The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a *trans*-acting factor

Paramjeet S. Bagga, Lance P. Ford, Fan Chen and Jeffrey Wilusz*

UMD-New Jersey Medical School, Department of Microbiology and Molecular Genetics, 185 South Orange Avenue, Newark, NJ 07103, USA

Received November 8, 1994; Revised and Accepted March 15, 1995

ABSTRACT

A downstream G-rich sequence (GRS), GGGGGAGGU-GUGGG, has been previously shown to influence the efficiency of 3' end processing of the SV40 late polyadenylation signal. We have now defined several important parameters for GRS-mediated polyadenylation. The ability of the GRS to influence 3' end processing efficiency was sensitive to individual and multiple point mutations within the element, as well as the position of the element in the downstream region. Competition analysis indicated that the GRS functioned through a titratable *trans*-acting factor. The GRS-specific DSEF-1 protein was found to be bound to the same population of RNAs as the 64 kDa protein of the general polyadenylation factor CstF, indicating that DSEF-1 is associated with RNA substrates undergoing 3' end processing. Furthermore, an association was obtained between the relative strength of DSEF-1 protein binding to GRS variants and the relative ability of the GRS variants to mediate efficient cleavage *in vitro*. Finally, mutations in the GRS affected the efficiency of cross-linking of the 64 kDa protein of CstF. These data define a novel class of auxiliary downstream element and suggest an important role for DSEF-1 in 3' end processing.

INTRODUCTION

Efficient cleavage and polyadenylation of the 3' end of most mammalian pre-mRNAs requires a series of *cis*-acting elements and *trans*-acting factors (reviewed in 1). A highly conserved AAUAAA motif, located within ~20 bases of the cleavage site, is absolutely required for 3' end formation (2,3). A general polyadenylation factor, CPSF, recognizes this hexanucleotide and mediates site-specific cleavage and polyadenylation (4,5). Additional elements which influence the efficiency of 3' end processing have been identified both upstream and downstream of the cleavage site in several polyadenylation signals (e.g. 6–10).

While downstream elements are faithfully recognized in cell free processing systems derived from HeLa nuclear extracts (11), the inability of *in vitro* systems derived from reconstituted components to consistently require specific downstream sequences (12–14) has slowed progress in developing a mechanistic understanding the role of this region in 3' end processing.

The region downstream of the cleavage site of most polyadenylation signals contains a loose GU or U-rich consensus (1). Further analysis of known polyadenylation signals has recently shown that over two-thirds of signals contain a four out of five base uridylyl tract within 30 bases of the cleavage site (15). This 'U-rich element' was shown to play a major role in determining the location and efficiency of cleavage, as well as serve as the binding site for the 64 kDa protein of CstF (16,17) on RNAs which are undergoing 3' end processing (9,15). These observations suggest that the U-rich element represents the major downstream element sequence in most polyadenylation signals.

The downstream element of the SV40 late (SVL) polyadenylation signal contains a bipartite structure which is unlike the redundancy noted in other downstream regions (18,19). The major U-rich element at +15 (8,9) serves as the cross-linking site for CstF (15). Deletion/substitution of a second element, located between +33 and +55 relative to the cleavage site, has also been shown to have a 3–4-fold effect on 3' end processing (7,10,20). We have identified this region to contain a functionally significant 14 base G-rich sequence (GRS), GGGGGAGGUGUGGG, which also serves as a binding site for a 50 kDa cellular protein called downstream element factor 1 (DSEF-1) (10). The role of DSEF-1, as well as the sequence requirements for the GRS in polyadenylation efficiency, remain to be elucidated.

In this study, we investigated several important parameters for the downstream GRS to positively influence 3' end processing of the SVL polyadenylation signal. The ability of the GRS to mediate efficient processing was sensitive to changes in sequence and position. Competition studies indicated that a GRS-specific *trans*-acting factor was required for efficient processing of the SVL polyadenylation signal. Finally, we observed an association between the relative strength of DSEF-1 protein binding to variants of the GRS and the relative ability of these GRS variants

* To whom correspondence should be addressed

to mediate efficient cleavage *in vitro*. Finally, we demonstrate that DSEF-1 protein is bound to the same RNAs as CstF and that GRS mutations affect the cross-linking efficiency of the 64 kDa protein of CstF to the RNA substrate. These data suggest that the GRS element may function by promoting the formation of stable complexes of general polyadenylation factors.

