MOLECULAR AND CELLULAR BIOLOGY, Oct. 1991, p. 5312–5320
0270-7306/91/105312-09$02.00/0

Vol. 11, No. 10

An RNA-Binding Protein Specifically Interacts with a Functionally Important Domain of the Downstream Element of the Simian Virus 40 Late Polyadenylation Signal

ZUWEI QIAN AND JEFFREY WILUSZ*

Department of Microbiology and Molecular Genetics, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, 185 South Orange Avenue, Newark, New Jersey 07103

Received 19 April 1991/Accepted 23 July 1991

We have identified an RNA-binding protein which interacts with the downstream element of the simian virus 40 late polyadenylation signal in a sequence-specific manner. A partially purified 50-kDa protein, which we have named DSEF-1, retains RNA-binding specificity as assayed by band shift and UV cross-linking analyses. RNA footprinting assays, using end-labeled RNA ladder fragments in conjunction with native gel electrophoresis, have identified the DSEF-1 binding site as 5′-GGGGGAGGUGUGGG-3′. This 14-base sequence serves as an efficient DSEF-1 binding site when placed within a GEM4 polylinker-derived RNA. Finally, the DSEF-1 binding site restored efficient in vitro 3′ end processing to derivatives of the simian virus 40 late polyadenylation signal in which it substituted for the entire downstream region. DSEF-1, therefore, may be a sequence-specific binding factor which regulates the efficiency of polyadenylation site usage.

Maturation of the 3′ end of most RNA polymerase II transcripts involves a site-specific endonucleolytic cleavage event followed by the polymerization of 150 to 200 adenylate residues (reviewed in references 19 and 42). Polyadenylation, which also plays a role in the termination of transcription (18, 41), provides a potential site for the posttranscriptional regulation of gene expression (17). The regulated use of poly(A) sites during viral infections (10, 23, 40) and the examples of interplay between poly(A) site and splice site choices (11, 27) provide evidence for the significance of the polyadenylation process in the pathway of gene expression.

Biochemical fractionation of activities involved in the enzymatic events of 3′ end processing has identified several *trans*-acting enzymatic and specificity factors (3, 13, 37, 38). *trans*-acting regulatory factors, however, remain largely undefined. Furthermore, since polyadenylation signals containing minimal amounts of sequence information are efficiently processed in fractionated in vitro systems (43), such regulatory factors may not be readily identified by this approach.

The *cis*-acting sequence elements which comprise the poly(A) signal are fairly well characterized. A highly conserved hexanucleotide, AAUAAA, located 5 to 30 bases upstream of the poly(A)/cleavage site (34, 46), confers specificity onto the polyadenylation process. Since the AAUAAA motif is a generic part of the poly(A) signal, it is unlikely that signal-specific factors involved in processing efficiency would directly act through this element. There are at least two additional sequence elements which influence poly(A) signal usage. First, additional upstream elements have recently been found for several poly(A) signals (2, 6, 30). Second, sequences located 5 to 55 bases downstream of the poly(A)/cleavage site have been identified for numerous signals. Deletion of this downstream element (DSE) results in a three- to fivefold reduction in processing efficiency (4, 5, 12, 15, 21, 31–33). DSEs do not contain a well-conserved consensus sequence, although GU- and U-rich regions have been noted (12, 22, 29). Fine deletion mapping and analysis of point mutations have been generally unsuccessful for the detailed mapping of DSE sequences (20, 50). This has led to the suggestion that a DSE is composed of a single large diffuse element or one that is reiterated.

Because of its complex and relatively unconserved sequence, the DSE may be a target for signal-specific, *trans*-acting factors which influence poly(A) site usage. We have previously reported that downstream sequences from several independent polyadenylation signals are required for efficient cross-linking to the heterogeneous nuclear ribonucleoprotein (hnRNP) C proteins (45). These UV cross-linking results may reflect the sequence-directed alignment of the C proteins of an hnRNP complex (reviewed in reference 9) or perhaps hnRNP phasing. We have recently identified the sequence UUUUU as an hnRNP C protein alignment element. This UUUUU element was able to mediate efficient polyadenylation when it substituted for the entire downstream region of three independent poly(A) signals (48). The hnRNP complex, therefore, may play a general role in 3′ end processing efficiency, perhaps by melting out RNA structures or interactions which are inhibitory to the efficient recognition of the polyadenylation signal by the 25S-50S precleavage complex (16, 26, 36, 49, 51).

In this report, we show that the UUUUU element of the DSE of the simian virus 40 (SV40) late polyadenylation signal (SVL) can be disrupted without affecting 3′ end processing efficiency. This result is consistent with the redundant nature of the DSE described previously. We have identified and purified a second factor which interacts with the 25-base SVL DSE. This 50-kDa protein, called DSEF-1 (for downstream element factor 1), interacts with the sequence 5′-GGGGGAGGUGUGGG-3′ in a highly specific manner. The DSEF-1 binding site increased 3′ end processing efficiency when placed in the downstream region of a polyadenylation signal. We propose that the 50-kDa protein may influence 3′ end processing through a sequence-specific protein-RNA interaction.

* Corresponding author.

MATERIALS AND METHODS

Plasmids and RNAs. All RNAs were transcribed in vitro by using SP6 polymerase and [^{32}P]UTP or [^{32}P]GTP and gel purified as described previously (48).

Transcripts were derived as follows. pSVL contains the *Bam*HI-*Bcl*I fragment of SV40 inserted into the *Bam*HI site of pSP65 (Promega). Transcription of *Dra*I- and *Fnu*4HI-linearized templates, respectively, yields RNAs of 224 bases (SVL) and 149 bases (SVL2). pSVL3 contains the SV40-specific *Alu*I-to-*Hin*dIII fragment of pSVL inserted into pGEM4 (Promega) at the *Hin*cII and *Hin*dIII sites. Transcription of *Dra*I-cut template yields a 122 base-RNA (SVL3). The construction of pSVL6, which contains a U-to-G transversion in the AAUAAA motif, has been described previously (45). Transcription of *Dra*I-cut DNA gives a 111-base RNA (SVL6). The construction of pSVL8, which contains downstream sequences from +8 to +55 relative to the cleavage site fused to the *Eco*RI site of pGEM4, has been previously described (45). Transcription of *Dra*I-cut DNA gives a 61-base RNA (SVL8). pSVL12 and pSVL14, in which SVL3- or SVL-specific sequences downstream of +29 are removed and substituted by vector-derived sequences, were constructed as described previously (45). Transcription of *Hin*fI-cleaved pSVL12 template yields a 123-base RNA (SVL12), while transcription of *Hin*fI-cleaved pSVL14 gives a 223-base RNA (SVL14). pSVL17, a derivative of pSVL which contains a U-to-G transversion at +52, was constructed in two steps. First, synthetic oligonucleotides containing the desired point mutation were inserted between the *Bsm*I and *Hin*dIII sites of pSVL3-S/H (45). Second, the *Hpa*I-to-*Hin*dIII fragment from this pSVL3 derivative was used to replace the *Hpa*I-to-*Hin*dIII fragment of pSVL. Transcription of *Dra*I-cut template yields a 223-base RNA containing the indicated point mutation (SVL17). pC220 contains a 1.38-kb *Eco*RI-to-*Hin*dIII fragment of bacteriophage lambda cloned into the same sites of pSP64 (Promega). Transcription of *Sca*I-cut DNA yields a 220-base control RNA (CON).

