RNA polymerase II terminates transcription *in vitro* in the SV40 origin region

David S.Grass, Richard Jove and James L.Manley

Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Received March 6, 1987; Revised and Accepted May 12, 1987

Accession no. Y00341

ABSTRACT

To begin to study signals on DNA that can cause mammalian RNA polymerase II to terminate transcription, we examined the RNA transcripts produced from a number of different DNA templates in a HeLa whole-cell extract. When transcripts initiating from the strong adenovirus late promoter and extending through the SV40 promoter-replication origin region were analyzed, it was observed that a significant fraction (~75%) were terminated within these SV40 sequences. These transcripts, the 3' ends of which were mapped to several sites within an approximately 200 base pair region, appeared to result from transcription termination, as judged by kinetic and pulse-chase experiments. The possible significance of these findings with respect to SV40 gene expression in particular and transcription termination in general are discussed.

INTRODUCTION

Transcription termination by *E. coli* RNA polymerase has been well studied (1-3). Two different mechanisms, ρ -independent and ρ -dependent, have been defined. ρ -independent terminators are characterized by a G-C rich stem-loop structure followed by a string of T residues on the sense strand, where termination occurs. In contrast, RNAs that are terminated in a ρ -dependent manner do not appear to share any common sequence element that is important for termination, and are characterized principally by their requirement for the protein factor ρ in the termination reaction (3). Furthermore, ρ -dependent termination can require upstream sequences as well as sequences at the termination site (3). In addition to forming the 3' ends of RNA molecules, termination is an important regulatory mechanism in bacteria, as exemplified by attenuation in the trp operon (2).

In eukaryotes the situation is more complex, in part because each of the three nuclear RNA polymerases terminates transcription in a different manner. Termination by RNA polymerase III may be the simplest, occurring within clusters of four or more T's, often surrounded by G-C rich sequences (4-5). Termination by RNAP III appears not to require any factors in addition to the

polymerase (5). Termination of RNA polymerase I transcription also appears to require clusters of T residues, but additional sequences are also necessary (6).

Much less is known about RNA polymerase II termination. Specific sequences that bring about termination have not been identified, nor have specific sites of termination (7). This may reflect the immense size of some RNAP II transcription units, which can extend for more than 100 kb. Furthermore, the 3' ends of most, if not all, mRNAs are formed not by transcription termination, but rather by an RNA processing event that entails an endonucleolytic cleavage of the pre mRNA (7). The fact that sequences downstream are apparently unstable has further complicated the search for transcription termination sites. However, despite these difficulties, several "regions" in which transcription of certain cellular genes appears to terminate have been identified (8-19). These studies, which involved primarily pulse-labeling of transcripts *in vivo* and/or *in vitro* nuclear run-on experiments, indicated that transcription terminates hundreds to even thousands of nucleotides downstream of the mRNA polyadenylation site. In addition, situations possibly related to the bacterial attenuation system have been proposed to exist in the late transcription units of the DNA tumor viruses SV40 (20) and adenovirus (21, 22).

The SV40 genome consists of early and late regions which are the templates for the viral mRNAs. Both the early and late promoters, as well as the origin of replication, are situated within a 300 bp region of the viral chromosome. The sites of transcription initiation and RNA processing have been well characterized (23-25). However, it has not been determined where either early or late SV40 transcription terminates. It is likely that termination does occur in this system, though, because greater than genome length transcripts of the circular DNA have only been rarely detected (eg., ref. 23). In fact, it has been speculated that SV40 early transcription might terminate in the origin region (24), thereby giving rise to approximately genome-length transcripts. In contrast, in cells infected with the related virus, polyoma, multiple genome length late (26), but not early (27), transcripts are readily detected, suggesting that polyoma late transcription does not terminate efficiently.

We previously constructed mutants of SV40 containing a fragment of adenovirus DNA encompassing the major late promoter (Ad2MLP) inserted in the SV40 late promoter region (28). Here we describe some of the properties of these DNAs when they are used as templates for transcription in HeLa whole-

cell extracts. Most notably, we show that when the Ad2MLP directs transcription of SV40 early-strand RNA, a significant fraction of transcripts initiating from the Ad2MLP terminate at several sites within the SV40 origin region.

