Regulation of polyadenylation in hepatitis B viruses: stimulation by the upstream activating signal PS1 is orientation-dependent, distance-independent, and additive

Roland H. Russnak

Department of Biochemistry, University of Rochester Medical Center, Rochester, NY 14642, USA

Received September 4, 1991; Revised and Accepted November 6, 1991

ABSTRACT

Hepatitis B viruses replicate by reverse transcription of a genomic RNA which harbors terminal redundancies. The synthesis of this RNA requires that transcription proceed twice through the polyadenylation (pA) site which, in mammalian strains, is flanked by the variant hexanucleotide UAUAAA and a T-rich downstream domain. These core elements are by themselves virtually defective in 3' end processing and require multiple upstream accessory elements which regulate pA site use. In ground squirrel hepatitis B virus (GSHV), one of these signals (PS1; –215 to –107 relative to UAUAAA) is transcribed only at the 3' end of genomic RNA and as such is analogous to retroviral U3 sequences. PS1 cooperates with other signals to enhance pA site use to very high levels and can be further sub-divided into two regions (A and B) which contribute equally to 3' end processing. Critical residues within PS1B have been localized to a 15 bp A/T-rich stretch which displays homology to other known upstream activating signals. A 15 bp segment within PS1A which has the identical A/T content but a divergent primary sequence plays a diminished role in processing. Furthermore, PS1 can activate GSHV core element usage autonomously. This stimulation has been shown to be additive since multiple copies of PS1 progressively increase polyadenylation, a phenomenon which also demands that PS1 exert its influence from a variety of distances from the hexanucleotide signal.

INTRODUCTION

An important step in the maturation of most newly transcribed eukaryotic mRNAs is the *de novo* synthesis of long poly(A) tracts at the 3' end. This involves cleavage of a longer precursor molecule and the tightly coupled addition of adenylate residues to the resulting 3' hydroxyl group. Poly(A) metabolism likely influences the functional half-life of mRNAs. In the cytoplasm, poly(A) tails are associated with poly(A)-binding protein (PABP). This nucleoprotein complex confers translational competency on the transcript (see 1,2 for reviews) and serves to protect the mRNA from nucleolytic attack (see 2,3 for reviews). A compelling argument for a translational role comes from yeast

genetics in which suppressors of a *ts* mutation of PABP have been localized to the gene for the 60S r-protein L46 (4) and to the gene encoding a putative rRNA helicase involved in the maturation of 25S rRNA (5).

Recruitment of mRNAs onto polysomes may be controlled by altering the length of poly(A). During *Xenopus* oocyte maturation, poly(A) is removed from cytoplasmic maternal mRNAs by a default reaction which, in certain transcripts, is overcome by active extension of the poly(A) tail and which is specified by both AAUAAA and U-rich sequences found in the 3' untranslated region (6,7,8; see 9 for review). A similar mechanism may be programmed into the mouse egg (10). These dynamic fluctuations in poly(A) tail length have also been documented in somatic cells. A notable example occurs in the suprachiasmatic nuclei of the brain where daily variation in the length of the vasopressin mRNA poly(A) tail may underlie the circadian rhythm of vasopressin peptide levels in cerebrospinal fluid (11).

In general, 3' end processing within the nuclei of higher eukaryotes requires two core sequence elements: the conserved AAUAAA hexanucleotide found 10–30 nt upstream of the RNA cleavage site and a more divergent GT- or T-rich domain downstream (see 12 for review). A synthetic oligonucleotide composed of only these elements and modelled after the beta-globin gene is sufficient to confer efficient polyadenylation on a heterologous transcript (13). AAUAAA is recognized by a specificity factor of high molecular weight which is composed of at least five distinct proteins (14), the largest of which (155–170 kd) contains an Sm epitope and can be UV-crosslinked to AAUAAA-containing substrates (14,15,16). Downstream elements are associated with another multicomponent factor, composed of at least three subunits (14,17), which forms a stable complex with pre-mRNA only in the presence of specificity factor (18,19). One component, a 64 kd protein, makes direct contact with RNA based on UV-crosslinking studies (14,15,19,20).

These trans-acting factors form the nucleus of a large processing complex (core complex), the stability of which is dependent on the interaction of these factors, thereby determining the efficiency with which cleavage and polyadenylation proceeds (18,21). Proper formation of this structure is dependent on correct spacing between the cis-acting core elements (16,21,22,23,24). Surprisingly, poly(A) polymerase contributes to the formation of the core complex and therefore is also required for the cleavage

reaction (16,25,26,27). Upon cleavage, the poly(A) tail is added in a biphasic reaction; the first 10 residues are incorporated in an AAUAAA-dependent fashion while full length extension requires only a short newly synthesized oligo(A) tract (28; see 29 for review).

