A Noncanonical Poly(A) Signal, UAUAUA, and Flanking Elements in Epstein–Barr Virus DNA Polymerase mRNA Function in Cleavage and Polyadenylation Assays

S. Catherine Silver Key*† and Joseph S. Pagano*‡†1

*Departments of Microbiology and Immunology, †Department of Medicine, and ‡UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill Chapel Hill, North Carolina 27599

Received April 2, 1997; returned to author for revision May 7, 1997; accepted May 30, 1997

Two forms of the Epstein–Barr virus DNA polymerase (pol) mRNA (3.7 and 5.1 kb) have been detected, neither of which contains a canonical poly(A) signal. The 5.1-kb pol mRNA, which contains a rare poly(A) signal, UAUAUA, studied only in transcripts of Hepadnaviridae and a plant pararetrovirus, was analyzed in cleavage and polyadenylation assays. Incubation of the pol transcript in cell extracts produced relatively low efficiency of cleavage (12 to 14%), which was improved by conversion of the poly(A) signal to AAUAAA. Deletion of the UAUAUA signal abolished cleavage and polyadenylation. An auxiliary element, UUUGUA, 3–8 nt upstream of the poly(A) signal and two downstream core elements, a GU-rich sequence 36–46 nt, and an AUUUGUGU sequence 47–53 nt downstream of the signal (8–19 nt and 20–28 nt downstream of cleavage site) facilitated processing of pol mRNA. Replacement of sequences near the cleavage/poly(A) site affected cleavage accuracy. Binding of the 64-kDa cleavage stimulatory factor to the U-rich as well as the GU-rich elements correlated with cleavage efficiency. Thus the UAUAUA hexanucleotide plus the other cis-acting elements are clearly functional in the native pol mRNA, but are relatively inefficient. Implications of the use of an anomalous poly(A) signal and its elements are discussed.

INTRODUCTION

Epstein–Barr virus (EBV) is the etiologic agent of both infectious mononucleosis, which involves primary infection of oropharyngeal epithelial cells and B lymphocytes, and hairy leukoplakia of the tongue (Wall and Raab-Traub, 1994; zur Hausen et al., 1978). The virus also causes lymphoproliferative conditions in immunocompromised persons, usually through reactivation of latently infected lymphocytes (Purtlio et al., 1978; Sugden, 1982; Edwards and Raab-Traub, 1994; Katz et al., 1989; Rea et al., 1994). In both primary infection and viral reactivation a virally encoded DNA polymerase (pol) together with a set of replication cofactors are required for viral replication (Fixman et al., 1992, 1995). Since in latent infection none of these early genes is expressed, the mechanisms through which they are suppressed and activated are important for an understanding of both viral latency and replication. We have selected the EBV DNA pol gene for detailed analysis of how expression of this key, presumably representative, early gene is regulated. Previous studies have addressed regulation of the pol gene at the promoter level (Furnari et al., 1992; Liu et al., 1996). In this work we examine unusual elements of the EBV pol mRNA that affect 3′ processing of the EBV pol mRNA.

The EBV pol is a 110-kDa protein encoded by the BALF5 ORF (Lin et al., 1991; Chen et al., 1995; Kiehl and Dorsky, 1995; Tsurumi et al., 1993). Initiation of pol mRNA transcription is at tightly clustered sites in conjunction with a TATA-less promoter that is constitutively inactive (Furnari et al., 1992). Activation of the EBV pol promoter requires the EBV immediate-early protein, BRLF-1 (R), which appears to act through at least two cellular factors, USF and E2F, in B lymphocytes (Liu et al., 1996).

Furnari et al. have reported that two transcripts of EBV DNA pol are detected, the 3.7- and 5.1-kb messages, both unspliced (Furnari et al., 1993). This report focuses on the 5.1-kb message because the 3.7-kb form was detected only in the B95-8 cell line, whereas the 5.1-kb mRNA was found in all other cell lines tested, including directly in hairy leukoplakia tissue. The 5.1-kb mRNA appears to be the generally authentic form; it is precisely cleaved and polyadenylated at a position corresponding to 183 nt downstream of map coordinate 152,012 in the prototype virus B95-8 (this region is deleted in B95-8) (Furnari et al., 1993; Parker et al., 1990). Cleavage occurs after a CA dinucleotide, 27 nucleotides downstream of the putative poly(A) signal. Since this sequence containing the cleavage and poly(A) site is deleted from the B95-8 genome, the pol mRNA in B95-8 is processed further upstream at map coordinate 153,215 resulting in the 3.7-kb message (Furnari et al., 1993). Both message forms are remarkable in that cleavage and polyadenylation appear to be accomplished without canonical poly(A)
signals. Although there is no recognizable poly(A) signal and the adenylation appears faulty, cleavage of the 3.7-kb mRNA is nevertheless precise. After the cleavage site a varying number of extra nucleotides, mainly A's and U's with occasional G's, none represented in the genome, are added before conventional polyadenylation (Furnari et al., 1993). In contrast, polyadenylation of the 5.1-kb pol mRNA appeared to proceed normally. However, a rarely documented noncanonical poly(A) signal, AAUAAA, rather than the canonical AAUAAA signal appeared to be used for processing of this mRNA.