MATERIALS AND METHODS

Plasmids

pSVAdE and pSVAdL were constructed as described previously (28) by inserting a fragment of adenovirus DNA containing the Ad2MLP from -66 to +33, with respect to the transcription start site, into the KpnI restriction site (SV40 nucleotide number 294; see ref. 23) of a plasmid containing the entire SV40 genome, pSVRI. pSVAddE was constructed in a similar manner, but contains two copies of the Ad2 fragment, both in the same orientation as the inserted fragment in pSVAdE. Figure 1 shows the structures of all plasmids used in this study.

R30 and R38 were constructed by inserting a promoter-origin containing fragment of SV40 into the BstXI site (SV40 nt 4759) of p ϕ 4SVAdE (29). The origin spanning fragment between the HindIII site (nt 5171) and the EcoRI site (nt 294) in pSVAdL was isolated, treated with the Klenow fragment of *E. coli* DNA polymerase I to render the ends blunt, and then ligated into p ϕ 4SVAdE at the filled-in BstXI site.

pSP65SV0 was constructed by cleaving both pSVAdL and SP65 (30) with EcoRI and HindIII and ligating the origin spanning fragment (nt 5171 to 294) into the SP65 vector.

In vitro transcription and RNA analysis

Whole-cell extract was prepared from HeLa cells as described previously (31, 32). Transcription reactions, RNA isolation, and RNA glyoxalation and electrophoresis through agarose gels were also performed by previously described methods (32, 33), except that treatment with DNase I was usually omitted. For pulse-chase analysis, reactions initially contained 5 μ M α -³²P GTP. To effect a chase, either the concentration of GTP was raised to 250 μ M with unlabeled GTP, or α -amanitin was added to the reaction mixtures at a concentration of 1 μ g/ml.

In vitro transcription with SP6 RNA polymerase was performed by a modification of the procedure used by Melton et al. (30). 1.25 μ g of linear DNA templates were transcribed in 40 mM Tris-Cl pH 7.5, 6 mM MgCl₂, 10 mM NaCl, 2 mM spermidine, 10 mM dithiothreitol, 1 U/ μ l RNasin, 100 μ M rNTPs, 1 mM 7^mGpppG (cap), and 5 units SP6 RNA polymerase at 37°C for 1 hour. Samples

were then treated with DNase I at a final concentration of 40 $\mu\text{g}/\mu\text{l}$ for 10 minutes at 37°C before phenol-chloroform extraction, ethanol precipitation in 1M ammonium acetate and washing with 80% ethanol.

Electrophoresis of RNA on 5% polyacrylamide-8 M urea gels was performed using DNA sequencing-type gels (34), with minor modifications. The gel running buffer was 0.045 M Tris-borate, pH 8.3, 1.25 mM EDTA (0.5 X TBE). RNA samples were resuspended in 8 M urea with marker dyes and heated to 75°C for 3 minutes before loading on the gel.

To determine the 3' ends of in vitro transcribed RNA, 5-10% of the RNA from an in vitro transcription reaction was hybridized to a HindIII-PvuII restriction fragment spanning SV40 nt 5171 to 270 and 3' end-labeled by incorporation of ^{32}P -deoxynucleotide triphosphates with T4 DNA polymerase. RNA from a DNase I-treated in vitro transcription reaction was mixed with this probe, precipitated with ethanol, and suspended in 15 μl of 0.4 M NaCl, 0.04 M PIPES pH 6.5, 1 mM EDTA, 80% formamide. The samples were heated to 75°C for 10 minutes and incubated at 49°C or 12 hours. S1 nuclease digestion (35) was at 625 U/ml at 0°C for 2 hours. Samples were then extracted with phenol-chloroform (1:1), precipitated with ethanol in 1 M ammonium acetate and washed with 80% ethanol prior to electrophoresis through a 5% polyacrylamide sequencing-type gel. To map the 5' ends of in vitro transcribed RNA, nuclease S1 analysis was performed as described previously (28), using a single-stranded BstNI fragment labeled at the 5' end (SV40 nt 358).