Transcription of *Hin*dIII-linearized pGEM4 gives a 65-base RNA (GEM4). pGEM50 was constructed by inserting the oligonucleotide 5′-GATCCGGGGGAGGTGTGGG-3′ and its appropriate complement into pGEM4 DNA cut with *Bam*HI and *Sal*I. The resulting plasmid (pGEM50-BS) was then cut with *Eco*RI and *Sma*I, treated with Klenow polymerase, and recircularized. Transcription of *Hin*dIII-cut DNA gives a 59-base RNA (GEM50). The identity of all plasmids derived by using synthetic oligonucleotides was confirmed by chemical sequence analysis (data not shown).

Polyadenylation signals containing substitutions for downstream sequences were constructed as follows. pSVL-GEM4 was prepared by inserting the *Eco*RI-to-*Hin*dIII fragment of pGEM4 into *Bsm*I-*Hin*dIII-cut pSVL. Transcription of *Hin*dIII-linearized template yields a 228-base RNA (SVL-GEM4). pSVL-GEM50 was constructed in a similar fashion except that pGEM50-BS was the source of the *Eco*RI-to-*Hin*dIII fragment. Transcription of *Hin*dIII-linearized DNA yields a 237-base RNA (SVL-GEM50).

The 24-base DSE-specific fragment was prepared from SVL8 RNA by complete digestion with RNase CL3. The resulting 24-base fragment was purified on an acrylamide gel prior to use.

In vitro polyadenylation/cleavage assays. Polyadenylation reactions were performed by using the in vitro system of Moore and Sharp (25) as described previously (45). Cleavage reactions were performed by adding the α,β-methylene analog of ATP ($AMPCH_2PP$) and EDTA to final concentrations of 1 mM to the reaction mixture. Purified RNA products were analyzed on 5% acrylamide gels containing 7 M urea. All reactions were performed with equimolar amounts of RNA and assayed in the linear range of product accumulation.

DSEF-1 purification. Nuclear salt wash extracts were prepared from 30 liters of HeLa spinner cells grown in 10% horse serum as previously described (8). Cleared, dialyzed extracts were loaded onto an S-Sepharose (Pharmacia) column in buffer A (100 mM NaCl, 10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.9], 10% glycerol, 0.5 mM phenylmethylsulfonyl fluoride, 0.5 mM dithiothreitol). DSEF-1 was detected in the flowthrough fraction by UV cross-linking analysis. This flowthrough fraction was loaded directly onto a DEAE-5PW HPLC column (Pharmacia). The flowthrough fraction from this anion-exchange column, which contained DSEF-1, was then loaded directly onto a poly(G) agarose column (Sigma). The column was washed with buffer A and eluted by the stepwise addition of buffer A containing 300 mM NaCl or 1 M NaCl. The 1 M NaCl fraction containing DSEF-1 was diluted to 100 mM NaCl with buffer A without salt and rechromatographed on the poly(G) agarose column as outlined above. The resulting 1 M NaCl fraction was collected, dialyzed against buffer A containing 20% glycerol, and stored at −80°C. DSEF-1 was the major species of approximately 10 silver-stained bands on sodium dodecyl sulfate (SDS)-acrylamide gels (data not shown). DSEF-1 preparations were stable under these storage conditions.

Analysis of protein-RNA interactions. UV cross-linking/label transfer analysis was performed as previously described (45). Band shift analysis was performed by incubating 10 to 50 fmol of ^{32}P-labeled RNA with partially purified DSEF-1 protein in the presence of 3% polyvinyl alcohol, 10 mM HEPES (pH 7.9), 20 mM phosphocreatine, 1 mM ATP, 50 mM NaCl, 0.1 mM EDTA, 0.25 mM dithiothreitol, and 10% glycerol for 5 min at 30°C. Heparin was added to a final concentration of 3 μg/μl and incubated for 5 min on ice to remove nonspecific protein-RNA interactions. One microliter of loading buffer (20% glycerol, 0.5% xylene cyanol, 0.5% bromophenol blue) was added, and samples were electrophoresed in a 4% acrylamide gel (1:40 or 1:80 bisacrylamide/acrylamide) in Tris-borate-EDTA. Gels were dried, and complexes were detected by autoradiography.

RNA footprinting. SVL8 RNA was labeled at its 5′ end with polynucleotide kinase or at its 3′ end with pCp and T4 RNA ligase (1) and purified on an acrylamide gel. RNA ladders were prepared by limited digestion with 100 U of RNase T_1 at 30°C for 3 min. Ladder fragments were incubated with DSEF-1, and native gel band shift assays were performed as described above. Protein-RNA complexes were detected by autoradiography, and the corresponding bands were excised and electroeluted. Recovered RNAs were purified, concentrated with ethanol, and run next to the original T_1 ladder fragments on an 8% (for 5′-labeled RNA) or 15% (for 3′-labeled RNA) acrylamide gel containing 7 M urea.