In addition to core elements, a third class of cis-acting sequences has recently been identified in a variety of animal and plant viral polyadenylation sites including SV40 late (30), adenovirus L1 (31), cauliflower mosaic virus (CaMV; 32), human immunodeficiency virus (HIV; 33,34,35,36,37), spleen necrosis virus (SNV; 33,38), and ground squirrel hepatitis B virus (GSHV; 33). These signals are located at varying distances upstream, and enhance the utilization of sub-optimal core signals which at first approximation conform to normal sequence requirements (GSHV is an exception; see below). These properties have been exploited to regulate the recognition of core signals in retroviral elements such as retroviruses (HIV) and hepadnaviruses (GSHV). In these viruses, synthesis of terminally redundant genomic RNA (the template for reverse transcription) demands that the closely spaced core elements be transcribed twice and as a result are present both at the extreme 5' end of the transcript as well as the 3' end where they are utilized almost exclusively (for reviews see 39,40,41). In HIV and GSHV, regulation of pA site use partially relies on the placement of polyadenylation enhancers upstream of the genomic RNA transcriptional start site (i.e., in the U3 region), ensuring that they be transcribed only at the 3' end of the transcript where they activate use of the core elements from a distance of at least 80–100 nt. The mechanism by which upstream signals exert their stimulatory effect and their prevalence in cellular genes is as yet unknown. GSHV is unique in that pA site usage relies entirely on upstream information given that a variant hexanucleotide UAUA AAA renders the core elements virtually non-functional (33). The same A to U nucleotide change has been shown to inactivate other pA sites *in vitro* (42,43).

To date, none of the factors associated with the core complex have been shown to directly alter the efficiency of pA site usage. Such an activity has been identified in extracts from herpes simplex virus-infected cells (44) but has not been characterized further. If upstream processing signals can influence the manner in which the core complex catalyzes 3' end formation, then it follows that factors binding to them, if any, may be prime candidates for the regulation of polyadenylation. As such, precise definition of the cis-acting sequences is obligatory to the identification of such factors. Previous deletion mapping in GSHV (33) has identified two regions which influence 3' end processing. Processing signal 2 (PS2), is a strong activator situated between the genomic RNA start sites (–130 and –110 relative to UAUA AAA) and the core elements, but is not recognized properly upon the first pass of the transcriptional machinery due to its proximity to the promoter (45). Processing signal 1 (PS1) has been originally defined by a 100 bp deletion from –229 to –135. Like retroviral U3 regions, PS1 overlaps the genomic RNA promoter and is transcribed only at the 3' end of genomic RNA.

The current study localizes PS1 to the region encompassing sequences from –215 to –105, the minimal domain which is sufficient for activating the core elements. PS1 demonstrates properties novel for a polyadenylation signal in that multiple copies placed in tandem results in progressively increasing pA site use. The cis-acting requirements for PS1 function are complex, involving the cooperative action of multiple determinants. In one case, a 15 bp A/T-rich stretch, which shares

homology with other upstream processing signals, was shown to be critical.

MATERIALS AND METHODS

Plasmid Constructions

SRC-8: The 1.6 kb *EcoRI-XhoI* fragment of pGC1-SRC (46), containing the entire coding region of chicken c-SRC, was cloned into pGEM-7Z (Promega) and re-excised with *HindIII-XhoI*. This fragment was subsequently used to replace all GSHV sequences upstream of the pA signal by digestion of Δ SL2 with the corresponding enzymes. Δ SL2 is a derivative of the wild type parental construct (pGSpA.wt, see legend to Fig. 3) in which sequences from –34 to –9 have been replaced with a unique *XhoI* site. The *HindIII* site of pGSpA.wt is located within a polylinker region between the SV40 promoter/enhancer and the GSHV sequences (–1805 to +835). The resulting plasmid, SRC-8, contained a new polylinker harboring unique *BamHI*, *BglII*, and *XhoI* restriction sites between the 3' end of the c-SRC DNA and GSHV sequences at –8. This polylinker also introduced *SalI* and *XbaI* sites which are not unique.

SRC-107/SRC-400: The c-SRC coding region was excised from pGC1-SRC with *NcoI-BglII* and repaired with Klenow. For SRC-107 this fragment was cloned into Δ 6.5 digested with *HindIII-BglII* and Klenow-repaired. The Δ 6.5 construct is a previously unreported exonuclease III deletion mutant of pGSpA.wt in which sequences from –170 to –107 have been replaced with a unique *BglII* linker. For SRC-400, the repaired *NcoI-BglII* fragment was cloned into pGSpA.wt digested with *HindIII-BstEII* (–400) and end-repaired. **SRC-400 Δ PS2:** A *BstEII-SalI* fragment from Δ PS2, a derivative of pGSpA.wt containing a deletion from –105 to –41, was repaired with Klenow and cloned into SRC-8 which had been digested with *SalI* and end-repaired. The *SalI* site common to both Δ PS2 and SRC-8 was situated within the polylinker sequence which separates the 3' end of the GSHV sequences (+830) from the downstream SV40 pA site.

SRC-215 Δ PS2: A unique *NdeI* site was introduced into Δ PS2 (see above) at position –215 by site-directed mutagenesis (47) (see Fig. 4A for exact base pair changes). The *NdeI-SalI* fragment was excised from this plasmid, repaired with Klenow, and cloned into SRC-8/*SalI* as above.