MATERIALS AND METHODS

RNA and plasmids

RNAs were transcribed *in vitro* using SP6 or T7 RNA polymerase in the presence of [³²P]UTP or [³²P]GTP and purified from 5% acrylamide/7 M urea gels as described previously (21). Cold competitor RNAs were transcribed using SP6 polymerase in 100 µl reactions containing trace levels (1:360 000) [³²P]UTP to aid in purification and quantitation.

Transcripts were derived as follows: pSVL contains the *Bam*HI–*Bcl*II fragment of SV40 inserted into the *Bam*HI site of pSP65. Transcription of *Dra*I linearized template yields RNA of 224 bases (SVL). A 220 base control RNA was derived from pC220 (which contained a 1.38 kb fragment of bacteriophage λ) cut with *Sca*I (20). The construction of pSVL-14, which contains a substitution of SVL downstream sequences beginning at +29 with vector-derived sequence, was described previously (20). Transcription of *Hinf*I-cleaved DNA generates a 223 base RNA (SVL-14). pSVL-20 was constructed by deleting the +50 to +55 UUUUUU tract from pSVL using synthetic oligonucleotides. Transcription of *Pst*I-cleaved DNA generates a 219 base RNA (SVL-20). pSVL-GRS36 and pSVL-GRS45 were constructed by a combination of conventional cloning and PCR approaches. A 207 bp region was amplified from the *Sph*I–*Xba*I fragment of pSVL using an SP6 promoter-specific primer (CATACGATTAGGTGACACTATAGAAGAGGGA GGTGTCAGTCTA) and a second primer derived from the SVL downstream region (+13 to +24) containing *Pvu*II and *Hinc*II sites at its 5' end (GCGTCGACCAGCTGAAACA-TAAAATG). Amplification reactions were performed in 100 µl using standard reaction mixtures for 35 cycles of 94°C (1 min), 49°C (1 min) and 72°C (1 min) and amplified DNA was purified on a 2% low melting point agarose gel prior to use. pSVL-GRS36 was constructed by digesting the PCR product with *Eco*RI and *Pvu*II and inserting the fragment between the *Eco*RI and *Sma*I sites of pGem50 (10). To construct pSVL-GRS45, the amplified DNA was cleaved with *Eco*RI and inserted between the *Eco*RI and *Sma*I sites of pGem50. Sequence analysis revealed that a two base deletion from the *Sma*I site of the vector occurred during cloning. DNA templates were linearized with *Hind*III prior to transcription using SP6 polymerase.

Site-specific mutations in the GRS of SVL were generated by a PCR approach. A 247 bp region from the beginning of the SP6 promoter to +55 downstream of the cleavage site was amplified from the *Nhe*I–*Hind*III fragment of pSVL. All amplification reactions contained the SP6 promoter-specific primer described above and one of the following primers to generate the desired mutant:

SVL-3U5'-AAAAAACCTCCACACCTCCACCTGAACGAA-3'
SVL-8U5'-AAAAAACCTCCACAACCTCCCCCTGAACCTGAA-3'
SVL-9G5'-AAAAAACCTCCACCCCTCCCCCTGAACCTGAA-3'
SVL-12U5'-AAAAAACCTCCAACACCTCCCCCTGAACCTGAA-3'
SVL-GRS12/135'-AAAAAACCTCTTACACCTCCCCCTGAACCTGAA-3'
SVL-GRS2/12/135'-AAAAAACCTCTGACACCTCCCTCTGAACCTGAA-3'

These latter primers are complementary to positions +23 to +55 relative to the polyadenylation site of pSVL (except for the desired mismatches). Amplification reactions were performed in 100 µl using standard reaction mixtures for 35 cycles of 94°C (1 min), 35°C (1 min) and 72°C (2 min). Amplified DNA was purified using Centricon filters (Amicon) prior to use. Transcription of PCR products yields RNAs of 224 bases.

Short RNAs containing the GRS, or a randomly chosen 14 base sequence, were prepared by SP6 transcription of synthetic DNA oligomers containing the desired sequences attached to an SP6 promoter. The sense strands of the oligomers used in these studies were as follows (underlined bases represent the portion transcribed to make the indicated RNA):

GRS-CATACGATTAGGTGACACTATAGAAGGGGGAGG TGTGGGTCTA
mtGRS-CATACGATTAGGTGACACTATAGAAGAGGGA GGTGTCAGTCTA
Control-CATACGATTAGGTGACACTATAGAAGCTGTCGAACCTAGTTCTA

Cell extracts and proteins

HeLa spinner cells were grown in JMEM media containing 10% horse serum. Nuclear salt wash extracts were prepared as described (22). Partially purified DSEF-1 protein was prepared from HeLa nuclear extracts by successive chromatography on S-Sepharose, DEAE and poly(G) agarose columns as described previously (10). Preparations contained ~10 silver stained bands, including a predominant species of 50 kDa (DSEF-1).