The canonical poly(A) signal, AAUAAA, which is located 10–30 nt upstream of the 3' cleavage site, is found in 90% of eukaryotic messages and has been studied in vivo (Fox et al., 1989; Paris and Richter, 1990; Sheets et al., 1995) and extensively in vitro (Sheets et al., 1990; Wigley et al., 1990; Wilusz et al., 1989). Some variation of this signal is tolerated: AAUAAA (EBV-BMRF2), AGUAAA, AAUACA (keratin), AAUAAA (Factor IX), AAUAAU (IGF-1), AAUGAA (rat TGFα), and others are functional (Sheets et al., 1990; Wilusz et al., 1989). However, in vitro alteration of a single nucleotide at several positions within the AAUAAA signal results in diminished processing efficiency. Specifically, transcripts with mutation to a UAUA poly(A) signal are processed much less efficiently (17 ± 3.0% of wild type) (Sheets et al., 1990; Wilusz et al., 1989). In vivo in Xenopus oocytes point mutation of the conserved hexanucleotide abolished mRNA processing (Fox et al., 1992; Sheets et al., 1994).

In the 5.1-kb pol mRNA, a AAUAAA sequence is positioned 27 nt upstream of the cleavage/poly(A) site. The rarity of this sequence as a poly(A) signal made it quite uncertain whether it was functional. AAUAAA as a poly(A) signal has been reported in only three other eukaryotic mRNAs: hepatitis B viruses including ground squirrel hepatitis B surface antigen and human hepatitis B X transcripts (Perfumo et al., 1992; Russnak and Ganem, 1990; Russnak, 1991; Simonsen and Levinson, 1983), and the figwort mosaic virus nopaline synthase mRNA (Sanfalcon, 1994). Examples of the AAUAAA poly(A) signal in the sequence data base are rare indeed (0.8% AAUAAA compared with 98% AAUAAA of 269 vertebrate cDNA sequences from the EMBL data bank (Sheets et al., 1990). Except for the GSHVsAg and the FMV nos processing signals, these signals have not been analyzed in their native context.

Elements flanking the AAUAAA and AAUAAA poly(A) signals may differ and are important for 3' processing. A G/U-rich or U-rich downstream element (DSE) is positioned 0–10 nt downstream of the cleavage site in most mammalian transcripts (for a review see Manley, 1988; Keller, 1995). Positioning of a U-rich element is important for cleavage precision and efficiency (Chen et al., 1995). Upstream elements (USE), usually AU-rich or U-rich, also contribute to 3' processing, including the consensus AUAUGUGRA found in SV40 and the HIV pentanucleotide and a consensus UUUGUA in plant viruses (Russnak and Ganem, 1990; Schek et al., 1992; Sanfalcon, 1994; Sanfalcon et al., 1991; Valsamakis et al., 1992, 1991).

In this paper we use extracts from a variety of cell lines including latently EBV-infected cells to analyze processing of the 5.1-kb EBV pol transcript. Analysis of mRNA processing in a cell-free system appears to reflect in vivo processing accurately (Fox et al., 1992; Moore and Sharp, 1984). This study is the first to analyze in detail the elements that govern 3' processing of a gammaherpesvirus transcript. We show that the AAUAAA poly(A) signal, while it is atypical, is used for cleavage and polyadenylation of the EBV pol mRNA, but is inefficient. We also identify other cis-acting elements, including a USE reported in plant viruses, that contribute to cleavage and processing. Finally, we identify where the cleavage stimulatory factor (CSF) 64-kDa protein, which is required for cleavage, binds to the pol mRNA.

MATERIALS AND METHODS

Cell lines

D98-HR1 (latently infected epithelial-lymphoid somatic-cell hybrid line created by fusion of the EBV-infected Burkitt lymphoblastoid cell line, P3J-HR-1, and a HeLa variant, D98), (Glaser and Nonoyama, 1974; Glaser and Rapp, 1972), and latently infected Burkitt's lymphoma (BL) lines: Akata, (Takada et al., 1991), Raji, Daudi, and B95-8, an EBV-producing marmoset lymphoid line, were propagated in RPMI-1640 with 10% fetal calf serum (FCS) and penicillin-streptomycin. HeLa S3, an epithelial-like line adapted for growth in suspension, was grown in suspension-minimum essential medium (S-MEM) with 5% horse serum and 5% FCS and antibiotics.

Nuclear extracts

HeLa, S3, and D98-HR1 cells were grown in spinner flasks to a density of 5–9 × 10⁶ cells/ml. Extracts were made using a modified Dignam procedure (Dignam et al., 1983). Cells were swollen in hypotonic buffer A for 10 min on ice. For lymphoblastoid cells 1 mM spermidine was added to protect nuclear membrane integrity. Epithelial cells were homogenized 10× and lymphoblastoid cells 5× in a 7-ml Dounce tube (Wheaton) with an A pestle. Nuclei were pelleted for 5 min at 2000 rpm with 8% sucrose added for lymphoblastoid cell nuclei. Buffer C was divided into NaCl low salt (100 mM) and NaCl high salt (420 mM). Nuclei were gently resuspended in low salt buffer C and brought up to 0.35 M with high salt buffer C. Extraction proceeded for 30 min with gentle stirring at 4°. Dialysis was not performed.
Constructs

The 193nt-wtPA construct contains 183 bp of DNA surrounding and including the UAUAAA sequence directionally cloned into the HindIII and EcoRI sites of the 3.2-kb pBS KS− vector (Stratagene). A 313-bp region of EBV DNA was amplified from induced B95-8 poly(A) selected mRNA using primers with appropriate restriction sites for directional cloning: the 5′ HindIII primer (5′-ATC ACG AGA AGC TTC TCG GCC TAC GAG GCC CGC GC-3') and the 3′ EcoRI primer (5′-ATC ACG AGG AAT TCT GCC CCT CGG CCC GTG CC-3'), and subcloned into pBS− to create p313wtPA (Furnari, 1993b). Removal of the HindIII/BglII fragment followed by blunt-end ligation generated the p193nt-wtPA construct. Sequence analysis was performed to verify the integrity of the sequence using the UNC-CH Automated DNA Sequencing Facility or the Sequenase Version 2.0 protocol (U.S. Biochemical Corp.).

