Nucleic Acids Research, 1998, Vol. 26, No. 23 **5343–5350**

DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation *in vitro*

Paramjeet S. Bagga, George K. Arhin and Jeffrey Wilusz*

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

Received August 25, 1998; Revised and Accepted October 9, 1998

ABSTRACT

DSEF-1 protein selectively binds to a G-rich auxiliary sequence element which influences the efficiency of processing of the SV40 late polyadenylation signal. We have obtained cDNA clones of DSEF-1 using sequence information from tryptic peptides isolated from DSEF-1 protein purified from HeLa cells. DSEF-1 protein contains three RNA-binding motifs and is a member of the hnRNP H family of RNA-binding proteins. Recombinant DSEF-1 protein stimulated the efficiency of cleavage and polyadenylation in an AAUAAA-dependent manner in *in vitro* reconstitution assays. DSEF-1 protein was shown to be able to interact with several poly(A) signals that lacked a G-rich binding site using a less stringent, low ionic strength gel band shift assay. Recombinant DSEF-1 protein specifically stimulated the processing of all of the poly(A) signals tested that contained a high affinity G-rich or low affinity binding site. DSEF-1 specifically increased the level of cross-linking of the 64 kDa protein of CstF to polyadenylation substrate RNAs. These observations suggest that DSEF-1 is an auxiliary factor that assists in the assembly of the general 3′-end processing factors onto the core elements of the polyadenylation signal.

INTRODUCTION

The cleavage and polyadenylation events which form the 3′-end of mammalian mRNAs reflect a concerted post-transcriptional process that involves the sequence-directed assembly of a multicomponent complex onto pre-mRNAs (1,2). Two conserved core elements which flank the cleavage site, an upstream AAUAAA motif and a downstream four-out-of-five base uridylate tract, direct the cooperative assembly of cleavage polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CstF) onto the RNA substrate (3–6). The assembly of these two factors on the core elements of the polyadenylation signal determines the efficiency and position of the 3′-end processing event (3). CPSF and CstF serve as a scaffold to direct the assembly of an endonuclease complex (CFI and CFII) and the template-independent poly(A) polymerase. In addition to their role in 3′-end formation, these elements/factors also influence transcription termination and splicing of the terminal intron (7–9).

Although the polyadenylation signal in most transcription units is a basal element in gene expression, the selection of alternative 3′-ends for a pre-mRNA can also be a regulated event (8). We have noted that many of the viral and cellular polyadenylation signals that undergo regulated processing contain sequence variations in one or both core elements (3; unpublished observations). It is likely, therefore, that auxiliary sequence elements influence the efficiency of 3′-end processing. In addition, regulation of poly(A) site selection may occur through these auxiliary sequences. Auxiliary elements can be located both upstream and downstream of the core elements (10–12). While the auxiliary region of many polyadenylation signals appears to function by maintaining the core elements in an unstructured conformation (13,14), several auxiliary elements have been identified which appear to function through protein–RNA interactions (15,16). These protein–RNA interactions may play an important role in the regulation of 3′-end processing or the interplay between polyadenylation and splicing/transcription.

We have previously identified the first auxiliary downstream element, a G-rich sequence (GRS) located downstream of the core U-rich element of the SV40 late polyadenylation signal (11,17). The GRS element stimulates the efficiency of 3′-end processing in a sequence- and position-dependent manner. Furthermore, the element requires a titratable *trans*-acting factor to stimulate processing. The GRS element specifically interacts with a 50 kDa nuclear protein we have named DSEF-1 (downstream element factor-1). Two lines of evidence suggest that a DSEF-1 protein–GRS element interaction may stimulate 3′-end processing efficiency. First, point mutations that affect the ability of the GRS element to mediate efficient 3′-end formation concomitantly affect its interaction with DSEF-1 in a similar fashion. Finally, DSEF-1 protein can be cross-linked to the same RNA as the 64 kDa protein of CstF, indicating that DSEF-1 is binding to RNA that is undergoing 3′-end processing. These data suggest a role for the DSEF-1 protein in efficient 3′-end formation, but a direct demonstration of the functional significance of DSEF-1 is currently lacking. A full appreciation of the role of this novel downstream auxiliary factor in 3′-end processing may provide insights into both polyadenylation signal recognition and selection.

In this study, we have obtained and characterized DSEF-1 cDNA clones using sequence information of tryptic peptides from purified DSEF-1 protein. The addition of recombinant

*To whom correspondence should be addressed. Tel: +1 973 972 5218; Fax: +1 973 972 3644; Email: wilusz@umdnj.edu

DSEF-1 protein to an *in vitro* 3′-end processing system specifically stimulated the efficiency of both cleavage and polyadenylation in a dose-dependent fashion. DSEF-1 protein increased processing efficiency by specifically stimulating the assembly of the general polyadenylation factors onto the RNA substrate as measured by UV cross-linking of the 64 kDa CstF subunit. These data define the DSEF-1 protein as a *bona fide* auxiliary factor which may play a key role in regulating the efficiency of 3′-end processing.

MATERIALS AND METHODS

Plasmids and RNAs

Capped RNAs were prepared by *in vitro* transcription reactions using SP6 or T7 RNA polymerase in the presence of [^{32}P]UTP or [^{32}P]GTP and purified from 5% acrylamide–7 M urea gels as described previously (18).