RESULTS

The DSE of SVL is redundant. We have previously mapped the SVL DSE in our in vitro system to a 25-base segment from +30 to +55 relative to the cleavage site (45). When this segment was replaced with plasmid-derived sequences, polyadenylation efficiency was decreased by four- to fivefold

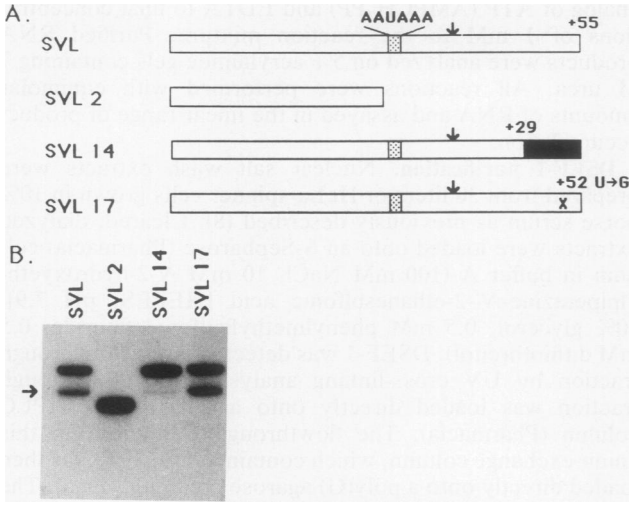

FIG. 1. Evidence that the SVL DSE consists of more than the UUUUU hnRNP cross-linking element. (A) Diagrammatic representation of the SVL and its derivatives. The AAUAAA is indicated by the light box. The arrow signifies the cleavage site. Nucleotide positions are given relative to the cleavage site. Shaded areas designate non-SVL, vector-derived sequences. (B) Cleavage assay. RNAs were incubated in the in vitro polyadenylation system to assess site-specific cleavage efficiency. Products were analyzed on a 5% acrylamide gel containing urea. Bands representing the 5′ cleavage product are indicated by the arrow.

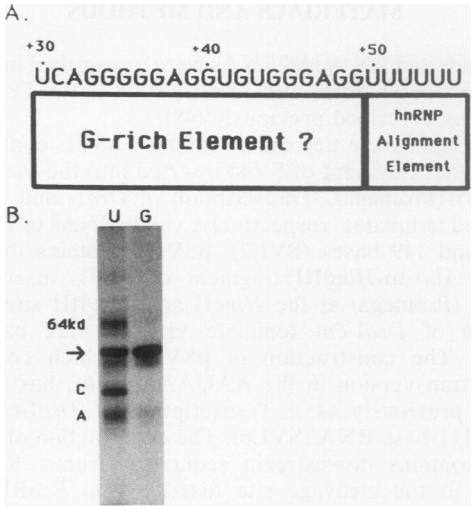

FIG. 2. Identification of a 50-kDa protein which cross-links to SVL RNA. (A) Sequence of the +30 to +55 SVL DSE, which can be divided into two domains based on sequence content. (B) Cross-linking assay. SVL3 RNA (diagrammed in Fig. 3) was radiolabeled at either U (lane U) or G (lane G) residues and incubated in the in vitro polyadenylation system. Protein-RNA interactions were assayed by UV cross-linking/label transfer analysis. Cross-linked proteins were identified on a 10% acrylamide gel with SDS. The positions of the AAUAAA-dependent 64-kDa protein and the hnRNP A and C proteins are indicated. The arrow shows the position of the 50-kDa protein.

(Fig. 1; SVL versus SVL14). Cleavage assays, which are performed with use of EDTA, were used since SVL variants with downstream substitutions are unstable in our extracts in the presence of Mg^{2+} (45, 48). We have recently identified a UUUUU element which lies within the SVL DSE at +50 to +55. This element can function as a DSE when placed within the context of polylinker-derived sequences located downstream of three independent polyadenylation signals (48). Furthermore, we have shown that the UUUUU element causes sequence-specific alignment of the C proteins within an hnRNP complex such that they become efficiently cross-linked by UV light. Analysis of deletion and point mutants indicates that the UUUUU element mediates efficient C protein cross-linking in the context of the SVL DSE. These data strongly suggest that one role of the DSE in polyadenylation efficiency may be mediated through a sequence-directed hnRNP complex-RNA interaction.

To evaluate the role of the UUUUU element in the natural context of the SVL DSE, we used site-directed mutagenesis to make a single U-to-G transversion at position +52 of the DSE. This point mutation changes the UUUUU element to UUGUU, effectively reducing its ability to serve as an hnRNP alignment element (48). On the basis of these observations, one might predict that the point mutation in the UUUUU element should render the SVL DSE ineffective, resulting in reduced processing efficiency compared with the wild type. As seen in Fig. 1B, however, the point mutant (SVL17) was processed as efficiently as wild-type SVL RNA. The functional domain of the SVL DSE, therefore, does not lie exclusively within the UUUUU element. In other words, within the 25-base SVL DSE are sequence elements which are functionally redundant. This result is consistent with earlier observations made with several independent DSEs (12, 20, 50).

Identification of second SVL DSE-specific RNA-protein interaction. The SVL DSE can be divided into two domains based on nucleotide content (Fig. 2A). The first domain, located from +33 to +49 relative to the cleavage site, is G rich (13 of 17 bases). This G-rich domain may correspond to the GU-rich consensus sequence found in other DSEs (29). The second domain is the UUUUU hnRNP C protein alignment element (+50 to +55), which may correspond to the U-rich consensus element. Since the SVL DSE appears to be functionally redundant (Fig. 1), we searched for a factor in addition to C proteins which could interact with the DSE, perhaps with the G-rich domain.

We performed UV cross-linking/label transfer analysis (45) using SVL3 RNA labeled only at G residues as a substrate. Since 45% of the G residues in the SVL3 transcript are in the DSE, this labeling bias would favor the detection of proteins which interacted with the G-rich domain. As shown in Fig. 2B, a 50-kDa protein was nearly exclusively detected by UV cross-linking analysis using G-labeled SVL3 RNA. The only other species labeled by this procedure was the hnRNP A proteins, which were shown previously to cross-link to RNAs in a nonspecific manner (48). The 50-kDa protein was also detected with U- and A-labeled SVL3 RNAs but not with C-labeled transcripts (Fig. 2B; data not shown). Thus, the 50-kDa protein is a potential DSE-binding factor.

The specificity of individual factors is difficult to assess when they are part of a large, multicomponent system like the 25S-50S poly(A) signal recognition complex (16, 26, 36, 51). To further characterize the interaction of the 50-kDa protein with the SVL poly(A) signal, therefore, purification of the factor was undertaken (Fig. 3A). The 50-kDa protein was present in the flowthrough of all ion-exchange columns used but bound relatively tightly to a poly(G) agarose column, requiring greater than 400 mM NaCl for elution.

