Polyadenylation at a cryptic site in the pBR322 portion of pSV2-neo: prevention of its utilization by the SV40 late poly(A) signal

Marco M.Kessler, Mark A.Westhafer, David D.Carson and Jeffrey L.Nordstrom*

Department of Biochemistry and Biophysics, Texas A&M University and the Texas Agricultural Experiment Station, College Station, TX 77843-2128, USA

Received August 21, 1986; Revised and Accepted December 5, 1986

ABSTRACT

Transcripts originating from the SV40 late promoter of pSV2-neo or pSV2-cat contain pBR322 sequences and are polyadenylated at the SV40 late poly(A) site, resulting in an RNA of 3500 nt. If the SV40(L) poly(A) signal is destroyed, late orientation transcripts are polyadenylated at a site within pBR322 sequences, yielding in an RNA of 2500 nt. This cryptic poly(A) site is located 42-46 nucleotides downstream from an AAUAAA. Utilization of the pBR322 poly(A) signal is undetectable in late orientation transcripts from pSV2-neo or pSV2-cat, although it is located 966 nucleotides upstream from the SV40(L) poly(A) signal. The pBR322 site is not utilized when the spacing between the two poly(A) signals is varied from 209 to 1913 nucleotides. The pBR322 poly(A) site was utilized only in constructs in which all or portions of the SV40(L) poly(A) signal were deleted, such as in a construct with a 7 bp deletion into the SV40(L) AATAAA and adjacent sequences.

INTRODUCTION

pSVN-3 is a variant of pSV2-neo in which a 66 bp fragment between the SV40 early and late poly(A) sites has been deleted and replaced with a 20 bp oligonucleotide that contains an AAUAAA (1). This deletion removes the two AAUAAA's that precede the SV40 early poly(A) site and the AAUAAA that precedes the SV40 late poly(A) site. Insertion of the oligonucleotide containing an AAUAAA in the early orientation regenerates the early orientation poly(A) signal, but the late orientation signal remains defective (Fig. 1).

S1 nuclease mapping of late orientation transcripts from pSVN-3 detected a polyadenylation site in the pBR322 portion of the transcript. No polyadenylation at this site was detected when late orientation transcripts from pSV2-neo were analyzed, even though the nucleotide sequence of pSVN-3 and pSV2-neo is identical in this region. The experiments described in this report were performed to characterize the polyadenylation site in the pBR322 sequence.

MATERIALS AND METHODS

Plasmids: pSV2-neo was constructed by Southern and Berg (2), pSVN-3 by Kessler et al. (1), and pSV2-cat by Gorman and Howard (3). To construct the 1634

Figure 1. Maps of pSV2-neo and pSVN-3: Sequences are derived from SV40 (open boxes), the neo gene of In5 (closed boxes), pBR322 (solid line) and oligonucleotides (small cross-hatched box). Transcripts that initiate from the early or late promoters are spliced (dotted line) and polyadenylated as indicated. pSVN-3 contains a regenerated SV40 early poly(A) signal, but lacks an SV40 late poly(A) signal (1). The arrow points to the location of the cryptic poly(A) site that is described in the text.

plasmid, the 1428 bp EcoRI-AvaI fragment of pBR322 (4) containing the tetracycline resistance gene, was inserted between the EcoRI and BamHI sites of pSV2-cat. The 1913 plasmid was constructed in the same way, except the pBR322 fragment was inserted between the EcoRI and ApaI sites of pSV2-cat. In these constructions, the AvaI, BamHI and ApaI ends were blunted by treatment with Klenow DNA polymerase. After transformation of *E. coli* RRI, colonies were screened for resistance to tetracycline and the plasmids identified by restriction mapping. The structures of the 1634 and 1913 plasmids are outlined in Fig. 4.