Linker-substitution mutants were generated with an oligonucleotide-directed method (Kunkel, 1985). The p193nt-U2, p193nt-U1, p193nt-D1, p193nt-D2, p193nt-D3, p193nt-D4, p193nt-D5, p193nt-GUD6 (D6*), and p193nt-D6 were generated by replacing 12-bp regions with the CiaI and HindIII restriction sites (ATCGAT and AAGCTT) as described in the T7-Gen In Vitro Mutagenesis kit (U.S. Biochemical Corporation) using the double-stranded wild-type plasmid with helper virus, M13KO7, (Bio-Rad) to generate single-stranded DNA. The oligonucleotides used for mutagenesis were as follows: p313wtPA-U5 (5′-CGG CGC CTG CGC AAG CCT ACT GAT TAA AGA TCG-3'), p313wtPA-U4 (5′-CCT GTA AAG ATG AAG CTT ACT GAT CCC GCT GCA TC-3'), p313wtPA-U3 (5′-TAA AGA TCT ACC GAT CCT ACT GAT TAA AGA TCG-3'), p313wtPA-U2/p193nt-U2 (5′-CCG CCT GCA TCT AAG CCT ACT GAT CCA GCT GCA TC-3'), p313wtPA-U1/p193nt-U1 (5′-TCC AGG TGC ACG AAG CTT ATC GAT TAA CTA TAA AGA TCG-3'), p193nt-D1 (5′-TGT GTC TCA YYY YYG TAG GTG CGG CTA-3'), p193nt-D2 (5′-AAG ACA AGC CTG ATC GAT AAG CTT CCT CAC CTC-3'), p193nt-D3 (5′-CTG TGC TCA AAT AAG CTT ATC GAT TGT CTC TGT CTC-3'), p313wtPA/p193nt-D4 (5′-CCT CAC GAC CTG AAG CTT ATC GAT TGT GATT GGT G-3'), p313wtPA/p193nt-D5 (5′-CTG TGT TGC TCA GAT GCT CCG CTT ATC TAT TGT GT GGT G-3'), p313wtPA/p193nt-D6 (5′-ATC ACG AGA AGC TTC TCG GCC TAC GAG GCC CGC GC-3').

Radioactive labeling of RNA templates

EBV (p193nt-wtPA and variants), HSV-2 (pSau5), and SV40 (pSVL) constructs were linearized with T3 RNA polymerase (for EBV transcripts) or SP6 RNA polymerase (for HSV-2 and SV40 transcripts) (Promega) in the presence of [32P]UTP (Amersham) and 7Me-GTP (Pharmacia) for HSV-2 and SV40 transcripts) (Promega) in the presence of [32P]UTP (Amersham) and 7Me-GTP (Pharmacia) at 30°C for 1 hr resulting in 193-nt EBV pol, 400-nt HSV-2 IE2, and 224-nt SV40 late transcript. Transcripts were partially purified through G-50 Sephadex columns (Boehringer-Mannheim). The first 10 nucleotides within the EBV transcript were derived from the pBS KS− vector (AACAAAGCT).

In vitro cleavage and polyadenylation assays

Labeled transcripts (50 to 200 cps/μl) were incubated at 30°C typically for 50 min in nuclear extracts in the presence or absence of 1 mM 3′-deoxyadenosine-5′-triphosphate (cordycepin; Boehringer Mannheim or Sigma) for the cleavage reaction. Adenosine triphosphate (1 mM) replaced cordycepin in the polyadenylation assays. Conditions for the assays have been described (McLauchlan et al., 1992). After incubation reaction mixtures were extracted once with phenol-chloroform-isooamyl alcohol and precipitated overnight at −70°C.

Quantitation

Products were separated through 7 M urea, 8% polyacrylamide gels, and detected by autoradiography or with a PhosphorImager (Molecular Dynamics). The efficiency of processing was estimated with a PhosphorImaging program, Image Quant 3, by dividing the counts per minute (cpm) of the products by the sum of cpm contained in the products and input transcript ×100. Cleavage and poly(A) product cpm measurements were multiplied by appropriate factors for each transcript variant to adjust
for the number of U residues found in the input transcript and products.

UV-crosslinking and immunoprecipitation

Transcripts were incubated in polyadenylation conditions for 10 min at 30°C, exposed to 312 nm light (Spectroline, model EB-280C) for 3 min at 4°C, and then treated with RNaseA for 15 min at 37°C. Complexes were immuno-

TABLE 1

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Cleavage efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeLa</td>
<td>14.5 ± 2.6%</td>
</tr>
<tr>
<td>D98-HR1</td>
<td>14.7 ± 1.8%</td>
</tr>
<tr>
<td>Raji</td>
<td>14.3 ± 3.4%</td>
</tr>
<tr>
<td>B95-8</td>
<td>12.5 ± 2.7%</td>
</tr>
<tr>
<td>Akata</td>
<td>14.2 ± 2.3%</td>
</tr>
<tr>
<td>Daudi</td>
<td>14.0 ± 3.9%</td>
</tr>
</tbody>
</table>

Values were assigned relative to the EBV input with the use of the ImageQuant 3 program after PhosphorImager analyses; averages of five independent reactions for each cell type.