DNA templates and transcripts were derived as follows. pGRS, which contains the 14 base GRS element inserted into pGEM3, has been described previously (11). Transcription of *Hin*dIII-digested plasmid DNA gave a 51 base transcript (GRS). pSVL contains the *Bam*HI–*Bcl*I fragment of SV40 inserted into the *Bam*HI site of pSP65. Transcription of *Dra*I-linearized template yields a 224 base RNA (SVL). A derivative of pSVL (pSVL-AAGAAA) that contains a single point mutation in the AAUAAA core element has been previously described (19). Transcription of *Dra*I-linearized templates gave a 224 base RNA (SVL-AAGAAA). pSVL3 contains the SV40-specific *Alu*I–*Hin*dIII fragment of pSVL inserted between the *Hin*cII and *Hin*dIII sites of pGEM4. Transcription of *Dra*I-cut template gives a 122 base RNA (SVL3). Transcription of *Hin*cII-linearized templates gave a pre-cleaved polyadenylation substrate RNA (SVL-pre). pAAV contains a 232 bp *Pst*I–*Xba*I fragment of adeno-associated virus inserted between the *Pst*I and *Xba*I sites of pGEM3. Transcription of *Eco*RI-linearized template gave a 286 base RNA (AAV). pGEM2μMPA has a 199 bp *Rsa*I fragment which contains the polyadenylation signal of the μ heavy chain inserted at the *Sma*I site of pGEM2 (20). Transcription templates were generated by PCR using an SP6 promoter primer and a downstream primer (5′-TATTCAGATAATCCTTGGGCTGTC) and yielded a 184 base transcript (μM RNA). IVA$_2$ RNA, which contains the polyadenylation signal of the IVA$_2$ gene of adenovirus, was transcribed from pIVA$_2$ as previously described (21).

DSEF-1 purification

Nuclear salt wash extracts were prepared as previously described from HeLa spinner cells grown in JMEM medium containing 10% horse serum (22). Nuclear extracts were loaded onto an S-Sepharose column in buffer D (20 mM HEPES, pH 7.9, 100 mM NaCl, 20% glycerol, 0.2 mM EDTA and 0.5 mM dithiothreitol). The flow-through of this column, which contains DSEF-1, was loaded onto a DEAE–Sepharose column equilibrated in buffer D. The flow-through fraction, which contained DSEF-1, was then loaded onto a poly(G)–agarose column equilibrated in buffer D. Following extensive washing with buffer D containing 300 mM NaCl, DSEF-1 protein was eluted from the affinity column using buffer D containing 1 M NaCl. Samples were dialyzed against buffer D and stored at –80 °C. DSEF-1 protein was detected throughout the purification by UV cross-linking to the GRS-containing SVL3 RNA which was radiolabeled using [^{32}P]UTP. Total protein in each fraction was analyzed by silver staining following separation on 10% acrylamide–SDS gels.

Partial peptide analysis

Partial peptide analysis was used to confirm the identity of the purified DSEF-1 protein as previously described (18,23). Briefly, GRS RNA which had been radiolabeled with [^{32}P]GTP was incubated with purified 50 kDa DSEF-1 protein or nuclear extracts under standard polyadenylation conditions (as described below) for 5 min at 30 °C to reconstitute protein–RNA complexes. Reaction mixtures were irradiated with a germicidal light for 10 min to covalently cross-link proteins to RNA. Mixtures were then incubated with RNase A and radiolabeled DSEF-1 proteins were isolated on 10% acrylamide–SDS gels. Proteins were then loaded onto a 15% acrylamide–SDS gel and incubated for 30 min with the indicated amounts of V8 protease in the stacking gel. Following completion of electrophoresis, radiolabeled protease digestion products were detected by autoradiography.

Amino acid sequencing

Purified DSEF-1 protein was concentrated by TCA precipitation, electrophoresed on a 10% acrylamide–SDS gel, transferred to a PVDF membrane (Immobilon-P) and sent to the Rockefeller University Protein/DNA Technology Center for sequencing of tryptic peptides. The sequence of the following four peptides was obtained: D1, RIGHRYIEIFK; D2, GPNSDDTANDXFVRLRXL; D3, RAEVRTHYDPPRK; D4, ETMGHRYVEVFK.

cDNA cloning, analysis and expression of recombinant proteins

Based on similarities to hnRNP proteins, the relative position of tryptic peptides was deduced and used to orient degenerate oligonucleotides for PCR amplification. A 243 bp DNA fragment was amplified from total HeLa cell RNA in a half nested RT–PCR reaction using the oligonucleotides 5′-TAYGTBGAGGTBTTY-AAG, 5′-YTTGAAKATYTCKATGTA and 5′-GTCGTTBGCB-GTGTCGTC derived from peptides D4, D1 and D2, respectively. This fragment was radiolabeled and used as a probe to screen a HeLa cDNA library (Clontech). The largest isolated clone contained a 2.4 kb insert and was selected for sequence analysis.

The Wisconsin Sequence Analysis Package (Genetics Computer Group Inc.) was used for all computer-based sequence analysis. Nucleic acid and protein databases were searched using the FASTA program. Secondary structure predictions were made using Pepplot and Peptidestructure. The Pileup program was used to assist in multiple sequence alignments, consensus sequences were derived using the Pretty program and sequence patterns were determined using the Motifs program.

A 1.35 kb fragment, containing the entire open reading frame of DSEF-1, was subcloned into pGEX2TZQ to form pDSEF-1 (24). Recombinant DSEF-1–GST fusion and GST proteins were induced by IPTG in *Escherichia coli* transformed with the pDSEF-1 or the vector alone, respectively, purified on glutathione–Sepharose columns and stored in elution buffer (50 mM Tris, pH 8.0, 10 mM reduced glutathione and 10% glycerol) at –80 °C.