Deletion of portions of the SV40 late poly(A) signal: HpaI-digested pSV2-neo DNA was treated with nuclease Bal-31, the ends made blunt with Klenow DNA polymerase, ligated in the presence of XhoI linkers (CCTCGAGG, Bethesda Research Laboratory), and used to transform *E. coli* RRI (5). Colonies were screened by restriction mapping and DNA sequencing (6). To create one-way deletions, the XhoI-EcoRI fragment from each chosen deletion was exchanged with the XhoI-EcoRI fragment of pSV2-neo that was modified by the addition of a XhoI linker at the HpaI site.

COS cell transfections and RNA analyses: COS cells (7) were transfected with the plasmids described above using DEAE dextran (8). Total cellular RNA was

isolated 48 hours after transfection and analyzed by S1 nuclease mapping with 3' end-labeled DNA probes. These procedures were performed as described previously (1, 9). Poly(A)⁺ and poly(A)⁻ RNA were separated by poly(U) sepharose chromatography (10).

RESULTS

Identification and characterization of the pBR322 poly(A) site

S1 nuclease mapping experiments were done to characterize the polyadenylation of late orientation transcripts from pSV2-neo and pSVN-3. Mapping pSV2-neo transcripts with a 3' labeled XmnI-RsaI fragment from pSV2-neo, resulted in the appearance of a 1287 nucleotide band that fractionated as poly(A)⁺ RNA and maps to the region of the SV40 late poly(A) site (Fig. 2A, lanes 1-4). We conclude that late orientation pSV2-neo transcripts are polyadenylated at the SV40 late poly(A) site. The level of late orientation RNAs from pSV2-neo not processed at the SV40 late poly(A) site is apparently negligible, since a fully protected probe band was not detected. In contrast, mapping pSVN-3 transcripts with the same XmnI probe resulted in the appearance of two bands, 1213 and 321 nucleotides in length, that were of approximately equal intensity and fractionated as poly(A)⁺ RNA (Fig. 2A, lanes 5-8). The 1213 nucleotide band (highlighted by the dots adjacent to lanes 5 and 8) maps the discontinuity between pSVN-3 RNA and the pSV2-neo probe. This band represents late orientation RNAs that were not processed in the 1213 nucleotide region tested. The 321 nucleotide band maps a processing site, located within the pBR322 region, that is approximately 80 nucleotides upstream from the EcoRI site. Inspection of the nucleotide sequence in this region shows that the processing site occurs approximately 40 nucleotides downstream from an AAUAAA, the most conserved feature of poly(A) signals (14). Many late orientation transcripts from pSVN-3, therefore, appear to be polyadenylated at a site in pBR322 sequence that is inactive for late orientation pSV2-neo transcripts.

To map the site of polyadenylation in pBR322 sequence more accurately, S1 nuclease mapping experiments were performed with a TaqI probe in which the site of labeling is more proximal to the site of polyadenylation. When late orientation transcripts from pSV2-neo are mapped with this probe, a single band of 345 nucleotides, representative of read-through, is observed (Fig. 2B, lane 1). No polyadenylation at the pBR322 site is observed. When late orientation transcripts from pSVN-3 are mapped with this probe, two bands of 267 and 345 nucleotides are observed (Fig. 2B, lane 2). The 267 nucleotide band maps the poly(A) site to the same region as the 321 nucleotide band from the XmnI

probe. The site of pBR322 polyadenylation is therefore localized to a site that is 42-46 nucleotides downstream from the AAUAAA (Fig. 3).