Cleavage and polyadenylation assays

Polyadenylation and cleavage reactions were performed using the *in vitro* system of Moore and Sharp (23) in 12.5 µl reaction volumes as described previously (21). RNA products were analyzed on 5 or 8% acrylamide gels containing 7 M urea. All reactions were performed with equimolar amounts of RNA and analyzed in the linear range of product accumulation. Competition experiments were performed by adding the indicated amounts of cold, gel purified competitor RNAs to the reaction mixtures prior to the addition of nuclear extract and incubation at 30°C. Quantitation was performed by excising bands of interest and determining the level of radioactivity by scintillation counting. All of the observations reported in this study were independently reproduced at least three times.

Analysis of RNA–protein interactions

RNAs (50–100 fmol; 200 000 c.p.m.) were reconstituted with partially purified DSEF-1 protein in reaction mixtures containing 3.25% polyvinyl alcohol, 10 mM HEPES (pH 7.9), 20 mM phosphocreatine, 1 mM ATP, 0.1 mM EDTA, 0.25 mM dithiothreitol and 10% glycerol. In some cases, phosphocreatine and ATP were substituted with 1 mM of the ATP analog AMPCH₂PP (Sigma) and 1 mM EDTA. Following incubation at 30°C for 10 min to allow protein–RNA complexes to form, heparin was added to a final concentration of 4 µg/ml and samples were incubated on ice for 5 min to remove non-specific complexes. Samples were electrophoresed in 4% native gels (40:1 acrylamide:bis-acrylamide) using 1 × TBE. Slower migrating protein–RNA complexes were visualized by autoradiography. For quantitation, bands were excised from gels and analyzed by scintillation counting in multiple experiments. All comparative experiments were performed using equimolar amounts of radiolabeled RNAs.

To analyze whether CstF and DSEF-1 were present on the same RNA molecules, SVL RNA radiolabeled at G residues was incubated in the *in vitro* polyadenylation system for 5 min to allow specific complexes to form. Reaction mixtures were then irradiated with a 15 W germicidal UV light to covalently cross-link closely associated proteins to the RNA. Reaction mixtures were resuspended in 400 μ l of RIPA buffer, pre-cleared and 1 μ l of RNasin was added. The monoclonal antibody 3A7, which is specific for the 64 kDa protein of CstF (16), was added as a hybridoma supernatant to the reaction mixture. Formalin-fixed *S.aureus* cells were added and precipitates were washed extensively in RIPA buffer. Covalent RNA-protein complexes were eluted by boiling. Selected samples were treated with RNase A. Samples were analyzed on a 10% gel containing SDS.

The effect of GRS mutations on the UV cross-linking of the 64 kDa protein of CstF was determined by immunoprecipitation as described above with two exceptions. First, RNAs were labeled at U residues. Secondly, reaction mixtures were treated with RNase A (and no RNase inhibitor was added) prior to pre-clearing and antibody addition.

RESULTS

Sequence requirements for efficient 3' end processing mediated by the GRS

Previous analyses using deletion and insertion variants have shown that the GRS is required for efficient processing of the SV40 late (SVL) polyadenylation signal (10). In order to explore the sequence requirements for GRS-mediated processing, we created a series of point mutations in the GRS using a PCR approach and analyzed the effect of these variations on the efficiency of SVL processing. Cleavage efficiencies were determined as the percentage of total radioactivity present in the 5' cleavage product relative to the recovered input RNA. The numbers presented are the average of four separate experiments. The overall efficiency of cleavage for wild-type SVL RNA ranged from 10 to 48%, depending on the extract used in these studies and the time of incubation. All extracts employed gave reproducible results.

As described previously, a complete substitution of the GRS resulted in a 75–80% decrease in cleavage efficiency relative to wild-type SVL RNA (10,24). Individual G→U transversions at positions 3, 8 or 12 of the GRS caused a modest but reproducible 15–20% decrease in cleavage efficiency (Fig. 1). Curiously, a U→G transversion at position 9 of the GRS caused a 50% increase in cleavage efficiency relative to the wild-type SVL construct. These data demonstrate that individual positions of the GRS have a functional impact on 3' end processing.