SRC-8 Insertion Series: The various PS1-containing fragments shown in Fig. 2 were purified from the SL1(*NdeI*,*MluI*) construct (see below), made blunt by treatment with Klenow, and cloned into SRC-8, which had been digested with *BamHI* and Klenow-repaired. In the case of SRCNH(2) and SRCNH(3), the second copy or a concatamer containing the second and third copy, respectively, was cloned into SRCNH(1) which had been digested with *XhoI* and filled-in with Klenow. SRCNH(3)R was constructed using the same strategy as SRCNH(3) except that the fragments were inserted in the opposite orientation. To create SRCNH(3)A1, two successive rounds of site-directed mutagenesis were carried out on SRCNH(3). The three identical A/T-rich stretches in site B were altered simultaneously using one mutagenic oligonucleotide. After identification of the desired mutant concatamer, simultaneous mutation of site A within all three copies of PS1 was carried out using a second oligonucleotide.

GSHV PS1 Mutations: The constructs Δ –229/–171, Δ –170/–136, and Δ PS1(Δ –229/–136) correspond to the deletion mutants pGSpA.5, pGSpA.6, and pGSpA.11,

respectively, which have been previously described (33). In SL1, sequences from -68 to -41 have been deleted and replaced with a unique *XhoI* site. In $\Delta-170/-107$, the same deletion was introduced into $\Delta 6.5$ (see SRC-107 construct described above) by oligo-directed mutagenesis. Site-directed mutagenesis of SL1 was used to create the triple mutant A2 which harbors a unique *AflIII* site within site A and a unique *BamHI* site within site B. Digestion of A2 with *AflIII*, followed by religation of mung bean nuclease recessed ends was carried out for construction of A3. For subsequent PS1 mutants, unique *NdeI* and *MluI* restriction sites were engineered into SL1 by site-directed mutagenesis at positions -215 and -134 , respectively (see Fig. 4A). This created the parental construct SL1(Nde,Mlu), from which all of the remaining PS1 mutations were derived. Δ PS1(SL1) was constructed by digesting SL1(Nde,Mlu) with *NdeI-MluI*, treating with Klenow to repair all ends, and religating. In a similar manner, Δ PS1A was generated by digestion of the same plasmid with *NdeI-BglIII* and religating subsequent to end-repair. To create $\Delta-134/-105$, a unique *NruI* site was first introduced at -105 , followed by the removal of sequences between *MluI* and *NruI* as described above. For Δ IVS, sequences from -186 to -160 were deleted by oligo-directed mutagenesis designed to create a unique *NruI* restriction site. All nucleotide substitutions within site B were done by mutagenesis of SL1(Nde,Mlu). Mutants B1 through B6 were the result of the deletion of sequences between *NdeI* and *BglIII* as described for Δ PS1A. For the three site A mutants (A1, A4, and A5), mutagenesis was carried out on Δ PS1B.

Cell Culture and RNA Analysis

Transfection of COS-7 cells and Northern analysis of cytoplasmic poly(A)⁺ RNA was performed as previously described (33). GSHV-specific RNA hybridization probes were generated by *in vitro* transcription of a pGSpA.wt derivative in which the SP6 promoter was juxtaposed with sequences at -400 (*BstEII*) followed by linearization with *SpeI* (-1235). For transcripts harboring SRC sequences, the cRNA probes corresponded to the 3' terminal 700 nt of the cDNA linearized at the internal *MluI* site. For quantification of RNA levels, radioactive filters were exposed on a PhosphorImager (Molecular Dynamics).

RESULTS

Identification of upstream processing signals

We have previously determined the efficiency of the GSHV core polyadenylation elements by placing them 3' to a heterologous coding region, c-SRC (33). These sequences, which include the variant hexanucleotide UAUAAA, the cleavage site, and downstream U-rich sequences known to be absolutely required for correct polyadenylation in the closely related human hepatitis B virus (48), conferred only 10% processing levels (see SRC-8, Fig. 1). In the assay system used, an efficient SV40 pA site is present downstream of GSHV sequences at $+830$ (relative to UAUAAA) to detect those transcripts not processed at the upstream site. Also, those GSHV sequences downstream of the pA site cannot confer a termination function since they are transcribed during genomic RNA synthesis.

Initially, I set out to investigate the contribution made by various upstream regions in activating the GSHV core elements. As shown in Fig. 1A, all GSHV sequences are situated downstream of a 1.6 kb c-SRC cDNA. Transcription initiated within SV40 sequences upstream of c-SRC, precluding any effects

on processing attributable to a proximal promoter. Also, the resulting transcript was similar in length to GSHV subgenomic RNA which is transcribed during viral infection from an internal promoter at -1900 and efficiently processed (33). Relevant to the synthesis of terminally redundant genomic RNA, in this assay system the upstream elements are encountered by the transcriptional as well as the polyadenylation machinery in a manner analogous to the second pass of the pA site.

As mentioned above, processing at the GSHV pA site was only 10% efficient in the absence of sequences 5' to -8 (SRC-8). Similar results were obtained with the inclusion of sequences to -43 (data not shown). Processing was enhanced to 60% in SRC-107 due to the presence of PS2 [-96 to -32 ; ref. 45]. Further addition of sequences extending to -400 (SRC-400; Fig. 1) promoted high levels of GSHV pA site use. This region, upstream of the genomic RNA start sites at -130 and -110 , functions also as a promoter in hepatocytes but was inactive in COS-7 cells. As shown in Fig. 1, in the absence of PS2 ($\Delta-105/-41$), sequences between -400 and -105 maintain polyadenylation at 50% levels (SRC-400 Δ PS2). This region can be further sub-divided into two domains. One domain from -215 to -105 was 30% effective in processing (see SRC-215 Δ PS2)