The UAUAAA poly(A) signal is essential for processing of the EBV DNA polymerase mRNA

To verify that the UAUAAA sequence is functional and needed, it was replaced with a HindIII (AAGCUU) restriction site. Both cleavage and polyadenylation were abolished (Fig. 2A, lane 3, and Fig. 2B, lane 4), whereas the wild-type transcript was processed as before (Fig. 2A, lane 2, and Fig. 2B, lane 3). Use of an oligo(dT) column (Qiagen Oligotex system) confirmed that poly(A) products were generated from the EBV pol 193-nt transcript. Thus the UAUAAA sequence functions as a cleavage and poly(A) signal in the context of the wild-type EBV DNA pol mRNA.

Conversion of the noncanonical poly(A) signal, UAUAAA, to a canonical signal, AAUAAA, increases processing efficiency.

Next we replaced the first uridine in the EBV signal with an adenosine. Two clones (AAUAAA.1 and .2) were generated, sequenced, and tested in HeLa nuclear extracts. The AAUAAA-substituted EBV 193-nt input transcripts yielded visibly increased amounts of both cleavage and poly(A) products (Fig. 3, compare lane 1 with 2 and 3 and lane 5 with 6 and 7). The 3′ cleavage product was again detected and was more abundant, presumably due to increased cleavage efficiency (lanes 6 and 7); some degradation of this product was visible in lanes 2 and 3. The wild-type cleavage and poly(A) products appear less abundant in Fig. 3 because of a shorter exposure used for quantitation (compare Fig. 1). Additionally, it is obvious that the EBV transcript is polyadenylated less efficiently than the SV40 transcript (Fig. 3, compare lanes 6 and 7 to 8), which suggests that other cis-acting elements are needed to achieve full efficiency.
FIG. 1. Cleavage and polyadenylation of an AAUAAA-containing HSV 2 transcript and a UAUAAA-containing EBV transcript in vitro. (A) Schematic of EBV-derived 193-nt transcript (see Materials and Methods) depicting poly(A) signal and surrounding elements. U and D indicate nucleotide positions upstream and downstream of the poly(A) signal. A 193-nt transcript results in a 93-nt 5' cleavage product. (B) HSV 2 (Sau5) template (lanes 1–5) and the EBV p193 nt-wtPA template (lanes 6–10) were transcribed in vitro in the presence of [32P]UTP with SP6 and T3 RNA polymerase, respectively. Transcripts were incubated with D98-HR1 nuclear extracts for 1 hr in the presence of cordycepin for cleavage (lanes 1 and 6); or for 30 min (lanes 3 and 8), 1 hr (lanes 4 and 9), or 2 hr (lanes 5 and 10) in the presence of ATP for polyadenylation. Nuclear extract was omitted in lanes 2 and 7.

elements in the EBV pol mRNA may contribute to inefficient processing.

The efficiency of cleavage and polyadenylation was calculated with the use of the Image Quant 3 program. A transcript containing SV40 processing signals including AAUAAA generated from pSVL was used as a control (gift of J. Wilusz and T. Shenk). Processing efficiencies calculated for the late SV40 transcripts were 73.1% for cleavage and 76.5% for polyadenylation (Fig. 3, lanes 4 and 8). For wild-type EBV transcript 11.0 ± 1.2% cleavage and 13.0 ± 1.3% polyadenylation efficiencies were calculated, whereas the AAUAAA-substituted EBV transcript resulted in 28.1 ± 6.5% cleavage and 33.8 ± 5.4% polyadenylation. Although the poly(A) products in the gel overlap the input mRNA region as determined by oligo(dT) selection analysis, only the adenylated products migrating above the input mRNA were used for determining polyadenylation efficiency, resulting in an underestimation. Thus, at least a twofold increase in processing efficiency resulted from the conversion.

Cis-acting elements upstream and downstream of the EBV DNA polymerase poly(A) signal are important for processing efficiency

The RNA sequence surrounding the defined cleavage and poly(A) site in the 5.1-kb pol mRNA contains three potential cis-acting elements: GU-rich, AUUUGUGA and AUUUGUA. Normally located 0–10 nt downstream of the cleavage site, a potential GU-rich element is located 8–19 nt downstream of the EBV pol cleavage site (36–46 nt downstream of the AAUAAA signal) (Fig. 4A). A sequence that matches the upstream SV40 AUUUGUGA element in 7 of 8 nucleotides (AUUUGUGU) was placed 20–28 nt downstream of the cleavage site (47–53 nt relative to the poly(A) signal). This EBV sequence may also represent a U-rich element (URE), which is a pentanucleotide with 4 of the 5 residues being uridines (Fig. 4A). Finally, the putative AUUUGUA element positioned 3–8 nt upstream of the AAUAAA signal in the pol transcript is similar to multiple elements upstream of the plant-pararetrovirus nopaline synthase (nos) mRNA poly(A) signal (Fig. 4A).

To test whether these three elements (shown in Fig. 4A) function in processing of the EBV pol mRNA, each was replaced with HindIII and ClaI restriction sites. Cleavage assays were performed in triplicate with D98-HR1 nuclear extracts and the wild-type and mutated transcripts. A typical assay is shown in Fig. 4B; all the results are summarized in Fig. 4C. Replacement of either the GU-rich (D4) or AUUUGUGA (D5) elements abolished the appearance of the cleavage product (Fig. 4B, lanes 3 and 4 compared with lane 2), indicating that both are needed for pro-
The UAUAAA sequence from the EBV 5.1-kb polymerase message is inefficiently used as a poly(A) signal in vitro. (A) The UAUAAA sequence was replaced with a HindIII restriction site (AAGCUU). The 193-nt input transcript was transcribed using the T3 RNA polymerase in the presence of [32P]UTP. Transcripts (1 μl, 50 cps/μl) were tested in vitro for cleavage (lanes 1 and 3) in HeLa cell nuclear extracts for 30 min. Control lanes (2 and 4) lacked exogenous nucleotides. (B) A similar experiment except that ATP was added to assess polyadenylation. The reaction in lane 2 lacks extract.