RESULTS

pSVAdE and pSVAdL each contain a fragment of Ad2 DNA encompassing the major late promoter from -66 to +33 with respect to the RNA initiation site, inserted into the late promoter region of SV40 (nt 294), but in opposite orientations (28). pSVAdDE is similar, except that it contains two copies of the Ad2 DNA fragment, both in the same orientation as the insert in pSVAdE. When pSVAdL and pSVAdE were introduced into permissive monkey cells, transcripts that had initiated not only at the adenovirus start site but also at all of the SV40 late start sites were detected (28). This finding suggested that the 100 base pair adenovirus DNA fragment inserted into the SV40 late promoter region did not influence the choice of SV40 late transcription start sites, which are located both upstream and downstream of the insertion.

Transcription initiation from adenovirus-SV40 recombinant plasmids

To study transcription from these chimeric promoters further, we decided to characterize the RNAs synthesized in vitro from these plasmids. Although the

Figure 1. **Plasmid constructions.** Structure of recombinant plasmids used, showing parent plasmids and position of inserts. (A) pSVRI, pSVAdE and pSVAdL. (B) R30 and R38. (C) SP65SVO. Thick lines denote adenoviral sequences and hatched lines denote SV40 sequences. Thin lines represent pBR322 sequences. SV40 early and late regions are labeled, as are approximate positions of SV40 early and late and adenovirus *in vitro* transcription start sites (arrows).

products of SV40 late transcription *in vitro* display much of the heterogeneity observed *in vivo* (eg. ref. 36), two transcripts predominate, one located upstream (nt 167) and one downstream (nt 325, the major *in vivo* start site) of the *Kpn*I restriction site at nt 294 (36-38). pSVRI and the three insertion mutants were linearized with *Bam*HI to product templates for run-off transcription and incubated in a HeLa whole-cell extract. The resulting transcription products were treated with glyoxal to denature the RNA and separated on a 1.2% agarose gel. The results show that, as expected, two predominant bands, representing 1.9 and 2.1 kb RNAs initiated from wild-type SV40 late promoter and elongated until the end of the molecule, could be

Figure 2. Characterization of in vitro transcripts from SV40-Ad2 recombinant plasmids. *Bam*HI-digested pSVRI (a), pSVAdE (b), pSVAddE (c), or pSVAdL (d) were transcribed in a HeLa whole-cell extract as described in Materials and Methods. RNA was isolated, glyoxalated and separated on a 1.2% agarose gel. Small lines point to RNAs initiating at SV40 early and late start sites and bold lines denote RNAs initiating at the inserted Ad2 major late initiation site. Lane M contains RNA size markers. Sizes are indicated in kilobases.

resolved (Figure 2, lane a). When 100 bp of Ad2 DNA was inserted between these two major SV40 late initiation sites, as in pSVAdE (lane b), the 1.9 kb SV40 late runoff transcript was unaffected, indicating that initiation downstream from the insert occurred at the same site. The 2.1 kb SV40 late runoff transcript, however, disappeared, and a new product 100 bp larger was detected, which is the expected result if transcription had initiated at the same position as in wild-type SV40. This data suggests that the insertion of 100 bp of Ad2 DNA had no effect on the utilization of SV40 start sites

Figure 3. Analysis of runoff transcripts produced from templates truncated at different sites. Runoff transcripts produced from pSVAdL (a, c, f), pSVAdE (b, d, g), or pSVRI (e) templates that had been digested with HindIII (a, b), BamHI (c, d, e) or EcoRI and PstI (f, g) were separated on a 1.2% agarose gel. (L) and (E) at the bottom of lanes represent pSVAdL and pSVAdE, respectively. The arrow points to runoff transcripts from the SV40 early promoter. Sizes of transcripts are indicated in kilobases. Lane M displays RNA size markers.