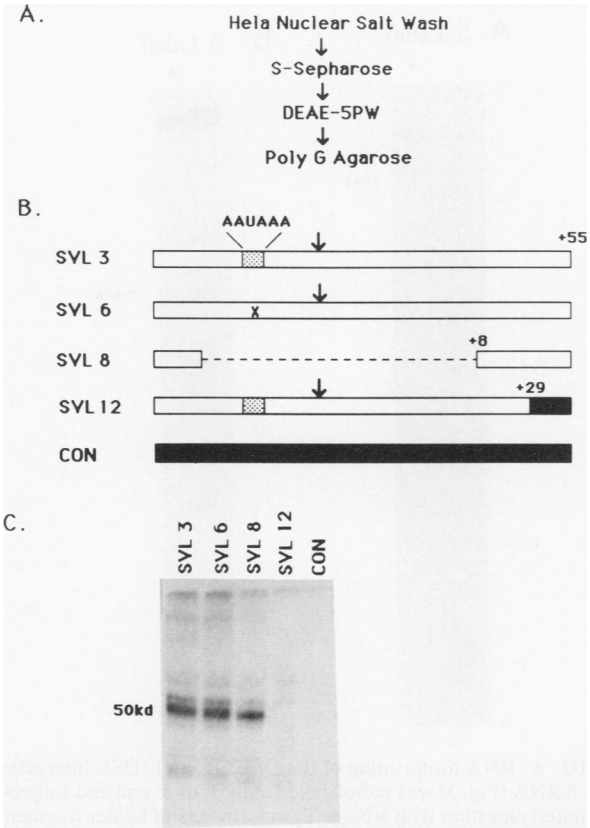

FIG. 3. Evidence that the partially purified 50-kDa protein requires SVL DSE-specific RNA sequences for efficient cross-linking. (A) Purification scheme used to isolate the 50-kDa protein. (B) Diagrammatic representation of derivatives of the SVL RNA. CON RNA is a control transcript derived from bacteriophage lambda sequences. Symbols and numbering are as described for Fig. 1. (C) UV cross-linking analysis using the partially purified 50-kDa protein and the indicated SVL derivatives radiolabeled at U residues. Cross-linked proteins were analyzed in a 10% acrylamide gel containing SDS.

This combination of conventional and affinity approaches provided a fast and efficient purification scheme. The 50-kDa protein was the major species of approximately 10 silver-stained bands when the poly(G) agarose fraction was assayed on an SDS-acrylamide gel (data not shown). While the possibility remains that some of our results reflect the properties of these minor components, all of the activity described below consistently copurified only with the 50-kDa protein.

As seen in Fig. 3C, the partially purified 50-kDa protein retained the ability to be efficiently cross-linked to specific RNAs (diagrammed in Fig. 3B). SVL derivatives containing an intact DSE (SVL3 and SVL6) and a 63-base RNA containing just the downstream region of SVL (SVL8) were efficiently cross-linked to the 50-kDa protein. SVL12 RNA, which contains plasmid-derived sequences in place of the 25-base DSE of SVL3 RNA, failed to be efficiently cross-linked to the 50-kDa protein, as was a 220-base RNA (CON) derived from bacteriophage lambda. The relative purity of the 50-kDa protein is also highlighted in this experiment, since cross-linking substrates were labeled at U residues (compare SVL3 in Fig. 2B and 3C). We conclude that the partially purified 50-kDa protein requires SVL DSE sequences to efficiently cross-link to RNA.

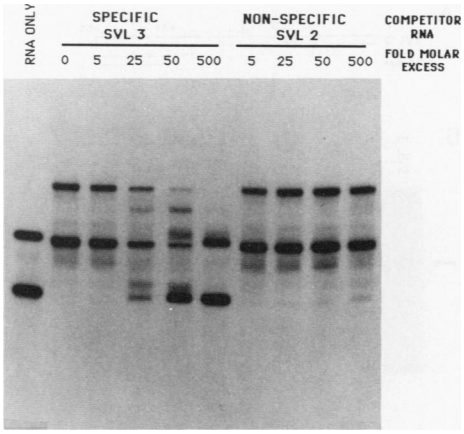

FIG. 4. Native gel band shift analysis showing the sequence specificity of the 50-kDa protein-RNA interaction. Radiolabeled SVL3 RNA (Fig. 3) was incubated with the 50-kDa protein in the presence of the indicated amounts of homologous (SVL3) or heterologous (SVL2; Fig. 1) cold competitor RNA. Heparin-resistant complexes were analyzed on a 4% native acrylamide gel. SVL3 RNA in the absence of protein runs as a doublet in this gel system (lane RNA Only).

The 50-kDa protein binds specifically to the G-rich domain of the SVL DSE. The UV cross-linking assay used to assess the specificity of the 50-kDa protein-RNA interaction in Fig. 3 is an indirect method which is not easily applied to quantitative analysis. In addition, since certain bases are more photoreactive than others (35), the technique may have some innate bias to the protein-RNA interactions that it detects. To directly assess the binding specificity of the 50-kDa protein, therefore, we used a native gel band shift approach. Most of single-strand nucleic acid binding activities (which make band shift assays using RNA and crude extracts difficult) have been removed during the purification of the 50-kDa protein.

Radiolabeled SVL3 RNA (diagrammed in Fig. 3) ran as a doublet on a 4% native acrylamide gel in the absence of protein (Fig. 4; lane RNA Only). If the RNA was denatured in urea prior to loading, it comigrated with the fast band of the doublet (data not shown). Undetermined higher-order RNA structure, therefore, probably caused the aberrant gel migration of SVL3 RNA. When the 50-kDa protein was incubated with SVL3 RNA, both RNA structural isomers shift to form slower-migrating, heparin-resistant complexes (Fig. 4, lane 0). These shifted bands were effectively competed for with cold, homologous SVL3 RNA as a competitor but not by a competitor RNA that lacks poly(A) signal sequences (SVL2; diagrammed in Fig. 1). RNAs derived from plasmid-specific sequences also failed to compete (data not shown). We conclude that the 50-kDa protein specifically binds to the SVL polyadenylation signal. The significance of the additional bands which appear during competition with cold SVL3 RNA are unclear. They may be additional SVL3 RNA structural isomers or a result of the association/dissociation of specific complexes during electrophoresis.

