and W.J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**: 5294–5299.
- Christofori, G. and W. Keller. 1988. 3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A)

- polymerase, a cleavage factor, and a snRNP. *Cell* **54**: 875–889.
- . 1989. Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro. *Mol. Cell. Biol.* **9**: 193–203.
- Conway, L. and M. Wickens. 1985. A sequence downstream of AAUAAA is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes. *Proc. Natl. Acad. Sci.* **82**: 3949–3953.
- Danner, D. and P. Leder. 1985. Role of an RNA cleavage/poly(A) addition site in the production of membrane-bound and secreted IgM mRNA. *Proc. Natl. Acad. Sci.* **82**: 8658–8662.
- DeZazzo, J.D. and M.J. Imperiale. 1989. Sequences upstream of the AAUAAA influence poly(A) site selection in a complex transcription unit. *Mol. Cell. Biol.* **9**: 4951–4961.
- Dougherty, J.P. and H.M. Temin. 1987. A promoterless retroviral vector indicates that there are sequences in U3 required for 3' RNA processing. *Proc. Natl. Acad. Sci.* **84**: 1197–1201.
- Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood. 1980. Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell* **20**: 313–319.
- Enders, G.H., D. Ganem, and H. Varmus. 1985. Mapping the major transcripts of ground squirrel hepatitis virus: The presumptive template for reverse transcriptase is terminally redundant. *Cell* **42**: 297–308.
- Falck-Pedersen, E. and J. Logan. 1989. Regulation of poly(A) site selection in adenovirus. *J. Virol.* **63**: 532–541.
- 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.
- Friedman, D.I., M.J. Imperiale, and S.L. Adhya. 1987. RNA 3' end formation in the control of gene expression. *Annu. Rev. Genet.* **21**: 453–488.
- Galli, G., J. Guise, P.W. Tucker, and J.R. Nevins. 1988. Poly(A) site choice rather than splice site choice governs the regulated production of IgM heavy-chain RNAs. *Proc. Natl. Acad. Sci.* **85**: 2439–2443.
- Ganem, D. and H.E. Varmus. 1987. The molecular biology of the hepatitis B viruses. *Annu. Rev. Biochem.* **56**: 651–694.
- Gil, A. and N.J. Proudfoot. 1984. A sequence downstream of AAUAAA is required for rabbit β -globin mRNA 3' end formation. *Nature* **312**: 473–474.
- . 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β -globin mRNA 3' end formation. *Cell* **49**: 399–406.
- Graham, F.L. and A.J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52**: 456–467.
- Hart, R.P., M.A. McDevitt, A. Haseeb, and J.R. Nevins. 1985. Definition of essential sequences and functional equivalence of elements downstream of the adenovirus E2A and the early simian virus 40 polyadenylation sites. *Mol. Cell. Biol.* **5**: 2975–2983.
- Helfman, D.M., S. Cheley, E. Kuismanen, L.A. Finn, and Y. Yamawaki-Kataoka. 1986. Nonmuscle and muscle tropomyosin isoforms are expressed from a single gene by alternative RNA splicing and polyadenylation. *Mol. Cell. Biol.* **6**: 3852–3859.
- Hernandez, N. and A.M. Weiner. 1986. Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell* **47**: 249–258.
- Hirsch, R., R. Colgrove, and D. Ganem. 1988. Replication of duck hepatitis B virus in two differentiated human hepatoma cell lines after transfection with cloned viral DNA. *Virology* **167**: 136–142.
- Kaufman, R.J. and P.A. Sharp. 1982. Construction of a modular dihydrofolate reductase cDNA gene: Analysis of signals utilized for efficient expression. *Mol. Cell. Biol.* **2**: 1304–1319.
- Kessler, M.M., R.C. Beckendorf, M.A. Westhafer, and J.L. Nordstrom. 1986. Requirement of AAUAAA and adjacent downstream sequences for SV40 early polyadenylation. *Nucleic Acids Res.* **14**: 4939–4952.
- Kmieciak, T.E. and D. Shalloway. 1987. Activation and suppression of pp60^{c-src} transforming ability by mutation of its primary sites of tyrosine phosphorylation. *Cell* **49**: 65–73.
- Kunkel, T.A., J.D. Roberts, and R.A. Zakour. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**: 376–382.
- Leff, S.E., R.M. Evans, and M.G. Rosenfeld. 1987. Splice commitment dictates neuron-specific alternative RNA processing in calcitonin/CGRP gene expression. *Cell* **48**: 517–524.
- Levitt, N., D. Briggs, A. Gil, and N.J. Proudfoot. 1989. Definition of an efficient poly(A) site. *Genes Dev.* **3**: 1019–1025.
- Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Manley, J. 1983. Accurate and specific polyadenylation of mRNA precursors in a soluble whole-cell lysate. *Cell* **33**: 595–605.
- Mason, P.J., J.A. Elkington, M.M. Lloyd, M.B. Jones, and J.G. Williams. 1986. Mutations downstream of the polyadenylation site of a Xenopus B-globin mRNA affect the positions but not the efficiency of 3' processing. *Cell* **46**: 263–270.
- McDevitt, M.A., M.J. Imperiale, H. Ali, and J.R. Nevins. 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. *Cell* **37**: 993–999.
- 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.
- McLauchlan, J., D. Gaffney, J.L. Whitton, and J.B. Clements. 1985. The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res.* **13**: 1347–1368.
- Melton, D.A., P.A. Krieg, M.R. Rebagliati, T. Maniatis, K. Zinn, and M.R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**: 7035–7056.
- Montell, C., E.F. Fisher, M.H. Caruthers, and A.J. Berk. 1983. Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' consensus sequence AAUAAA. *Nature* **305**: 600–629.
- Moore, C. and P. Sharp. 1984. Site-specific polyadenylation in a cell-free reaction. *Cell* **36**: 581–591.
- . 1985. Accurate cleavage and polyadenylation of an exogenous RNA substrate. *Cell* **41**: 845–855.
- Neuman de Vegvar, H.E., E. Lund, and J.E. Dahlberg. 1986. 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell* **47**: 259–266.
- Nevins, J.R. and M.C. Wilson. 1981. Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. *Nature* **290**: 113–118.
- Panganiban, A.T. and H.M. Temin. 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. *Cell* **36**: 673–679.
- Peterson, M.L. and R.P. Perry. 1989. The regulated production of μ m and μ s mRNA is dependent on the relative efficiencies of μ s poly(A) site usage and the C μ 4-to-M1 splice.