The RNAs from pSVN-3 that are not processed at the pBR322 poly(A) site, represented by the 1213 nucleotide band in Fig. 2A (lanes 5-8), fractionate as poly (A)⁺ RNA. Since the SV40 late poly(A) signal has been destroyed in pSVN-3, another "cryptic" polyadenylation site may exist outside the region spanned by the probe. Alternatively, the larger polyadenylated RNAs may be

Figure 2. Mapping the 3' ends of late orientation transcripts from pSV2-neo and pSVN-3: RNA from transfected COS cells was analyzed by S1 nuclease mapping using 3' end-labeled DNA probes. Abbreviations above the gel lanes are T (total RNA), Y (yeast RNA), A⁺ (polyA⁺ RNA), A⁻ (polyA⁻ RNA), M (markers) and P (probe). **A. XmnI probe:** The 1435 bp XmnI-RsaI fragment from pSV2-neo was 3' labeled at the XmnI site located at 399. Hybridizations were performed at 45°C. RNA from pSV2-neo transfected cells was analyzed in lanes 1-4, and RNA from pSVN-3 transfected cells in lanes 5-8. Lanes 1 and 5: 20 µg of total RNA. Lanes 2 and 6: 10 µg of poly(A)⁺ RNA. Lanes 3 and 7: 2 µg of poly(A)⁺ RNA. Lanes 4 and 8: 20 µg of poly(A)⁻ RNA. Lane 9: markers (Øx174 DNA cut with HaeIII and 3' end-labeled). Lane 10: probe. A small amount of fully protected probe is observable in lane 3. The 1213 nucleotide bands in lanes 5-8 (highlighted by the dots) are not to be confused with the faint sub-bands in lanes 1-4, that probably result from S1 nuclease digestion artifacts. **B. TaqI probe:** The 642 bp TaqI-PvuII fragment from pUC8 was 3' labeled at the TaqI site. This probe is continuous with the sequence of pSV2-neo and pSVN-3 between the TaqI site of 345 and the EcoRI site at 0. The discontinuous portion is indicated by the wavy line in the diagram. Lane 1: 20 µg of total RNA from pSV2-neo transfected cells. Lane 2: 20 µg of total RNA from pSVN-3 transfected cells. Lane 3: 20 µg of RNA from cells transfected with an equimolar mixture of pSV2-neo and pSVN-3. Lane 4: markers (pBR322 DNA cut with MspI and 3' end labeled). The faint bands larger than 345 nucleotides are irrelevant because the region to which they map is beyond the point of discontinuity between the probe and the RNA.

Since utilization of the poly(A) site in pBR322 sequence appears to be prevented from being utilized by cis-acting sequences, the nucleotide sequence and

Figure 4. Variation in the distance between the pBR322 poly(A) site and the SV40 late poly(A) site: In pSV2-cat, the pBR322 and SV40 late poly(A) sites are 966 bp apart. The 1428 bp EcoRI-AvaI fragment of pBR322, containing the tetracycline resistance gene, was inserted between the EcoRI and BamHI sites of pSV2-cat or between the EcoRI and ApaI sites of pSV2-cat, resulting in clones in which the pBR322 and SV40 late poly(A) sites are separated by 1624 and 1913 bp, respectively. RNA from COS cells transfected with these plasmids was subjected to S1 mapping using the 3' labeled TaqI probe of Fig. 2B. Lane 1: 20 µg of total RNA from pSV2-cat transfected cells. Lane 2: 20 µg of total RNA from cells transfected with the 1634 clone. Lane 3: 20 µg of total RNA from cells transfected with the 1913 clone. Lane 4: size markers.

Figure 5. Utilization of the pBR322 poly(A) site in pSV2-neo variants with altered SV40 late poly(A) signals: The Hpa I site is 6 bases downstream from the 3' end of the AAUAAA and 7 bases upstream from the site of SV40 late cleavage/polyadenylation. One way Bal-31 deletions upstream from the Hpa I site from the Hpa I site of pSV2-neo were constructed. The lengths of the deletions are indicated. An 8 bp XhoI linker (CCTCGAGG) was added at the point of deletion. In each deletion, sequences downstream from the Hpa I site, including the poly(A) site and the downstream element identified by Conway and Wickens (13), was left intact. The 3' labeled TaqI-PvuII probe of Fig. 2B was hybridized to total RNA from cells transfected with pSV2-neo and each of the variants. Lane 1: markers. Lane 2: 25 μ g of yeast RNA. Lane 3: 10 μ g of RNA from pSV2-neo transfected cells. Lane 4: 10 μ g of RNA from $\Delta 7$ transfected cells. Lane 5: 10 μ g of RNA from $\Delta 8$ transfected cells. Lane 6: 10 μ g of RNA from $\Delta 13$ transfected cells.