We next determined if multiple point mutations in the GRS would further debilitate its ability to function in cleavage efficiency. As seen in Figure 2, two G→A transitions at positions 12 and 13 (SVL-GRS 12/13) decreased cleavage efficiency by nearly 50% of wild-type levels. A variant containing a G→C transversion at position 12 and G→A substitutions at positions 2 and 13 resulted in a poor cleavage efficiency very similar to variants containing deletions in the GRS (compare SVL 14 with SVL GRS 2/12/13). Multiple mutations in the GRS, therefore, significantly decrease the ability of the element to influence 3' end processing.

The SVL 20 variant shown in Figure 2 contains a deletion of the six base U tract from +50 to +55, just downstream of the GRS.

Figure 1. Point mutations in the GRS affect its ability to influence the efficiency of 3' end RNA processing. SVL RNA or derivatives containing a single transversion mutation at the indicated position (e.g. 3U is a G→U change at position 3 of the GRS) were incubated in the *in vitro* cleavage system for 5 min. Products were analyzed on an 8% acrylamide gel containing 7 M urea. A representative gel is shown in (A). The arrow indicates the 5' cleavage product. (B) shows the average cleavage efficiency of the indicated RNA relative to the cleavage efficiency of wild-type SVL RNA. The sequence of the GRS is shown above the graph with individual positions numbered.

This region had been hypothesized to be important for efficient processing of the SVL polyadenylation signal (7). SVL 20, however, was processed slightly more efficiently than the wild-type transcript. Removal of this U tract may make the signal more efficient by removing an element that may be competing for CstF binding with the U-rich element located at +15 (15).

In conclusion, these data demonstrate that the processing of the SVL polyadenylation signal is sensitive to both transition and transversion mutations at a variety of positions in the GRS.

The position of the GRS influences its ability to mediate efficient 3' end processing

Previous studies have demonstrated a positional limit for the major U-rich downstream element to within ~30 bases from the cleavage site (9). We wished to determine whether the auxiliary GRS element also had a positional requirement to mediate efficient 3' end processing. A pair of SVL variants were constructed which contained the GRS placed at positions +36 and +45, three and 12 bases respectively downstream from the wild-type location at +33. As seen in Figure 3, movement of the GRS three bases downstream resulted in a 23% decrease in cleavage efficiency compared to wild-type levels, while movement of the element 12 bases downstream caused a 44% decrease in efficiency. Considering that the complete absence of the GRS

Figure 2. Multiple point mutations in the GRS have a significant effect on the ability of the element to mediate efficient 3' end processing. The RNAs diagrammed in (A), containing the indicated mutations in lower case, were incubated in the *in vitro* cleavage system. Reaction products were analyzed on a 5% acrylamide gel containing 7 M urea. A representative gel showing the positions of the input RNA and 169 base 5' cleavage product is shown in (B). The radioactivity was determined in each of these bands and the cleavage efficiency of the various RNAs was expressed as a percentage of the efficiency obtained for wild-type SVL RNA (C).

results in an 80% decrease in cleavage efficiency, shifting the position of the GRS further downstream had a significant negative impact on its ability to function. In contrast to observations made by shifting the U-rich element further downstream (9,15), moving the GRS did not alter the site of cleavage (data not shown). Finally, placement of the GRS element at +8, 5' of a U-rich element located at +25, resulted in the inhibition of *in vitro* cleavage (data not shown). We conclude that the GRS must be positioned downstream of the U-rich element and that the relative location of the GRS affects its ability to function as a positive mediator of 3' end processing.

GRS-mediated 3' end processing requires a *trans*-acting factor

The next question we addressed was whether a specific *trans*-acting factor was required for efficient processing of the SVL polyadenylation signal or whether the GRS was acting to influence 3' end processing solely as an RNA sequence (perhaps by influencing secondary structure). We prepared small 21 base RNAs which contained an intact 14 base GRS, a debilitated GRS containing the three mutations described in Figure 2 for GRS 2/12/13, or control, randomly selected sequences. The effect of increasing concentrations of these RNAs on the polyadenylation of SVL RNA in HeLa nuclear extracts was assessed in