To further identify the RNA sequences required for the 50-kDa protein interaction, we used CL3 nuclease to isolate a 24-base RNA fragment from SVL3 RNA. This 24-mer is composed exclusively of DSE sequences (+31 to +55). The 24-mer efficiently bound to the partially purified 50-kDa protein as assayed by UV cross-linking analysis (Fig. 5B) or

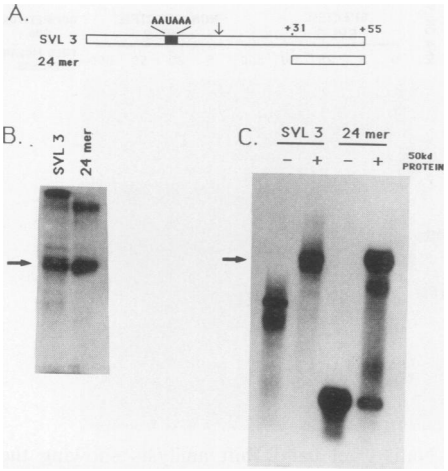

FIG. 5. Evidence that the SVL +31 to +55 DSE-specific RNA segment is sufficient to bind DSEF-1. (A) Diagram of RNAs used. The 24-mer (+31 to +55) was derived from SVL3 RNA by nuclease digestion and gel purification. UV cross-linking analysis (B) and band shift experiments (C) using DSEF-1 and the indicated RNAs were performed. Arrows indicate the result of the specific DSEF-1 protein-RNA interaction obtained in each assay.

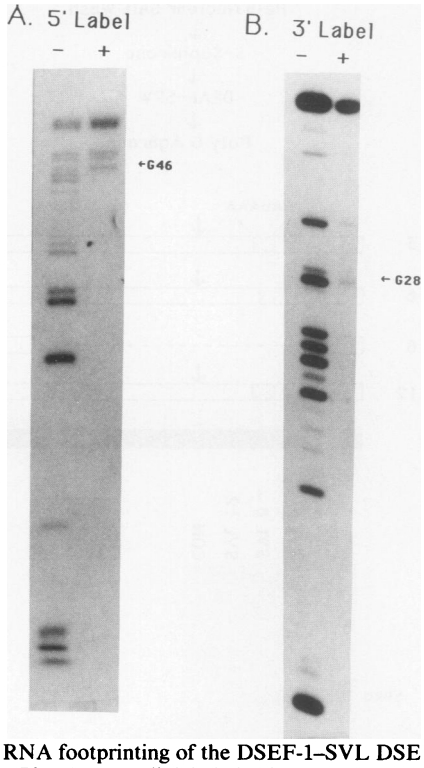

FIG. 6. RNA footprinting of the DSEF-1–SVL DSE interaction. SVL8 RNA (Fig. 3) was radiolabeled at its 3′ or 5′ end and subjected to limited digestion with RNase T_1 to form sets of ladder fragments. RNA ladders were incubated with DSEF-1, and complexes were identified by band shift experiments. DSEF-1-specific complexes were excised, and the bound RNA fragments were eluted and analyzed (lanes +) next to the input ladder fragments (lanes −) on acrylamide gels containing 7 M urea. Positions of the 3′ limit (G46) and the 5′ limit (G28) of the binding site as determined by this assay are indicated.

by native gel band shift electrophoresis (Fig. 5C). A different bisacrylamide/acrylamide ratio was used in this native gel (1:40) than for the gel shown in Fig. 4 (1:80), causing a difference in the migration of SVL3 RNA in the two gel systems. In addition, mutations within the G-rich region dramatically affect the interaction of the 50-kDa protein with the SVL RNA (data not shown). We conclude that the 24-base SVL DSE is sufficient to bind to the 50-kDa protein. We will subsequently refer to the 50-kDa protein as DSEF-1.

Fine mapping of the DSEF-1 binding site. The 3′ and 5′ limits of the binding site for DSEF-1 on SVL3 RNA were deduced by a form of RNA footprinting which took advantage of the observation that the migration of DSEF-1-specific complexes on 4% native acrylamide gels was independent of the size of the RNA substrate (Fig. 5C).

SVL3 RNA was labeled at the 5′ end by using [γ-^{32}P]ATP and T4 polynucleotide kinase or at the 3′ end by using pCp and RNA ligase (1). Ladders were prepared from these end-labeled transcripts by limited digestion with RNase T_1. The ladder fragments were incubated with DSEF-1, and complexes were separated by native gel electrophoresis. RNAs which formed a complex with DSEF-1 were eluted from the native gel and electrophoresed in a denaturing acrylamide gel next to the original T_1 ladder fragments. DSEF-1 bound fragments from either the 5′- or 3′-end-labeled SVL3 RNA ladder would, therefore, respectively identify the 3′ or 5′ limit of the DSEF-1 binding site. As seen in Fig. 6, the DSEF-1 binding site spans from G28 to G46.

Due to potential end effects on binding, this technique may not allow the precise definition of a minimal binding site. From comparison with our previous deletion analyses (Fig. 5; data not shown), we conclude that the binding site for DSEF-1 involves from +33 to +46 relative to the SVL cleavage site, or 5′-GGGGGAGGUGUGGG-3′.

To directly test the ability of this 14-base G-rich sequence to bind DSEF-1, we placed it into the RNA expression vector pGEM4 between the *Eco*RI and *Sal*I sites. Short RNA transcripts containing the putative DSEF-1 binding site (GEM50) efficiently bound to DSEF-1 as assayed by band shift analysis (Fig. 7). Curiously, the GEM50 RNA itself ran as a doublet in the native gel. Denatured GEM50 RNA runs as a single band which comigrates with the fast form of the GEM50 doublet (Fig. 7). These results may indicate that the 14-base DSEF-1 binding site possesses higher-order structure. Control RNAs derived from the GEM4 vector failed to interact with DSEF-1 and ran as a single band of the expected size in native gels (Fig. 7). Finally, GEM50 RNA, but not GEM4 RNA, was efficiently cross-linked to DSEF-1 when incubated with HeLa nuclear extract in the in vitro polyadenylation system (data not shown). We conclude that DSEF-1 binds to the 14-base G-rich element by recognizing either the primary sequence or some structural determinant.