Figure 1. Determination of the contribution made by upstream domains in the activation of the defective GSHV pA signal. (A) Illustration of chimeric SRC-GSHV constructs containing varying upstream regions (denoted by numbers relative to UAUAAA). Transcription initiated at an SV40 promoter 5' to the 1.6 kb SRC cDNA. In each case, the efficiency with which nascent polII transcripts were processed at the GSHV pA site is listed to the right as a percentage based on direct comparison to readthrough transcripts processed at an SV40 pA site located 0.9 kb downstream. (B) Quantification of transcripts expressed from the indicated plasmids, the 3' ends of which correspond to processing at either the GSHV pA site or the downstream SV40 pA site as indicated by arrows. All Northern blot analyses in this study were carried out using cytoplasmic poly(A)⁺ RNA purified from COS-7 cells 48 hr post-transfection. Plasmids used were high copy number vectors containing an SV40 origin of replication. See Materials and Methods for protocols and a description of the hybridization probes which were used.

and defines PS1. Comparison of SRC-215 Δ PS2 with SRC-400 Δ PS2 demonstrates that sequences between -400 and -215 also had a stimulatory effect. This region has been previously designated PS3 (33) and will be discussed below. Although quite efficient in processing, SRC-400 still allowed substantial (10%) readthrough, a phenotype which was distinct from wild type (lane 1, Fig. 3B) where readthrough was not observed.

PS1 function is orientation-dependent, distance-independent, and additive

SRC-8 contains three unique restriction sites just upstream of the UAUAAA element, thus allowing for the insertion of various PS1-derived DNA fragments. In SRCNH(1) (Fig. 2), sequences from -215 to -107 were placed into the polylinker of SRC-8 approximately 30 bp from the GSHV hexanucleotide signal. Activation from a 10% background to 30% was observed, consistent with SRC-215 Δ PS2 (Fig. 1) in which the same region was situated 41 bp from the pA signal. A progressive increase in pA site use was observed as additional copies of PS1 were introduced to create SRCNH(2) (50% processing) and SRCNH(3) (70%). Activation by PS1 displayed strict orientation-dependence since three copies in the reverse direction [SRCNH(3)R] had processing levels barely above background (Fig. 2). Based on these data, PS1 seemingly conveys its full

stimulatory potential from distances of 30–40 nt, 150 nt (second copy), and 260 nt (third copy).

The high degree of differential processing observed with multiple copies of PS1 compared to background (70% and 10%, respectively) provides an effective system for analyzing PS1 mutations in the absence of other signals. For example, in SRCNH(3)A1 (Fig. 2), all three copies of PS1 contain the A/T-rich site lesions analyzed in mutant A1 (see below and Fig. 4). As a result, processing is reduced to 35% (less than half of the activation observed with wild type PS1), but is not abolished.

I explored also the possibility that smaller segments containing the major determinants of PS1, as determined by deletion analysis and dinucleotide substitutions described below, would display similar activation. As shown in Fig. 2, construct SRCNM(2) contains two copies of the sequences from -215 to -134 while SRCBM(4) contains four copies of the region from -170 to -134. In both cases, stimulation of polyadenylation was not comparable to that observed with two copies of the *NdeI-HinfI* fragment which to date represents the minimal sequence requirement for activation.

Figure 2. Activation of GSHV pA site usage by PS1. (A) Description and (B) Northern blot analysis of SRC-8 constructs harboring various PS1 insertions upstream of the UAUAAA signal. Sites A and B refer to the regions within PS1 which were analyzed in detail. The arrows shown in SRCNH(3)A1 denote lesions within these regions as defined by the mutant A1 (see Fig. 4). Numbering refers to the natural location of these sequences within the context of the GSHV genome, not within the recombinant forms shown.

Figure 3. Disruption of PS2 facilitates the fine-structure analysis of PS1. (A) The wild type construct is equivalent to pGSpA.wt (33) in which GSHV sequences from -1800 to +830 were placed between an SV40 promoter and an SV40 pA signal. This construct contains a *BglII* site which resulted from the insertion of a 10 bp linker into the *ClaI* site of GSHV. See Materials and Methods for the creation of the restriction sites shown in parentheses. The sequences of Δ IVS and A1 are shown in Fig. 4A. (B) Transcripts processed at either the GSHV pA or SV40 pA sites are 0.7 kb and 1.6 kb in length, respectively [excluding poly(A) tracts]. This is due to an efficient (>90%) splice of 1.1 kb in COS-7 cells, the boundaries of which have been determined by PCR analysis (33) and are indicated in (A). The resulting altered profile of mRNA produced from these plasmids is due to the most slowly migrating species which is polyadenylated at the GSHV pA site but remains unspliced.

Detailed Analysis of PS1

The existence of multiple elements within PS1 was strongly implied by the earlier analysis of deletions which are shown again in Fig. 3. The smaller deletions (Δ -229/-171 and Δ -170/-136) had minimal effects on processing compared to a combination of the two (Δ PS1). Inspection of the sequence in this region, displayed in Fig. 4A, revealed the presence of two 15 bp A/T-rich stretches. These motifs were chosen as candidates for repetitive determinants and subjected to detailed analyses as described below. The regions immediately surrounding the A/T-rich stretches and defined by the deletions discussed above will

be referred to as PS1A (-229 to -171) and PS1B (-170 to -136).