upstream or downstream from the site of insertion. When an additional 100 bp was introduced into this region (pSVAdE), the 2.1 kb runoff transcript was replaced by a transcript 200 nt larger, while the 1.9 kb runoff transcript was unaffected (Figure 2, lane c), indicating that this 200 bp insert also had no effect on the utilization of SV40 late initiation sites in vitro. Because SV40 DNA sequences upstream of the KpnI site are required for transcription initiation in vitro from the downstream 325 site (36, 39), these results indicate that insertion of as much as 200 bp of DNA between a required promoter element and transcription start site does not interfere with initiation. Whether this is because the upstream SV40 late promoter sequences function in a manner similar to an enhancer, and we have reproduced this phenomenon in vitro, or because the mechanism of initiation from the TATA box-

lacking SV40 late promoter is different from that utilized by other promoters (see ref. 40 for review), is unclear.

Efficiency of runoff transcript production from pSVAdE and pSVAdL

As expected, the Ad2MLPs in all three plasmids were very active. If the efficiency of transcription initiation from the Ad2MLP was independent of its orientation in these plasmids, the amount of ^{32}P -GTP incorporated into runoff transcripts would be predicted to be strictly a function of the size of the runoff transcript, because the fraction of guanosine residues is similar for both transcripts in either orientation (eg, ref. 41). However, during the process of characterizing transcription initiation from the Ad2MLP in pSVAdE and pSVAdL templates, it became apparent that significantly more runoff transcripts were produced from the Ad2MLP in pSVAdL than in pSVAdE (eg. Fig. 2). This can be observed regardless of the size of the runoff RNAs. For example, Figure 3 shows the results of agarose gel electrophoresis of runoff transcripts synthesized and glyoxalated as above from both pSVAdE and pSVAdL templates that had been digested with HindIII (lanes a,b), BamHI (lanes c,d,e) or EcoRI and PstI (lanes f,g). Transcription from the Ad2MLP in pSVAdL incorporated considerably more radioactivity into runoff transcripts than did transcription from the Ad2MLP in pSVAdE. That this effect was not due to differences in DNA concentration, RNA recovery, etc. can be deduced by the observation that the transcripts produced from the SV40 early promoter in both plasmids (Figure 3, lanes f and g) are present in equal amounts. By comparing the intensities of the corresponding run-off transcripts and correcting for differences in the sizes of the RNAs, we estimate that 3-4 fold more run-off transcripts were produced from the Ad2MLP in pSVAdL than in pSVAdE.

Efficiency of Ad2MLP utilization in pSVAdE and pSVAdL

The above results do not provide an explanation for the unequal accumulation of run-off RNAs transcribed from pSVAdE and pSVAdL. Because cleavage with EcoRI removes sequences upstream of the Ad2 DNA inserts in pSVAdL and pSVAdE, the possibility that SV40 upstream sequences influenced the efficiency of initiation could be discounted. To determine directly whether initiation of transcription from the Ad2MLP occurred with the same efficiency in pSVAdE and pSVAdL, S1 nuclease analysis of the 5' ends of the RNA synthesized in vitro from HindIII digested templates was performed using homologous single-stranded BstNI probes labeled at their 5' ends to the same specific activities. The lengths of the DNA fragments protected by RNA transcribed from SVAdL and pSVAdE templates should be 93 and 91 nt,

Figure 4. 5' end analysis of in vitro transcripts. RNA transcribed in whole-cell extract from HindIII-digested pSVAdL (L) or pSVAdE (E) was hybridized to homologous single-stranded BstNI fragments covering SV40 nt 232 to 358, labeled at the 5' end at nt 358 for pSVAdL and 5' end at nt 232 for pSVAdE. S1 resistant DNA fragments were resolved on a 5% polyacrylamide-8 M urea gel. Lanes (a) contained products from the indicated transcription reaction, while lanes (b) contained no RNA in hybridization reactions. Marker (M) is a HpaII digest of pBR322.

respectively. Figure 4 shows that fragments of the expected sizes could be detected, and suggests that, if anything, somewhat more Ad2MLP-specific RNA was initiated from pSVAdE than from pSVAdL. However, quantitation of these results, as well as of additional similar experiments (not shown), indicate that there was no significant difference in the level of transcription initiation from these two promoters. Thus it appears that the difference in accumulation of runoff transcripts was the result of an event occurring after transcription initiation.