Cleavage and polyadenylation assays

Cleavage and polyadenylation reactions were performed using equimolar amounts of capped RNAs labeled to the same specific

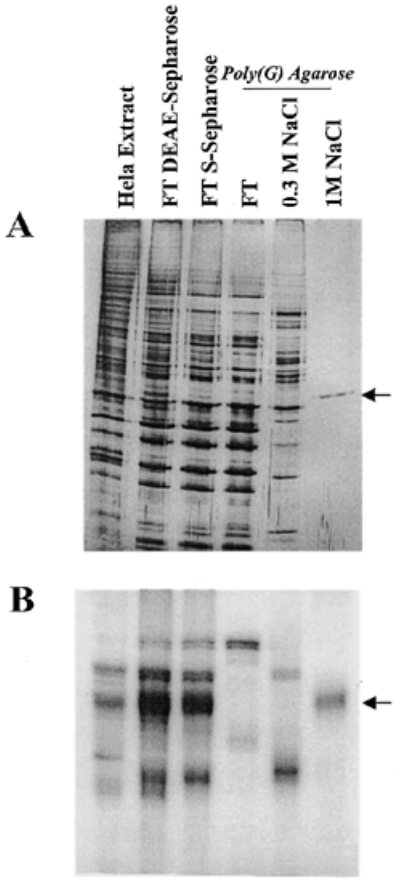

Figure 1. Purification of DSEF-1 protein from HeLa nuclear extracts. DSEF-1 protein was isolated using the indicated ion exchange and affinity columns. (**A**) Silver stained 10% acrylamide–SDS gel of selected column fractions obtained during the purification of DSEF-1. (**B**) UV cross-linking assays were performed using the fractions indicated above in (A) and radiolabeled GRS RNA. Cross-linked proteins were analyzed on a 10% acrylamide gel containing SDS. The arrow at the right of each panel indicates the position of the DSEF-1 protein.

activity in the *in vitro* system originally described by Moore and Sharp (25) as previously described (18). The ATP analog α,β-methylene-ATP and EDTA were used instead of ATP and phosphocreatine in cleavage reactions. Reaction products were analyzed on 5% acrylamide gels containing 7 M urea.

Endogenous DSEF-1 protein present in nuclear extracts was depleted by incubation with a biotinylated RNA oligonucleotide that contained the high affinity GRS binding site (5′-biotin-GGGGGA-GGUGUGGG) for 5 min at 30°C followed by the addition of streptavidin–agarose beads. DSEF-1 complexes with the biotinylated RNA were removed by centrifugation. Supernatants were determined to be free of DSEF-1 protein as assayed by cross-linking to the SVL RNA and used directly in *in vitro* cleavage reactions.

UV cross-linking, band shift assays and immunoprecipitations

RNA–protein interactions were analyzed by UV cross-linking/label transfer analysis as previously described (18). Band shift assays were performed as described previously (17), except that gels were run in a 50 mM Tris–borate–EDTA (TBE) buffer instead of 100 mM TBE. The 3A7 monoclonal antibody, which is specific for the

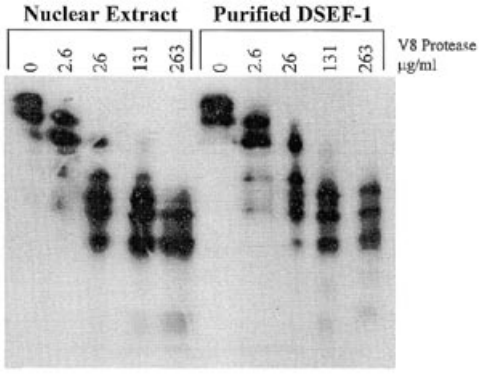

Figure 2. The pattern of partial protease digestion products demonstrate that the purified 50 kDa protein is DSEF-1. Radiolabeled GRS RNA was incubated with HeLa nuclear extract or purified DSEF-1 protein and UV cross-linking/label transfer analysis was performed. Cross-linked 50 kDa proteins from each reaction were excised from the gel and incubated with the indicated amount of V8 protease. Digestion products were analyzed on a 15% acrylamide gel containing SDS.

64 kDa subunit of CstF (26), was used in immunoprecipitation reactions as previously described (17). Cross-linked proteins were analyzed on 10% acrylamide gels containing 0.1% SDS.

RESULTS

Purification of the DSEF-1 protein

The isolation of DSEF-1 protein, the first putative auxiliary downstream element binding factor, is a key step in determining its functional significance to 3′-end processing. DSEF-1 protein was purified to near homogeneity using a combination of ion exchange (DEAE and S-Sepharose) and poly(G)–agarose affinity chromatography steps. A silver stained SDS–acrylamide gel of DSEF-1-containing fractions is shown in Figure 1A. DSEF-1 protein was followed through purification by sequence-specific UV cross-linking to an RNA (SVL3) containing a GRS sequence element (Fig. 1B). DSEF-1 protein was present in the flow-through fractions of both anion and cation exchange columns and was selectively eluted with 1 M NaCl from the poly(G) affinity column. Silver staining of this fraction showed a single band of 50 kDa (Fig. 1A, lane 1 M NaCl) which specifically cross-linked to the GRS-containing SVL3 RNA (Fig 1B, lane 1 M NaCl). The heterogeneity of cross-linked DSEF-1 protein seen on the gel is due to inefficient cutting of the cross-linked radiolabeled GRS element by RNase A in the assay.

Although the data in Figure 1 is highly suggestive that we have indeed purified DSEF-1 protein to near homogeneity, it was important to verify the identity of the 50 kDa band. This is especially true given the large number of poly(G)-binding proteins of approximately this molecular weight (24,27,28). We assessed the identity of the purified protein by comparing the partial protease products of the cross-linked purified 50 kDa species with the protease map of DSEF-1 cross-linked to RNA from nuclear extracts. Since UV cross-linking is very likely to occur at a single site on an RNA-binding protein (29,30), identical proteins should have identical patterns of cleavage products in this assay. The purified cross-linked 50 kDa protein, along with authentic cross-linked DSEF-1 from unfractionated nuclear extracts, were treated with increasing amounts of V8