Given that the complete removal of these regions results in a phenotype difficult to distinguish from wild type (80% versus >95%) and that mutations within each region would inevitably produce partial phenotypes (not to mention the possibility of an up-regulatory effect), further characterization of PS1 in a wild type background would be insufficient. I therefore took advantage of the following observation. SL1 (Fig. 3) contains a deletion from -68 to -41 which partially inactivated PS2 (45) but still retained high processing levels of 80%. Placement of unique *Nde*I and *Mlu*I sites at -215 and -134 had no effect (lane 4, Fig. 3B),

A		GSHV PROC
SL1(Nde,Mlu)	-215 cATATgAGAGATCAATTATTAACCTTTATGGGAGGAGGGTATCATCgcagatctgcGATCCTAGGCTGAAATTATTGTATTAGGAGGCTGTAcGCgT Nde -202 -188 BglII linker -158 -144 Mlu -134	80%
Δ IVS	cATATgAGAGATCAATTATTAACCTTTATG-----tcgcga-----TGAAATTATTGTATTAGGAGGCTGTAcGCgT	65%
Δ PS1(SL1)	cATA-----cGCgT	25%
SITE B MUTANTS		
Δ PS1A	cATA-----gatctgcGATCCTAGGCTGAAATTATTGTATTAGGAGGCTGTAcGCgT	50%
B1	cATA-----gatctgcGATCCTAGGCTGaggTTATTGTATTAGGAGGCTGTAcGCgT	50%
B2	cATA-----gatctgcGATCCTAGGCTGAAaggATTGTATTAGGAGGCTGTAcGCgT	30%
B3	cATA-----gatctgcGATCCTAGGCTGAAATTgcTTGTATTAGGAGGCTGTAcGCgT	50%
B4	cATA-----gatctgcGATCCTAGGCTGAAATTATccGTATTAGGAGGCTGTAcGCgT	30%
B5	cATA-----gatctgcGATCCTAGGCTGAAATTATTGccTTAGGAGGCTGTAcGCgT	30%
B6	cATA-----gatctgcGATCCTAGGCTGAAaggATccGTATTAGGAGGCTGTAcGCgT	25%
SITE A MUTANTS		
Δ PS1B	cATATgAGAGATCAATTATTAACCTTTATGGGAGGAGGGTATCATCgcagatctgcGATCCTAGGCTGAAaggATccGTATTAGGAGGCTGTAcGCgT	50%
Δ -170/-107	cATATgAGAGATCAATTATTAACCTTTATGGGAGGAGGGTATCATCgcagatctgc-----to -107	50%
A1	cATATgAGAGATCAA-----TGGGAGGAGGGTATCATCgcagatctgcGATCCTAGGCTGAAaggATccGTATTAGGAGGCTGTAcGCgT	40%
A2	TATATAAGAGATCAcTTAagAACTTTATGGGAGGAGGGTATCATCgcagatctgcGATCCTAGGCTGAAaggATccGTATTAGGAGGCTGTAGGCAT	50%
A3	TATATAAGAGATCAc---gAACTTTATGGGAGGAGGGTATCATCgcagatctgcGATCCTAGGCTGAAaggATccGTATTAGGAGGCTGTAGGCAT	40%
A4	cATATgAGAGATCAATTggggAACTTTATGGGAGGAGGGTATCATCgcagatctgcGATCCTAGGCTGAAaggATccGTATTAGGAGGCTGTAcGCgT	65%
A5	cATATgAGAGATCAATTATTAgtagcTGGGAGGAGGGTATCATCgcagatctgcGATCCTAGGCTGAAaggATccGTATTAGGAGGCTGTAcGCgT	50%

B

Figure 4. Detailed analysis of the two A/T-rich stretches (PS1A and PS1B) located within the PS1 domain. (A) For each mutant, the region from -215 to -134 is displayed. All plasmids contain, in addition, a deletion from -68 to -41 (SL1) which largely inactivates PS2 function. The 15 bp A/T-rich regions are shown in bold type while all nucleotide changes are indicated by lower case letters. Dashed lines specify deletions. The processing efficiencies listed to the right were based on the transcript analyses shown in (B).

		REFERENCE
PS1B	<u>GCTGAAATTATTGTATTAGGA</u>	this study
CaMV	<u>TTAGTATGTATTGTATTGTA</u>	(32)
CYC1-512-C	<u>TAGATATTATGTAAT</u>	(51)

Figure 5. Comparison of PS1B with known functional upstream polyadenylation elements from widely divergent cell types. In plants, the CaMV sequences shown are sufficient to partially activate a defective cellular pA site. Also shown is a sequence derived from a *CYC1-512* revertant which may confer efficient 3' end processing at a site approximately 100 nt downstream. This sequence matches a consensus motif (underlined) which is found at the 3' ends of many yeast mRNAs.

however, deletion of the region between them [Δ PS1(SL1)] resulted in a dramatic decrease in pA site use to 25%. This relatively large operating range (80% to 25%) was sufficient to differentiate between a variety of phenotypes and provided the rationale for conducting the investigation of PS1 in an SL1 background.

