MOLECULAR AND CELLULAR BIOLOGY, Oct. 1988, p. 4477–4483
0270-7306/88/104477-07$02.00/0

Vol. 8, No. 10

The C Proteins of Heterogeneous Nuclear Ribonucleoprotein Complexes Interact with RNA Sequences Downstream of Polyadenylation Cleavage Sites

JEFFREY WILUSZ, DANIEL I. FEIG, AND THOMAS SHENK*

Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544

Received 20 April 1988/Accepted 22 July 1988

The heterogeneous nuclear ribonucleoprotein C1 and C2 proteins were preferentially cross-linked by treatment with UV light in nuclear extracts to RNAs containing six different polyadenylation signals. The domain required for the interaction was located downstream of the poly(A) cleavage site, since deletion of this segment from several polyadenylation substrate RNAs greatly reduced cross-linking efficiency. In addition, RNAs containing only downstream sequences were efficiently cross-linked to C proteins, while fully processed, polyadenylated RNAs were not. Analysis of mutated variants of the simian virus 40 late polyadenylation signal showed that uridylate-rich sequences located in the region between 30 and 55 nucleotides downstream of the cleavage site were required for efficient cross-linking of C proteins. This downstream domain of the simian virus 40 late poly(A) addition signal has been shown to influence the efficiency of the polyadenylation reaction. However, there was not a strict correlation between cross-linking of C proteins and the efficiency of polyadenylation.

Maturation of the 3′ end of most pre-mRNAs involves a site-specific endonucleolytic cleavage followed by polymerization of 150 to 200 adenylate residues at the newly formed 3′ terminus (3, 24). Two sequence elements are required for efficient cleavage and polyadenylation. A highly conserved hexanucleotide, AAUAAA, located 10 to 30 bases upstream from the site of cleavage (25) has been shown to be absolutely required for polyadenylation of most RNA transcripts (12, 17, 20, 34). A second element, located 3′ to the cleavage site, has been identified in several genes and shown to enhance the efficiency of polyadenylation (7, 8, 13, 26, 28). Although U or GU residues are highly represented, a consensus sequence for the downstream element is unclear (1, 14, 15, 19).

The cleavage and poly(A) addition reactions have been accurately and efficiently reproduced in vitro (22). A complex is formed in association with the substrate RNA during in vitro processing (18, 23, 29, 33, 37, 38). Formation of specific complexes requires the AAUAAA sequence and the downstream element or, alternatively, a 3′ end in proximity to the AAUAAA. These complexes may include small nuclear ribonucleoproteins (snRNPs), since autoimmune sera containing antibodies specific for the La and U1-RNP antigenic determinants, as well as a monoclonal antibody specific for the Sm determinant, can inhibit in vitro polyadenylation (21, 22). Furthermore, antibodies specific for the Sm antigen or trimethyl cap of snRNAs can precipitate T1 nuclease-resistant RNA fragments containing the AAUAAA sequence (16, 18) and trimethyl cap-specific antibodies can precipitate substrate RNA present in a polyadenylation-specific 50S complex (23).

We have previously used photocross-linking to transfer radioactivity from ^{32}P-labeled substrate RNAs to associated proteins within a cell-free polyadenylation extract (35). This procedure identified a 64-kilodalton (kDa) polypeptide that specifically interacted with all polyadenylation substrates tested. Its binding was shown to be dependent on the AAUAAA sequence, since an RNA carrying an altered AAGAAA sequence failed to interact with the polypeptide.

In this report we demonstrate the association of heterogeneous nuclear ribonucleoprotein (hnRNP) C proteins with several RNAs that contain a polyadenylation signal. The C1 and C2 proteins are constituents of the 30S to 50S hnRNP monoparticle (reviewed in reference 10). They are antigenically related, evolutionarily conserved 32-kDa phosphoproteins (4, 5, 11, 32) which have been shown to play a role in RNA splicing (6). The C proteins can be efficiently cross-linked to sequences downstream of the cleavage site in RNA molecules containing polyadenylation signals.

MATERIALS AND METHODS

Plasmids and transcripts. Linearized templates were transcribed in vitro by using bacteriophage RNA polymerases (SP6 or T7) in the presence of [α-^{32}P]UTP and ^{7m}GpppG (35). The plasmid constructions and transcripts produced are as follows. pC220 (New England Biolabs, Inc.) contains the 1.38-kilobase (kb) *Eco*RI–*Hin*dIII fragment of bacteriophage lambda (sequence 26104 to 27479) cloned into the equivalent sites of pSP64. Transcription of *Sca*I-cut DNA yields a 220-base RNA (termed λC220). pTPL contains the *Hha*I-*Sal*I fragment of pDF4-15 (Logan and Shenk, unpublished), including the adenovirus type 5 (Ad5) tripartite leader cDNA (originally derived from pJAW43 [42]) inserted between the *Hin*dIII and *Sal*I sites of pSP64. Transcription of *Sal*I-cut DNA yields a 243-base RNA (TPL). pGEM4 (Promega Biotech, Inc.) was transcribed after cutting with *Pvu*II to produce a 107-base RNA (GEM). pSP64 (Promega Biotech, Inc.) was transcribed after linearization with *Pvu*II to yield a 232-base RNA (termed 64). pAd5L3 contains a 260-base-pair (bp) *Kpn*I-*Dra*I fragment of Ad5 inserted between the *Kpn*I and *Hin*cII cleavage sites of pGEM4. Transcription of *Hin*dIII- or *Ava*I-cut templates yields a 298-base or 240-base RNA (AdL3 and AdL3-1, respectively). pAd5E1B contains an Ad5 fragment (sequence 3943 to 4122) cloned into the *Bam*HI site of pSP65. Transcription of *Xba*I- or *Hin*fI-cut DNAs generates a 224-base or 144-base RNA (AdE1B and

