A Multicomponent Complex Is Required for the AAUAAA-Dependent Cross-Linking of a 64-Kilodalton Protein to Polyadenylation Substrates

JEFFREY WILUSZ,^{1,*} THOMAS SHENK,¹ YOSHIO TAKAGAKI,² AND JAMES L. MANLEY²

Howard Hughes Medical Institute, Department of Biology, Princeton University, Princeton, New Jersey 08544,¹ and Department of Biological Sciences, Columbia University, New York, New York 10027²

Received 13 September 1989/Accepted 16 November 1989

A 64-kilodalton (kDa) polypeptide that is cross-linked by UV light specifically to polyadenylation substrate RNAs containing a functional AAUAAA element has been identified previously. Fractionated HeLa nuclear components that can be combined to regenerate efficient and accurate polyadenylation in vitro have now been screened for the presence of the 64-kDa protein. None of the individual components contained an activity which could generate the 64-kDa species upon UV cross-linking in the presence of substrate RNA. It was necessary to mix two components, cleavage stimulation factor and specificity factor, to reconstitute 64-kDa protein-RNA cross-linking. The addition of cleavage factors to this mixture very efficiently reconstituted the AAUAAA-specific 64-kDa protein-RNA interaction. The 64-kDa protein, therefore, is present in highly purified, reconstituted polyadenylation reactions. However, it is necessary to form a multicomponent complex to efficiently cross-link the protein to a substrate RNA.

The maturation of the 3' ends of most RNA polymerase II transcripts involves an endonucleolytic cleavage event followed by the polymerization of 150 to 200 adenylate residues onto the newly formed terminus (reviewed in references 2 and 16). This cleavage-polyadenylation event is mediated through two *cis*-acting sequence elements: a highly conserved hexanucleotide, AAUAAA, located 5 to 30 bases upstream of the cleavage site (8, 12, 22, 27, 35, 37), and a less conserved U- or GU-rich stretch located downstream of the cleavage site (6, 7, 10, 17, 19, 20, 28, 30, 31). The selective recognition of alternative polyadenylation sites may play a role in the posttranscriptional regulation of gene expression (reviewed in reference 14).

Progress in understanding the mechanism of polyadenylation has been facilitated greatly by the development of accurate and faithful in vitro systems (15, 24). The recognition of the polyadenylation signal occurs through the formation of a large, multicomponent complex. Native gel and density gradient analyses have shown that a 40 to 50S complex forms in an AAUAAA-dependent manner on substrate RNAs (13, 25, 32, 33, 39, 40). Immunological and biochemical studies suggest that small nuclear RNAs play some role in this complex (1, 11, 13, 24-26). Label transfer experiments with UV light have identified proteins of 64 to 68 and 155 kilodaltons (kDa) which interact with RNAs that contain a polyadenylation signal (23, 38). Additionally, biochemical fractionation of the nuclear extract has identified several independent components which are required for reconstitution of in vitro processing (3, 4, 9, 18, 21, 34, 34a).

We have now assayed partially purified polyadenylation components (34, 34a) for the presence of the 64-kDa protein by UV cross-linking. None of the individual fractions contained an activity that could be cross-linked to RNA to generate the 64-kDa species. However, a mixture of at least

two components which have been shown previously to play roles in the specificity and efficiency of the cleavage-polyadenylation reaction (34, 34a) was required to regenerate the 64-kDa cross-linked species. These results support a functional role for the 64-kDa protein in polyadenylation and suggest that the 64-kDa protein cannot bind independently to substrate RNAs in a manner in which it can be cross-linked by UV light.