The DSEF-1 binding site mediates efficient 3′ end processing. To assess the role of the sequence-specific interaction of DSEF-1 in polyadenylation efficiency, we constructed an SVL derivative which had downstream sequences from +8 to +55 removed and substituted with a polylinker-derived segment (SVL-GEM4 [48]). This substitution effectively removed all possible redundancy from the downstream region so that the functional significance of individual elements could be tested. As described previously, SVL derivatives containing downstream substitutions are unstable in our HeLa nuclear extracts in a Mg^{2+}-dependent fashion (45, 48). This makes it impossible to use polyadenylation assays to compare the 3′ end processing efficiency of SVL deriva-

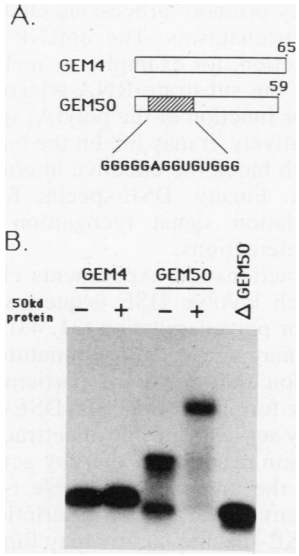

FIG. 7. Evidence that a 14-base RNA segment is sufficient for DSEF-1 binding. (A) Diagram of RNAs derived from the pGEM4 polylinker region (GEM4) or one containing the DSEF-1 binding site in place of the *Eco*RI-to-*Sal*I sites of the polylinker (GEM50). (B) Native gel band shift analysis of the interaction of DSEF-1 with the indicated RNAs. The migration of heat-denatured GEM50 RNA in the absence of protein is demonstrated in lane ΔGEM50.

tives with DSE substitutions. Cleavage assays, which can be performed in the presence of EDTA, therefore, were used instead to fairly compare processing efficiency. As seen in Fig. 8B, the cleavage efficiency of the SVL-GEM4 RNA was much lower than that of wild-type SVL RNA.

When the binding site for DSEF-1 was inserted into the polylinker-derived downstream region of SVL-GEM4 (SVL-GEM50), 3′ end processing efficiency was dramatically increased compared with that of the SVL-GEM4 RNA (Fig. 8B). In all assays performed, processing of the SVL-GEM50 RNA was always somewhat less efficient than that of the wild type but clearly increased in relation to the level for SVL-GEM4 RNA. Furthermore, SVL-GEM50 RNA, but not SVL-GEM4 RNA, was efficiently cross-linked to DSEF-1 when incubated with nuclear extract in the in vitro polyadenylation system (data not shown). The effect of the DSEF-1 binding site in the cross-linking assay was specific, as cross-linking to no other proteins was influenced by its presence (data not shown). We conclude that the DSEF-1 binding site can enhance 3′ end processing efficiency independently of other SVL downstream sequences.

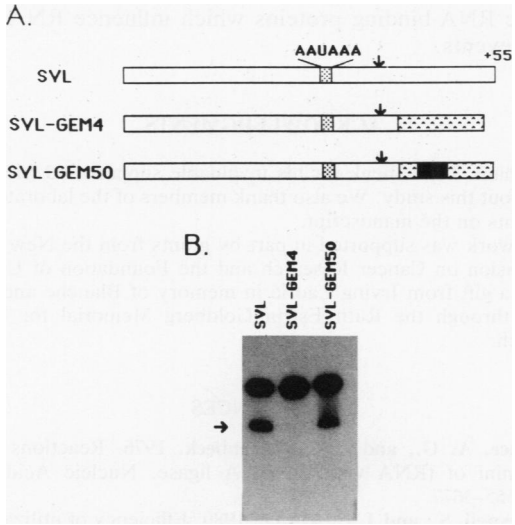

FIG. 8. Evidence that the DSEF-1 binding site promotes efficient 3′ end processing. (A) SVL downstream sequences were replaced by polylinker-specific sequences (SVL-GEM4) or by polylinker-derived sequences containing the DSEF-1 binding site (SVL-GEM50). (B) Equimolar amounts of the indicated RNAs were incubated in the in vitro polyadenylation system to measure cleavage efficiency. RNA products of the processing reaction were analyzed on a 5% acrylamide gel containing urea. The arrow indicates the position of the 5′ cleavage product.

DISCUSSION

We have identified and partially purified a 50-kDa protein (DSEF-1) which specifically binds to a 14-base, G-rich domain of the SVL DSE. Insertion of the binding site for DSEF-1 downstream of a polyadenylation signal significantly enhances the efficiency at which that signal is processed in vitro. Preliminary experiments suggest that our mutations and substitutions affect 3′ end formation in vivo in a manner consistent with the in vitro effects described here (data not shown).

These observations provide a strong correlation for the suggestion that DSEF-1 is a protein which directly influences the efficiency of 3′ end processing. Experiments which would provide direct evidence for this model, such as antibody depletion studies or the addition of DSEF-1 to a fractionated system, are currently hindered by the lack of reagents or the availability of a fractionated in vitro system that reproducibly responds to SVL DSE influences. We do have several other lines of data, however, which provide additional correlations for the functional significance of DSEF-1 to 3′ end processing. First, cross-linked DSEF-1 is coimmunoprecipitated by antibodies against the AAUAAA-dependent 64-kDa protein and the hnRNP C proteins (37, 47; data not shown), suggesting that the proteins are cross-linked to the same RNA substrate molecule. Second, cross-linked DSEF-1 can also be seen in native gels to be a component of the 25S-50S precleavage complex which forms on polyadenylation substrate RNAs (data not shown). Finally, poly(G) will effectively inhibit in vitro polyadenylation of the SVL substrate at levels of between 0.1 and 1 μg. The cross-linking of DSEF-1 to the SVL RNA substrate is also selectively inhibited at the same concentrations of poly(G) (data not shown). These data make it very likely, but do not prove, that DSEF-1 plays an important role in 3′ end processing. Experiments are currently in progress to definitively address the role of DSEF-1 in mediating polyadenylation efficiency.

DSEF-1 may, therefore, be an SVL-specific processing factor. We have not detected a stable interaction of DSEF-1 with any of our other model polyadenylation signals (data not shown). Placement of the DSEF-1 binding site downstream of a heterologous polyadenylation signal derived form the SV40 early transcription unit does, however, effectively substitute for the DSE of that signal in vitro (data not shown). This result indicates that the function of DSEF-1 is not restricted to the SVL poly(A) signal. Computer-aided searches of viral sequence data bases show that only the MN isolate of human immunodeficiency virus (14) has a highly similar DSEF-1 binding site. Curiously, this putative

DSEF-1 element is located upstream of the polyadenylation signal in the 3′ noncoding region of the mRNA. This region has recently been suspected to contain an upstream element which plays a role in the selective use of retroviral long terminal repeat-derived polyadenylation signals (7, 39). Experiments are in progress to directly test the role of the DSEF-1 element in human immunodeficiency virus 3′ end processing.

With respect to data base searches for DSEF-1 binding sites, however, it should be pointed out that we do not yet know how much flexibility there is in the DSEF-1 binding site. Preliminary mutagenesis studies indicate that binding is relatively unaffected by point mutations which interrupt the G tracts within the element (data not shown). DSEF-1 is not, therefore, simply a poly(G)-binding protein. DSEF-1 binding sites, because of their potential flexibility, may be more prevalent than stringent homology searches would suggest.