Although deletion of sequences from -215 to -134 (encompassing PS1A and PS1B) in Δ PS1(SL1) altered processing dramatically, it has already been demonstrated that they were insufficient for GSHV pA site activation in the SRC fusion experiments [SRCNM(2), Fig. 2] and required sequences extending to -107 [SRCNH(2), Fig. 2]. Deletion of the latter sequences alone (Δ -134/-105, Fig. 3) caused a reduction in processing to 60%, consistent with their apparent role in 3' end processing.

The parental construct for analyzing PS1 mutations, SL1, contains a 10 bp *Bgl*II linker inserted into the 29 bp intervening region between the A/T-rich segments of PS1A and PS1B (see Fig. 4A). This had no effect on GSHV pA site function and prompted me to determine whether the intervening region was dispensable. In Δ IVS (Figs. 3, 4A), 26 bp of this region was removed and replaced with a 6 bp *Nru*I site. The observed decrease in processing to 65% was minor compared to the further deletion of the A/T-rich regions [Δ PS1(SL1); Figs. 3, 4A], or to A1, which bears specific mutations within the A/T-rich regions.

The A/T-rich Regions of PS1A and PS1B are not Equivalent

To further facilitate the analysis of PS1 shown in Fig. 4, I chose to analyze PS1B mutants in the absence of PS1A and visa versa. PS1A was removed by a deletion from -215 to -171 (Δ PS1A) which resulted in a decrease of polyadenylation efficiency to 50%. I then set out to disrupt the A/T sequences of PS1B with G/C dinucleotide substitutions. Two mutations, B1 and B3, had no effect. In three mutants (B2, B4, and B5), processing levels decreased to 30% which approached the phenotype observed with a deletion of the entire region [Δ PS1(SL1); 25% processing]. A processing level of 25% was observed also with the introduction of four G/C nucleotide substitutions (mutant B6). In the presence of PS1A, the same four G/C substitutions decreased pA site use to 50% (see Δ PS1B). Importantly, Δ PS1B had a phenotype as severe as the 65 bp deletion in Δ -170/-107, and thus the further elimination of sequences from -134 to -107 did not influence processing in contrast to that observed for Δ -134/-105 (Fig. 3) with PS1B intact.

Δ PS1B was subsequently used for analyzing PS1A mutations (Fig. 4). The substitution of two (A2) or three (A5) G/C base pairs caused no discernible change in processing efficiency. A modest reduction to 40% was observed with the deletion of four bases in A2 to create A3, and was not further affected by a

virtually complete deletion of the A/T-rich stretch (A1). These values do not approach the phenotype observed with the larger 40 bp deletion (B6) and thus the A/T-rich stretch in PS1A, in contrast to PS1B, is not a major determinant of GSHV pA site function. Furthermore, the substitution of four G residues into this region (A4) actually increased processing to 65%, indicative of a sub-optimal signal.

DISCUSSION

I have more precisely characterized those GSHV upstream processing signals which are present only at the 3' end of genomic RNA (analogous to the U3 region of retroviruses). The results demonstrate that sequences from -400 to -107 augment the effects of PS2 (contained within the terminal repeats) to ensure high levels of cleavage/polyadenylation. In the absence of PS2, this region stimulates basal processing from 10% to 50%, reflecting the combined effect of PS1 (-215 to -107), which on its own confers 30% processing, and sequences extending upstream to -400. This confirms earlier findings (33) that the region immediately upstream of PS1 (termed PS3) contributes to GSHV pA site use. It is noteworthy that this region has not yet been shown to influence 3' end formation in the absence of PS1 and may simply be an extension of the diffuse PS1A element. For this reason I presently hesitate in designating the PS3 region as a distinct processing signal.

To add to the existing complexity, it is possible that sequences upstream of -400 either stimulate pA site use directly or enhance the effect of other accessory signals. This is based on the observation that the wild type construct displays more efficient processing than its counterpart in which sequences from -1800 to -400 have been replaced with a c-SRC cDNA (compare SRC-400, Fig. 1 to wild type, Fig. 3). SRC-containing transcripts remain unspliced in contrast to those containing only GSHV sequences and so it is possible that the removal of intervening sequences may further stimulate polyadenylation, a phenomenon which has been reported both *in vitro* (49) as well as *in vivo* (50). Alternatively, either c-SRC may harbor inhibitory sequences, or the SRC-GSHV chimeric transcripts may fold in a manner which is not optimal for the proper recognition of the processing signals.

The multiple processing signals of GSHV are clearly redundant in function. This explains the differential effects of deleting PS1 from wild type (>95% to 80%) as opposed to deleting the same region with PS2 disrupted (80% to 25%) (see Fig. 3). Also, PS2 can autonomously increase pA site use to 60% from 10%, although its removal in the presence of all other upstream signals results in a small decrease in processing to 80% (45).