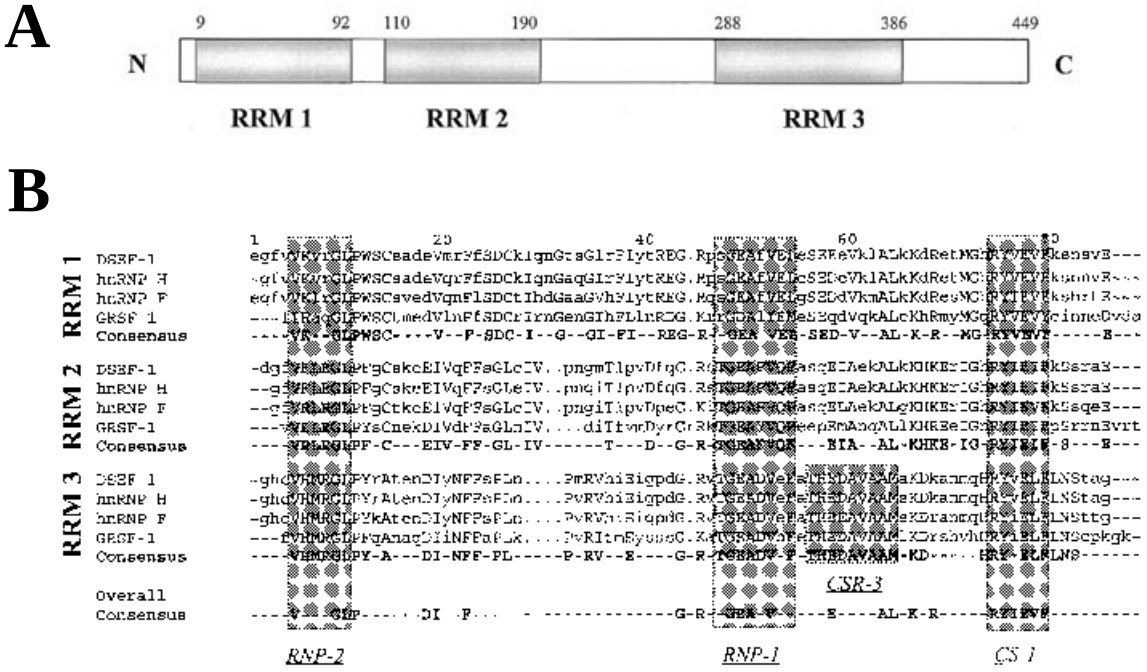

Figure 3. Molecular characteristics of DSEF-1 protein. (**A**) Diagrammatic representation of DSEF-1 protein showing the relative positions of the RNA recognition motifs (RRMs). The numbers above the diagram refer to amino acid positions relative to the N-terminus. (**B**) Alignment of RNA binding domains of proteins with three RRMs which bind G-rich sequences with high affinity. Regions of significant similarity are shaded.

protease and digestion products were analyzed on a 15% acrylamide gel containing SDS. As seen in Figure 2, authentic cross-linked DSEF-1 protein and the purified cross-linked 50 kDa protein produced an identical pattern of partial protease products. We conclude, therefore, that we have purified the DSEF-1 protein to near homogeneity.

Molecular characterization of DSEF-1

In order to obtain cDNA clones which encode the DSEF-1 protein, the sequence of four tryptic peptides was obtained (Materials and Methods). These tryptic peptides displayed significant similarity with the RNA-binding domains of several hnRNP proteins which bind to poly(G) with high affinity (Fig. 3B). We took advantage of this similarity to determine the probable relative alignment of the DSEF-1 tryptic peptides and prepared degenerate primers for PCR amplification. A 243 bp DSEF-1-specific fragment was amplified from total HeLa cell RNA. This fragment was used to probe a λ cDNA library from HeLa cells by plaque hybridization. We obtained and characterized a 2.4 kb cDNA clone containing the entire open reading frame of DSEF-1. This cDNA clone is identical to hnRNP H′, a variant of hnRNP H protein (27). The DSEF-1/hnRNP H′ gene has been mapped to two chromosomes, Xq21.3–q22 and 6q25.3–q26 (27,31). The very high degree of similarity between hnRNP H and H′ (96.2% identity), as well as the fact that the sequence of the four isolated DSEF-1 tryptic peptides is shared by both proteins, suggests that the identity of DSEF-1 is consistent with either hnRNP H or H′. A monoclonal antibody obtained from Gideon Dreyfuss which recognizes hnRNP F and H proteins (28), however, failed to immunoprecipitate cross-linked DSEF-1 protein from nuclear extracts (data not shown). DSEF-1, therefore, is a member of the hnRNP H family of proteins and most likely identical to hnRNP H′.

As seen in Figure 3, DSEF-1/hnRNP H′ is 449 amino acids and contains three RNA recognition motifs (RRMs) arranged in a manner similar to other poly(G)-binding proteins (24,27,28). The protein contains two glycine-rich auxiliary domains: a 98 amino acid region located between RRM2 and RRM3 and a C-terminal domain. DSEF-1/hnRNP H′ is the fourth member of a family of three-RRM-containing RNA-binding proteins which recognize G-rich sequences. All four proteins have a similar organization of these three RRMs and differ mostly by the number of amino acids at the N- or C-terminal ends (24,27). The RRMs of poly(G)-binding proteins are slightly different than those found in other RNA-binding proteins. The RNP-1 and RNP-2 sub-motifs deviate slightly from the known consensus sequence (27,28). The sequences of the three individual RNA-binding domains of these four proteins have been aligned in Figure 3B. In addition to the previously described RNP-1 and RNP-2 consensus sequences (32), the binding domains of these G-rich RNA-binding proteins contain another conserved domain, CS-1 (RYIEVF), located 18 residues downstream of the RNP-1 motif. A fourth conserved domain, CSR-3 (THEDAVAAM), is located on the C-terminal side of RNP-1 in RRM3 of all four proteins. The highly conserved CS-1 and CSR-3 domains appear to be unique to proteins that interact with G-rich sequences and may contribute to the RNA-binding specificity of this family of proteins.

DSEF-1 specifically stimulates 3′-end processing *in vitro*

Previous work has shown that DSEF-1 interacts specifically with the GRS element, a motif which stimulates cleavage efficiency (11,17). These data provided suggestive evidence for a role for the DSEF-1

protein in efficient 3′-end formation, but a direct demonstration of the functional significance of DSEF-1 is clearly warranted.