Figure 3. The position of the GRS influences its ability to mediate efficient 3' end processing. The GRS was moved three (SVL-GRS36) or 12 (SVL-GRS45) bases downstream from its normal position in the SVL polyadenylation signal. RNAs were incubated in the *in vitro* cleavage system and reaction products were analyzed by gel electrophoresis. (A) Representative gel. The arrow indicates the 5' cleavage product. (B) The relative cleavage efficiency of the indicated RNAs was determined from three independent experiments and expressed as a percent of wild-type SVL cleavage efficiency.

competition assays. As seen in Figure 4, the small RNA containing an intact GRS decreased SVL polyadenylation to ~30% of wild-type levels. Small RNAs containing a GRS with three mutations were much less efficient competitors for efficient processing of the SVL polyadenylation signal. Based on the slopes derived from the linear range of the competition curves, the GRS 2/12/13 RNA was an ~4-fold weaker competitor than wild-type GRS. Finally, the control RNA had no significant effect on the polyadenylation of the SVL RNA substrate at the concentrations employed in these assays. We have minimized the possibility that the observed competition was due to hybridization with the SVL substrate RNA by making the competitor RNAs short, containing only the 14 base GRS flanked by three to four bases to reduce possible end effects.

We conclude, therefore, that a titratable *trans*-acting factor is involved in GRS-mediated processing efficiency. The observation that processing efficiency did not fall <30% in these competition assays indicates that the GRS-specific factor is not absolutely required for 3' end processing to occur, but rather plays a role in the efficiency of the reaction. This is exactly what one would predict for a factor which binds to an auxiliary downstream element.

DSEF-1 protein-RNA interactions are tightly associated with GRS-mediated efficient 3' end processing

We have previously shown that a 50 kDa nuclear protein which we have called DSEF-1 specifically binds to the GRS (10). We next wished to determine whether the interaction of DSEF-1 with

Figure 4. A titratable *trans*-acting factor is required for GRS-mediated efficient processing of the SVL polyadenylation signal. Radiolabeled SVL RNA was incubated in the *in vitro* polyadenylation system in the presence of increasing amounts of the indicated cold competitor RNA. Reaction products were analyzed on 8% acrylamide gels containing 7 M urea and bands were quantitated by excision and scintillation counting. Polyadenylation efficiency is expressed relative to parallel reactions containing no competitor RNA. Circles: non-specific competitor RNA (GAAUGUCCGAACUAGUUCUA); diamonds: competitor RNA containing a mutated GRS element (GAAGAGG-GAGGUGUCAGUCUA); squares: competitor RNAs containing the wild-type GRS (GAAGGGGGAGGUGUGGGUCUA).

the GRS was associated with GRS-mediated efficient 3' end processing. While DSEF-1 protein can be efficiently cross-linked to SVL RNAs in the *in vitro* polyadenylation system (11), it is not clear whether this interaction was occurring on functional SVL substrate RNAs or if DSEF-1 was associated with a population of input transcript which does not get processed.

Functional SVL substrate RNAs may be defined as that population which is in a complex with the general polyadenylation factors. SVL RNA radiolabeled at G residues was incubated in the *in vitro* processing system and cross-linked with UV light to closely associated proteins (21). Intact RNAs cross-linked to the 64 kDa protein of the U-rich element specific CstF complex were then isolated by immunoprecipitation using 64 kDa protein specific monoclonal antibodies in RIPA buffer. Isolated complexes (Fig. 5, lane -) were then treated with RNase A and cross-linked proteins with short radioactive RNA oligomers were analyzed by SDS-PAGE (Fig. 5, lane +). DSEF-1 protein was specifically co-precipitated with the 64 kDa protein on the intact substrate RNA under denaturing conditions, indicating that both factors were cross-linked to the same RNA molecule. RNase treatment prior to immunoprecipitation, or the use of control monoclonal antibodies, failed to precipitate any DSEF-1 protein (data not shown).

We conclude from this data that DSEF-1 protein is found on the same RNAs that are stably associated with the 64 kDa subunit of CstF, suggesting that DSEF-1 is bound to functional RNA substrates. Since UV cross-linking is a relatively low efficiency technique (3–5% of possible complexes are routinely cross-linked in our assays), the observation that DSEF-1 was co-precipitated on intact RNAs with the 64 kDa protein suggests that a significant percentage of RNA substrates must contain both proteins bound at sites adjacent to each other.

