Selection of sequence elements that substitute for the standard AATAAA motif which signals 3' processing and polyadenylation of late simian virus 40 mRNAs

---

Candace Swimmer¹ and Thomas Shenk^{2*}

Department of Microbiology, Health Science Center, State University of New York, Stony Brook, NY 11794, USA

Received 19 August 1985; Revised 14 October 1985; Accepted 15 October 1985

ABSTRACT

A method is described which allows selection of sequences which can substitute for the normal AATAAA hexanucleotide involved in polyadenylation of SV40 late mRNAs. Plaques were generated from viral DNA lacking the motif, forcing acquisition of substitute sequences. Four variants were characterized. All displayed wild-type growth kinetics and produced normal levels of late mRNAs and proteins. Two variants had reacquired AATAAA elements and one acquired an ATTAAA sequence. The last variant carried an ATTTTTTAAA segment, suggesting this novel sequence, or some portion of it, can also signal poly A addition.

INTRODUCTION

In eukaryotes the 3' termini of mature mRNAs arise by a cleavage event followed by poly A addition (1). This has been demonstrated for both cellular and viral genes (2-6). A conserved hexanucleotide, AATAAA (7), which usually lies 11 to 30 base pairs upstream from the poly A addition site is essential for cleavage and polyadenylation of mRNAs (8-11). Some variation in this sequence has been found. The sequence ATTAAA has been identified in a number of cases (12-14), and rarely TATAAA, AATATA, or AGTAAA (15-18). These sequence elements alone do not direct 3' processing of mRNAs since cleavage and polyadenylation are not observed at all sites which contain them. In fact there is good evidence that sequences downstream from the poly A addition site serve as part of the processing signal (19-23). A second consensus sequence, CAYTG (Y=pyrimidine), often found upstream or downstream of the poly A addition site may also play a role in efficient polyadenylation (24).

We have constructed a recombinant plasmid which carries an SV40 genome lacking a 16 base-pair segment which includes the hexanucleotide required for processing late mRNAs. This variant, as well as a similar recombinant clone lacking a somewhat larger segment, was used to transfect monkey kidney cells. Infectious virus was recovered at high frequency which carried alterations within the 3' noncoding region of the late transcription unit. In three cases, rearrangements restored a common processing signal, i.e. AATAAA or ATTAAA. In one case a novel element, ATTTTTTAAA, was introduced. This novel element functioned at good efficiency to signal poly A addition.

MATERIALS AND METHODS

Plasmids, Viruses and Cells

The plasmid pSVL contains the replication origin plus late region of SV40 (map position 4739 to 2770) in pML2 (25) and pSV15L is a derivative of pSVL which lacks a 16 base-pair segment that includes the hexanucleotide element required for processing late mRNAs (map position 2570 to 2586). Plasmid pSVWT contains the entire SV40 chromosome inserted at its unique BamH1 site into the BamH1 cleavage site of pBR322, and pSV15 is a derivative of pSVWT which lacks the same 16 base-pair segment that was deleted in pSV15L. Plasmid pd11141 was a gift from Dr. J. Pipas. It carries a similar deletion to that in pSV15 and has been described (26).

Wild-type SV40 (wt830) is a plaque-purified derivative of the SVS strain (27). CV-1P and COS-1 (28) cells were maintained in medium supplemented with 10% calf serum. CV-1P cells were transfected with viral DNAs using the DEAE-dextran method (29) and COS-1 cells were transfected using the chloroquine modification of the DEAE-dextran method (30).

Preparation and Analysis of RNA

To monitor steady-state RNA levels, cytoplasmic, polyadenylated RNA was prepared (31) either from CV-1P cells 48-60 hrs after infection at a multiplicity of 10 plaque-forming units/cell or from COS-1 cells at 48 hrs after transfection. For Northern-type analysis, RNA was subjected to electrophoresis in formaldehyde-containing agarose gels (32) and blotted to nitrocellulose (33). A subclone of SV40 carrying the 874 to 2138 nucleotide region was labeled by nick translation (34) and used to probe specifically for late viral mRNAs. For S1 (35) or exonuclease VII (36) analyses, probe DNA segments were end-labeled by repair of a BamH1-cleaved terminus with the Klenow fragment of DNA polymerase I (α - ^{32} P-CTP, 3000Ci/mole, 50uCi/10ug DNA), and strands were separated by electrophoresis.

To determine transcription rates, CV-1P cells were labeled for 15 min with ^{3} H-uridine (50Ci/mole, 200uCi/ml) 48 hrs after infection at a multiplicity of 10 plaque-forming units/cell. Nuclear RNA was prepared (37) and hybridized to single-stranded probe DNAs (carried in recombinant M13 phages) bound to nitrocellulose filters (38, 39).