Previous studies using *in vitro* systems derived from HeLa nuclear extracts have demonstrated a faithful reproduction of *in vivo* observations related to the effect of downstream sequences on 3′-end processing (33). *In vitro* cleavage and polyadenylation systems using purified components, on the other hand, may be lacking *in vivo* regulatory influences such as proteins which form the RNP substrate, competing proteins, the proper titrations of available general factors, etc. In order to assess the functional significance of DSEF-1 protein to 3′-end processing efficiency, therefore, we focused our efforts on a nuclear extract-based *in vitro* system.

HeLa nuclear extracts were diluted ~2-fold from the levels we ordinarily use to observe efficient 3′-end processing. The rationale behind this was to determine if DSEF-1 could stimulate processing efficiency of a sub-optimal system. As seen in Figure 4A, lanes 0, substrate RNAs containing the SVL polyadenylation signal were processed at a detectable but albeit inefficient level in this diluted system. Adding increasing amounts of a GST–DSEF-1 fusion protein to this system dramatically increased (≥4-fold) the cleavage efficiency of the SVL substrate RNA in a linear, dose-dependent fashion. The addition of similar amounts of the GST protein alone had no effect on processing efficiency. Similar results were obtained with several independent nuclear extract preparations (Fig. 4B and data not shown), as well as with undiluted nuclear extracts (Fig. 6). We conclude that DSEF-1 is an auxiliary factor capable of specifically stimulating cleavage efficiency *in vitro*.

We next assessed whether recombinant DSEF-1 fusion protein could also stimulate polyadenylation efficiency. As seen in Figure 4B, DSEF-1 specifically stimulated polyadenylation of the SVL RNA substrate in the *in vitro* system. Stimulation of 3′-end processing efficiency by DSEF-1 also required the core elements of the polyadenylation signal. As seen in Figure 4B, mutation of the AAUAAA of the SVL substrate RNA to AAGAAA abrogated the ability of DSEF-1 to stimulate either cleavage or polyadenylation. We conclude that DSEF-1 protein stimulates 3′-end processing which is directed by the core elements.

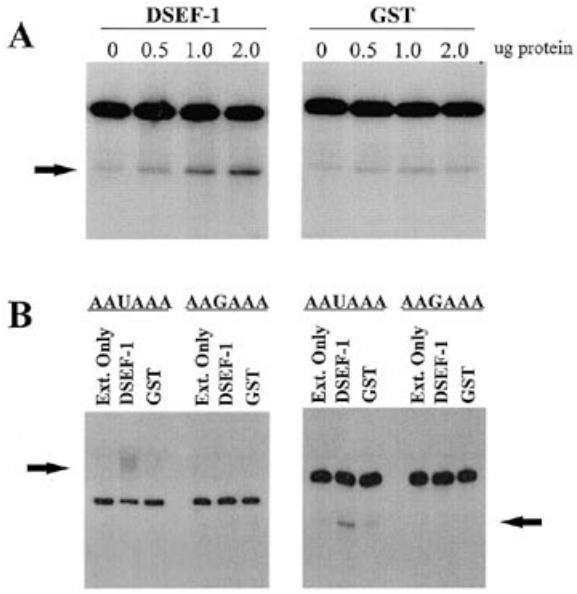

Figure 4. DSEF-1 specifically stimulates 3′-end cleavage and polyadenylation in an AAUAAA-dependent fashion. (**A**) Increasing amounts of a GST–DSEF-1 fusion protein or GST protein alone were added with an RNA substrate containing the SVL polyadenylation signal to an *in vitro* cleavage system containing sub-optimal amounts of nuclear extract. Cleavage products were analyzed on a 5% acrylamide gel containing urea. The 5′ cleavage product is indicated by the arrow. (**B**) An RNA substrate containing a wild-type SVL polyadenylation signal (AAUAAA) or one containing a U→G transversion in the canonical hexanucleotide motif (AAGAAA) was incubated in an *in vitro* processing system containing sub-optimal amounts of nuclear extract in the presence of DSEF-1 fusion protein (lanes DSEF-1), GST protein (lanes GST) or no added protein (lanes Ext. Only). The left side shows the results under polyadenylation conditions, while the right side shows the results of cleavage assays. Products were analyzed on 5% acrylamide–urea gels. The arrows indicate the polyadenylated products (left) or 5′ cleavage product (right).

DSEF-1 protein can stimulate 3′-end processing of polyadenylation signals which lack a GRS element

hnRNP proteins are found on the majority of pre-mRNAs in the nucleus (32). Since DSEF-1 protein is a member of the hnRNP H family of RNA-binding proteins, it is likely, therefore, to interact with many different pre-mRNAs. Previous surveys analyzing the frequency of the high affinity DSEF-1 binding site (the GRS element) on RNAs, however, showed that only a few polyadenylation signals contained the well-characterized GRS binding site (17). It is possible, however, that DSEF-1 may be capable of interacting with many polyadenylation signals due to low affinity RNA–protein interactions that may be stabilized by protein–protein interactions with other members of the hnRNP complex.

Previous assays for DSEF-1 binding using a high stringency band shift assay demonstrated its strong affinity for the GRS (17). In order to search for low affinity binding sites, we tested the ability of DSEF-1 protein to interact with RNAs using a low ionic strength band shift assay. This assay allows detection of DSEF-1 RNA interactions that are less stable than the DSEF-1 protein–GRS interaction. As seen in Figure 5, in addition to forming a complex with the GRS-containing SVL poly(A) signal, DSEF-1 protein also formed complexes with the adenoviral IVA_2 poly(A) signal and the cellular Ig μM signal (both of which lack a GRS). DSEF-1 protein did not, however, form stable complexes with the adeno-associated virus (AAV) poly(A) signal (lanes AAV). DSEF-1 protein failed to form complexes with the IVA_2 or Ig μM RNAs in band shift assays using higher ionic strength gels (data not shown). We conclude that DSEF-1 protein can interact with sequences in addition to the G-rich element and with poly(A) signals in addition to SVL.