In this report I have modified our original definition of PS1 to contain sequences from -215 to -107, the smallest fragment shown to activate polyadenylation in the described assay system (Fig. 2). PS1 is further comprised of two regions (PS1A and PS1B) which harbor 15 bp A/T-rich tracts, and a third region (between -134 and -107) which I will refer to as PS1C. The nature of PS1C is unclear. Deletion of this region alone results in a decrease in processing from 80% to 60% (Δ -134/-105, Fig. 3), consistent with its requirement in the reconstitution experiments. If PS1B is rendered non-functional, however, this region becomes completely dispensable (compare Δ -170/-107 with Δ PS1B, Fig. 4). One explanation is that PS1C serves a spacer function which separates PS1B from downstream elements such as PS2 in the wild type context, UAUAAA in SRCNH(1) (Fig. 2), or PS1A in the multimeric constructs

SRCNH(2) and SRCNH(3). However, attribution of this general interference observed with three different signals to improper spacing seems unlikely. An alternative model is that PS1B and PS1C display strict synergy in that one cannot function properly without the other. Thus the critical residues identified within the A/T-rich stretch could either (1) interact directly with sequences in PS1C to form a functional RNA structure or (2) be involved in the binding of a protein factor which interacts also with PS1C or another factor associated with PS1C.

The importance of the A/T-rich region in PS1B is indicated by the number of relatively subtle mutations which lead to severely hampered function. The situation with PS1A is quite distinct, whereby complete deletion of this stretch is only partially disruptive. This suggests a supporting or cooperative role for the surrounding sequences and, in fact, a 9 bp deletion within the polypurine tract immediately 3' decreases processing slightly (data not shown). This polypurine tract was also removed in Δ IVS (Fig. 3) implying that the observed decrease in processing from 80% to 65% cannot be totally due to the juxtaposition of the two A/T-rich sections. Whatever the contribution made by the A/T-rich sequences in PS1A, its high A/T content is not a significant factor since the substitution of four adjacent G/C base pairs increased processing (mutant A4, Fig. 4).

Had the two 15 bp A/T-rich stretches displayed similar involvement in 3' end processing, a common mechanism could have been proposed. Because they possess dissimilar primary sequences, this would have partially reconciled the apparent lack of sequence identity between upstream elements and would have implicated structural determinants for activation. This simple notion, however, does not seem to be the case. The results point to a divergent method of action for PS1A and PS1B/C. The extrapolation of this would be that PS2 and retroviral U3 regions, which show no sequence homology, utilize different pathways for the activation of polyadenylation. The enhancement of polyadenylation by a variety of means is underscored by the possibility that splicing influences GSHV 3' end formation.

Given the diversity of upstream elements within single viral genomes and a variety of animal viruses, coupled with the lack of definition of core elements in plants, it is remarkable that the A/T-rich sequences of PS1B share extensive homology to a region found immediately upstream of the pA site in the plant reovirus element CaMV. A 22 bp oligonucleotide corresponding to this region (shown in Fig. 5) can activate a completely defective heterologous plant pA site with 30% efficiency (32). These sequences, in conjunction with others farther upstream are responsible for efficient use of the CaMV pA site at the 3' end of genomic RNA. Also shown in Fig. 5 is the intriguing similarity between the upstream elements of GSHV and CaMV with one of the putative polyadenylation signals in *S. cerevisiae* derived from a *CYC1-512* revertant (51). In all cases the homologous sequences are characterized by a high A/T content with no C, and a limited number of G residues.

Basal processing at the GSHV pA site in the absence of upstream signals is very low (10%). This allowed me to test PS1, which by itself only partially activated processing, for the ability to stimulate polyadenylation using tandem multimeric constructs. Since PS1 is contained within 100 bp, use of multiple copies demands that the signals function at drastically altered distances from the core elements and further requires that they do not interfere with each other. These properties were met using three copies of PS1, the demonstration of which is novel for upstream elements, and is reminiscent of the action of transcriptional

activation by multiple promoter/enhancer elements. In contrast to transcriptional enhancers, PS1 functions in only one direction consistent with its recognition at the level of RNA.

The demonstration that high levels of 3' end processing can be obtained with multiple copies of an upstream element such as PS1 will greatly facilitate the analysis of cis-acting sequences in isolation as has been shown with SRCNH(3)A1 (Fig. 2). In addition, the properties displayed by PS1 suggest cooperative binding of some trans-acting processing factor(s), although a strictly RNA-mediated mechanism cannot be dismissed. In this case, the use of multimeric constructs will also aid in the identification and characterization of these factors, which should prove to be an interesting area of future research.

ACKNOWLEDGMENTS

I am grateful to the members of the Finnegan's Wake Social Club for their support and helpful discussions. I would also like to thank Marlys Koschinsky, Marty Petkovich, and Chris Mueller for their critical review of the manuscript. Special thanks to Don Ganem in whose laboratory the present work was carried out. This work was supported by the Medical Research Council of Canada.