A flow chart of the biochemical purification steps performed is shown in Fig. 1. Details of the purification steps as well as the functional assays performed have been described previously (34, 34a). Briefly, the nuclear extract used in the in vitro polyadenylation system was fractionated into two components by gel filtration chromatography following selective ammonium sulfate precipitation. This step separated poly(A) polymerase components from those factors required for cleavage and specificity. Cleavage specificity factor (CSF) is a fraction required to confer AAUAAA dependence on the polymerase and also contains specific cleavage activity (34). CSF was further separated into three individual fractions by Mono Q fast protein liquid chromatography. Cleavage stimulation factor (CstF), an activity present in flowthrough fractions which stimulates endonucleolytic cleavage severalfold, was further purified on a heparin-agarose column. Cleavage factor (CF), which is required for endonucleolytic cleavage, eluted from the Mono Q column at low salt and was further fractionated into two activities (CFI and CFII) by Mono S fast protein liquid chromatography. The specificity factor (SF), which is required for both cleavage and AAUAAA-dependent polyadenylation, eluted at high salt and was further purified by hydrophobic interaction fast protein liquid chromatography. These four activities, along with highly purified poly(A) polymerase (29), will reconstitute accurate and efficient cleavage and polyadenylation (34a).

UV cross-linking and label transfer analysis was done with RNAs ³²P labeled at uridylate residues and with the chromatography fractions described above. Briefly, RNAs were incubated with column fractions for 5 min followed by

* Corresponding author.

† Present address: Department of Microbiology and Molecular Genetics, UMDNJ-New Jersey Medical School, Newark, NJ 07103-2757.

FIG. 1. Flow diagram of the biochemical fractionation performed. Details of the chromatographic procedures as well as activity assays can be found elsewhere (34, 34a).

irradiation with a germicidal lamp (38). Proteins cross-linked to RNAs were identified by sodium dodecyl sulfate-gel electrophoresis after RNase treatment and boiling of the irradiated samples in loading buffer.

Several proteins (Fig. 2) can be cross-linked to the simian virus 40 late polyadenylation signal containing RNA (SVL) by using CSF fractions from the Superose 6 chromatography step outlined in Fig. 1. Numerous proteins were also cross-linked to an SVL derivative which contains a U-to-G transversion in the conserved hexanucleotide motif of the polyadenylation signal (Fig. 2B). By comparing the SVL wild-type (AAUAAA) and mutant (AAGAAA) label transfer patterns, two proteins were cross-linked in an AAUAAA-dependent manner: a 64-kDa protein present in fractions 3 to 5 as well as a 150-kDa protein present in fractions 3 to 4. A nonspecific protein of about 64 kDa which was responsible for the residual cross-linked product when the AAGAAA mutant was employed in the assay was also cross-linked to the SVL RNAs in these fractions (Fig. 2B) (see below). Partial protease analysis of this cross-linked peptide gives a different pattern from the AAUAAA-specific 64-kDa protein (data not shown). Similar cross-linking results were obtained with either the L3 (Fig. 3A) or the simian virus 40 early polyadenylation signal (data not shown). Analysis of partial V8 protease digestion products of the cross-linked 64-kDa protein obtained from fraction 4 of the Superose 6 column or crude nuclear extracts gave similar patterns (Fig. 3B). This finding indicates that the protein identified in the column fractions is identical to the 64-kDa protein previously described (38).

From these data and activity assays performed, we conclude that the 64-kDa protein copurified on the Superose 6 column with the CSF activity. The heterogeneous nuclear ribonucleoprotein C proteins shown previously to be cross-linked to polyadenylation substrates (36) were found throughout the column, making it impossible to assess their contribution to any specific fractionated activity. The 150-

FIG. 2. The 64-kDa protein copurifies with CSF. UV cross-linking and label transfer analysis was done with the SVL and Superose 6 column fractions described in Fig. 1. RNA (10 fmol) labeled with [32 P]UTP to the same specific activity was incubated with the indicated fractions followed by irradiation with UV light. Samples were treated with RNase A, boiled in sample loading buffer, and separated by electrophoresis in a sodium dodecyl sulfate-10% acrylamide gel. Positions of the 64- and 150-kDa (kd) proteins are indicated. (A) Assay performed with the SVL poly(A) signal (SVL 3 [38]). (B) Assay performed with an SVL derivative containing an AAUAAA-to-AAGAAA point mutation (SVL 6 [38]). Positions of the CSF (lanes 3 to 5) and poly(A) polymerase (lanes 8 and 9) activities as determined in independent assays are noted.

kDa protein detected in the CSF-specific fractions may be the same as the 155-kDa protein described by Moore et al. (23). These authors reported, however, that an SVL RNA was not cross-linked to this polypeptide. In addition, the 150-kDa protein was not consistently observed in every Superose 6 column run (data not shown), nor did a cross-linked species appear in crude or ammonium sulfate-fractionated extracts (Fig. 2 to 4). The significance of the 150-kDa protein to polyadenylation is, therefore, difficult to assess.