The effect of DSEF-1 in the processing of poly(A) signals that lack a GRS was assessed. We first tested whether DSEF-1 protein is important for determining the basal level of processing of these poly(A) signals in nuclear extracts. RNAs containing the SVL, IVA_2, AAV and Ig μM polyadenylation signals were incubated in an *in vitro* cleavage system using nuclear extracts depleted of endogenous DSEF-1 protein by incubation with a biotinylated GRS RNA oligonucleotide and streptavidin–agarose. These extracts were depleted of DSEF-1 protein as judged by the absence of a DSEF-1 cross-linking to the GRS-containing SVL RNA (data not shown). As seen in Figure 6, depletion of DSEF-1 protein reduced the processing of the SVL poly(A) signal to ~30% of the level seen with mock-depleted extracts. The depletion of endogenous DSEF-1 protein had only minimal

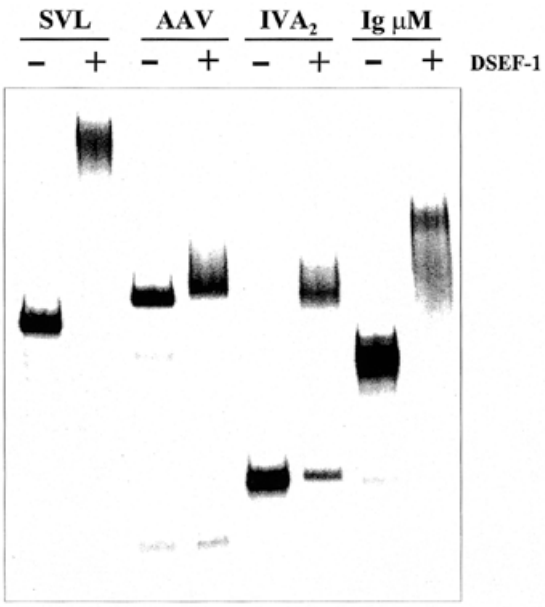

Figure 5. DSEF-1 can interact with some polyadenylation signals that lack a GRS high affinity binding site. RNA substrates containing the SVL, adenovirus IVA_2, immunoglobulin μM (Ig μM) or adeno-associated virus (AAV) polyadenylation signals were incubated with recombinant DSEF-1 protein (+ lanes) and protein–RNA complexes were analyzed on a non-denaturing 5% acrylamide gel. (–) lanes indicate the input RNA.

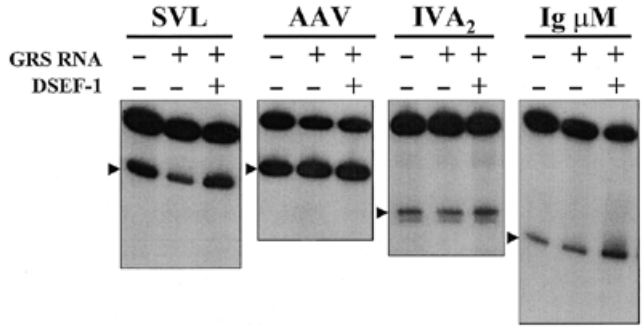

Figure 6. DSEF-1 stimulates processing of polyadenylation signals that contain a GRS element or a low affinity binding site. RNA substrates containing the SVL, adenovirus IVA_2, immunoglobulin μM (Ig μM) or adeno-associated virus (AAV) polyadenylation signals were incubated in the *in vitro* cleavage system using mock-depleted nuclear extracts (lanes GRS RNA –, DSEF-1 –), extracts that were depleted of DSEF-1 (lanes GRS RNA +, DSEF-1 –) or DSEF-1-depleted extracts supplemented with recombinant DSEF-1 protein (lanes GRS RNA +, DSEF-1 +). Reaction products were analyzed on 5% acrylamide gels containing urea. The positions of the 5′ cleavage products are indicated by the arrowheads.

effects on the processing of the IVA_2, AAV and Ig μM poly(A) signals (Fig. 6). We conclude that unlike the SVL signal, none of the other three polyadenylation signals tested rely on DSEF-1 as an auxiliary factor to determine the basal level of processing efficiency observed in HeLa nuclear extracts. DSEF-1 protein, however, may only be available in limiting amounts to bind RNA substrates in HeLa nuclear extracts. In fact, RNAs like IVA_2 and Ig μM that lack a high affinity GRS binding site do not appear to be capable of interacting with DSEF-1 protein in HeLa extracts as judged by UV cross-linking (data not shown). If added in higher amounts to nuclear extracts, however, DSEF-1 protein might still be able to enhance the processing efficiency of poly(A) signals that lack a GRS element. As seen in Figure 6, the addition of recombinant DSEF-1 protein to depleted extracts stimulated the processing of all of the polyadenylation signals that DSEF-1 can interact with (SVL, 4.1 ± 0.6-fold; IVA_2, 1.8 ± 0.6-fold; Ig μM, 2.1 ± 0.4-fold), but did not stimulate the processing of the AAV signal, which it failed to form a stable complex with in band shift assays (Fig. 5). We conclude that DSEF-1 protein is an auxiliary factor that can stimulate the processing efficiency of a variety of cellular and viral polyadenylation signals.

DSEF-1 stimulates 3′-end processing by increasing the efficiency of CstF assembly on polyadenylation substrate RNAs

The observation in Figure 4B that DSEF-1 requires core elements of the polyadenylation signal to stimulate 3′-end processing suggests that the auxiliary factor may function by promoting the assembly of the general polyadenylation factors. In order to test this hypothesis, we determined the influence of DSEF-1 protein on the level of CstF cross-linking to the U-rich downstream element of the RNA substrate. This is an excellent measure of overall complex assembly, since efficient CstF cross-linking also requires the cooperative binding of CPSF to the AAUAAA element (34).

Radiolabeled SVL RNA was incubated in the *in vitro* processing system in the presence of DSEF-1, GST or no recombinant protein. Reaction mixtures were incubated briefly to allow the assembly of polyadenylation complexes onto the RNA substrate. Mixtures were irradiated with UV light to covalently cross-link proteins to the RNA substrate. Following RNase treatment, cross-linked 64 kDa subunit of CstF (CstF-64) was specifically immunoprecipitated using the 3A7 monoclonal antibody (26). As seen in Figure 7, the addition of DSEF-1 protein to the *in vitro* system specifically increased the level of CstF-64 protein cross-linking to the SVL substrate RNA. This strongly suggests that DSEF-1 protein increases the efficiency of 3′-end processing by promoting the assembly of the general polyadenylation factors onto the RNA substrate.