Quantitation of results of three experiments disclosed cleavage efficiencies of 0.8 ± 0.1% (D4) and 1.8 ± 0.2% (D5) compared with 13.3 ± 1.4% for the wild type (WT, Fig. 4C). Replacement of the UUUGUA (U1) immediately upstream of the poly(A) signal also reduced cleavage efficiency (4.0 ± 0.1%) (Fig. 4B, lane 1, and Fig. 4C).

To confirm that the absence of the elements and not the introduced sequence caused the differences in cleavage efficiency, a downstream region that should not be important for processing was targeted. Four clones were tested; results with two are shown (Fig. 4B, lanes 5 and 6). One of the mutations (D6) did not affect cleavage efficiency, 12.5 ± 0.8%, as expected. However, the transcript from the other clone (D6*) yielded less cleavage product, 4.0 ± 0.5%. Sequencing of the two clones explained this discrepancy. The D6* template contained a point mutation in the GU-rich element, U to G 18 nt downstream from the cleavage and poly(A) site (45 nt relative to the poly(A) signal), which was presumably responsible for the threefold drop in cleavage efficiency (Fig. 4C, D6* column). This point mutant confirmed the importance of the GU-rich element. Each transcript was also tested for polyadenylation efficiency that paralleled the effects on cleavage (data not shown). Thus three elements flanking the UAUAAA poly(A) signal upstream and downstream may contribute to cleavage/polyadenylation efficiency: the GU-rich element, which is conven-
FIG. 4. Mutational analyses of EBV DNA polymerase transcript in the in vitro cleavage assay. (A) Schematic diagram illustrating the cis-acting elements in the p193nt-wPA transcript. The substituted sequences are shown in bold letters. The ClaI and HindIII linker-substitutions are below the appropriate sequences (italics). The mutants are designated: U1, D4, D5, and D6. The position of the single point mutation is also shown above the EBV sequence (*). U indicates the sequence is upstream of the UAUAAA poly(A) signal, and D, downstream; arrow, cleavage site. (B) Linker-substitution mutants were transcribed in the presence of [32P]UTP and tested for cleavage efficiency: mRNA (1 μl, 200 cps/μl) was incubated in D98-HR1 nuclear extracts for 1 hr at 30°C. The last lane shows the wild-type transcript from p193nt-wPA in the absence of cordycepin or ATP (WT−). (C) Averages from three cleavage assays.

The downstream AUUUGURA sequence functions as a U-rich element

Introduction of single and multiple point mutations into the D5 sequence suggested that it might function like the U-rich element in the late SV40 transcript (Chen et al., 1995). Several mutations, shown in Table 2, that diminished the number of U’s decreased cleavage efficiency (Fig. 5A, lanes 1, 5, and 6; Fig. 5B, lanes 2 and 3), and the mutation that increased the number of U’s, DU16, enhanced cleavage efficiency (Fig. 5A, compare lane 3 to 7). The smaller product near the bottom of the gel is nonspecific, since it is present when no extract is added (data not shown). Most of the mutations affected cleavage as expected for a URE, but a few did not. Mutants DU18 and DU21 decreased cleavage efficiency, but not as dramatically as similar mutations, for example, DU2 and DU14 (Table 2). Sequence analysis verified that no other elements had been altered. When the first nucleotide of the AUUUGUGU sequence was changed to C and one GU was deleted, DU11, there was a dramatic reduction in cleavage efficiency (Fig. 5A, lane 4). One other nucleotide change was detected in the DU11 mutant, but it is far downstream of the cleavage site and should not affect interaction of the processing complex.

<table>
<thead>
<tr>
<th>DNA sequence</th>
<th>Mutant designation</th>
<th>Cleavage efficiency</th>
<th>Binding of CSF 64-kDa protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTTGT</td>
<td>WT</td>
<td>12.1 ± 4.6%</td>
<td>++</td>
</tr>
<tr>
<td>ATCCTC</td>
<td>DU2</td>
<td>0.9 ± 1.7%</td>
<td>−</td>
</tr>
<tr>
<td>ATCTCT</td>
<td>DU3</td>
<td>12.4 ± 2.7%</td>
<td>++</td>
</tr>
<tr>
<td>ATTTCT</td>
<td>DU6</td>
<td>9.5 ± 2.0%</td>
<td>++</td>
</tr>
<tr>
<td>CTTTGT</td>
<td>DU11</td>
<td>1.5 ± 0.55%</td>
<td>−</td>
</tr>
<tr>
<td>ACCCTTC</td>
<td>DU14</td>
<td>1.0 ± 2.2%</td>
<td>−</td>
</tr>
<tr>
<td>ACCCCCT</td>
<td>DU15</td>
<td>1.7 ± 4.1%</td>
<td>−</td>
</tr>
<tr>
<td>ATTTTT</td>
<td>DU16</td>
<td>15.7 ± 4.6%</td>
<td>++</td>
</tr>
<tr>
<td>ATCCCT</td>
<td>DU18</td>
<td>4.8 ± 2.7%</td>
<td>+</td>
</tr>
<tr>
<td>ATCCCT</td>
<td>DU21</td>
<td>5.5 ± 5.2%</td>
<td>+/-</td>
</tr>
<tr>
<td>ATCTTT</td>
<td>DU32</td>
<td>6.9 ± 2.7%</td>
<td>+/−</td>
</tr>
</tbody>
</table>

* Bold letters indicate mutated nucleotides.