DSEF-1 may recognize a primary RNA sequence determinant or, alternatively, a structural determinant formed by its 14-base binding site. The aberrant migration of the GEM50 RNA on a 4% native acrylamide gel (Fig. 7) supports the notion that there may be higher-order structure contained within the DSEF-1 binding site. Since the site contains no C's, 1 A, 2 U's, and 11 G's, it is unlikely that it could form a structure composed of standard Watson/Crick-type hydrogen bonding. The interaction of DSEF-1 with RNA shows a linear relationship, even at low concentrations of RNA (data not shown). This makes it unlikely that intermolecular RNA hybrids play a significant role in DSEF-1 binding. The relationship of binding site structure to function is currently being pursued by a combination of mutagenesis and biophysical approaches.

All of the independent poly(A) signals that we have previously tested require downstream sequences for efficient cross-linking to the hnRNP C protein (45). The sequence-directed interaction of an hnRNP complex with downstream sequences of the polyadenylation signal, therefore, appears to be a common property of a polyadenylation signal. We have previously reported that a UUUUU element, located adjacent to the DSEF-1 binding site in the SVL DSE, can also mediate efficient 3′ end processing when it substituted for the downstream sequences of several polyadenylation signals (48). The UUUUU element is also required for efficient UV cross-linking of the C proteins of an hnRNP complex to the substrate RNA. In conjunction with the data on DSEF-1 presented in this report, therefore, we have provided a possible explanation for the apparent redundancy of DSEs seen previously by mutagenesis experiments (20, 50). This redundancy of the SVL downstream region may have important implications to SV40 biology, allowing the expression of the late viral transcription unit in a variety of cell types and intracellular conditions.

The relationship of DSEF-1 to the C proteins of the hnRNP complex is currently unclear. Although DSEF-1 is clearly not a major hnRNP protein, it may be one of the relatively uncharacterized 20 to 30 auxiliary hnRNP-associated proteins which are currently under investigation (28, 44). DSEF-1 is not coimmunoprecipitated by monoclonal antibodies specific for the hnRNP C or A protein after UV cross-linking and RNase treatment (48). This result implies that there is not a tight association between DSEF-1 and the major hnRNP proteins. Alternatively, DSEF-1 could be an independent factor.

The mechanism by which DSEF-1 and the hnRNP complex may mediate efficient 3′ end processing is also unclear. These factors may promote processing efficiency by similar or independent mechanisms. The hnRNP complex and/or DSEF-1 may function, for example, by melting out secondary structures in the substrate RNA which inhibit the efficient formation or function of the poly(A) signal recognition complex. Alternatively, it may inhibit the binding of nonspecific factors which block the effective interaction of polyadenylation factors. Finally, DSE-specific factors may promote polyadenylation signal recognition through direct protein-protein interactions.

Biochemical fractionation experiments clearly show that interactions which involve DSE sequences are not absolutely required for polyadenylation (31, 43). The role of the DSE, therefore, may be regulatory in nature. Previous data support the notion that the DSE participates in efficient 25S-50S complex formation (13, 50). DSE-specific factors, like DSEF-1, may serve to specifically attract components of the polyadenylation machinery, thereby actively promoting the assembly of the complex. This role is similar to that proposed for many eukaryotic transcription factors (24). Alternatively, DSE-specific factors may function to prepare a substrate RNA free of inhibitory structure or factors, analogous to a nucleosome-free region of DNA in the vicinity of active promoter regions. The requirement for the DSE may be lost, therefore, if purified processing factors which lack these inhibitory factors are used in fractionated systems.

In conclusion, we have described a sequence-specific RNA-binding protein which interacts with a sequence element that plays a role in polyadenylation efficiency. By providing a molecular explanation for the apparent redundancy of the SVL DSE, the identification of DSEF-1 and the UUUUU C protein cross-linking element may remove some of the confusion concerning the effect of downstream sequences on polyadenylation efficiency. DSEF-1 may also play a role in transcription termination, since this process is known to require an active polyadenylation signal (18, 41). Finally, DSEF-1 may prove to be a model for sequence-specific RNA-binding proteins which influence RNA processing events.

ACKNOWLEDGMENTS

We thank Tom Shenk for his invaluable support and friendship throughout this study. We also thank members of the laboratory for comments on the manuscript.

This work was supported in part by grants from the New Jersey Commission on Cancer Research and the Foundation of UMDNJ and by a gift from Irving Laurie in memory of Blanche and Edith Laurie through the Ruth Estrin Goldberg Memorial for Cancer Research.

REFERENCES

1. **Bruce, A. G., and O. C. Uhlenbeck.** 1976. Reactions at the termini of tRNA with T4 RNA ligase. Nucleic Acids Res. **5**:3665–3677.
2. **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.
3. **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.
4. **Cole, C. N., and T. P. Stacy.** 1985. Identification of sequences in the herpes simplex virus thymidine kinase gene required for efficient processing and polyadenylation. Mol. Cell. Biol. **5**:2104–2113.