REFERENCES

1. Munroe, D. and Jacobson, A. (1990) *Gene* 91, 151–158.
2. Jackson, R.J. and Standart, N. (1990) *Cell* 62, 15–24.
3. Bernstein, P. and Ross, J. (1989) *Trends Biochem. Sci.* 14, 373–377.
4. Sachs, A.B. and Davis, R.W. (1989) *Cell* 58, 857–867.
5. Sachs, A.B. and Davis, R.W. (1990) *Science* 247, 1077–1079.
6. Varnum, S.M. and Wormington, W.M. (1990) *Genes Dev.* 4, 2278–2286.
7. Fox, C.A. and Wickens, M. (1990) *Genes Dev.* 4, 2287–2288.
8. McGrew, L.L. and Richter, J.D. (1990) *EMBO J.* 9, 3743–3751.
9. Wickens, M. (1990) *Trends Biochem. Sci.* 15, 320–324.
10. Vassalli, J.-D., Huarte, J., Belin, D., Gubler, P., Vassalli, A., O'Connell, M.L., Parton, L.A., Rickles, R.J. and Strickland, S. (1989) *Genes Dev.* 3, 2163–2171.
11. Robinson, B.G., Frim, D.M., Schwartz, W.J. and Majzoub, J.A. (1988) *Science* 241, 342–344.
12. Friedman, D.L., Imperiale, M.J. and Adhya, S.L. (1987) *Annu. Rev. Genet.* 21, 453–488.
13. Levitt, N., Briggs, D., Gil, A. and Proudfoot, N.J. (1989) *Genes Dev.* 3, 1019–1025.
14. Gilmartin, G. and Nevins, J.R. (1991) *Mol. Cell. Biol.* 11, 2432–2438.
15. Moore, C.L., Chen, J. and Whoriskey, J. (1988) *EMBO J.* 7, 3159–3169.
16. Gilmartin, G. and Nevins, J.R. (1989) *Genes Dev.* 3, 2180–2189.
17. Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J. and Shenk, T. (1990) *Genes Dev.* 4, 2112–2120.
18. Weiss, E.A., Gilmartin, G.M. and Nevins, J.R. (1991) *EMBO J.* 10, 215–219.
19. Wilusz, J., Shenk, T., Takagaki, Y. and Manley, J.L. (1990) *Mol. Cell. Biol.* 10, 1244–1248.
20. Wilusz, J. and Shenk, T. (1988) *Cell* 52, 221–228.
21. Ahmed, Y.F., Gilmartin, G.M., Hanly, S.M., Nevins, J.R. and Greene, W.C. (1991) *Cell* 64, 727–737.
22. McDevitt, M.A., Hart, R.P., Wong, W.W. and Nevins, J.R. (1986) *EMBO J.* 5, 2907–2913.
23. Gil, A. and Proudfoot, N.J. (1987) *Cell* 49, 399–406.
24. Heath, C.V., Denome, R.M. and Cole, C.N. (1990) *J. Biol. Chem.* 265, 9098–9104.
25. Takagaki, Y., Ryner, L.C. and Manley, J.L. (1988) *Cell* 52, 731–742.
26. Christafari, G. and Keller, W. (1989) *Mol. Cell. Biol.* 9, 193–203.
27. Terns, M.P. and Jacob, S.T. (1989) *Mol. Cell. Biol.* 9, 1435–1444.
28. Sheets, M.D. and Wickens, M. (1989) *Genes Dev.* 3, 1401–1412.
29. Wickens, M. (1990) *Trends Biochem. Sci.* 15, 277–281.
30. Carswell, S. and Alwine, J.C. (1989) *Mol. Cell. Biol.* 9, 4248–4258.
31. DeZazzo, J.D. and Imperiale, M.J. (1989) *Mol. Cell. Biol.* 9, 4951–4961.
32. Sanfaçon, H., Brodmann, P. and Hohn, T. (1991) *Genes Dev.* 5, 141–149.
33. Russnak, R. and Ganem, D. (1990) *Genes Dev.* 4, 764–776.

34. DeZazzo, J.D., Kilpatrick, J.E. and Imperiale, M.J. (1991) *Mol. Cell. Biol.* 11, 1624–1630.
35. Valsamakis, A., Zeichner, S., Carswell, S. and Alwine, J.C. (1991) *Proc. Natl. Acad. Sci. USA* 88, 2108–2112.
36. Brown, P.H., Tiley, L.S. and Cullen, B.R. (1991) *J. Virol.* 65, 3340–3343.
37. Cherrington, J. and Ganem, D., unpublished.
38. Dougherty, J.P. and Temin, H.M. (1987) *Proc. Natl. Acad. Sci. USA* 84, 1197–1201.
39. Coffin, J.M. and Moore, C. (1990) *Trends Genet.* 6, 276–277.
40. Proudfoot, N. (1991) *Cell* 64, 671–674.
41. Imperiale, M.J. and DeZazzo, J.D. (1991) *New Biol.* 3, 531–537.
42. Sheets, M., Ogg, S.C. and Wickens, M.P. (1990) *Nucleic Acids Res.* 18, 5799–5805.
43. Wilusz, J., Pettine, S.M. and Shenk, T. (1989) *Nucleic Acids Res.* 17, 3899–3908.
44. McLauchlan, J., Simpson, S. and Clements, B.J. (1989) *Cell* 59, 1093–1105.
45. Cherrington, J., Russnak, R. and Ganem, D., unpublished.
46. Hirai, H. and Varmus, H.E. (1990) *Mol. Cell. Biol.* 10, 1307–1318.
47. Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) *Methods Enzymol.* 154, 376–382.
48. Simonsen, C.C. and Levinson, A.D. (1983) *Mol. Cell. Biol.* 3, 2250–2258.
49. Niwa, M., Rose, S.D. and Berget, S.M. (1990) *Genes Dev.* 4, 1552–1559.
50. Huang, M.T.F. and Gorman, C.M. (1990) *Nucleic Acids Res.* 18, 937–947.
51. Russo, P., Li, W.-Z., Hampsey, D.M., Zaret, K.S. and Sherman, F. (1991) *EMBO J.* 10, 563–571.