* Corresponding author.

AdE1B-1, respectively). pAd5E1B-2 contains the 70-bp *Hin*fI-*Bam*HI fragment of pAd5E1B inserted into pGEM3 DNA that was cleaved with *Hin*cII and *Bam*HI. Transcription of *Bam*HI-cut DNA yields a 104-base RNA (AdE1B-2). pAd5IVa2 contains the 155-bp *Bam*HI-*Pvu*II fragment of pAd5E1B cloned into pGEM4 DNA that was cut with *Hin*cII and *Bam*HI. Templates cut with the following enzymes yield the indicated RNAs: *Hin*dIII, 210 bases (AdIVa2); *Bgl*I, 156 bases (AdIVa2-1); and *Mnl*I, 139 bases (AdIVa2-2). pAd 5IVa2-4 contains the *Hin*dIII-*Hin*fI fragment of pAd5E1B inserted into *Hin*cII-cut pGEM3 DNA. Transcription of *Bgl*I-cut DNA yields a 96-base RNA (AdIVa2-4). pSVE contains the 241-bp *Bam*HI-*Bcl*I fragment of simian virus 40 (SV40) (sequence positions 2533 to 2770) inserted into the *Bam*HI site of pSP65. Transcription of *Bam*HI-cut DNA yields a 269-base RNA (SVE). pSVL contains the *Bam*HI-*Bcl*I fragment of SV40 inserted into the *Bam*HI site of pSP65 in the opposite orientation from that in pSVE. Transcription of a *Dra*I-cut plasmid gives a 224-base RNA (SVL), and *Fnu*4HI-linearized template yields a 149-base RNA (SVL-2). pSVL-7 was produced by inserting the *Hin*cII-*Xba*I fragment of pSVL into pGEM4 DNA cleaved with *Sma*I and *Xba*I. Transcription of *Dra*I-cut DNA gives an 89-base RNA (SVL-7). The *Bsm*I-*Eco*RI fragment of pSVL-7 was excised, and the remaining large fragment was blunted with Klenow polymerase and recircularized with DNA ligase to produce pSVL-8. Transcription of *Dra*I-cut DNA gives a 61-base RNA (SVL-8). To make pSVL-12, the DNA oligonucleotide 5′-AGCTTACCTGAAACATAAAATGAATGCA-3′ with its complement was inserted into pSVL3 DNA cleaved with *Bsm*I and *Hin*dIII. pSVL-14 was then constructed by inserting the *Hin*dIII-*Hin*cII fragment of pSVL DNA, which contains sequences upstream of the polyadenylation signal, into the appropriate sites of pSVL-12. Transcription of *Hin*fI-cleaved DNA generates a 223-base RNA (SVL-14). To make pSVL-15 and pSVL16, pSVL3 (which contains the SV40-specific *Alu*I-*Hin*dIII fragment of pSVL inserted into pGEM4 DNA digested with *Hin*cII and *Hin*dIII) was cut at its *Sal*I and *Hin*dIII sites located downstream of the insert, the overhanging ends were blunted, and the plasmid was recircularized to make pSVL-3-S/H. Oligonucleotides containing mutations were inserted between the *Bsm*I and *Hin*dIII sites of pSVL3-S/H. The *Hin*dIII-*Hin*cII fragment of pSVL DNA was then inserted into the appropriate site of the new constructs. Transcription of *Dra*I-cut DNAs yields 224-base RNAs (SVL-15 and SVL-16), with the base changes indicated.

In vitro polyadenylation. Nuclear extracts were prepared from HeLa spinner culture cells grown in 5% calf serum as described by Dignam et al. (10). Polyadenylation reaction mixes contained final concentrations of 3% polyvinyl alcohol, 1 mM ATP, 20 mM phosphocreatine, 12 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid, pH 7.9), 12% glycerol, 60 mM KCl, 0.12 mM EDTA, 0.3 mM dithiothreitol, and 60% (vol/vol) nuclear extract. Reactions were performed at 30°C.