Russnak and Ganem

- Mol. Cell. Biol.* **9**: 726–738.
- Proudfoot, N.J. and G.G. Brownlee. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**: 211–214.
- Sadofsky, M., S. Connelly, J.L. Manley, and J. Alwine. 1985. Identification of a sequence element on the 3' side of AAUAAA which is necessary for simian virus 40 late mRNA 3'-end processing. *Mol. Cell. Biol.* **5**: 2713–2719.
- Seeger, C., D. Ganem, and H.E. Varmus. 1984. The cloned genome of ground squirrel hepatitis virus is infectious in the animal. *Proc. Natl. Acad. Sci.* **81**: 5849–5852.
- Simonsen, C.C. and A.D. Levinson. 1983. Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. *Mol. Cell. Biol.* **3**: 2250–2258.
- Sompayrac, L.M. and K.J. Danna. 1981. Efficient infection of monkey cells with DNA of simian virus 40. *Proc. Natl. Acad. Sci.* **78**: 7575–7578.
- Stacy, T.P. and C.N. Cole. 1985. Identification of sequences in the herpes simplex virus thymidine kinase gene required for efficient processing and polyadenylation. *Mol. Cell. Biol.* **5**: 2104–2113.
- Takagaki, Y., L.C. Ryner, and J.L. Manley. 1988. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* **52**: 731–742.
- Terms, M.P. and S.T. Jacob. 1989. Role of poly(A) polymerase in the cleavage and polyadenylation of mRNA precursor. *Mol. Cell. Biol.* **9**: 1435–1444.
- Wickens, M. and P. Stephenson. 1984. Role of the conserved AAUAAA sequence: Four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science* **226**: 1045–1051.
- Woychik, R.P., R.H. Lyons, L. Post, and F.M. Rottman. 1984. Requirement for the 3' flanking region of the bovine growth hormone gene for accurate polyadenylation. *Proc. Natl. Acad. Sci.* **81**: 3944–3948.
- Zhang, F., R.M. Denome, and C.N. Cole. 1986. Fine-structure analysis of the processing and polyadenylation region of the herpes simplex virus type 1 thymidine kinase gene by using linker scanning, internal deletion, and insertion mutations. *Mol. Cell. Biol.* **6**: 4611–4623.

Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses.

R Russnak and D Ganem

Genes Dev. 1990, **4**:

Access the most recent version at doi:[10.1101/gad.4.5.764](https://doi.org/10.1101/gad.4.5.764)

References

This article cites 55 articles, 24 of which can be accessed free at:
<http://genesdev.cshlp.org/content/4/5/764.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](#).