Volume 17 Number 10 1989 Nucleic Acids Research

Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal

Jeffrey Wilusz, Stefan M.Pettine and Thomas Shenk

Princeton University, Department of Biology, Princeton, NJ 08544, USA

Received December 20, 1988; Revised and Accepted March 27, 1989

ABSTRACT

We have constructed 14 independent point mutations in the conserved AAUAAA element of the SV40 late polyadenylation signal in order to study the recognition and function of alternative polyadenylation signals. A variant RNA containing an AUUAAA was polyadenylated at 20% the level of wild-type substrate RNA, while all other derivatives tested were not functional *in vitro*. The AUUAAA variant RNA formed specific complexes in native polyacrylamide gels and crosslinked to the AAUAAA-specific 64kd polypeptide, but at a lower efficiency than wild-type substrate RNA.

INTRODUCTION

The maturation of the 3' end of most messenger RNAs of higher eukaryotes involves a site-specific endonucleolytic cleavage event followed by the addition of 150 to 200 adenylate residues (reviewed in 1). This polyadenylation event requires at least two signals contained in the precursor RNA. A highly conserved hexanucleotide, AAUAAA, is located 5 to 30 bases upstream of the processing site (2). Mutational analyses have shown this element to be required for polyadenylation both *in vivo* and *in vitro* (3-7). A second element, located downstream of the cleavage site, appears to be of limited sequence complexity, consisting of GU, U, and/or G rich sequences and may be reiterated (8-14). The downstream element is not absolutely required for polyadenylation, but plays a role in the efficiency of the processing event (15-20).

A variety of experimental evidence suggests that the elements of the polyadenylation signal are recognized by a multicomponent complex. Native gel and gradient analyses indicate a large, 40 - 50S complex forms on functional polyadenylation substrates (21-26). These complexes require an intact AAUAAA and downstream element for efficient formation. Polyadenylation *in vitro* is inhibited by U1-RNP, Sm, La, and tri-methyl cap specific antibodies (27, 28). In addition, monoclonal Sm and tri-methyl cap antibodies can precipitate RNase T1 fragments containing elements of the polyadenylation signal (21, 25, 29). Biochemical fractionation studies reveal the complex nature of the polyadenylation machinery, as several fractions

are required for the efficient reconstitution of polyadenylation/cleavage activity *in vitro* (30-33). Finally, UV crosslinking analyses have identified two factors, a 64kd protein which requires an intact AAUAAA for efficient crosslinking, and an association of the downstream element with the hnRNP C protein (34, 35). Protein factors, snRNPs, and hnRNP particles may, therefore, all play a role in the formation of the polyadenylation signal recognition complex.

A survey of the Genbank data base reveals that naturally occurring AAUAAA variants comprise approximately 10% of the putative mRNA polyadenylation signals. These variations are primarily single base differences, the most common of which is AUUAAA. Several putative signals, such as the HTLV 1 LTR (36) and DHFR genes (37), bear little resemblance to the conserved hexanucleotide element. This apparent degeneracy of polyadenylation signals may reflect signal specific factors which allow recognition of AAUAAA variants, a general flexibility in the components of the polyadenylation signal recognition complex, or alternative pathways for 3' end maturation. One such alternative pathway has been extensively described for the non- polyadenylated histone mRNAs (1).

In order to better understand the mechanism by which variant polyadenylation signals are used, we have constructed fourteen independent point mutants in the AAUAAA element of the polyadenylation signal of the SV40 late transcription unit. This well chararacterized signal was chosen to standardize the contribution from downstream sequences to polyadenylation efficiency, thus allowing a meaningful comparison of the variations generated. Results indicate that aside from the wild-type hexanucleotide, only the AUUAAA variant was functional and able to form polyadenylation-specific complexes *in vitro*.

MATERIALS AND METHODS

Generation of AAUAAA Variants.