Protein Analysis

CV-1P cells were labeled for 3 hrs with ^{35} S-methionine (600Ci/mole, 50uCi/ml) 48 hrs after infection at a multiplicity of 0.1 plaque-forming unit/cell. Nuclear proteins were prepared (40), subjected to electrophoresis on an SDS-containing, 15% polyacrylamide gel (41), and radioactive bands were detected by fluorography (42).

RESULTS

Deletion of the AATAAA Segment Prevents Polyadenylation

By transfecting monkey kidney cells with SV40 DNA lacking the normal AATAAA element within the late viral transcription unit (pSV15 and pd11141, Fig. 1) and selecting for viable viruses, we hoped to force acquisition of

Figure 1. Diagram of the SV40 chromosome, 3' non-coding regions of early and late mRNAs, and deletions which remove the late mRNA AATAAA motif. Early and late mRNA exons are indicated by arrows and introns by spaces on the circular SV40 genome. The expanded region details the early and late poly A addition sites. Solid rectangles locate mutations which delete the late unit AATAAA sequence.

novel sequences which could signal poly A addition. However, it was first necessary to establish that such a mutated SV40 DNA does not utilize an alternative polyadenylation site to generate functional late mRNAs. Accordingly, COS-1 cells were transfected with either pSVL which carries an intact SV40 late transcription unit or pSV15L which lacks a 16 base-pair segment including the AATAAA segment. Both plasmids contain an SV40 origin of DNA replication, and therefore are amplified within COS-1 cells which contain and express the SV40 T antigen gene (Fig. 2A). Only pSVL expresses the 16 and 19S late mRNAs (Fig. 2B). No late mRNAs were detected in cells transfected with pSV15L which lacks the AATAAA motif.

Isolation of Viral Variants with Altered 3' Noncoding Regions

Two mutant plasmids were utilized for the selection of viruses carrying novel substitutions for the AATAAA element. pSV15 and pd11141 each carry the entire SV40 genome but lack 16 and 34 base-pair segments, respectively, removing the late AATAAA signal (Fig. 1). Full length, linear SV40 DNA was released from both plasmids by cleavage with *Bam*HI and CV-1P cells were transfected with the linear, mutant DNA. Both DNAs gave rise to surprisingly high numbers of plaques (Table 1), producing about one percent the number obtained by transfection with wild-type DNA. Curiously, when the assay was performed at elevated temperature (39° instead of 37°C), the number of plaques generated by mutant DNA increased 10-fold while the number of

Figure 2. Replication and expression of late SV40 mRNAs in COS-1 cells by pSVL and pSV15L. A. DNA blot analysis of plasmid DNAs after replication in transfected cells. DNA was prepared at the times indicated, subjected to electrophoresis in an agarose gel, blotted to nitrocellulose and hybridized to ^{32} P-labeled pSVL probe DNA. DNA forms I, II and III are designated. B. RNA blot analysis of late viral mRNAs produced 48 hr after transfection of cells with pSVL, pSV15L or wild-type SV40 DNA (labeled 830). An identical amount of each polyadenylated RNA preparation was analyzed using excess probe DNA to permit accurate quantitation. 16 and 19S RNA species are labeled.

wild-type plaques remained constant (data not shown). The reason for the temperature effect is unclear. Perhaps higher temperature favors recombination events which give rise to rearranged viral chromosomes.

Virus from several plaques was used to infect CV-1P cells and the viral DNA produced was heterogeneous. Therefore, the primary isolates were plaque-purified a second time. DNA from these plaque isolates was homogeneous. The alterations within four isolates were subjected to nucleotide sequence analysis (43). Sequences of parents and variants are displayed in figure 3. Two variants (d11466 and in1467) were derived from transfection with pSV15 and two (d11468 and in1469) were from pd11141. The mutants acquired small additions of unknown origin or deletions near the normal 3' end of late mRNAs. The sequence AATAAA was generated across the junction of the insertion in in1467, and the hexanucleotide, AATAAA, was inserted in in1469. A deletion of 66 base pairs was found in d11466 and the new junction generated the sequence ATTAAA. d11468 contained a new deletion of 39 base pairs. This variant does not contain an obvious match to the hexanucleotide consensus, but its deletion junction generates the sequence ATTTTTTAAA.

The Variants Efficiently Signal Poly A Addition

The growth kinetics of these variants were examined (Fig. 4). All grew as well as wild-type virus suggesting that the altered sequences were functioning at good efficiency to signal polyadenylation.