* Calculated as described under Materials and Methods. Average of three cleavage reactions.

* Binding reactions as described in the legend to Fig. 8.
FIG. 5. Multiple or single point mutations introduced in the AUUUGURA downstream element affect cleavage. 1 µl (200 cps/µl) of each 313-nt transcript was incubated with HeLa nuclear extracts for 1 hr. (A) Reaction products with DU2, DU3, DU16, DU11, DU14, DU15, WT, or the D5 mutant transcripts were separated on an 8% 7 M urea polyacrylamide gel. (B) Reactions with DU6, DU18, DU21, DU32, WT, or D5 transcripts.

with the EBV pol transcript since this portion can be deleted without changing cleavage efficiency (data not shown). Since the DU11 mutation did not alter the URE-like portion of the sequence, the 5' flanking context of the URE may be important for function of the element.

Further mutational analysis of sequences flanking the UAUAAA poly(A) signal

To complete the analysis additional upstream and downstream sequences that bore no resemblance to re-

FIG. 6. Replacement of two downstream sequences affects cleavage precision. (A) Replaced EBV sequences are shown above the line in bold capital letters. The positions of the ClaI and HindIII linker-substitutions are shown below the line in their positions relative to the EBV pol sequences (italics). Mutant designations U5, U4, U3, U2, U1, D1, D2, and D3, are placed above the first of the replaced EBV nucleotides. The large arrow indicates the cleavage site. (B) Linker-substituted labeled transcripts (1 µl at 50 cps/µl) were incubated for 30 min at 30°C in HeLa nuclear extracts in the presence of cordycepin. (C) Average of three separate experiments with the linker mutants (U2, D1, D2, and D3) compared with wild type, WT.
ported cis-acting elements were targeted for mutation (Fig. 6A). Also, the cleavage/poly(A) site was replaced with a linker sequence. Mutation of the cleavage site (D3) or a position immediately 3' to the poly(A) signal (D1) affected the accuracy (Figs. 6B and 6C, lanes 1 and 3), but not efficiency of cleavage significantly: 10.4 ± 5.2% and 12.2 ± 6.0%, respectively. Cleavage product sizes were estimated using a DNA sequencing ladder and labeled run-off transcripts of 90 and 97 nt (data not shown). Instead of the 93-nt cleavage product, a faster migrating (~89 nt) and possibly a more slowly migrating cleavage product (~96 nt) were detected for D3. Because cleavage typically occurs after an A residue (Chen et al., 1995), we searched the sequence data for potential cleavage sites. The cleavage products correspond well to cleavage occurring after the first A in the Clai restriction site (89-nt product) and the second A in the HindIII restriction site (96-nt product). The normal 93-nt product as well as an 87-nt cleavage product were detected with D1 corresponding to the last A in the triplet which lies 6 nt upstream of the cleavage site in the wild-type sequence. A partial cleavage shift in the D1 variant was unexpected because the cleavage site was unaltered, and the distance between the site and the cis-acting elements also remained constant; this result has been obtained repeatedly. Therefore, the context immediately surrounding the cleavage site in EBV pol mRNA may help to determine precision of cleavage.

Mutations in positions U2 and D2 yielded an average threefold reduction in processing efficiency, 4.2 ± 1.8% and 4.4 ± 0.4%, compared with wild type: 14.1 ± 5.7% (Figs. 6C and 6B). No obvious elements can be recognized at these positions (Fig. 6A). A mutation here could change the secondary structure, or the processing proteins may require these flanking sequences for optimal binding.

A DNA template, p313wtPA, which contained more sequence 5' of the TATAAA element than p193nt-wtPA, was used to generate mutations of sequences 30-66 bp upstream (Fig. 6A). The 313wtPA transcript was effectively cleaved and polyadenylated similar to its smaller counterpart, 193nt-wtPA (data not shown). The set of mutations created confirmed that the upstream UUUGUA element contributed to efficient cleavage of the polymerase transcript. Further, it again suggested that other sequences, namely at the U2 and U3 positions, may play an auxiliary role in processing (Fig. 7A, lanes 3 and 4). Although lane 3 was under-loaded in this experiment, the average cleavage efficiency with U3 from three experiments was lower (5.7 ± 3.7%) than with the wild-type transcript (12.1 ± 3.7%) (Fig. 7B). However, sequences at positions U4 (12.5 ± 3.9%) and U5 (8.6 ± 3.9%) did not appear to affect cleavage efficiency significantly (Fig. 7A, lanes 1 and 2, and Fig. 7B, WT versus U4 and U5). Thus, sequences that do not correspond to known cis-elements may contribute to interaction of processing factors or to RNA secondary structure needed to ensure accuracy and efficiency of cleavage.