UV cross-linking and label transfer analysis. Gel-purified RNAs (10 fmol), ^{32}P-labeled to the same specific activity, were incubated in the in vitro polyadenylation system for 10 min. *Escherichia coli* tRNA (5 μg) was added, and the sample was irradiated for 10 min at 4°C with a germicidal light (Sylvania G15T8) placed 4 cm from the sample. The UV light generates covalent cross-links between RNA bases and closely associated proteins (30). RNase A was added to a final concentration of 1 mg/ml, and samples were incubated at 37°C for 15 min. An equal volume of protein gel loading buffer containing sodium dodecyl sulfate (SDS) and β-mercaptoethanol was added, and samples were heated to 100°C for 5 min prior to electrophoresis in polyacrylamide gels containing 0.2% SDS.

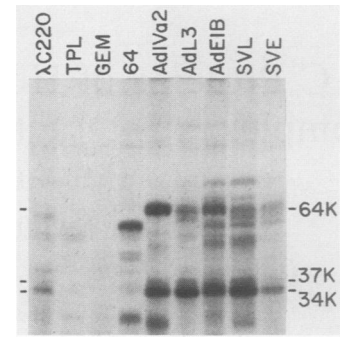

FIG. 1. Label transfer analysis of polypeptides bound to RNAs in polyadenylation extracts. ^{32}P-labeled polyadenylation substrates (AdIVa2, AdL3, AdE1B, SVL, and SVE) or nonspecific RNAs (λC220, TPL, GEM, and 64) of the same specific activity were added to polyadenylation extracts, incubated at 30°C for 10 min, digested with RNase A, and subjected to electrophoresis in a 10% polyacrylamide gel containing 0.1% SDS. The sizes (in kilodaltons) of key polypeptides determined relative to marker proteins are indicated.

Immunoprecipitation. Following UV cross-linking, samples were diluted fivefold with buffer A (50 mM Tris [pH 7.5], 75 mM NaCl, 0.05% Nonidet P-40) and centrifuged for 2 min in a microcentrifuge at 4°C. Antibody was added to supernatants, and samples were incubated at 4°C for 30 min. Rabbit anti-mouse immunoglobulin G (Cappel Laboratories Inc.) was added and incubated for 15 min on ice. Formalin-fixed *Staphylococcus aureus* cells (Pansorbin; Calbiochem, Inc.) were added, and after incubation for 15 min on ice, the pellets were washed five times with buffer A. Loading buffer was added, and samples were boiled for 5 min prior to electrophoresis in SDS-containing polyacrylamide gels.

RESULTS

hnRNP C proteins interact with RNAs that contain a polyadenylation signal. A label transfer experiment in which five different polyadenylation substrates and four control RNAs were examined is displayed in Fig. 1. The polyadenylation substrate RNAs were derived from the SV40 early (SVE), SV40 late (SVL), Ad5 E1B (AdE1B), Ad5 L3 (AdL3), and Ad5 IVa2 (AdIVa2) transcription units. RNAs were labeled with [α-^{32}P]UTP and incubated in a polyadenylation extract. All substrates were efficiently cleaved and polyadenylated in the extracts (data not shown). After 10 min, 5 μg of tRNA was added, and the mixture was UV irradiated for 10 min at 4°C. The cross-linked mixture of protein and RNA was then digested with RNase A, and proteins were analyzed by gel electrophoresis. All of the substrates transferred ^{32}P label to a variety of proteins, including a 64-kDa polypeptide (35) and a protein doublet that migrated at 34 to 37 kDa relative to markers (Fig. 1). Similar results were obtained with an RNA carrying the Ad5 E2A polyadenylation signal (data not shown). None of four control RNAs (λC220, TPL, GEM, and 64) was cleaved or polyadenylated in the extracts (data not shown). The control RNAs failed to efficiently transfer label to the doublet of 34 to 37 kDa (34-37K doublet) although a low level of transfer was detectable in the case of the λC220 RNA and possibly the 64 RNA (Fig. 1). Thus, the 34-37K polypeptides can be

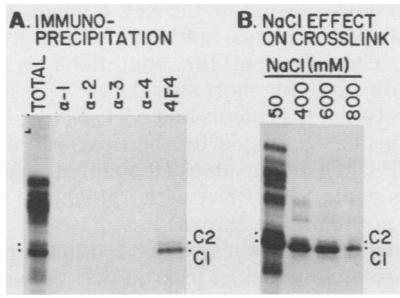

FIG. 2. Identification of the hnRNP C proteins as major ^{32}P-labeled products in label transfer experiments. (A) Immunoprecipitation with a C protein-specific antibody. After cross-linking of ^{32}P-labeled SVL RNA and RNase A treatment, immunoprecipitations were carried out with nonspecific (α1 to α4) or C protein-specific (4F4) monoclonal antibodies. (B) Effect of NaCl on the C protein-RNA interaction. Reaction mixes were adjusted to the indicated concentration of NaCl prior to UV irradiation.

cross-linked to all polyadenylation substrates tested, but their specificity is not absolute.

Additional proteins were cross-linked to each of the polyadenylation substrates (Fig. 1). These proteins were not cross-linked to all of the RNAs tested, and their significance is unclear.