We next tested if the ability of DSEF-1 to bind RNA substrates could be directly associated with efficient 3' end processing. The

Figure 5. The 64 kDa protein of CstF and DSEF-1 can be cross-linked to the same RNA molecule. SVL RNA labeled at G residues was incubated in the *in vitro* polyadenylation system and reaction mixtures were irradiated with UV light. Covalent complexes between the 64 kDa protein and the substrate RNA were isolated by immunoprecipitation using the 3A7 monoclonal antibody which is specific for 64 kDa (lane -). Complexes were then treated with RNase A to release proteins with small radioactive RNA oligomers covalently attached (lane +). Proteins were analyzed on a 10% acrylamide gel containing SDS.

various point mutants used in the functional assays shown in Figures 1 and 2 were tested for their ability to bind DSEF-1. RNAs were mixed with partially purified DSEF-1 protein (10) and heparin-resistant complexes were identified by band shift analysis on native gels. G→U point mutations at positions 3, 8 and 12 all resulted in an ~20% decrease, while a U→G transversion at position 9 caused an ~40% increase in the relative percentage of stable DSEF-1 complexes formed (Fig. 6). While the changes in DSEF-1 binding efficiencies are relatively modest, the efficiencies of DSEF-1 binding to the GRS variants correlate very well with the effect these GRS variations have on the efficiency of 3' end processing (Fig. 1). As seen in Figure 7, the interaction of DSEF-1 protein with the double and triple point mutants was also significantly reduced. The efficiency of these variants for DSEF-1 binding was calculated in competition assays for DSEF-1 complex formation using radiolabeled SVL RNA and increasing amounts of the SVL variants. SVL-GRS 11/12 was a 1.8-fold poorer competitor than SVL, while SVL GRS 2/12/13 was a 4-fold poorer competitor (data not shown). These numbers correlate very well with the decrease in processing efficiency caused by these GRS variation shown in Figures 2 and 4. These data suggest an association between the ability of the GRS to bind DSEF-1 protein, and the ability of the GRS to mediate efficient 3' end processing.

The GRS influences the formation of a stable complex between CstF and SVL RNA

The GRS may influence polyadenylation efficiency by stabilizing the assembly of the general polyadenylation factors on the SVL RNA substrate. The 64 kDa protein of CstF has recently been shown to cross-link to a U-rich downstream element of the SVL polyadenylation signal located at +15 (15). The stable association of CstF with polyadenylation substrate RNAs has been shown to be reflected by the level of UV cross-linking of its 64 kDa

Figure 6. Individual point mutants in the GRS significantly affect the ability of the RNA to mediate DSEF-1 binding. Band shift analysis showing the DSEF-1 interaction with equimolar amounts of the indicated RNAs (described in Fig. 1). Lane designations refer to incubations performed with (+) or without (-) DSEF-1 protein. Reconstituted protein-RNA complexes (indicated by an arrow) were identified on a 4% native acrylamide gel.

Figure 7. Multiple point mutations in the GRS significantly affect the ability of the RNA to mediate DSEF-1 binding. Band shift analysis showing the DSEF-1 interaction with equimolar amounts of the indicated RNAs (diagrammed in Fig. 2). Reconstituted protein-RNA complexes (indicated by an arrow) were identified on a 4% native acrylamide gel.

component (3,25). We decided, therefore, to test whether mutations in the GRS affect the efficiency of 64 kDa protein cross-linking to the SVL polyadenylation signal.

Equimolar amounts of wild-type SVL RNA, the SVL-GRS variants described in Figure 2, or a 220 base control RNA were incubated in the *in vitro* polyadenylation system. Reaction mixtures were irradiated with UV light and treated with RNase A. Cross-linked 64 kDa protein was specifically immunoprecipitated using a monoclonal antibody (16) and analyzed on an SDS-polyacrylamide gel. As seen in Figure 8, SVL variants containing multiple point mutations in the GRS showed a significantly reduced level of cross-linking to the 64 kDa protein compared to wild-type SVL RNA. The control RNA failed to interact with the 64 kDa protein. These data suggest that the GRS may stabilize the association of the general polyadenylation factors with substrate RNAs.