The CSF 64-kDa protein binds the pol mRNA in UV-crosslinking and immunoprecipitation assays

Since eukaryotic cleavage reactions require the cleavage stimulatory factor (CSF) complex, and the CSF 64-kDa protein appears to bind weakly to a UAUAAA-containing transcript (Wilusz et al., 1989), we wanted to determine whether and where the CSF 64-kDa protein binds to the pol mRNA. Therefore, poly(A) reaction mixtures were UV-crosslinked and immunoprecipitated with the monoclonal antibody to the 64-kDa factor (3A7, kindly provided by C. MacDonald) with both wild-type and mutated EBV pol transcripts. Assays were standardized using the transcript made from the pSVL construct (data not shown) (Chen et al., 1995). CSF protein–RNA complex was readily detected when the wild-type EBV transcript was used as the RNA target (Fig. 8A, WT). No complex was detected with vector alone (data not shown). Mutants D6 and D2 formed complexes well with the 64-kDa protein (Fig. 8B). D6*, which contains a point mutation in the GU-rich element, appeared to complex with the 64-kDa protein less well. Interestingly, little CSF–RNA complex was detected with the D3 mutant, which alters the cleavage site and yields two products in cleavage assays. Also, there was little CSF–RNA complex formation with the D4 mutant, which replaces the downstream GU-rich element (Fig. 8A, lane D4), and little to no complex with the D5 mutant, which replaces the U-rich element (Fig. 8B, lane D5, and Fig. 8A, lane D5). Thus, in EBV pol mRNA the putative U-rich element (URE) appears to be needed for and the GU-rich element contributes to binding of the CSF 64-kDa protein.

Since the protein appeared to bind preferentially to the URE, we next introduced point and multiple mutations into the putative URE 21–26 nt downstream of the EBV pol cleavage site targeting this element and then compared binding of the protein and cleavage efficiencies (Figs. 5 and 8B, and summarized in Table 2). The binding of the 64-kDa protein correlates well with the ability of the RNA to be cleaved in vitro. Basically, if the URE remains intact, the 64-kDa protein binds and cleavage occurs. However, the DU11 mutation, which alters the nucleotide immediately proceeding the URE, but retains the UUUGUA sequence, is reproducibly unable to interact with the 64-kDa protein (Fig. 8C, lane DU11); nor is cleavage produced (Fig. 5A, lane 4). Thus, the results again indicate that the GU-rich element is important for interaction of the 64-kDa factor. In summary, CSF does bind to this native UAUAAA-containing transcript, especially to the URE, but also to the GU-rich element, and is presumably required for cleavage.
DISCUSSION

Usage of a UAUAAA poly(A) signal is exceedingly rare in eukaryotic mRNAs. Changing the initial A to U in SV40 reduced efficiency of processing of mRNA by 80% or more (Sheets et al., 1990). Here we show that the UAUAAA signal in EBV DNA polymerase mRNA is functional and essential for cleavage and polyadenylation, but inefficient as predicted by earlier in vitro mutations (Sheets et al., 1990). Identified also are cis-acting elements flanking the UAUAAA signal: a UUUGUA upstream element (USE) and downstream elements (DSE) including a GU-rich and a U-rich element (URE). The upstream UUUGUA has been detected previously only in plant virus mRNAs. Also accuracy of cleavage is affected by sequences immediately downstream of the poly(A) signal. Conversion by mutation of a canonical EBV signal in the context of three elements improves cleavage efficiency.

Altering the poly(A) signal from UAUAAA to AAUAAA in GSHV in an SV40 chimeric construct with the GU-rich element intact produced a modest improvement in processing shown by Northern analysis (Russnak and Ganem, 1990). With the EBV transcript, replacement of the initial U with A produced a significant increase (two-fold). This increase may be due to the ability of the 160-kDa factor of the Cleavage and Polyadenylation Specificity Complex (CPSF) to bind with higher affinity to the AAUAAA than the UAUAAA poly(A) signal (Sheets et al., 1990; Wilusz et al., 1989; Jenny and Keller, 1995). The Cleavage Stimulatory Factor (CSF) 64-kDa protein also binds to transcripts containing the conserved hexanucleotide (Wilusz et al., 1989), preferring to bind to a URE 10–30 nt downstream of the poly(A) site (MacDonald et al., 1994). CSF and CPSF complexes are thought to interact; the positioning and sequence of the elements surrounding the AAUAAA signal influence the degree of communication between CSF and CPSF and hence help to determine which set of sequences constitute a 3' processing site (Gilmartin and Nevins, 1989). mRNA secondary structure is also likely to affect processing (Sittler et al., 1995; Gravely et al., 1996). Binding of CSF to the WT pol transcript was detected reproducibly, was abolished by linker-substitution of the UAUAAA, and was enhanced by conversion to AAUAAA (data not shown).

Upstream elements similar to the EBV UUUGUA are critical for processing UAUAAA-containing transcripts in the GSHV and FMV systems. In addition to the downstream G/U-rich stretch and UAUAAA signal the

---

**FIG. 7.** Linker-scanner mutational analysis of sequences further upstream of the UAUAAA signal. (A) Polyacrylamide gel analysis of cleavage reactions using mutant transcripts (1 μl at 200 cps/μl) incubated in HeLa nuclear extracts for 1 hr. p313wtPA (lane 9) was incubated under cleavage conditions in the absence of nuclear extract (WT*). (B) Summary of the average cleavage efficiency of each transcript from three data sets.