Figure 4A and B shows the results obtained by UV cross-linking to the SVL transcript with the more-purified

FIG. 3. Characterization of the 64-kDa (kd) protein from the Superose 6 column fractions. (A) UV cross-linking and label transfer analysis of Superose 6 column fractions with an L3 polyadenylation signal containing RNA. Assay was performed as described in the legend to Fig. 2. (B) Analysis of partial digestion products of 32 P-labeled cross-linked 64-kDa proteins with V8 protease. UV cross-linking was performed as described above, the 64-kDa bands were excised, and proteins were digested with the indicated amounts of V8 protease (5) before electrophoresis in a sodium dodecyl sulfate-12.5% acrylamide gel. Headings over lanes indicate nanograms of V8 protease added to each sample.

Mono Q column fractions. The flowthrough (CstF), low-salt (CF), and high-salt (SF) fractions of this column, which contain activities required for efficient and specific pre-mRNA 3' cleavage, failed to cross-link to the 64-kDa protein when assayed individually (Fig. 4A). The 64-kDa protein was also not detected when any of the other fractions of the column were used in the assay (data not shown). This result suggested that multiple fractions are required to reconstitute the cross-linking of the 64-kDa protein to RNA. Therefore, various fractions were combined and assayed for their ability to reconstitute 64-kDa protein-RNA cross-linking. The cross-linking of an AAUAAA-specific 64-kDa protein was reconstituted to a modest extent when the flowthrough (CstF) and high-salt (SF) fractions were combined (Fig. 4B). When the CstF, SF, and low-salt (CF) fractions were mixed, however, the cross-linking of the 64-kDa protein was very efficiently reconstituted.

We conclude that the 64-kDa protein was present in either or both of the CstF and SF activities. The addition of

FIG. 4. Multiple factors are required for efficient 64-kDa cross-linking. UV cross-linking analysis was done with Mono Q column fractions assayed individually (A) or in various combinations (B). (C) UV cross-linking analysis with more-purified column fractions (Fig. 1). Assays were performed with an SVL wild-type RNA (SVL 3) or an AAGAAA mutant RNA (SVL 6) as described in the legend to Fig. 2 and analyzed on sodium dodecyl sulfate-10% acrylamide gels. Fractions used were prepared as described in Fig. 1. AS, Ammonium sulfate; kd, kilodaltons.

cleavage factor to these fractions resulted in a more prominent cross-linked 64-kDa species, indicating that the presence of cleavage factor may lead to a more stable complex or to one in which the 64-kDa protein is more optimally oriented for cross-linking.

Cross-linking of the 64-kDa polypeptide was also seen with a reconstituted mixture of more purified CstF, CFI and CFII, and SF (Fig. 4C). The more purified fractions again failed to cross-link to the 64-kDa protein when assayed individually (data not shown). It is interesting that the heparin-agarose-purified CstF activity contained a 66-kDa protein that cross-linked to RNA in an AAUAAA-independent manner (data not shown). This activity could be the same nonspecific cross-linking activity noted above in the CSF fraction of the Superose 6 column.

The data presented here argue strongly for the functional significance of the 64-kDa protein which is cross-linked to polyadenylation substrate RNAs in an AAUAAA-dependent fashion. The assembly of a multicomponent complex of cleavage-polyadenylation factors on the substrate RNA is a prerequisite for the formation of a 64-kDa protein-RNA interaction that can be cross-linked with UV light. Studies with native gels also support this notion (37; data not shown). The addition of highly purified poly(A) polymerase

had no effect on the cross-linking of the 64-kDa protein to RNA substrates, nor did purified poly(A) polymerase independently cross-link to RNA (data not shown). Collectively, these observations indicate that the UV cross-linking approach interfaces well with biochemical fractionation to dissect components of the polyadenylation machinery.