distance between the pBR322 and SV40 late poly(A) sites was varied. In pSV2-cat and pSV2-neo, the SV40(L) poly(A) site is located 966 nucleotides downstream from the pBR322 poly(A) site. Late orientation transcripts from pSV2-cat, like those from pSV2-neo, are not polyadenylated at the pBR322 poly(A) site (Fig. 4, lane 1). Increasing the distance between the pBR322 and SV40(L) poly(A) sites to 1634 and 1913 by insertion of sequences from the tetracycline resistance gene failed to cause the pBR322 poly(A) site to be utilized (Fig. 4, lanes 2 and 3). Similarly, the pBR322 site was not utilized when the distance between

it and the SV40(L) site was decreased to 216 nucleotides by deletion of the BamHI-EcoRI fragment from pSV2-neo (data not shown). Therefore, the pBR322 poly(A) signal remains inactive even though the nucleotide sequence and the distance between it and the SV40(L) poly(A) signal is varied considerably.

Small deletions in the SV40(L) signal of pSV2-neo were analyzed to test whether the functional state of the SV40(L) poly(A) site, in contrast to its absence, influences the utilization of the pBR322 poly(A) site. Bal-31 deletions were constructed from the unique HpaI site of pSV2-neo, which is located 6 bases downstream from the AAUAAA and 7 bases upstream from the late site of cleavage/polyadenylation. Constructs in which deletions ended 7, 8 or 13 bases upstream from the HpaI site were chosen for analysis. The pBR322 poly(A) site was utilized in RNAs transcribed from all three deletions (Fig. 5B, lanes 4-6), as evidenced by the 267 nucleotide band. In each case, the ratio of read-through to polyadenylated RNAs was approximately 1:1. The efficiency of utilization of the pBR322 poly(A) site in transcripts from these deletions therefore appears to be comparable to that of transcripts from pSVN-3.

The data show that utilization of the pBR322 poly(A) site in pSV2-neo is dependent upon the integrity of the SV40(L) poly(A) signal. The pBR322 poly(A) site is not utilized at all if the SV40(L) signal is functional, even though the pBR322 site may be as much as 1913 nucleotides upstream. However, if the SV40(L) late poly(A) signal is altered or deleted, the pBR322 poly(A) site is utilized. Therefore, the SV40(L) poly(A) signal appears to totally out-compete the pBR322 poly(A) signal, even when it is located 1-2 kilobases downstream.

DISCUSSION

RNA processing at a cryptic site in pBR322 has been observed in variants of pSV2-neo that have defective or deleted SV40(L) poly(A) signals. The cryptic RNA processing site is located 74-78 nucleotides upstream from the EcoRI cleavage site of pBR322 (Fig. 3). Five lines of evidence indicate that the cryptic processing site is a polyadenylation site: 1. The site is detected with 3' labeled probes. 2. RNAs processed at this site fractionate as poly(A)⁺ RNA. 3. The processing site is 42-46 nucleotides downstream from an AAUAAA. 4. The sequence CAUUUUUAUGACAUU, located approximately 14 nucleotides downstream from the processing site, is homologous to the functionally defined downstream elements of the adenovirus E2A and SV40(L) poly(A) signals (11-13) (Fig. 3). 5. Processing at the cryptic site is inversely dependent upon the functional

state of the cis-located SV40(L) poly(A) signal, suggesting that the two signals are in competition with each other.

The spacing of 42-46 nucleotides between the AAUAAA and the site of polyadenylation is unusually long. Most sites of polyadenylation are located 10-30 nucleotides from the 3' end of the hexanucleotide (14). The abnormal spacing may explain why the pBR322 poly(A) signal appears to be relatively inefficient.