DISCUSSION

In this report, we have defined the requirements for the unique G-rich downstream sequence present in the SV40 late polyadenylation signal which serves as an auxiliary element to positively influence the efficiency of 3' end formation. The GRS was

Figure 8. Mutations in the GRS decrease the interaction of the SVL polyadenylation signal with CstF. Equimolar amounts of the indicated variants of the SVL polyadenylation signal (diagrammed in Fig. 2) or a similarly sized control RNA were incubated in the *in vitro* polyadenylation system and subjected to UV cross-linking. Following RNase treatment, cross-linked CstF-specific 64 kDa protein was isolated by immunoprecipitation and analyzed on a 10% SDS gel.

sensitive to point and cluster mutations and showed a position-dependent effect on its ability to mediate efficient processing. A titratable *trans*-acting factor was shown to be required for the GRS to mediate efficient processing. Finally, two lines of evidence were presented to support the hypothesis that the DSEF-1 protein may be this GRS-specific *trans*-acting factor. First, DSEF-1 was shown to be associated with the same substrate RNAs as the 64 kDa protein of CstF, suggesting that DSEF-1 was associated with RNAs that were actively undergoing 3' end processing. Secondly, the sequence requirements for DSEF-1 binding to the GRS were well associated with the sequence requirements for the GRS to positively influence 3' end processing. Mutations that decreased the affinity of DSEF-1 for the GRS also resulted in a less active polyadenylation substrate. The U→G mutation at position 9 of the GRS which increased the level of DSEF-1 binding, on the other hand, increased the efficiency of 3' end processing. Finally, mutations in the GRS element which reduced processing efficiency also reduced the level of cross-linking of the 64 kDa protein of CstF to the SVL polyadenylation signal. Further exploration of the role of DSEF-1 in GRS-mediated polyadenylation efficiency will require the development of a reconstituted *in vitro* system that faithfully reproduces the effect of the auxiliary GRS element on SVL RNA 3' end processing.

The requirement for a *trans*-acting factor in GRS-mediated polyadenylation, in conjunction with the observation of a position dependence of the element to influence processing efficiency, suggests that the GRS specific factor may mediate efficient 3' end processing through protein-protein interactions with other polyadenylation factors. Due to its binding to the downstream U-rich element which lies just 5' of the GRS, the likely target for these interactions may be CstF (16,17). The data in Figure 5 which demonstrate that the 64 kDa component of CstF and the GRS-specific DSEF-1 protein are on the same RNA also raises the possibility that these two proteins may interact. The interac-

tion of DSEF-1 (or other GRS-specific factor) with CstF may help to stabilize the weak interaction of the 64 kDa protein on the RNA substrate (25), allowing for a more efficient assembly of the complex of general polyadenylation factors, and therefore, more efficient polyadenylation. The observation that mutations in the GRS result in a decrease in CstF cross-linking (Fig. 8) are consistent with such a model.

Previous studies have suggested that the SVL downstream region contains a functionally significant AGAUUUUUU element which lies just downstream of the GRS at positions +47 to +55 relative to the cleavage site (7). This conclusion was deduced from the observation that deletion of the sequence from +8 to +46 has a small effect on 3' end processing, but extending this deletion from +8 to +55 resulted in a significant decrease in processing. It should be noted, however, that the +8 to +46 deletion, while removing the major U-rich element and the GRS, brings the +50 to +55 UUUUUU tract to position +11. As we have recently shown, the insertion of a five base U tract at this position will restore efficient processing to the SVL signal (9). U tracts positioned beyond ~+30, however, do not mediate efficient 3' end processing. This explanation, in conjunction with the data shown in Figure 2 that deletion from +50 to +55 resulted in a slight increase in SVL processing efficiency (SVL-20), argues that the region from +47 to +55 does not play a role in 3' end processing. The decreases in polyadenylation efficiency previously seen in the analysis of deletion mutations downstream of +33 in the SVL polyadenylation signal can best be explained, therefore, by disruption of the GRS.

Another important question which arises from this study is whether auxiliary elements such as GRS exist for other polyadenylation signals. In a preliminary survey of the primate database, we have identified similarities to the GRS in the downstream region of seven other polyadenylation signals. These GRS similarities include the mouse Hox 2.3 gene which contains the sequence GGGGGAGGUGUaaG which is identical to our G12/13 mutant which has an intermediate effect on SVL processing (Fig. 2). While these data suggest that the GRS located in the Hox 2.3 gene may be functionally significant, we have been unable to test this directly due to the instability of the Hox 2.3 polyadenylation signal in our *in vitro* system (data not shown). GRS similarities were also found in the region upstream of the cleavage site of 12 polyadenylation signals, including a segment of the HIV-1 upstream region (-150) which has been shown by Alwine and colleagues to have a minor effect on the accumulation of poly(A)⁺ RNA *in vivo* (26). We have shown that the HIV similarity to the GRS can indeed bind DSEF-1 (data not shown). The generalization of the GRS to other polyadenylation signals, as well as the search for additional auxiliary elements that influence 3' end processing, is currently being pursued in our laboratory.