The identity of the 34-37K polypeptides was probed by using antibodies. They were precipitated from cross-linked mixtures containing SVL RNA by using a monoclonal antibody (4F4 [5, 11]) specific for the hnRNP C1 and C2 proteins (Fig. 2A). They were precipitated after RNase digestion, demonstrating that the 34-37K doublet was indeed the moiety recognized by the antibody. If the immunoprecipitation step preceded the RNase treatment, the C proteins, the 64-kDa polypeptide, and the additional cross-linked proteins were precipitated (data not shown), indicating that these proteins can interact with the same molecule of RNA. Immunoprecipitations were performed with the other polyadenylation substrates, and in each case the 4F4 antibody reacted with the 34-37K doublet (data not shown). The association of the 34-37K proteins with SVL RNA was resistant to 600 mM NaCl (Fig. 2B), consistent with previous analyses of C protein-RNA interactions (2).

The 34-37K polypeptides could be precipitated with the 4F4 antibody, and their interaction with RNA was resistant to salt. Furthermore, the 34-37K species were cross-linked to RNAs in a ratio (C1 > C2) similar to that observed for C proteins in vivo (11), and their apparent size corresponded well with that derived from analysis of C protein-specific cDNAs (31.9 kDa [36]). We conclude that the 34-37K doublet represents the hnRNP C1 and C2 proteins.

C proteins interact downstream of the polyadenylation cleavage site. To ascertain which regions of the substrate RNAs interacted with the C proteins, deletion derivatives were prepared and tested for their ability to transfer label. The RNAs used in these experiments are diagrammed in Fig. 3A. Positions on the RNA substrates are designated relative to the cleavage site used during polyadenylation: the first nucleotide upstream is −1 and the first downstream nucleotide is +1. SVL-2 contains only sequences upstream of the cleavage site (−142 to −20), and its ability to transfer label to the C proteins, as well as to other proteins normally labeled in transfer experiments by SVL, was substantially reduced (Fig. 3B). In contrast, two SVL derivatives containing downstream sequences (SVL-7, −6 to +55; SVL-8, +8 to +55) transferred label to C proteins with an efficiency similar

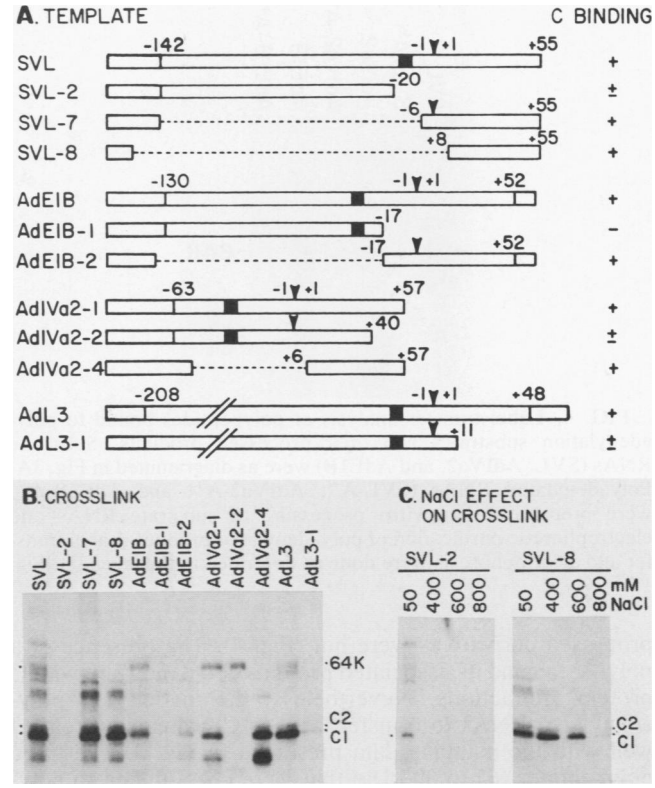

FIG. 3. Localization of the domain on polyadenylation substrates required for interaction with C proteins. (A) Diagram of RNAs used in the analysis. Sequences present in an RNA are represented by rectangles, while those absent are indicated by dashed lines. The AAUAAA sequence is marked by a solid portion of the rectangle. The cleavage site is designated by an arrowhead, and plasmid-specific sequences are shaded. The first nucleotide upstream of the cleavage site is −1, and the first downstream nucleotide is +1. The ability of each RNA to interact with C proteins as assayed by label transfer analysis is indicated to the right. (B) Label transfer analysis of polypeptides bound to RNAs in polyadenylation extracts. Experimental procedures were as described in the legend to Fig. 1. (C) Effect of NaCl concentration on the C protein-RNA interaction. Label transfer reaction mixes were adjusted to the indicated concentrations of NaCl prior to UV irradiation.

to that of the entire SVL transcript (Fig. 3B). Furthermore, the C protein interaction with SVL-8 was resistant to high salt, while the low-level interaction with SVL-2 was completely inhibited by 400 mM NaCl (Fig. 3C). These results indicate that the C proteins preferentially interact with the downstream domain of SVL and that the interaction is independent of the presence of an AAUAAA sequence or its attendant 64-kDa binding protein. These findings were generalized by testing cross-linking to derivatives of the AdE1B, AdIVa2, and AdL3 RNAs. In each case, efficient label transfer mapped to the domain downstream of the polyadenylation cleavage site (Fig. 3B). Similar results were obtained with truncated SVE and AdE2A RNAs (data not shown).