While the combination of partially purified SF and CstF was minimally required for the cross-linking of the 64-kDa polypeptide, the addition of purified cleavage factors was necessary for optimal label transfer (Fig. 4). It is therefore unclear whether the 64-kDa protein is part of the stimulation or specificity activity, or both. Recent studies have shown that the most purified CstF (native molecular weight of 200,000 [34a]) contains only three major proteins, one of which is about 64 kDa (data not shown). Since UV cross-linking analysis adds approximately 1 to 2 kDa to the apparent molecular size of the protein species analyzed, this protein most likely corresponds to the nonspecific 64- to 66-kDa cross-linked species discussed above. UV cross-linking analysis of highly purified CstF fractions does indeed give a nonspecific 66-kDa species (data not shown). The observed difference in gel mobility of this species and the AAUAAA-dependent 64-kDa cross-linked polypeptide makes it unlikely that the 64- to 66-kDa cross-linked protein is a component of CstF that binds nonspecifically to RNA in the absence of SF but specifically to AAUAAA-containing RNAs in its presence.

We favor the view that the 64-kDa protein is a component of SF (native molecular weight of 290,000 [34a]). This is perhaps the most logical location for the protein, because SF appears to recognize the AAUAAA element during both cleavage and polyadenylation. Our model thus suggests that the function of CstF is to induce a more stable or optimally aligned interaction between SF and the pre-mRNA, thereby enhancing the efficiency of the cleavage reaction and the cross-linking of the 64-kDa polypeptide. We are currently attempting to produce antibodies against the cross-linked 64-kDa protein to identify the fraction that contains the polypeptide and to serve as reagents for the further characterization of this polypeptide and its role in site-specific polyadenylation.

This work was supported by Public Health Service grants GM-28983 to J.L.M. and CA-38965 to T.S. from the National Institutes of Health. J.W. received postdoctoral support from the American Cancer Society and Howard Hughes Medical Institute. T.S. is an American Cancer Society professor.