The parent construct contained the 241 base pair *BamH*I-*Bcl* I fragment of SV40 inserted into the *BamH*I site of pSP65 (Promega Biotech, Inc.). pSVL3 (34), which contains the SV40 specific *Alu* I- *Hind* III fragment of pSVL inserted into the *Hinc* II and *Hind* III sites of pGem4, was cut at its *Sal* I and *Hind* III sites (located downstream from the insert) and re-circularized to make pSVL3-S/H. Synthetic oligonucleotides containing a random mixture of the three variant nucleotides at a single position in the AAUAAA element were inserted between the *Bsm* I and *Hind* III sites of pSVL3-S/H. Plasmids containing the desired mutations were determined by chemical DNA sequencing (38). Finally, a *Pst* I-*Eco* RI fragment of pSVE2, which contains the *BamH*I to *Alu* I fragment of pSVL at the *Bam* Hl -

Hinc II sites of pGEM4, was inserted into pSVL3-S/H and its derivatives at the *Eco RI* and *Sma* I sites upstream of the insert.

In Vitro Transcription.

DNA templates were linearized with *Dra* I and transcribed *in vitro* by using bacteriophage SP6 RNA polymerase in the presence of [α-^{32}P] UTP and ^{7m}GpppG (39). RNAs were purified on 5% acrylamide gels containing 7M urea prior to use.

In Vitro Polyadenylation.

Nuclear extracts were prepared from HeLa spinner cells grown in 10% horse serum as described (40). Polyadenylation reaction mixes contained final concentrations of 3% polyvinyl alcohol, 1 mM ATP, 20 mM phosphocreatine, 12 mM HEPES, pH 7.9, 12% glycerol, 60 mM KCl, 0.12 mM EDTA, 0.3 mM dithiotreitol, and 60% (vol/vol) nuclear extract. Reactions were performed at 30°C.

Native Gel Electrophoresis.

RNAs were incubated at 30°C for 5 min in the *in vitro* polyadenylation system. Reactions were then adjusted to 4 ug/ul heparin (Sigma Chemical Co.) and incubated on ice for 5 min. Loading buffer (50% glycerol, 0.05% xylene cyanol) (1 ul) was added and samples were electrophoresed in a 4% bis-acrylamide:acrylamide (1:80) gel as described (41).

UV Crosslinking.

[^{32}P]labeled, gel purified RNAs (10 fmole) of the same specific activity were incubated in the *in vitro* polyadenylation system for 10 min. *Escherichia coli* tRNA (5ug) was added, and the sample was irradiated for 10 min at 4°C with a G15T8 light (Sylvania) placed 4 cm from the sample. Ribonuclease A was added to a final concentration of 1mg/ml, and samples were incubated at 37°C for 15 min. An equal volume of protein gel loading buffer containing SDS and β-mercaptoethanol was added, and samples were heated to 100°C for 5 min prior to electrophoresis in polyacrylamide gels containing 0.1% SDS.

RESULTS

In Vitro Polyadenylation of Variant RNAs.

The identification of AAUAAA variants indicates a limited degree of flexibility in the recognition of signals required for 3' end maturation of RNA polymerase II synthesized transcripts. We have analyzed the flexibility of the AAUAAA motif in the *in vitro* polyadenylation system (28,29) by creating point mutants of the SV40 late polyadenylation signal (SVL). These mutants were constructed through the use of synthetic oligonucleotides containing random combinations of the three variant nucleotides at the desired positions as described in Materials and Methods. This approach

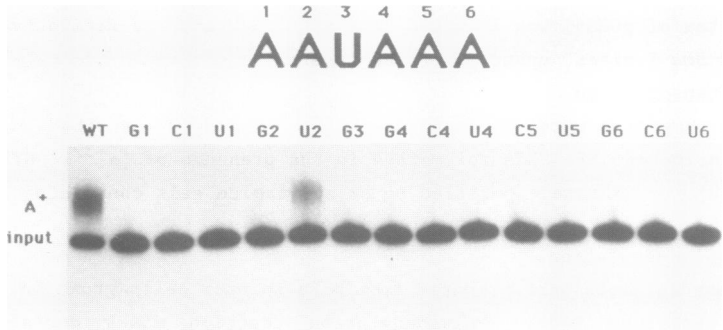

Figure 1: Effect of point mutations in the AAUAAA of the SV40 late polyadenylation signal on 3' end processing *in vitro*. RNAs were incubated in the *in vitro* polyadenylation system for 1 hr and then analyzed by electrophoresis on a 5% acrylamide gel containing 7M urea. The AAUAAA variants are identified by the numbering system described in the heading.