Transcripts smaller than runoff size are synthesized from pSVAdE template.

Using templates digested with HindIII, a number of bands representing RNAs

Figure 5. Small RNAs are synthesized from pSVAdE, but not pSVAdL. (A) Runoff RNAs transcribed from *Hind*III-digested pSVAdE (E) and pSVAdL (L) templates were separated on a 1.4% agarose gel. Thin lines point to RNAs smaller than runoff size transcribed from pSVAdE. Lanes M contain RNA size markers. (B) Transcripts produced from a circular pSVAdE template. RNA transcripts transcribed from *Hind*III-digested pSVAdE (a), pSVAdE (b), pSVAdL (c) and closed circular pSVAdE (d) were electrophoresed through a 1.4% agarose gel. Dark arrows point to the runoff transcript from the Ad2MLP. Open arrows point to small RNAs.

smaller than the Ad2MLP runoff transcripts could be detected when pSVAdE, but not pSVAdL, was transcribed. Figure 5A shows a 1.4% agarose gel of gloxalated RNAs transcribed from either *Hind*III digested pSVAdE or *Hind*III digested pSVAdL. The small transcripts (<350 nt) found below the runoff RNA initiated from the Ad2MLP in pSVAdE did not correspond to the expected size runoff RNAs from either the SV40 early or late promoters. Studies with the inhibitor α -amanatin, however, indicated that they were transcribed by RNA polymerase II (data not shown). Additional analyses of RNA transcribed from pSVAdE indicated that these small RNAs could be detected regardless of which

Figure 6. Analysis of accumulation and stability of small RNAs. A. Time course of runoff transcript and small RNA accumulation. *Hind*III digested pSVAdE template was incubated in whole-cell extract for 10, 20, 30, 40, 50 or 60 minutes before transcription products were separated on a 1.2% agarose gel as described in Materials and Methods. B. Chase with either α -amanatin or cold rNTPs for 30 (b) or 60 (c) minutes after pulse labeling for 15 minutes (a). Reactions were as described in Materials and Methods. Dark arrows point to the runoff transcript from the Ad2MLP in pSVAdE. Open arrows point to small RNAs.

restriction enzymes were used to cleave the template. Figure 5B shows that these transcripts can be detected even when a circular DNA template was transcribed. These data suggest that these RNA species were produced by transcription initiation by RNA polymerase II followed by either termination, pausing, or RNA processing/degradation. In addition, since no such RNAs have ever been observed with wild-type SV40 templates (data not shown), it is likely that these transcripts were produced by transcription initiation at the inserted Ad2MLP. Likewise, because such transcripts have never been detected from other Ad2MLP-containing plasmids (eg. pSVAdL, Figure 5A, lane L; Figure 5B, lane C; see also refs. 32, 41, 42), it appears that they are the result of SV40 sequences in pSVAdE lying downstream of the inserted Ad2MLP.

To investigate whether the small RNAs were generated by transcription termination or an RNA processing event, a time course experiment was performed to determine whether the runoff RNA from the Ad2MLP and the smaller RNAs accumulated with the same kinetics. *Hind*III digested pSVAdE template was

Figure 7. Polyacrylamide gel electrophoresis of in vitro transcripts. RNA transcription products were separated on a 5% polyacrylamide gel after transcription of HindIII-digested SP65SV0 with SP6 RNA polymerase (a), and undigested pSVAdE (b) or HindIII-digested pSVAdE (c) in whole cell extract. The markers are HpaII-digested pBR322 (1) and RNA size markers (2). Predominant RNAs with 3' ends in the origin region are indicated by dots or brackets.

incubated for various times in HeLa whole-cell extract, and the RNA products were resolved on a 1.2% agarose gel. Figure 6A shows that the runoff transcript and the smaller RNAs accumulated with essentially the same kinetics, supporting the notion that they arose from termination rather than post-transcriptional processing. A pulse-chase experiment, in which transcription reactions were incubated for 15 minutes and then chased for 30 or 60 minutes with either cold GTP or α -amanatin, was also performed. The

results (Figure 6B) are in agreement with the time course experiment, as no runoff transcript was chased into the smaller transcripts. In addition, the fact that RNA was not chased from the smaller species to a larger one argues against the possibility of transcriptional pausing without termination.