The mapping data predict that C proteins might not interact with the polyadenylated products of the substrate RNAs. To test this prediction, RNAs that were cleaved and polyadenylated in vitro were subjected to label transfer analysis in cell extracts. While SVL, AdIVa2, and AdE1B RNAs were efficiently cross-linked to C proteins, their

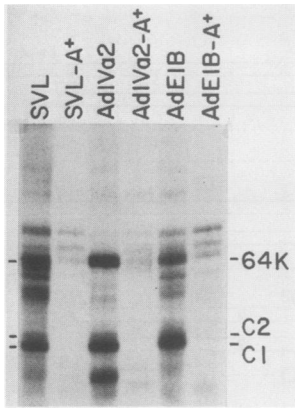

FIG. 4. Label transfer analysis of polypeptides bound to polyadenylation substrates and their processed products. Substrate RNAs (SVL, AdIVa2, and AdE1B) were as diagrammed in Fig. 3A. Polyadenylated RNAs (SVL-A^+, AdIVa2-A^+, and AdE1B-A^+) were prepared by in vitro processing of substrate RNAs and electrophoretic purification of polyadenylated product. Label transfer and electrophoresis were done as described in the legend to Fig. 1.

processed derivatives were not (Fig. 4). The presence of a poly(A) tail and its associated proteins could interfere with C protein interactions. Nevertheless, the inability of polyadenylated RNAs to transfer label to C proteins correlated well with the mapping data presented in Fig. 3. The three polypeptides (65 to 90 kDa) that were cross-linked to polyadenylated RNAs might be part of a complex associated with polyadenylated RNAs that has been described by Humphrey et al. (18).

Fine mapping and functional analysis of the C protein site of interaction. The fact that the C protein-binding domain mapped to the downstream side of the poly(A) cleavage site raised the possibility that it coincided with the downstream element required for efficient polyadenylation. Accordingly, SVL was chosen for further analysis, since its downstream element has been extensively studied (9, 26, 27).

To better define the sequence required for interaction of C proteins with SVL RNA, additional derivatives of the RNA were prepared. SVL-14 RNA (contains nucleotides from −142 to +29, Fig. 5A) was not efficiently cross-linked to C proteins (Fig. 5B). Possible 3′-end effects on the C protein interaction were ruled out by including plasmid-specific sequences downstream of +29 in SVL-14 (Fig. 5A). Since SVL-8 (contains nucleotides from +8 to +55, Fig. 3A) was efficiently cross-linked to C proteins (Fig. 3B), we can conclude that at least part of the sequence required for the interaction lies between nucleotides +30 and +55 (underlined sequence, Fig. 5E).

SVL-14 RNA failed to transfer label to several polypeptides in addition to the C proteins (Fig. 5B). The significance of the additional altered polypeptides in the SVL-14 pattern to polyadenylation is unclear, since they were not efficiently cross-linked to all polyadenylation signals tested (Fig. 1 and 3).

Recently, Swanson and Dreyfuss (31) used affinity chromatography to implicate polyuridylate stretches in the binding of C proteins to RNA. The domain required for cross-linking of C proteins to SVL RNA contained two U stretches, two and six residues in length (Fig. 5E). SVL-16 RNA, which contains U to G transversions at +30 and +52 that disrupt these U stretches, transferred label to C proteins with much reduced efficiency (Fig. 5A and B). SVL-15 RNA, with transversions outside of the key domain, generated a pattern of labeled proteins indistinguishable from that of SVL RNA. The effect of the alterations in SVL-16 is consistent with a role for short stretches of U residues in the interaction between C proteins and SVL RNA. A role for U residues is further suggested by the observation that SVL RNA with ^{32}P-labeled U residues efficiently transferred label to C proteins while SVL RNA with ^{32}P-labeled G, C, or A residues did not (data not shown).

Our mapping of SVL indicates that a domain required for C protein cross-linking lies at least in part between nucleotides +29 and +55 and that U residues at either +30 or +52 are important. This mapping coincides with the downstream domain (nucleotides +47 to +55, overlined in Fig. 5E) required for efficient polyadenylation at the SVL site within transfected cells (27). Therefore, the in vitro polyadenylation efficiency of the SVL derivatives was assessed (Fig. 5C). Curiously, SVL RNAs lacking sequences downstream of +29 were rapidly degraded in polyadenylation extracts (SVL-2 and SVL-14, Fig. 5C and data not shown). The RNAs were stabilized in the presence of EDTA, allowing the efficiency of in vitro cleavage at the poly(A) addition site to be monitored. Identical cross-linking results were obtained in the presence of EDTA (data not shown). SVL-14 RNA, which was cross-linked inefficiently to C proteins, was cleaved three- to fourfold less efficiently than the wild-type substrate (Fig. 5D). However, SVL-16 RNA also cross-linked poorly to C proteins (Fig. 5B) but was cleaved (Fig. 5D) and polyadenylated (Fig. 5C) normally. Perhaps C proteins functionally interact with SVL-16 RNA but fail to be cross-linked. Alternatively, C proteins might have no role in the polyadenylation reaction. The functional species could be one of the additional polypeptides that fail to interact with SVL-14 RNA. Inefficient cleavage and polyadenylation do not always correlate with the inability to transfer label to C proteins.