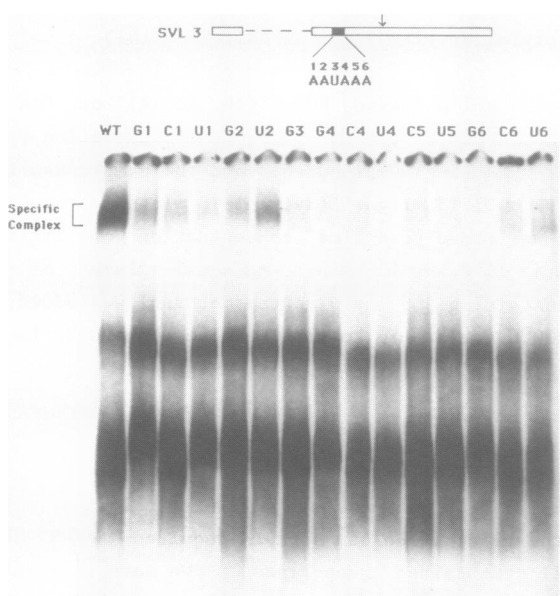

Figure 2: Effect of point mutations in the AAUAAA on complex formation. RNAs were incubated in the *in vitro* polyadenylation system for 5 min, heparin was added to remove non-specific RNA -protein interactions, and complexes were analyzed on 4% native acrylamide gels. Free RNA was run off the bottom of the gel to maximize the resolution of large complexes.

allowed a rigorous assessment of AAUAAA variants in the context of a single polyadenylation signal, as the contribution from downstream sequence elements or other bases remained constant throughout the study. Fourteen of the 18 possible mutants were obtained and characterized.

RNAs were transcribed *in vitro* using SP6 RNA polymerase in the presence of [α^{32}P UTP] and purified on denaturing acrylamide gels prior to use. Equivalent amounts of wild type and variant RNAs were incubated in the *in vitro* polyadenylation system described previously (28) and processed products were analyzed by gel electrophoresis. As seen in figure 1, the wild-type SVL RNA was efficiently cleaved and polyadenylated at the appropriate *in vivo* site in the cell free system. Of the 14 variants tested, only the AUUAAA signal was polyadenylated at a detectable level. Based on samples taken when poly $(A)^+$ product was accumulating in a linear fashion, SVL derivatives containing the AUUAAA variation were processed at about 20% of the wild type efficiency. This was determined by comparing the levels of poly $(A)^+$ product RNA to substrate RNA in six independent experiments by excising the bands from the gel and liquid scintillation counting. This observation correlates well with the fact that the AUUAAA variant is the most common alternative hexanucleotide found in mRNA polyadenylation signals. Similar results were obtained when the assay was performed in the presence of α-β methylene adenosine 5'-triphosphate. This ATP analog uncouples the cleavage reaction from poly (A) addition (28). The efficiency of cleavage of the AUUAAA variant RNA was reduced compared to wild-type and the other variant substrates were not processed (data not shown).

Formation of the Polyadenylation Signal Recognition Complex by Variant RNAs.

Several groups have recently identified a large complex which is specifically associated with polyadenylation signal containing RNAs (21-26). In addition, we have recently identified a 64 kD protein which requires an intact AAUAAA for UV crosslinking to polyadenylation signals (34). In order to further delineate the sequence requirements for complex formation and for 64kD protein crosslinking, as well as to correlate these interactions with functional polyadenylation signals, we analyzed the protein-RNA associations of the SVL AAUAAA variants by both native gel electrophoresis and UV crosslinking/label transfer analysis.