5. **Conway, L., and M. P. 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. USA **82**:3949–3953.
6. **DeZazzo, J. D., and M. J. Imperiale.** 1989. Sequences upstream of AAUAAA influence poly(A) site selection in a complex transcription unit. Mol. Cell. Biol. **9**:4951–4961.
7. **DeZazzo, J. D., J. E. Kilpatrick, and M. J. Imperiale.** 1991. Involvement of long terminal repeat U3 sequences overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3′ end formation. Mol. Cell. Biol. **11**:1624–1630.
8. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475–1489.
9. **Dreyfuss, G.** 1986. Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. Annu. Rev. Cell Biol. **2**:459–498.
10. **Falck-Pedersen, E., and J. Logan.** 1989. Regulation of poly(A) site selection in adenovirus. J. Virol. **63**:531–541.
11. **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. USA **85**:2439–2443.
12. **Gil, A., and N. J. Proudfoot.** 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3′ end formation. Cell **49**:399–406.
13. **Gilmartin, G. M., and J. R. Nevins.** 1989. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. Genes Dev. **3**:2180–2189.
14. **Gurgo, C., H.-G. Guo, G. Franchini, A. Aldovini, E. Collalti, K. Farrell, F. Wong-Staal, R. C. Gallo, and M. S. Reitz, Jr.** 1988. Envelope sequences of two new united states HIV-1 isolates. Virology **164**:531–536.
15. **Hart, R. P., M. A. McDevitt, and J. R. Nevins.** 1985. Poly(A) site cleavage in HeLa nuclear extract is dependent on downstream sequences. Cell **43**:677–683.
16. **Humphrey, T., G. Christofori, V. Lucijanic, and W. Keller.** 1987. Cleavage and polyadenylation of messenger RNA precursors in vitro occurs within large and specific 3′ processing complexes. EMBO J. **6**:4159–4168.
17. **Leff, S. E., M. G. Rosenfeld, and R. M. Evans.** 1986. Complex transcriptional units: diversity in gene expression by alternative mRNA processing. Annu. Rev. Biochem. **55**:1091–1117.
18. **Logan, J., E. Falck-Pedersen, J. E. Darnell, and T. Shenk.** 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta-globin gene. Proc. Natl. Acad. Sci. USA **84**:8306–8310.
19. **Manley, J. L.** 1988. Polyadenylation of mRNA precursors. Biochim. Biophys. Acta **950**:1–12.
20. **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.
21. **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.
22. **McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AAUAAA signal is required for efficient formation of mRNA 3′ termini. Nucleic Acids Res. **13**:1347–1367.
23. **McLauchlan, J., S. Simpson, and J. B. Clements.** 1989. Herpes simplex virus induces a processing factor that stimulates poly(A) site usage. Cell **59**:1093–1105.
24. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**:371–378.
25. **Moore, C. L., and P. A. Sharp.** 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell **41**:845–855.
26. **Moore, C. L., H. Skolnick-David, and P. A. Sharp.** 1988. Sedimentation analysis of polyadenylation-specific complexes. Mol. Cell. Biol. **8**:226–233.
27. **Niwa, M., S. D. Rose, and S. M. Berget.** 1990. In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes Dev. **4**:1552–1559.
28. **Pinol-Roma, S., Y. D. Choi, M. J. Matunis, and G. Dreyfuss.** 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. Genes Dev. **2**:215–227.
29. **Proudfoot, N. J.** 1991. Poly(A) signals. Cell **64**:671–674.
30. **Russnak, R., and D. Ganem.** 1990. Sequences 5′ to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. Genes Dev. **4**:764–776.
31. **Ryner, L. C., Y. Takagaki, and J. L. Manley.** 1989. Sequences downstream of AAUAAA signals affect pre-mRNA cleavage and polyadenylation in vitro both directly and indirectly. Mol. Cell. Biol. **9**:1759–1771.
32. **Sadofsky, M., and J. C. Alwine.** 1984. Sequences on the 3′ side of hexanucleotide AAUAAA affect efficiency of cleavage at the polyadenylation site. Mol. Cell. Biol. **4**:1460–1468.
33. **Sadofsky, M., S. Connelly, J. L. Manley, and J. C. 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.
34. **Sheets, M. D., S. C. Ogg, and M. P. Wickens.** 1990. Point mutations in AAUAAA and the poly(A) addition site: effects of the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic Acids Res. **18**:5799–5805.
35. **Smith, K. C.** 1976. The radiation-induced addition of proteins and other molecules to nucleic acids, p. 187–218. *In* S. Y. Wang (ed.), Photochemistry and photobiology of nucleic acids, vol. 2. Academic Press, Inc., New York.
36. **Stefano, J. E., and D. E. Adams.** 1988. Assembly of a polyadenylation-specific 25S ribonucleoprotein complex in vitro. Mol. Cell. Biol. **8**:2052–2062.
37. **Takagaki, Y., J. L. Manley, C. C. MacDonald, J. Wilusz, and T. Shenk.** 1990. A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. Genes Dev. **4**:2112–2120.
38. **Takagaki, Y., L. C. Ryner, and J. L. Manley.** 1989. Four factors are required for 3′ end cleavage of pre-mRNAs. Genes Dev. **3**:1711–1724.
39. **Valsamakis, A., S. Zeichner, S. Carswell, and J. C. Alwine.** 1991. The human immunodeficiency virus type 1 polyadenylation signal: a 3′ long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation. Proc. Natl. Acad. Sci. USA **88**:2108–2112.
40. **Weichs an der Glon, C., J. Monks, and N. J. Proudfoot.** 1991. Occlusion of the HIV poly(A) site. Genes Dev. **5**:244–253.
41. **Whitelaw, E., and N. J. Proudfoot.** 1986. Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human globin gene. EMBO J. **5**:2915–2922.
42. **Wickens, M. P.** 1990. How the messenger got its tail: addition of poly(A) in the nucleus. Trends Biochem. Sci. **15**:277–281.
43. **Wigley, P. L., M. D. Sheets, D. A. Zarkower, M. E. Whitmer, and M. Wickens.** 1990. Polyadenylation of mRNA: minimal substrates and a requirement for the 2′ hydroxyl of the U in AAUAAA. Mol. Cell. Biol. **10**:1705–1713.
44. **Wilk, H. E., H. Werr, D. Friedrich, H. H. Kiltz, and K. P. Schaefer.** 1985. The core proteins of 35S hnRNP complexes: characterization of nine different species. Eur. J. Biochem. **146**:71–81.
45. **Wilusz, J., D. I. Feig, and T. Shenk.** 1988. The C proteins of heterogeneous nuclear ribonucleoprotein complexes interact with RNA sequences downstream of the polyadenylation cleavage site. Mol. Cell. Biol. **8**:4477–4483.
46. **Wilusz, J., S. M. Pettine, and T. Shenk.** 1989. Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. Nucleic Acids Res. **17**:3899–3908.
47. **Wilusz, J., and T. Shenk.** 1988. A 64kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. Cell **52**:221–228.

48. **Wilusz, J., and T. Shenk.** 1990. A uridylate tract mediates efficient heterogeneous nuclear ribonucleoprotein C protein-RNA cross-linking and functionally substitutes for the downstream element of the polyadenylation signal. Mol. Cell. Biol. **10**:6397–6407.
49. **Zarkower, D. A., and M. Wickens.** 1987. Specific pre-cleavage and post-cleavage complexes involved in the formation of SV40 late mRNA 3′ termini in vitro. EMBO J. **6**:4185–4192.
50. **Zarkower, D. A., and M. Wickens.** 1988. A functionally redundant downstream sequence in SV40 late pre-mRNA is required for mRNA 3′-end formation and for assembly of a precleavage complex in vitro. J. Biol. Chem. **263**:5780–5788.
51. **Zhang, F., and C. N. Cole.** 1987. Identification of a complex associated with processing and polyadenylation in vitro of herpes simplex virus type I thymidine kinase precursor RNA. Mol. Cell. Biol. **7**:3277–3286.