Our recent identification of four cytoplasmic proteins which specifically bind the GRS (27), as well as the relative abundance of the GRS in GenBank entries, suggest that the GRS may be a

general element which may influence many aspects of RNA metabolism. The insights that are gained in the elucidation of the role of the GRS in 3' end processing, therefore, may provide a general model for GRS-mediated processes.

ACKNOWLEDGEMENTS

We wish to thank Dr Zuwei Qian for helpful discussions and the initial characterization of the effect of point mutations in the GRS on DSEF-1 binding. This work was supported in part by grants from the National Institutes of Health (AI 31165) and the Council for Tobacco Research USA, Inc. to JW. FC was supported in part by training grant CA 09663-01. JW is a Pew Scholar in the Biomedical Sciences.

REFERENCES

- 1 Wahle, E., and Keller, W. (1992) *Annu. Rev. Biochem.* **61**, 419-440.
- 2 Sheets, M.D., Ogg, S.C., and Wickens, M.P. (1990) *Nucleic Acids Res.* **18**, 5799-5805.
- 3 Wilusz, J., Pettine, S.M., and Shenk, T. (1989) *Nucleic Acids Res.* **17**, 3899-3908.
- 4 Keller, W., Bienroth, S., Lang, K.M., and Christofori, G. (1991) *EMBO J.* **10**, 4241-4249.
- 5 Murthy, K.G.K., and Manley, J.L. (1992) *J. Biol. Chem.* **267**, 14804-14811.
- 6 Carswell, S., and Alwine, J.C. (1989) *Mol. Cell. Biol.* **9**, 4248-4258.
- 7 Sadofsky, M., Connelly, S., Manley, J.L., and Alwine, J.C. (1985) *Mol. Cell. Biol.* **5**, 2713-2719.
- 8 Zarkower, D.A., and Wickens, M.P. (1988) *J. Biol. Chem.* **263**, 5780-5788.
- 9 Chou, Z.F., Chen, F., and Wilusz, J. (1994) *Nucleic Acids Res.* **22**, 2525-2531.
- 10 Qian, Z., and Wilusz, J. (1991) *Mol. Cell. Biol.* **11**, 5312-5320.
- 11 Green, T.L., and Hart, R.P. (1989) *Mol. Cell. Biol.* **8**, 1839-1841.
- 12 Weiss, E.A., Gilmartin, G.M., and Nevins, J.R. (1991) *EMBO J.* **10**, 215-219.
- 13 Ryner, L.C., Takagaki, Y., and Manley, J.L. (1989) *Mol. Cell. Biol.* **9**, 1759-1771.
- 14 Wigley, P.L., Sheets, M.D., Zarkower, D.A., Whitmer, M.E., and Wickens, M.P. (1990) *Mol. Cell. Biol.* **10**, 1705-1713.
- 15 MacDonald, C.C., Wilusz, J., and Shenk, T. (1994) *Mol. Cell. Biol.* **14**, 6647-6654.
- 16 Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J., and Shenk, T. (1990) *Genes Dev.* **3**, 2112-2120.
- 17 Takagaki, Y., MacDonald, C.C., Shenk, T., and Manley, J.L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1403-1407.
- 18 Chen, J.S., and Nordstrom, J.L. (1992) *Nucleic Acids Res.* **20**, 2565-2572.
- 19 Gil, A., and Proudfoot, N.J. (1987) *Cell* **49**, 399-406.
- 20 Wilusz, J., Feig, D.I., and Shenk, T. (1988) *Mol. Cell. Biol.* **8**, 4477-4483.
- 21 Wilusz, J., and Shenk, T. (1988) *Cell* **52**, 221-228.
- 22 Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) *Nucleic Acids Res.* **11**, 1475-1489.
- 23 Moore, C.L., and Sharp, P.A. (1985) *Cell* **41**, 845-855.
- 24 Wilusz, J., and Shenk, T. (1990) *Mol. Cell. Biol.* **10**, 6397-6407.
- 25 Wilusz, J., Shenk, T., Takagaki, Y., and Manley, J.L. (1990) *Mol. Cell. Biol.* **10**, 1244-1248.
- 26 Valsamakis, A., Zeichner, S., Carswell, S., and Alwine, J.C. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2108-2112.
- 27 Qian, Z., and Wilusz, J. (1994) *Nucleic Acids Res.* **22**, 2334-2343.