DISCUSSION

The main conclusion of this work is that the hnRNP C proteins can be cross-linked to sequences downstream of the cleavage site on a variety of RNAs that serve as polyadenylation substrates (Fig. 1, 2, and 3). RNA sequence downstream of the cleavage site is sufficient to mediate the interaction, since deleted substrates including only the downstream domain (SVL-8 and AdIVa2-4, Fig. 3) can efficiently transfer label to C proteins. Thus, there is no dependence on the canonical AAUAAA polyadenylation signal or its attendant 64-kDa binding protein (35) for interaction of C proteins with the downstream element. Conversely, the 64-kDa moiety can bind to polyadenylation substrates lacking the preferred C protein-binding site (35).

The specificity of the C protein interaction with the downstream domain of polyadenylation substrates is not absolute. Substrate RNAs lacking the downstream domain (e.g., SVL-2, Fig. 3) and even RNAs with no known function in eucaryotic cells (e.g., λC220, Fig. 1) can be cross-linked to C proteins. However, the efficiency of cross-linking is much reduced (Fig. 1 and 3) and the interaction is more readily inhibited by high salt (Fig. 3 and data not shown) for these RNAs compared with the downstream domains of polyadenylation substrates. Thus, C protein cross-linked to a nonspecific RNA probably represents a different type of interaction than occurs when C proteins are cross-linked to the downstream domain of poly(A) addition signals. Perhaps the downstream domains contain a high-affinity sequence

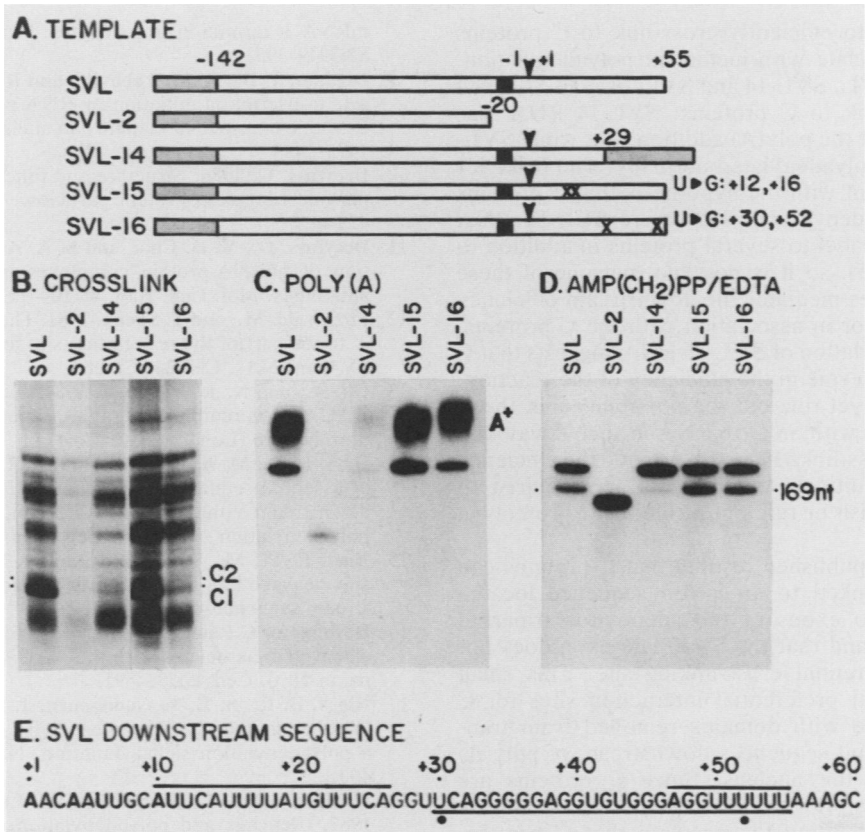

FIG. 5. Localization of the domain required for preferential C protein interaction with SVL RNA and the effect of its loss on cleavage and polyadenylation. (A) Diagram of substrate RNAs. Symbols and numbering are as described in the legend to Fig. 3A. (B) Label transfer analysis of polypeptides bound to RNAs in polyadenylation extracts. Label transfer and electrophoresis were as described in the legend to Fig. 1. (C) Accumulation of polyadenylated product. Substrate RNAs were incubated in poly(A) addition extracts for 30 min and then analyzed by electrophoresis on a 5% acrylamide gel containing 7 M urea. Bands representing polyadenylated product are designated A^+. (D) Accumulation of 5′ cleavage products. RNAs were incubated as described above, but the α,β-methylene analog of ATP [AMP(CH_2)PP] and EDTA were added instead of ATP and phosphocreatine. Products were analyzed by electrophoresis on a 10% acrylamide gel containing 7 M urea. nt, Nucleotides. (E) Sequence of the SVL downstream domain. The underlined segment was defined by deletion analysis to include at least part of the site of C protein interaction. Solid circles designate U-to-G mutations in SVL-16. The overlined sequences mark the region (+10 to +26) required for efficient cleavage at the poly(A) addition site within injected frog oocytes (8) and the sequence (+47 to +55) required for efficient polyadenylation in transfected COS cells (27).

motif recognized directly by C proteins. Alternatively, the interaction of C proteins with polyadenylation substrates might be stabilized by an additional, unidentified factor or factors that also interact with the RNA. If additional factors are involved, they might provide the preference for binding downstream of poly(A) cleavage sites. Thus, we cannot yet conclude that C proteins themselves exhibit the sequence specificity.