As seen in figure 2, the wild-type SVL RNA formed a large complex when incubated with HeLa cell nuclear extract in the *in vitro* polyadenylation system. With the exception of the AUUAAA variant, none of the point mutations tested permitted the formation of this complex to a significant level. The AUUAAA variant formed the polyadenylation signal-specific complex, but at a reduced efficiency compared to wild type. This correlates well with the ability of these RNAs to be polyadenylated *in vitro* (figure 1) and is

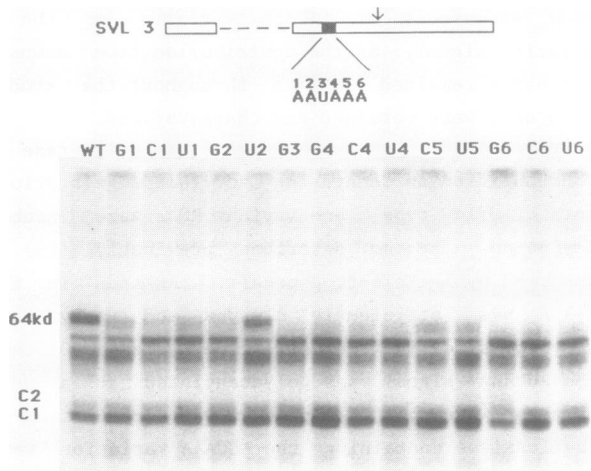

Figure 3: UV crosslinking/label transfer analysis of proteins bound to RNAs in the polyadenylation system. [^{32}P]labeled wild-type and AAUAAA variant RNAs were incubated in the *in vitro* polyadenylation system for 10 min and crosslinked to proteins with UV light. Samples were digested with RNase A and subjected to electrophoresis in a 10% acrylamide gel containing 0.1% SDS. The positions of key proteins are indicated.

consistent with the notion that the formation of a 40 - 50S complex is a requirement for polyadenylation (21-26).

The protein-RNA associations of the SVL variants were further tested by UV crosslinking analysis as described previously (34). Briefly, [^{32}P] labeled RNAs were incubated for a short period of time in the *in vitro* polyadenylation system to allow formation of specific complexes. The reactions were then irradiated on ice using a germicidal light for 10 min., followed by treatment with ribonuclease. Proteins labeled with [^{32}P] through covalent associations with small ribonucleotides were then separated in denaturing acrylamide gels and visualized by autoradiography.

As seen in figure 3, only RNAs containing AAUAAA or AUUAAA polyadenylation elements were efficiently crosslinked to the 64kd polypeptide. Crosslinking of the 64kd protein to RNA containing the AUUAAA signal was slightly less efficient as compared to wild-type substrate RNA. The additional [^{32}P]labeled uridylate residue available for crosslinking in the variant hexanucleotide, however, makes a quantitative comparison of the two RNAs difficult. UV crosslinking of additional proteins, including the hnRNP C proteins (35), to the SVL variants was not affected. This is consistent with previous observations indicating that downstream sequences were necessary and sufficient for the association of SVL

RNA with these additional proteins (35). Taken together, these data further correlate the 64kd protein with specific polyadenylation complex formation, as well as with the functional nature of a polyadenylation signal (figure 1).

DISCUSSION

We have constructed 14 of the 18 possible point mutations in the conserved AAUAAA hexanucleotide of the polyadenylation signal of the SV40 late transcription unit. Only RNAs containing the wild-type AAUAAA and the AUUAAA variant were processed at detectable levels in the *in vitro* polyadenylation system. In addition, these two RNAs also efficiently formed a specific complex as analyzed by native gel electrophoresis and were crosslinked to the 64kd polypeptide shown previously to be associated with polyadenylation signal containing RNAs (34). These data correlate both the formation of a large complex as well as 64kd protein crosslinking with polyadenylation signals containing functional hexanucleotide elements.

The observations presented here extend previous studies involving the functional analysis of variants in the AAUAAA element of the SV40 late polyadenylation signal (4, 6, 7, 30, 34). In these studies, all variants tested were processed at less than 12% of wild type levels.
Several laboratories have shown a dramatic effect of AAUAAA mutations on cleavage efficiency *in vivo* with only minor effects on polyadenylation efficiency (3, 4, 5). Recent experiments using reconstituted *in vitro* systems, however, have clearly demonstrated the existence of a specificity factor which renders the poly(A) polymerase dependent on the presence of an AAUAAA motif (30, 32, 33). Our data are consistent with the latter observations which imply the AAUAAA affects both the cleavage and polyadenylation events.