Table 1. Plaque-Forming efficiencies of pSV15 and pd11141 DNA

<u>Bam</u> H1-cut DNA	Plaque-forming units/ug
pSVWT	5×10^5
pSV15	2×10^3
<u>pd11141</u>	7×10^3

1×10^{-4} ug BamH1-cleaved DNA was used to transfect CV-1P cells. Plaque-formation was monitored after incubation at 37°C for 14 days.

Mutant d11468 was examined in greater detail since it carried a novel sequence presumed to be responsible for 3' processing. Steady state levels of late d11468 mRNAs were examined by Northern-type analysis of cytoplasmic, polyadenylated RNA, and found to be indistinguishable from wild-type virus-coded RNA at 48 hr after infection (Fig. 5A). Similarly, synthesis of a late viral polypeptide, VP-1, was also normal in mutant-infected cells (Fig. 5B). To rule out the possibility that the novel sequence in d11468 signals

Figure 3. Rearrangements within the 3' non-coding regions of d11466, in1467, d11468 and in1469. Deletions are depicted by boxed areas and insertions by triangles. The sequence around the junction of each rearrangement is shown. pSV15 is the parent to d11466 and in1467, while d11468 and in1469 are derived from pd11141.

Figure 4. Growth kinetics of wild-type and mutant viruses in CV-1P cells. Infections were at a multiplicity of 5 plaque-forming units/cell. Cultures were harvested at the indicated times and infectious virus measured by plaque assay on CV-1P cells.

Figure 5. Production of late viral mRNAs and protein in cells infected with dl1468 or wt830. A. Blot analysis of late mRNAs. Cytoplasmic, polyadenylated RNA was prepared at 60 hr after infection at a multiplicity of 0.1 plaque-forming unit/cell and analyzed using a probe DNA specific for late viral mRNAs. 16 and 19S species are labeled. B. Electrophoretic analysis of late nuclear proteins. CV-1P cells were labeled with ³⁵S-methionine for 3 hr beginning 48 hr after infection at a multiplicity of 0.1 plaque-forming unit/cell. Electrophoresis was on a 15% polyacrylamide gel containing SDS.

Table 2. Transcription rates preceding and following d11468 and wt830 late mRNA poly A addition sites

Virus	CPM/base pair	
	Upstream probe	Downstream probe
<u>wt830</u>	3.2	6.2
<u>d11468</u>	8.2	14.1

M13 probe DNAs contained either a PstI to HpaI site fragment (upstream probe=sequence position 1988-2666) or BclII to HindIII site fragment (downstream probe=sequence position 2770-3476). The wild-type hexanucleotide motif is located at position 2657 and the novel junction in d11468 resides at position 2684.

termination of transcription and not polyadenylation-specific processing, transcription rates were measured by annealing pulse-labeled RNA to single-stranded probe DNAs located immediately preceding or following the ATTTTAA sequence (Table 2). No significant difference in rate was evident at the two locations, indicating that RNA polymerase transcribes past the novel sequence.

To further implicate the novel junction in d11468 in polyadenylation, the poly A sites were mapped for late mRNAs coded by the variant. Cytoplasmic, polyadenylated RNA was prepared from mutant or wild-type virus-infected cells and hybridized to a single-stranded 3' end-labeled probe DNA (35). The probe DNAs extended from the labeled BamHI site (position 2533). The annealed mixture was treated with S1 endonuclease and the size of protected probes determined by electrophoresis (Fig. 6A). RNA from wild-type virus-infected cells protected a 141/142 nucleotide fragment. This mapped precisely to the known late mRNA poly A site at sequence position 2674, 2675 (141, 142 nucleotides downstream from the labeled BamHI site). RNA from d11468-infected cells protected probe fragments ranging in size from 140-152 nucleotides as well as two larger fragments (188 and 370 nucleotides). To confirm that these fragments most likely map poly A sites and not 5' splice sites, annealed mixtures were digested with exonuclease VII. If a splice site were involved, this nuclease would generate larger bands than S1 endonuclease, provided the 3' splice site was located upstream of the probe's 5' end. As shown in figure 6A, the endo and exonucleases generated nearly identical fragments arguing strongly that they have detected poly A sites. Thus, the late mRNAs from d11468-infected cells are polyadenylated 14 to 26 nucleotides downstream (140-152 nucleotide fragments) of the sequence ATTTTAA with the major site falling at 25/26 nucleotides (Fig. 6B). A minor site is present 62 nucleotides downstream (188 nucleotide fragment). The 370 nucleotide protected fragment could result from polyadenylation downstream of an ATTTTAA sequence found at sequence position 2880.