When the SV40(L) poly(A) signal is present in various downstream locations, the pBR322 poly(A) signal is not utilized at all. The pBR322 poly(A) signal is observed to function only if the SV40(L) poly(A) signal has been altered or removed. Polyadenylation at the pBR322 site was not inhibited when the SV40(L) signal was present in trans, indicating that the SV40(L) poly(A) signal interferes with the utilization of the pBR322 poly(A) signal only when present on the same transcript.

The mechanism by which this interference occurs is not known. It appears, however, to be related to the relative strengths of the two poly(A) signals. Read-through past the SV40(L) poly(A) site was not detected but read-through past the pBR322 poly(A) site was considerable (Fig. 2A). If read-through can be interpreted as a measure of poly(A) signal strength, the pBR322 signal appears to be a much weaker signal than the SV40 late signal. This is consistent with the observation that the SV40(L) poly(A) signal appears to completely out-compete the pBR322 poly(A) signal. The lack of inhibition in trans indicates that the two poly(A) signals are not competing for factors that are limiting in concentration. Thus, the exclusive utilization of the SV40(L) poly(A) site when present as far as 2 kilobases downstream may be symptomatic of the process by which poly(A) sites on a given transcript are selected.

When duplicate poly(A) signals are placed in tandem, the one located upstream may be selected exclusively (15,16) or both may be selected approximately equally (17, 18). However, no examples of exclusive selection of the downstream copy of a duplicated signal has been reported. Many genes have transcripts that are polyadenylated at multiple sites (3,19-23), indicating that neither the site that is furthest upstream nor the one that has the highest efficiency is necessarily selected exclusively. The exclusive selection of downstream poly(A) sites has been observed previously in regulated systems in which alternative poly(A) signals are selected in a developmental or tissue-specific manner (24-26). These cases are complex, however, because one of the alternative poly(A) signals is in competition with an intron, and cell-specific trans-acting factors appear to be involved in the selection

process (27). Exclusive selection of a downstream poly(A) site in a non-regulated cell system has also been observed previously. When the SV40 early poly(A) signal was placed downstream from the rat preproinsulin poly(A) signals and tested in COS cells, the SV40 early site was selected exclusively (16). In this case, and in the case described in this report, the lack of detection of even a small amount of processing at the upstream site is puzzling. We suspect that it relates to a particularly large difference in efficiency between the two competing poly(A) signals.

The lack of utilization of the pBR322 polyadenylation site in pSV2-neo transcripts suggests that pre-mRNAs can be cleaved and polyadenylated only once. Otherwise, the RNAs cleaved at the SV40 late site should get cleaved again and polyadenylated at the upstream site, even if cleavage at the second site is less efficient. The mechanism by which this restriction occurs is not known. One possibility is that cleavage or the presence of a poly(A) tail, perhaps via the poly(A) binding protein, may sufficiently alter the structure of the pre-mRNA or the ribonucleoprotein complex to prevent other poly(A) signals from being recognized. A second possibility is that polyadenylated RNAs may be rapidly moved to a location, perhaps even within the nucleus, that lacks factors required for further cleavage/polyadenylation. Experiments on the selection of poly(A) sites in vitro should reveal insights into the validity of some of these possibilities.

The poly(A) signal described in this report is the second poly(A) signal that has been found in the pBR322 portion of pSV2-neo. The other site is located at nucleotide 200 ± 2 (28-30). It functions in early orientation transcripts, but is detectable only when the upstream SV40(E) poly(A) signal has been deleted or altered. Polyadenylation at this site is relatively inefficient, occurring downstream from a cluster of variant hexanucleotides. The presence of multiple cryptic polyadenylation sites in pBR322 sequence suggests that the structural requirements for polyadenylation are highly flexible.