Whether or not additional factors are involved, the RNA sequence must contain a recognition signal that mediates an efficient C protein interaction. Perhaps, at least in some cases, the recognition sequence is as simple as an accessible stretch of U residues. Swanson and Dreyfuss (31) have recently shown that C proteins display high binding affinities for poly(U) homopolymers. Consistent with their observation, the SVL downstream domain contains a stretch of six uridylate residues, and their disruption by a single U to G transversion inhibited transfer of ^{32}P label to C protein (Fig. 5). Furthermore, only SVL RNA with ^{32}P-labeled U residues transferred label at high efficiency. ^{32}P label transfer to C proteins with RNAs containing radioactive G, C, or A residues was much less efficient (data not shown).

Recent work has identified a large 40S to 50S complex which specifically associates with polyadenylation substrate RNAs (18, 23, 29, 37, 38). It is likely that the 64-kDa protein described previously (35) as well as the hnRNP C proteins described in this report are components of this complex. Additional cross-linked proteins, although they were not detected with all of the polyadenylation substrates tested, also require downstream sequence elements for their interaction with RNA (Fig. 3 and 5). These may reflect generic components of the 40S to 50S complex which vary in their ability to cross-link due to RNA topology, or they may represent interactions specific to individual polyadenylation signals.

The region downstream of poly(A) addition sites has been shown in a number of instances to play a role in the efficiency of polyadenylation both in vivo and in vitro (7, 8, 13, 15, 26, 28, 38). Is the interaction with C proteins responsible for this effect? As yet, our data are inconclusive on this point. The 26-nucleotide interval to which the C protein interaction has been mapped on SVL RNA includes a 9-nucleotide stretch shown by Sadofsky et al. (27) to be required for efficient polyadenylation of the SVL transcript within transfected COS cells. In fact, the 9-nucleotide segment identified in the transfection assay includes the run of six uridylate residues which plays a role in the transfer of label from SVL RNA to C proteins (Fig. 5). However, the

inability of an RNA to efficiently cross-link to C proteins does not always correlate with inefficient polyadenylation. Two derivatives of SVL, SVL-14 and SVL-16 (Fig. 5), failed to efficiently cross-link to C proteins. SVL-14 RNA was inefficiently cleaved at the poly(A) addition site, while SVL-16 was cleaved and polyadenylated normally. The behavior of SVL-14 is consistent with the hypothesis that C proteins play a role in polyadenylation. However, SVL-14 RNA failed to transfer ^{32}P label to several proteins in addition to the C proteins (Fig. 5), so it is possible that one of these additional polypeptides mediates the downstream efficiency function either alone or in association with the C proteins. The normal polyadenylation of SVL-16 RNA suggests that C proteins play no direct role in the efficiency of the reaction. However, we cannot yet rule out the possibility that the C proteins can associate with SVL-16 RNA in such a way that they are no longer cross-linked but still provide the efficiency function. In sum, additional experiments are required to unambiguously establish or rule out a role for C proteins in polyadenylation.

We have found (unpublished results) that C proteins can be efficiently cross-linked to an intron sequence located between the first two exons of the adenovirus tripartite leader sequence (36) and that the SV40 late exon does not contain such a preferential cross-linking site. This small sampling suggests that preferential interaction sites for C proteins may coincide with domains removed from transcripts, i.e., introns and sequences downstream of poly(A) addition sites within the nucleus. Since C proteins are removed from mRNAs prior to their transport from nucleus to cytoplasm, it is tempting to speculate that C protein-binding sites mark RNA segments for retention in the nucleus.

ACKNOWLEDGMENTS

We thank Gideon Dreyfuss for the generous gift of 4F4 antibody.

This work was supported by Public Health Service grant CA-38965 from the National Institutes of Health. J. Wilusz is an American Cancer Society Postdoctoral Fellow (PF 2706), and T. Shenk is an American Cancer Society Research Professor.