DISCUSSION

Deletion of the AATAAA motif from the SV40 late transcription unit followed by selection for plaque formation proved to force acquisition of sequence elements able to supply the deleted function. This protocol makes it possible to screen for diverse elements capable of substituting for the normal motif. Variants able to produce late mRNAs arose at exceptionally high frequency (Table 1). Several factors likely contribute to this observation. First, the SV40 chromosome can undergo illegitimate recombination events at high efficiency (44). Second, in order to form a plaque, a recombination event need not occur until long after the onset of viral DNA replication. Thus, the key recombination event which activates production of late mRNAs can occur late after infection when the SV40 chromosome copy number within the original infected cell is very high. It should be possible to use this scheme to identify functional homologs of additional SV40 late unit components such as promoter elements and downstream sequences involved in poly A addition.

Of the four variants analyzed, two re-acquired the standard AATAAA element (7), one the less commonly utilized ATTAAA sequence (12-14), and the final mutant, d11468, contained a novel element, ATTTTATAAA (Figure 3). A number of observations indicate that this novel element or some portion of its sequence signals 3' processing and poly A addition. First, d11468-infected cells contained normal levels of late mRNAs (Fig. 5). Second, transcription did not terminate at this sequence (Table 2) as has been suggested for a related sequence (TTTTTATA) in yeast (45, 46). Third, polyadenylation occurred within 14 to 26 nucleotides of the novel element (Fig. 6), consistent with the normal spacial relationship of AATAAA motif and poly A site (see ref. 8). Thus, it appears reasonable to conclude that the ATTTTATAAA element can functionally substitute for the more common versions of the AATAAA motif.

The dihydrofolate reductase gene encodes multiple mRNAs which utilize different poly A addition sites. One of these mRNAs does not contain a standard hexanucleotide motif preceding its poly A site, and it was suggested

Figure 6. Late mRNA poly A addition sites coded by d11468 and its parent, wt830. A. S1 endonuclease and exonuclease VII analysis of late viral mRNAs using 3' end-labeled probe DNAs. Cytoplasmic, poly A-plus RNA was prepared 48 hr after infection of CV-1P cells at a multiplicity of 10 plaque-forming units/cell. The DNA probes were prepared from mutant or wild-type DNA and extended from a BamHI to PvuII cleavage site (sequence position 2533 to 3506). The fragments were 3' end-labeled at the BamHI site and strand separated. Single-stranded, 3' end-labeled probe DNA was hybridized to polyadenylated RNA and then digested with S1 endonuclease (left panel experiment) or exonuclease VII (right panel). Protected fragments were sized on a denaturing 8% polyacrylamide relative to sequencing markers (left panel) or relative to previously sized S1-generated fragments for the exonuclease VII experiment (right panel). The sizes of protected fragments are indicated.

B. Poly A addition sites utilized for late viral mRNAs coded by d11468 and wt830. Sites determined by the above analyses are designated by asterisks above the SV40 sequence. The signal motifs are indicated by lines above the sequence.

that the sequence ATAA could serve as the signal for this mRNA (47). Interestingly, the message also contains an ATTTTTTATA element which is closely related to the novel signal in d11468.

In addition to sites in the vicinity of the ATTTTTTAAA sequence, d11468 appears to generate a class of mRNAs polyadenylated about 220 nucleotides further downstream (Fig. 6). These 3' ends map to the vicinity of an ATTAAA element which is not detectably utilized in the wild-type virus, but is used when the normal poly A site is compromised by mutation and functions at somewhat reduced efficiency (22).

Finally, the two deletion variants (d11466 and 1468) lack DNA segments which code the extreme carboxy-terminus of the SV40 large T antigen. Nevertheless, they grow like wild-type virus. There is precedent for viable SV40 mutants which carry deletions in the carboxy-terminal region of large T antigen (48).

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Cancer Institute (CA28919).

¹Present address: The Biological Laboratories, Harvard University, Cambridge, MA 02138, USA