ACKNOWLEDGEMENTS

We are indebted to George W. Stearns for his excellent technical assistance. This work was supported in part by the National Institute of General Medical Sciences (Grant GM 33410), the Robert A. Welch Foundation (Grant A-903) and the Texas Agricultural Experiment Station (Hatch Project H-6521).

*To whom correspondence should be addressed

REFERENCES

1. Kessler, M.M., Beckendorf, R.C., Westhafer, M.A. and Nordstrom, J.L. (1986) *Nucleic Acids Res.* 14, 4939-4952.
2. Southern, P.J. and Berg, P. (1982) *J. Molec. Appl. Genet.* 1, 327-341.
3. Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell. Biol.* 2, 1044-1051.
4. Sutcliffe, J.G. (1978). *Cold Spring Harbor Symp. Quant. Biol.* 43, 77-90.
5. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 113-127, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
6. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
7. Gluzman, Y. (1981) *Cell* 23, 175-182.
8. Luthman, H. and Magnusson, G. (1983) *Nucl. Acids Res.* 11, 1295-1308.
9. Nordstrom, J.L., Hall, S.L. and Kessler, M.M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1094-1098.
10. Nevins, J.R. (1980) *Methods Enzymol.* 65, 768-785.
11. McDevitt, M.A., Imperiale, M.J., Ali, H. and Nevins, J.R. (1984) *Cell* 37, 993-999.
12. Hart, R.P., McDevitt, M.A., Ali, H. and Nevins, J.R. (1985) *Mol. Cell. Biol.* 5, 2975-2983.
13. Conway, L., and Wickens, M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 3949-3953.
14. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) *Cell* 41, 349-359.
15. Gil, A. and Proudfoot, N.J. (1984) *Nature* 312, 473-474.
16. Horowitz, M., Cepko, C.L. and Sharp, P.A. (1982) in *Eukaryotic Viral Vectors*, Y. Gluzman, Ed., pp. 47-54, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. Fitzgerald, M. and Shenk, T. (1981) *Cell* 23, 251-260.
18. Lanoix, J., Tseng, R.W. and Acheson, N.H. (1986) *J. Virol.* 58, 733-742.
19. Setzer, D.R., McGrogan, M. and Schimke, R.T. (1982) *J. Biol. Chem.* 257, 5143-5147.
20. Tosi, M., Young, R.A., Hagenbuchle, O. and Shiber, U. (1981) *Nucl. Acids. Res.* 9, 2313-2323.
21. Gerlinger, P., Krust, A., LeMeur, M., Perrin, F., Cochet, M., Gannon, F., Dupret, D. and Chambon, P. (1982) *J. Mol. Biol.* 162, 345-364.
22. Parnes, J.R., Robinson, R.R. and Seidman, J.G. (1983) *Nature* 302, 449-452.
23. Reynolds, G.A., Basu, S.K., Osborne, R.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984) *Cell* 38, 275-285.
24. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) *Cell* 20, 313-319.
25. Rosenfeld, M.G., Mermod, J.-J., Amara, S.G., Swanson, L.W., Sawchenko, P.E., Rivier, J., Vale, W.W. and Evans, R.M. (1983) *Nature* 304, 129-135.
26. Capetanaki, Y.G. Nagai, J., Flytzanis, C.N. and Lazarides, E. (1983) *Cell* 35, 411-420.
27. Nelson, K.J., Hsiniomich, J. and Perry, R.I. (1983) *Mol. Cell. Biol.* 3, 1317-1332.
28. Woychik, R.P., Lyons, R.H., Post, L. and Rottman, R.M. (1984) *Proc. Natl. Acad. Sci. USA* 81, 3944-3948.
29. Cole, C.N. and Stacy, T.P. (1985) *Mol. Cell. Biol.* 5, 2104-2113.
30. Nordstrom, J.L. and Westhafer, M.A. (1986) *Biochim. Biophys. Acta* 867, 152-162.