LITERATURE CITED

1. **Berget, S. M.** 1984. Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature (London) **309**:179–182.
2. **Beyer, A. L., M. E. Christensen, B. W. Walker, and W. M. LeStourgeon.** 1977. Identification and characterization of the packaging proteins of core 40S hnRNP particles. Cell **11**:127–138.
3. **Birnstiel, M. L., M. Bussinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41**: 349–359.
4. **Choi, Y. D., and G. Dreyfuss.** 1984. Isolation of the heterogeneous nuclear RNA-ribonucleoprotein complex (hnRNP): a unique supramolecular assembly. Proc. Natl. Acad. Sci. USA **81**:7471–7475.
5. **Choi, Y. D., and G. Dreyfuss.** 1984. Monoclonal antibody characterization of the C proteins of heterogeneous nuclear ribonucleoprotein complexes in vertebrate cells. J. Cell. Biol. **99**:1997–2004.
6. **Choi, Y. D., P. J. Grabowski, P. A. Sharp, and G. Dreyfuss.** 1986. Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. Science **231**:1534–1539.
7. **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.
8. **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.
9. **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.
10. **Dreyfuss, G.** 1986. Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. Annu. Rev. Cell Biol. **2**: 459–498.
11. **Dreyfuss, G., Y. D. Choi, and S. A. Adam.** 1984. Characterization of hnRNA-protein complexes in vivo with monoclonal antibodies. Mol. Cell. Biol. **4**:1104–1114.
12. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell **24**:251–260.
13. **Gil, A., and N. J. Proudfoot.** 1984. A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3′-end formation. Nature (London) **312**:473–474.
14. **Hart, R. P., M. A. McDevitt, and J. R. Nevins.** 1985. Definition of essential sequences and functional equivalence of elements downstream of the adenovirus E2A and the early simian virus 40 polyadenylation sites. Mol. Cell. Biol. **5**:2975–2983.
15. **Hart, R. P., M. A. McDevitt, and J. R. Nevins.** 1985. Poly(A) site cleavage in a HeLa nuclear extract is dependent on downstream sequences. Cell **43**:677–683.
16. **Hashimoto, C., and J. A. Steitz.** 1986. A small nuclear ribonucleoprotein associates with the AAUAAA polyadenylation signal in vitro. Cell **45**:581–591.
17. **Higgs, D. R., S. E. Y. Goodbourn, J. Lamb, J. B. Clegg, D. J. Weatherall, and N. J. Proudfoot.** 1983. α-Thalassemia caused by a polyadenylation signal mutation. Nature (London) **306**:398–400.
18. **Humphrey, T., G. Christofori, V. Lucijanic, and W. Keller.** 1987. Cleavage and polyadenylation of mRNA precursors *in vitro* occurs within large and specific 3′ processing complexes. EMBO J. **6**:4159–4168.
19. **McLauchlin, 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.
20. **Montell, C., E. F. Fisher, M. H. Caruthers, and A. J. Berk.** 1983. Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3′ consensus sequence AAUAAA. Nature (London) **305**:600–605.
21. **Moore, C. L., and P. A. Sharp.** 1984. Site-specific polyadenylation in a cell free reaction. Cell **36**:581–591.
22. **Moore, C. L., and P. A. Sharp.** 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell **41**:845–855.
23. **Moore, C. L., H. Skolnick-David, and P. A. Sharp.** 1988. Sedimentation analysis of polyadenylation-specific complexes. Mol. Cell. Biol. **8**:226–233.
24. **Nevins, J. R., and J. E. Darnell.** 1978. Steps in the processing of Ad2 mRNA: poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell **15**:1477–1493.
25. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ Noncoding region sequences in eukaryotic messenger RNA. Nature (London) **263**: 211–214.
26. **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.
27. **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.
28. **Simonsen, C. C., and A. D. Levinson.** 1983. An analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. Mol. Cell. Biol. **3**:2250–2258.
29. **Skolnik-David, H., C. L. Moore, and P. A. Sharp.** 1987. Electrophoretic separation of polyadenylation-specific complexes. Genes Dev. **1**:672–682.
30. **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, New York.

31. **Swanson, M. S., and G. Dreyfuss.** 1988. Classification and purification of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. Mol. Cell. Biol. **8**:2237–2241.
32. **Swanson, M. S., T. Y. Nakagawa, K. LeVan, and G. Dreyfuss.** 1987. Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. Mol. Cell. Biol. **7**:1731–1739.
33. **Takagaki, Y., L. C. Ryner, and J. L. Manley.** 1988. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell **52**:731–742.
34. **Wickens, M., and P. Stephenson.** 1984. The role of the conserved AAUAAA sequence in mRNA maturation: four AAUAAA point mutants prevent 3′ end formation. Science **226**: 1045–1051.
35. **Wilusz, J., and T. Shenk.** 1988. A 64kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. Cell **52**:221–228.
36. **Zain, S., J. Sambrook, R. J. Roberts, W. Keller, M. Fried, and A. R. Dunn.** 1979. Nucleotide sequence analyses of the leader segments in a cloned copy of adenovirus 2 fiber mRNA. Cell **16**: 851–861.
37. **Zarkower, D., and M. Wickens.** 1987. Formation of mRNA 3′ termini: stability and dissociation of a complex involving the AAUAAA sequence. EMBO J. **6**:177–186.
38. **Zhang, F., and C. N. Cole.** 1987. Identification of a complex associated with processing and polyadenylation in vitro of herpes simplex virus type 1 thymidine kinase precursor RNA. Mol. Cell. Biol. **7**:3277–3286.