DISCUSSION

In this study we report the purification, molecular characterization and the functional significance of the first auxiliary factor specific for a downstream auxiliary element of a mammalian polyadenylation signal. These data demonstrate that proteins which interact with downstream sequences other than the core U-rich element can play a role in determining processing efficiency. Since downstream regions are not generally conserved among polyadenylation signals, this study illustrates a potential for the presence of other downstream binding proteins which could affect polyadenylation in a tissue- or developmental-specific fashion. The DSEF-1 protein is an obvious candidate for such a regulatory protein, especially given the observation that it can stimulate processing efficiency in the absence of a GRS element. In conjunction with

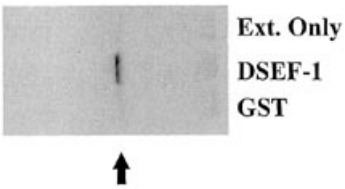

Figure 7. DSEF-1 specifically stimulates cross-linking of the 64 kDa subunit of CstF to polyadenylation substrate RNAs. SVL RNA was incubated in the *in vitro* polyadenylation system alone (lane Ext. Only), in the presence of DSEF-1 (lane DSEF-1) or GST protein (lane GST). Reaction mixtures were irradiated with UV light, treated with RNase and the 64 kDa protein of CstF was isolated by immunoprecipitation with the 3A7 monoclonal antibody. The level of cross-linked 64 kDa protein (indicated by the arrow) was analyzed on a 10% acrylamide gel containing SDS.

previous data concerning the stimulation of processing by U1A protein through an auxiliary upstream element (16,35), this study provides an explanation why the SVL polyadenylation signal is processed so efficiently both *in vivo* and *in vitro*. Perhaps a highly efficient polyadenylation signal is vital to allow SV40 virus to express its late genes at a time post-infection when general polyadenylation factors may be limiting due to virus-mediated shutdown/appropriation of cellular processes.

DSEF-1/hnRNP H′ protein is very similar to poly(G)-binding hnRNP F protein (75.4% identity) and is nearly identical to the hnRNP H protein (96.2% identity) (28). GRSF-1, another RNA-binding protein that specifically recognizes G-rich sequences (24), exhibits 43.4% identity (53.3% similarity) to DSEF-1. Alignment of the RRMs of this group of G-rich RNA-binding proteins demonstrates two regions (in addition to the canonical RNP-1 and RNP-2 domains) that appear to be unique to this family and conserved with respect to position and sequence (Fig. 3). These regions, which we have called CS-1 and CSR-3, may contribute to the affinity and specificity of this protein family for G-rich tracts. Mutagenesis experiments are currently underway to test this hypothesis.

In addition to forming a vital part of pre-mRNP complexes, several hnRNP proteins that bind to G-rich sequences have been implicated in the regulation of post-transcriptional events (36–38). These include the formation of neural-specific alternative splicing complexes containing hnRNP F (39). The data presented in this study with DSEF-1 protein demonstrates that members of the hnRNP family of proteins can also influence the efficiency of 3′-end formation, the other major nuclear RNA processing event.

The hnRNP F protein has also recently been implicated in the stimulation of RNA splicing by the nuclear cap binding complex through a direct protein–protein interaction (40). The presence of a cap structure and cap-binding proteins is also required for efficient 3′-end processing (41,42) and recent data has shown that the 5′- and 3′-ends of a polyadenylation substrate RNA are associated with each other during processing (42). Based on its similarity to hnRNP F, we hypothesized that the DSEF-1 protein may be responsible for cap-dependent stimulation of polyadenylation efficiency. In order to address this model, we tested the effect of DSEF-1 protein on the processing of uncapped RNA substrates in our *in vitro* system. Removal of the 5′-cap from RNA substrates reduced overall 3′-end processing to undetectable levels (data not shown). Addition of DSEF-1 protein failed to stimulate processing of these uncapped RNAs (data not shown). Since uncapped transcripts failed to be processed at detectable levels in the absence of DSEF-1 protein, however, we cannot determine if the failure of DSEF-1 protein to stimulate processing suggests a direct interaction with the cap complex. Alternatively, DSEF-1 protein may simply not be able to overcome the debilitation of overall 3′-end processing caused by the removal of the cap from the polyadenylation signal.

The data in Figure 7 shows that DSEF-1 stimulates 3′-end formation by promoting the assembly of the general polyadenylation factors on the RNA substrate. We propose two models for how DSEF-1 may be directly stimulating this assembly. First, DSEF-1 may interact with the RNA substrate and affect its structure, specifically by unwinding secondary structures which may impede the assembly of the general polyadenylation factors. Recent studies have suggested that an unstructured core region is very efficiently recognized by the general factors (13,14,43). In our second model for DSEF-1 function, DSEF-1 interacts with the RNA substrate and may promote the assembly of CstF by direct protein–protein interactions. The observation that the upstream auxiliary factor U1A may promote assembly of CPSF by direct interaction with the 160 kDa subunit (16) provides precedence for auxiliary factors acting in this fashion. Furthermore, the observation that placement of the GRS element farther away from the core U-rich element reduced its ability to mediate efficient processing (17) may be consistent with a requirement for a DSEF-1–CstF interaction.

ACKNOWLEDGEMENTS

We wish to thank Lance Ford for critical comments on the manuscript. This work was supported by a grant from the National Institutes of Health (GM56434) and an award from the Pew Charitable Trusts to JW.