LITERATURE CITED

- Berget, S. M., and B. L. Robberson. 1986. U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for *in vitro* splicing but not polyadenylation. *Cell* **46**:691-696.
- Birnstiel, M. L., M. Busslinger, and K. Strub. 1985. Transcription termination and 3' processing: the end is in site! *Cell* **41**:349-359.
- Christofori, G., and W. Keller. 1988. 3' Cleavage and polyadenylation of mRNA precursors *in vitro* requires a poly(A) polymerase, a cleavage factor, and a snRNP. *Cell* **54**:875-889.
- Christofori, G., and W. Keller. 1989. Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA *in vitro*. *Mol. Cell. Biol.* **9**:193-203.
- Cleveland, D. W., S. G. Fischer, M. W. Kirschner, and U. K. Laemmli. 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* **252**:1102-1106.
- 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.
- 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.
- 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.
- Gilmartin, G. M., M. A. McDevitt, and J. R. Nevins. 1988. Multiple factors are required for specific RNA cleavage at a poly(A) addition site. *Genes Dev.* **2**:578-587.
- 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.
- Hashimoto, C., and J. A. Steitz. 1986. A small nuclear ribonucleoprotein associates with the AAUAAA polyadenylation signal *in vitro*. *Cell* **45**:581-591.
- 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.
- Humphrey, T., G. Christofori, V. Lucijanic, and W. Keller. 1987. Cleavage and polyadenylation of messenger RNA precursors *in vitro* occurs within large and specific 3' processing complexes. *EMBO J.* **6**:4159-4168.
- Leff, S. E., M. G. Rosenfeld, and R. M. Evans. 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annu. Rev. Biochem.* **55**:1091-1117.
- Manley, J. L. 1983. Accurate and specific polyadenylation of mRNA precursors in a soluble whole-cell lysate. *Cell* **33**:595-605.
- Manley, J. L. 1988. Polyadenylation of mRNA precursors. *Biochim. Biophys. Acta* **950**:1-12.
- Mason, P. J., J. A. Elkington, M. M. Lloyd, M. B. Jones, and J. G. Williams. 1986. Mutations downstream of the polyadenylation site of a *Xenopus* β -globin mRNA affect the position but not the efficiency of 3' processing. *Cell* **46**:263-270.
- McDevitt, M. A., G. M. Gilmartin, W. H. Reeves, and J. R. Nevins. 1988. Multiple factors are required for poly(A) addition to a mRNA 3' end. *Genes Dev.* **2**:588-597.
- McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins. 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. *Cell* **37**:993-999.
- McLauchlin, J., D. Gaffney, J. L. Whitton, and J. B. Clements. 1985. The consensus sequence YGTGTTY located downstream from the AAUAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res.* **13**:1347-1367.
- McLauchlin, J., C. L. Moore, S. Simpson, and J. B. Clements. 1988. Components required for *in vitro* cleavage and polyadenylation of eukaryotic mRNA. *Nucleic Acids Res.* **16**:5323-5344.
- 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 the mRNA 3' consensus sequence AAUAAA. *Nature (London)* **305**:600-605.
- Moore, C. L., J. Chen, and J. Whoriskey. 1988. Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding. *EMBO J.* **7**:3159-3169.
- Moore, C. L., and P. A. Sharp. 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. *Cell* **41**:845-855.
- Moore, C. L., H. Skolnik-David, and P. A. Sharp. 1988. Sedimentation analysis of polyadenylation-specific complexes. *Mol. Cell. Biol.* **8**:226-233.
- Mowry, K. L., and J. A. Steitz. 1988. snRNP mediators of 3' end processing: functional fossils? *Trends Biochem. Sci.* **13**:447-451.
- Proudfoot, N. J., and G. G. Brownlee. 1976. 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature (London)* **263**:211-214.
- Ryner, L. C., Y. Takagaki, and J. L. Manley. 1989. Sequences

- downstream of AAUAAA signals affect pre-RNA cleavage and polyadenylation in vitro both directly and indirectly. *Mol. Cell. Biol.* **9**:1759–1771.
29. Ryner, L. C., Y. Takagaki, and J. L. Manley. 1989. Multiple forms of poly(A) polymerases purified from HeLa cells function in specific mRNA 3'-end formation. *Mol. Cell. Biol.* **9**:4229–4238.
30. 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.
31. 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.
32. Skolnik-David, H., C. L. Moore, and P. A. Sharp. 1987. Electrophoretic separation of polyadenylation-specific complexes. *Genes Dev.* **1**:672–682.
33. Stefano, J. E., and D. E. Adams. 1988. Assembly of a polyadenylation-specific 25S ribonucleoprotein complex in vitro. *Mol. Cell. Biol.* **8**:2052–2062.
34. 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.
- 34a. Takagaki, Y., L. C. Ryner, and J. L. Manley. 1989. Four factors are required for 3' end cleavage of pre-mRNAs. *Genes Dev.* **3**:1711–1724.
35. 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.
36. Wilusz, J., D. I. Feig, and T. Shenk. 1988. The C proteins of heterogeneous nuclear ribonucleoprotein complexes interact with RNA sequences downstream of polyadenylation cleavage sites. *Mol. Cell. Biol.* **8**:4477–4483.
37. Wilusz, J., S. M. Pettine, and T. Shenk. 1989. Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. *Nucleic Acids Res.* **17**:3899–3908.
38. Wilusz, J., and T. Shenk. 1988. A 64kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. *Cell* **52**:221–228.
39. Zarkower, D., and M. Wickens. 1987. Specific pre-cleavage and post-cleavage complexes involved in the formation of SV40 late mRNA 3' termini *in vitro*. *EMBO J.* **6**:4185–4192.
40. Zhang, F., and C. N. Cole. 1987. Identification of complex associated with processing and polyadenylation in vitro of herpes simplex virus type 1 thymidine kinase precursor RNA. *Mol. Cell. Biol.* **7**:3277–3286.