Approximately 10% of mRNAs surveyed contain putative polyadenylation signals which have variations in the AAUAAA element (GENBANK, 1985). Sixty percent of these variants were of the sequence AUUAAA. The remainder includes several point mutations similar to those constructed in this study (i.e. AGUAAA (42), AAUACA (43), AAUUAA (44), and AAUAAU (45)). In addition, the 3' proximal region of several mRNAs, such as DHFR (37) and HTLV-1 (36), bear only a superficial resemblance to the conserved hexanucleotide element. While these variants are presumably functional in their natural context, they are used very inefficiently, if at all, in the context of the SV40 late polyadenylation signal *in vitro*.

There are several possible explanations for this observation. Additional sequence elements may be associated with polyadenylation signals containing variant hexanucleotide motifs to allow efficient assembly of the

signal recognition complex. A requirement of downstream sequence elements for efficient poly (A) site use has been demonstrated for several polyadenylation signals (15-20). We have recently found that downstream sequences were required for crosslinking of the hnRNP C proteins to six independent polyadenylation signals (35). Additional proteins showing apparent specificity for downstream sequences were also noted (35), although without functional correlations the significance of these findings is unclear.

Multiple mechanisms exist for 3' end maturation. Non-polyadenylated histone mRNAs, for example, require different conserved elements as well as the U7 snRNP for 3' end maturation (reviewed in 1). In addition, yeast do not contain 3' processing signals similar to those found in higher eukaryotes. An 82 base pair region of the CYC1 gene, for example, has recently been implicated in 3' end formation in *Saccharomyces cerevisiae* (46). Putative polyadenylation signals, such as DHFR or HTLV-1 LTR, which bear little resemblance to the AAUAAA element, might be processed through alternative pathways.

Finally, additional factors not present in HeLa cell nuclear salt wash extracts may be required for the processing of alternative polyadenylation signals. An HPV-11 specific transcript, containing a putative AGUAAA polyadenylation signal element, was not detectably processed in our *in vitro* system (data not shown). We are currently undertaking further experiments to test the models described above in order to identify additional influences on polyadenylation which may play a role in the regulation of gene expression.

ACKNOWLEDGEMENTS

We thank R. Dickinson and M. Flocco for synthesis of oligonucleotides, J. Schaack and E. Vakalopoulou for critical discussions and E. Chiarchiaro for typing this manuscript.

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.

REFERENCES

1. Birnstiel,M.L., Busslinger,M., and Strub,K. (1985) Cell **41**, 349-359.
2. Proudfoot, N.J. and Brownlee, G.G. (1976). Nature **263**, 211-214.
3. Higgs,D.R., Goodbourn,S.E.Y., Lamb,J., Clegg,J.B.,Weatherall,D.J., and Proudfoot,N.J. (1983) Nature **306**, 398-400.
4. Fitzgerald,M. and Shenk,T. (1981) Cell **24**, 251-260.
5. Montell,C., Fischer,E.F., Caruthers,M.H., and Berk,A.J. (1983) Nature **305**, 600-605.
6. Wickens,M.P. and Stephenson,P. (1984) Science **226**, 1045-1051.
7. Zarkower,D., Stephenson,P., Sheets,M., and Wickens,M. (1986) Mol. Cell. Biol. **6**, 2317-2323.