In order to separate the small transcripts more effectively and to determine the sites at which these RNAs were terminated, RNAs from transcription reactions that contained either circular (Figure 7, lane b) or Hind III digested (lane c) pSVAdE were resolved on a 5% polyacrylamide-8 M urea gel. A number of low molecular weight species are apparent, with predominant species between 355 and 200 nt.

3' end analysis of small RNAs synthesized from pSVAdE

To confirm that these small RNAs do have 3' termini in the SV40 origin region, nuclease S1 analysis was performed using a 3' end labelled DNA probe labelled at SV40 nt 270 (PvuII) and extending 342 nucleotides across the SV40 origin region to a HindIII site at SV40 nt 5171. When this probe was hybridized with pSVAdE transcribed RNA (from either HindIII digested or circular templates) and the DNA fragments protected from S1 nuclease digestion separated on a 5% polyacrylamide-8M urea gel, the resulting pattern of bands (Figure 8, lanes a and b) was strikingly similar to that observed when the RNA products themselves were resolved (Figure 7). These RNAs were not detected when nuclease S1 analysis was performed with RNA transcribed from pSVAdL (lane c). One predominant band, which could represent an RNA terminated 10-15 nt before the end of the DNA, was detected when the pSVAdE template was digested with HindIII (lane b), but was only a minor species when a circular template was transcribed (lane a). Because the major product detected when these RNAs were analyzed directly (Figure 7, lane c) was also about 10 to 15 nucleotides smaller than expected, we conclude that the majority of RNA polymerase II molecules terminated well before reaching the end of this DNA template. Although we do not know the reason for this, it has not been commonly observed with other fragments transcribed in vitro (eg, ref. 42).

The sequence within the SV40 origin region contains three very A-T rich stretches at which S1 nuclease might cleave DNA-RNA hybrids artifactually. To provide a control for this, SV40 sequences between SV40 nt 294 and 5171 were inserted into the plasmid SP65 (30), so that RNA could be synthesized by the bacteriophage SP6 RNA polymerase, with the SP6 promoter directing transcription across the early strand. HindIII runoff transcripts were synthesized from this plasmid and gel purified to insure that only full-length

Figure 8. Analysis of the 3' ends of *in vitro* transcribed RNA. RNA transcribed from undigested pSVAdE (a), *Hind*III-digested pSVAdE (b), *Hind*III-digested pSVAdL (c), *Bam*HI-digested R30 (d) or *Bam*HI-digested R38 (e) was isolated and hybridized to a *Hind*III (SV40 nt 5171) to *Pvu*II (nt 270) fragment of SV40 DNA labeled at the 3' end (nt 270). RNA transcribed from *Hind*III-digested pSP65SVO by SP6 RNA polymerase was also analyzed, using the same probe, before (f) and after (g) gel purification of the RNA. Following S1 nuclease digestion, protected DNA fragments were resolved on a 5% polyacrylamide-8M urea gel. The marker (M) is a *Hpa*II digest of pBR322. Predominant DNA fragments representing RNAs with 3' ends in the origin region are indicated by dots or brackets.

RNA was present (Figure 7). When S1 analysis was performed using this RNA and the above mentioned 3' end labeled probe, essentially only full-length hybrids were detected (Figure 8, lane g).