**FIG. 8.** The cleavage stimulatory factor 64-kDa protein - EBV pol RNA complex is immunoprecipitated from UV-crosslinked poly(A) reactions. After immunoprecipitation with mAb 3A7, 64-kDa – RNA complexes of linker-substitution (A) or point and multiple mutations in the UUUGUA sequence (B) were separated using 10% SDS - PAGE and visualized by autoradiography. WT, D4, D5, D6, and D6* are as described in the legend of Fig. 4A. DU2, DU3, DU6, DU11, DU14, DU15, DU16, DU18, DU21, and DU32 are as shown in Table 2.
GSHVsAg transcript requires two AU-rich upstream elements, designated processing signal I (PSI) A and PSI B (Russnak, 1991). These elements serve a specialized function in HBV. Transcription of the HBV virion-associated DNA polymerase mRNA begins upstream of the UAAAAA signal, but the message is not processed until the entire genome is transcribed and the PSI elements are encountered. In the FMV nos transcript, which does not contain a DSE such as the GU-U rich element, processing appears to rely entirely on the presence of seven repeats of the UUGUA element upstream of the UAUAAA (Sanfacon, 1994). When the single perfect copy of the UUGUA element was mutated in the EBV DNA poly A transcript, a threefold reduction in processing efficiency resulted. Although the EBV upstream element resembles the HSV PSI and the plant pararetrovirus repeat elements, the UUGUA in the EBV poly A transcript appears to play an auxiliary rather than a critical role in regulating processing.

Typically, the poly(A) signal works coordinately with a core element downstream of the cleavage/poly(A) site. McDevitt et al. reported that sequences 35 nt distal to the poly(A) addition site were sufficient for 3' end formation of the adenovirus E2A mRNA which contains an AAUAAA poly(A) signal (McDevitt et al., 1984). M. Lauchlin et al. identified a consensus for the downstream element, YGTGTTYY, in herpes simplex virus 1 (HSV-1) and 2 (HSV-2) IE-5 transcript that was present in 67% of the then known mammalian 3’ termini (M. Lauchlan et al., 1985). EBV pol contains a GU-rich element located 8–19 nt downstream of the cleavage/poly(A) site (36–46 nt downstream of UAAAAA), which we show is essential for processing, and has some requirement for the U residue 18 nt downstream of the cleavage site since point mutation of that U greatly diminished cleavage efficiency. However, the EBV sequence does not conform to the consensus described for HSV, nor has this consensus sequence been confirmed in other work (Goodwin and Rottman, 1992; Mason et al., 1986).

Recently, another downstream element has been identified. The U-rich element (URE) is generally located 10–30 nt past the cleavage/poly(A) site and is defined as a stretch of five U residues (Chen et al., 1995). This element is reported to serve as a recognition site for binding of the 64-kDa member of the cleavage stimulatory factor (CSF) complex, an essential step in cleavage (Gilmartin and Nevins, 1989; Keller, 1995; Manley and Takagaki, 1996; Proudfoot, 1996). Interestingly, the downstream UUUUGUGU in EBV has a U-rich stretch, and total replacement of this element significantly reduces processing efficiency. Introduction of single and multiple point mutations in the downstream EBV AUUUGUGU sequence gave results expected of a mutated URE (Chou et al., 1994). Indeed the U residues in the downstream AUUUGUGA fall 21 to 25 nt downstream of the cleavage site (48–52 nt downstream of UAAAAA) in perfect agreement with the model proposed for UREs (Chen et al., 1995). However, both the reduction in cleavage efficiency and the inability of the 64-kDa protein to interact with D6* and DU11 may suggest that the GU-rich region is also required for the 64-kDa protein to bind. In the EBV transcript, both elements appear to cooperate.

Cleavage and polyadenylation normally occur after a CA (Sheets et al., 1990). Mutations at and near the conserved adenosine at the poly(A) site affect accuracy but not efficiency of 3' processing (Sheets et al., 1990). Cleavage occurs no closer than 11 nt and no farther than 23 nt downstream of the poly(A) signal with few exceptions (Chen et al., 1995; Wigley et al., 1990). In EBV pol the cleavage site, which is preceded by a CA dinucleotide, is 27 nt downstream of the UAAAAA poly(A) signal. Mutations of the cleavage site produced two products, neither of which represented the original cleavage product size, as expected. Interestingly, when the 10-nt sequence immediately downstream of the UAAAAA signal was replaced, a partial shift in cleavage-site usage was observed although the distance between the poly(A) signal and the cleavage/poly(A) site was not altered. Introduced or removed secondary structure may account for this result.

The question arises why the anomalous poly(A) signal and accompanying elements that function inefficiently are found in a transcript of a critical replicative gene. Since ordinarily all vertebrate mRNAs are efficiently processed by constitutive cellular factors, our findings open the possibility that virus infection may compensate for this inefficiency. In studies in progress we are determining whether early viral proteins enhance the efficiency of processing. Likely candidates are the products of the EBV BSLF2/BMLF1 (SM/M) early genes. These proteins have potential RNA-binding domains in their N-terminus and have been implicated in posttranscriptional regulation (Buisson et al., 1989; Cook et al., 1994; Furnari, 1993a; Kenney et al., 1989). Additionally, the M genes have homology to the HSV ICP27 (IE63) gene which has also been implicated in regulating 3' processing (Markovitz et al., 1989; McLauchlan et al., 1992; Sandri-Goldin and Mendoza, 1992; McGregor et al., 1996). Studies in progress should show whether SM/M proteins, by interacting with EBV pol mRNA and cellular processing proteins, may stabilize their interaction and improve efficiency of polyadenylation and cleavage dictated by the UAAAAA signal and its surrounding elements.

ACKNOWLEDGMENTS

This work was supported by National Institute of Health Grant 5-R01-CA56695 from the National Cancer Institute. We thank Zbigniew Dominski, Ryszard Kole, and William Marzluff for advice on preparing nuclear extracts and for providing reagents. We are grateful to Clinton MacDonald, Jeffrey Wilusz, Tom Shenk, and Barklie Clements for providing reagents.
REFERENCES