²Present address: Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

*To whom correspondence should be addressed

REFERENCES

1. Nevins, J.R. and Darnell Jr., J.E. (1978) *Cell* 15, 1477-1493.
2. Acheson, N.H. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4754-4758.
3. Ford, J.P. and Hsu, M.T. (1978) *J. Virol.* 28, 795-801.
4. Fraser, N.W., Nevins, J.R., Ziff, E., Darnell, J.E. (1979) *J. Mol. Biol.* 129, 643-656.
5. Hagenbuchle, O., Bovey, R. and Young, R.A. (1980) *Cell* 21, 179-187.
6. Hofer, E. and Darnell, J.E. (1981) *Cell* 23, 585-593.
7. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214.
8. Fitzgerald, M. and Shenk, T. (1981) *Cell* 24, 251-260.
9. Higgs, D.R., Goodburn, S.E., Lamb, J., Cleary, J.B., Weatherall, J.J. and Proudfoot, N.J. (1983) *Nature* 306, 398-400.
10. Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1983) *Nature* 305, 600-605.
11. Wickens, M. and Stephenson, P. (1984) *Science* 226, 1045-1051.
12. Jung, A., Sippel, A.E., Grez, M. and Schutz, G. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5759-5763.
13. Ahmed, C.M., Chanda, R., Stow, N. and Zain, B.S. (1982) *Gene* 19, 297-301.
14. Fornwald, J., Kuncio, G., Peng, I. and Ordahl, C.P. (1982) *Nucl. Acids Res.* 10, 3861-3876.
15. Donehower, L., Huang, A.L., Hager, G.L. (1981) *J. Virol.* 37, 226-238.
16. Tosi, M., Young, R.A., Hagenbuchle, O. and Schibler, U. (1981) *Nucl. Acids Res.* 9, 2313-2323.
17. Capon, D.J., Chen, E.Y., Levinson, A.D., Seeburg, P.H. and Goeddel, D.V. (1983) *Nature* 302, 33-37.
18. Simonsen, C.C. and Levinson, A.D. (1983), *Mol. Cell. Biol.* 3, 2250-2258.
19. Gil, A. and Proudfoot, N.J. (1984) *Nature* 312, 473-474.

20. McDevitt, M.A., Imperiale, M.J., Ali, H. and Nevins, J.R. (1984) *Cell* 37, 993-999.
21. Woychik, R.P., Lyons, R.B., Post, L. and Rottman, F.M. (1984) *Proc. Natl. Acad. Sci. USA* 81, 3944-3948.
22. Sadofsky, M. and Alwine, J.C. (1984) *Mol. Cell. Biol.* 4, 1460-1468.
23. Conway, L. and Wickens, M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 3949-3953.
24. Berget, S.M. (1984) *Nature* 309, 179-181.
25. Lusky, M. and Botchan M. (1981) *Nature* 293, 79-81.
26. Pipas, J.M., Peden, K.W.C. and Nathans, D. (1983) *Mol. Cell. Biol.* 3, 203-213.
27. Takemoto, K.K., Kirchstein, R.L. and Habel, K. (1964) *J. Bact.* 92, 990-994.
28. Gluzman, Y. (1981) *Cell* 23, 175-182.
29. Mertz, J.E. and Berg, P. (1974) *Viol.* 62, 112-124.
30. Luthman, H. and Magnusson, G. (1983) *Nucl. Acids Res.* 11, 1295-1308.
31. McGrogan, M., Spector, D.J., Goldenberg, C., Halbert, D.N. and Raskas, H.J. (1979) *Nucl. Acids Res.* 6, 593-608.
32. Rave, N., Crkvenjakov, R. and Boedtker, H. (1979) *Nucl. Acids Res.* 6, 3559-3567.
33. Alwine, J.C., Kemp, D.J. and Stark, G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5350-5354.
34. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
35. Weaver, R.F. and Weissman, C. (1979) *Nucl. Acids Res.* 7, 1175-1183.
36. Berk, A.J., Sharp, P.A. (1978) *Cell* 14, 695-711.
37. Chirgwin, J.M., Przybla, A.E., McDonald, R.J. and Rutter, W.T. (1979) *Biochemistry* 18, 5294-5299.
38. Nevins, J.R. (1980) *Methods Enzymol.* 65, 768-785.
39. Babich, A. and Nevins, J.R. (1981) *Cell* 26, 371-379.
40. Tegtmeyer, P., Robb, J.A., Widmer, C. and Ozer, H.L. (1974) *J. Virol.* 14, 997-1007.
41. Laemmli, U.K. (1970) *Nature* 227, 680-685.
42. Bonner, W.M. and Laskey, R.A. (1974) *Eur. J. Biochem.* 46, 83-88.
43. Maxam, A. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
44. Wake, C.T. and Wilson, J.H. (1979) *Proc. Natl. Acad. Sci. USA* 76, 2876-2880.
45. Zaret, K.S. and Sherman, F. (1982) *Cell* 28, 563-573.
46. Henikoff, S., Kelly, J.D. and Cohen, E.H. (1983) *Cell* 33, 607-614.
47. Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schmieke, R.T. (1980) *Cell* 22, 361-370.
48. Cole, C.N., Crawford, L.V. and Berg, P. (1979) *J. Virol.* 30, 683-691.