8. Berget,S.M. (1984) Nature **309**, 179-182.
9. Hart,R.P., McDevitt,M.A., and Nevins,J.R. (1985) Mol. Cell. Biol. **5**, 2975-2983.
10. McLaughlin,J., Gaffney,D., Whitton,J.L., and Clements,J.B. (1985) Nucl. Acids Res. **13**, 1347-1367.
11. McDevitt,M.A., Imperiale,M.J., Ali,H., and Nevins,J.R. (1984) Cell **37**, 993-999.
12. Mason,P.J., Elkington,J.A., Lloyd,M.M., Jones,M.B., and Williams,J. (1986) Cell **46**, 263-270.
13. McDevitt,M.A., Hart,R.P., Wong,W.W., and Nevins,J.R. (1986) EMBO J. **5**, 2907-2913.
14. Zhang,F., Denome,R.M., and Cole,C.N. (1986) Mol. Cell. Biol. **6**, 4611-4623.
15. Simonsen,C.C. and Levinson,A.D. (1983) Mol. Cell. Biol. **3**, 2250-2258.
16. Gil,A. and Proudfoot,N.J. (1984) Nature **312**, 473-474.
17. Sadofsky,M. and Alwine,J.C. (1984) Mol. Cell. Biol. **4**, 1460-1468.
18. Sadofsky,M., Connelly,S., Manley,J.L., and Alwine,J.C. (1985) Mol. Cell. Biol. **5**, 2713-2719.
19. Cole,C.N. and Stacy,T.P. (1985) Mol. Cell. Biol. **5**, 2104-2113.
20. Conway,L. and Wickens,M.P. (1985) Proc. Natl. Acad. Sci. USA **82**, 3949-3953.
21. Humphrey,T., Christofori,G., Lucijanic,V. and Keller,W. (1987) EMBO J. **6**, 4159-4168.
22. Skolnick-David,H., Moore,C.L., and Sharp,P.A. (1987) Genes Dev. **1**, 672-682.
23. Zarkower,D. and Wickens,M. (1987) EMBO J. **6**, 4177-4186.
24. Zhang,F. and Cole,C.N. (1987) Mol. Cell. Biol. **7**, 3277-3286.
25. Moore,C.L., Skolnick-David,H., and Sharp,P.A. (1988) Mol. Cell. Biol. **8**, 226-233.
26. Stefano,J.E. and Adams,D.E. (1988) Mol. Cell. Biol. **8**, 2052-2062.
27. Moore,C.L. and Sharp,P.A. (1984) Cell **36**, 581-591.
28. Moore,C.L. and Sharp,P.A. (1985) Cell **41**, 845-855.
29. Hashimoto,C. and Steitz,J.A. (1986) Cell **45**, 581-591.
30. Takagaki,Y., Ryner,L.C., and Manley,J.L. (1988) Cell **52**, 731-742.
31. Gilmartin,G.M., McDevitt,M.A., and Nevins,J.R. (1988) Genes Dev. **2**, 578-587.
32. McDevitt,M.A., Gilmartin,G.M., Reeves,W.H., and Nevins,J.R. (1988) Genes Dev. **2**, 588-597.
33. Christofori,G. and Keller,W. (1988) Cell **54**, 875-889.
34. Wilusz,J. and Shenk,T. (1988) Cell **52**, 221-228.
35. Wilusz,J., Feig,D.I., and Shenk,T. (1988) Mol. Cell. Biol. **8**, 4477-4483.
36. Seiki,M., Hattori,S., Hirayama,Y., and Yoshida,M. (1983) Proc. Natl. Acad. Sci. USA **80**, 3618-3622.
37. Setzer,D.R., McGrogan,M., Nunberg,J.H., and Schimke,R.T. (1980) Cell **22**, 361-370.
38. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol. **65**, 499-560.
39. Melton,D., Krieg,P.A., Regliagliati,M.R., Maniatis,T., Zinn,K., and Green,M.R. (1984) Nucl. Acids Res. **12**, 7035-7056.
40. Dignam,J.D., Lebovitz,R.M., and Roeder,R.G. (1983) Nucl. Acids Res. **11**, 1475-1488.
41. Konarska,M.M. and Sharp,P.A. (1986) Cell **46**, 845-855.
42. Tamura,T., Noda,M., and Takano,T. (1981) Nucl. Acids Res. **9**, 6615-6626.
43. Hanukoglu,I. and Fuchs,E. (1982) Cell **31**, 243-252.

44. Kurachi,K. and Davie,E.W. (1982) Proc. Natl. Acad. Sci. USA **79**, 6461-6464.
45. Jansen,M., van Schaik,M.A., Ricker,A.T., Bullock,B., Woods,D.E., Gabbay,K.H., Nussbaum,A.L., Sussenbach,J.S. and Van den Brande,J.L. (1983) Nature **306**, 609-611.
46. Ruohola,H., Baker,S.M., Parker,R., and Platt,T. (1988) Proc. Natl. Acad. Sci. USA **85**, 5041-5045.