REFERENCES

1 Proudfoot,N. (1996) *Cell*, **87**, 779–781.
2 Wahle,E. and Keller,W. (1996) *Trends Biochem. Sci.*, **21**, 247–250.
3 Chen,F., MacDonald,C.C. and Wilusz,J. (1995) *Nucleic Acids Res.*, **23**, 2614–2620.
4 Gilmartin,G.M. and Nevins,J.R. (1991) *Mol. Cell. Biol.*, **11**, 2432–2438.
5 MacDonald,C.C., Wilusz,J. and Shenk,T. (1994) *Mol. Cell. Biol.*, **14**, 6647–6654.
6 Murthy,K.G.K. and Manley,J.L. (1995) *Genes Dev.*, **9**, 2672–2683.
7 Edwalds-Gilbert,G., Prescott,J. and Falck-Pedersen,E. (1993) *Mol. Cell. Biol.*, **13**, 3472–3480.
8 Edwalds-Gilbert,G., Veraldi,K.L. and Milcarek,C. (1997) *Nucleic Acids Res.*, **25**, 2547–2561.
9 Niwa,M. and Berget,S.M. (1991) *Genes Dev.*, **5**, 2086–2095.
10 Moreira,A., Wollerton,M., Monks,J. and Proudfoot,N.J. (1995) *EMBO J.*, **14**, 3809–3819.
11 Qian,Z. and Wilusz,J. (1991) *Mol. Cell. Biol.*, **11**, 5312–5320.
12 Valsamakis,A., Schek,N. and Alwine,J.C. (1992) *Mol. Cell. Biol.*, **12**, 3699–3705.
13 Graveley,B.R., Fleming,E.S. and Gilmartin,G.M. (1996) *Mol. Cell. Biol.*, **16**, 4942–4951.
14 Graveley,B.R., Fleming,E.S. and Gilmartin,G.M. (1996) *J. Biol. Chem.*, **271**, 33654–33663.
15 Gilmartin,G.M., Fleming,E.S., Oetjen,J. and Graveley,B.R. (1995) *Genes Dev.*, **9**, 72–83.
16 Lutz,C.S., Murthy,K.G.K., Schek,N., O'Conner,J.P., Manley,J.L. and Alwine,J.C. (1996) *Genes Dev.*, **10**, 325–337.
17 Bagga,P.S., Ford,L.F., Chen,F. and Wilusz,J. (1995) *Nucleic Acids Res.*, **23**, 1625–1631.
18 Wilusz,J. and Shenk,T. (1988) *Cell*, **52**, 221–228.
19 Sheets,M.D., Ogg,S.C. and Wickens,M.P. (1990) *Nucleic Acids Res.*, **18**, 5799–5805.

20 Peterson,M.L. and Perry,R.P. (1989) *Mol. Cell. Biol*., **9**, 726–738.
21 Wilusz,J., Feig,D.I. and Shenk,T. (1988) *Mol. Cell. Biol*., **8**, 4477–4483.
22 Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res*., **11**, 1475–1489.
23 Cleveland,D.W., Fischer,S.G., Kirschner,M.W. and Laemmli,U.K. (1977) *J. Biol. Chem*., **252**, 1102–1106.
24 Qian,Z. and Wilusz,J. (1994) *Nucleic Acids Res*., **22**, 2334–2343.
25 Moore,C.L. and Sharp,P.A. (1985) *Cell*, **41**, 845–855.
26 Takagaki,Y., Manley,J.L., MacDonald,C.C., Wilusz,J. and Shenk,T. (1990) *Genes Dev*., **4**, 2112–2120.
27 Honore,B., Rasnussen,H.H., Vorum,H., Dejgaard,K., Liu,X., Gromov,P., Madsen,P., Gesser,B., Tommerup,N. and Celis,J.E. (1995) *J. Biol. Chem*., **270**, 28780–28789.
28 Matunis,M.J., Xing,J. and Dreyfuss,G. (1994) *Nucleic Acids Res*., **22**, 1059–1067.
29 Amrute,S.B., Abdul-Manan,Z., Pandey,V., Williams,K.R. and Modak,M.J. (1994) *Biochemistry*, **33**, 8282–8291.
30 Merrill,B.M., Stone,K.L., Cobianchi,F., Wilson,S.H. and Williams,K.R. (1988) *J. Biol. Chem*., **263**, 3307–3313.
31 Vorechovsky,I., Vetrie,D., Holland,J., Bently,D.R., Thomas,K., Zhou,J., Notarangelo,L.D., Plebani,A., Fontan,G., Ochs,H.D., Hammarstrom,L., Sideras,P. and Smith,C.I.E. (1994) *Genomics*, **21**, 517–524.
32 Dreyfuss,G., Matunis,M.J., Pinol-Roma,S. and Bird,C.G. (1993) *Annu. Rev. Biochem*., **62**, 289–321.
33 Green,T.L. and Hart,R.P. (1988) *Mol. Cell. Biol*., **8**, 1839–1841.
34 Wilusz,J., Shenk,T., Takagaki,Y. and Manley,J.L. (1990) *Mol. Cell. Biol*., **10**, 1244–1248.
35 Lutz,C.S. and Alwine,J.C. (1994) *Genes Dev*., **8**, 576–586.
36 Ashiya,M. and Grabowski,P.J. (1997) *RNA*, **3**, 996–1015.
37 Chan,R.C. and Black,D.L. (1997) *Mol. Cell. Biol*., **17**, 4667–4676.
38 Horowitz,D.S. and Krainer,A.R. (1994) *Trends Genet*., **10**, 100–106.
39 Min,H., Chan,R.C. and Black,D.L. (1995) *Genes Dev*., **9**, 2659–2671.
40 Gamberi,C., Izaurralde,E., Beisel,C. and Mattaj,I.W. (1997) *Mol. Cell. Biol*., **17**, 2587–2597.
41 Cooke,C. and Alwine,J.C. (1996) *Mol. Cell. Biol*., **16**, 2579–2584.
42 Flaherty,S.M., Fortes,P., Izaurralde,E., Mattaj,I.W. and Gilmartin,G.M. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 11893–11898.
43 Chen,F. and Wilusz,J. (1998) *Nucleic Acids Res*., **26**, 2891–2898.