In order to determine whether the region between the *Kpn*I site at nt 294 and the *Hind*III site at nt 5171 can cause termination of *in vitro* transcription in a different DNA sequence context, transcripts synthesized *in*

Figure 9. Diagram showing the nucleotide sequence of the SV40 promoter-origin region. SP1 binding sites and stretches of T residues are underlined. Numbers above diagram are SV40 coordinates (see ref. 20). Brackets above sequence indicate regions in which termination occurs. Direction of transcription is indicated by arrow.

vitro from the plasmids R30 and R38 were analyzed. These plasmids contain the SV40 sequences between nt 294 and 5171 inserted, in either orientation, into the BstXI restriction site of p ϕ SVAdE (see Figure 1), in which the Ad2MLP directs transcription of the SV40 early region. When Bam HI-digested R30 and R38 were incubated in a HeLa whole-cell extract and the resulting RNAs analyzed by S1 nuclease analysis similar to that performed above, transcripts terminated at the same sites as those synthesized from pSVAdE were detected from R30, although they were present in significantly lower quantities (Figure 8, lane d). Thus, the efficiency of transcription termination was reduced when the SV40 origin region was situated in a different DNA context.

DISCUSSION

In this paper, we have provided evidence that transcription termination of SV40 early RNAs can occur in vitro at several sites in the replication origin-promoter region of the viral genome. This conclusion is supported by several lines of evidence. First, the production of full-length runoff transcripts from the Ad2MLP in the plasmid pSVAdE was significantly lower than expected, although the efficiency of transcription initiation was not. Second, several small RNAs accumulated during transcription of pSVAdE but not other Ad2MLP-containing plasmids, and the 3' ends of these transcripts mapped in the SV40 origin region. Finally, these small RNAs appear to have arisen by transcription termination rather than by pausing, RNA processing or degradation. From a comparison of the relative amounts of runoff RNAs produced from pSVAdE and pSVAdL, we estimate that approximately 75% of the pSVAdE transcripts were terminated in the origin region.

The approximate sites at which transcription terminates in vitro are

indicated in Figure 9. It is apparent that termination does not occur at a unique sequence, but rather at many sites scattered over a region of approximately 200 base pairs. Although contrasting with termination sites identified for other polymerases, both prokaryotic and eukaryotic, which give rise to unique termination products, our findings are consistent with recent observations of others, who have studied transcription termination in vivo of other RNA polymerase II-transcribed genes and obtained evidence that termination occurs at multiple, heterogeneous sites. These include the mouse β -globin gene (8), the mouse α -amylase gene (13), and the sea urchin H2A histone gene (18).

To date, there is only limited information available on the nature of the sequences that might be important for RNA polymerase II transcription termination. Although no consensus sequence has emerged, several lines of evidence suggest that thymidine rich sequences may be important in the termination process. In yeast, termination in the adenine 8 and CYC1 genes appears to require an A-T rich sequence (43, 44). In higher eukaryotes, the termination signals of the human α -globin gene (19) and the sea urchin histone H2A gene (18) also appear to include A-T rich sequences, while an oligothymidylate sequence has been shown to be an element involved in transcription termination of the human gastrin gene, both in vivo (17) and in vitro (45).

Interestingly, there are three strings of four or more T residues on the early strand of SV40 DNA in origin region, one consisting of eight nucleotides (nt 21-28) directly upstream of the early promoter TATA box, and two shorter stretches further downstream (nt 5203-5198 and 5182-5185). Each of these T stretches is associated with a termination product and, taken together with the examples mentioned above, suggest that RNA polymerase II termination signals may include this factor common to terminators for other RNA polymerases. Also of potential significance are G-C rich sequences located in this region. Upstream of all three T stretches are six copies of the sequence CCGCCC, which is in fact the binding site for the transcription factor SP1 (46). A G-C rich stem-loop structure can also be drawn in this region (see Figure 9). Indeed, it was previously speculated (24) that transcription of SV40 early strand RNA might terminate in this region of the viral genome because of the presence of this stem-loop structure and the T stretch located immediately downstream of it, a sequence similar to those involved in ρ -independent termination in prokaryotes. Further support for the idea that these sequences may be important in signaling termination are previous

experiments which showed that RNA polymerase II contained in HeLa whole-cell extract will terminate transcription at a prokaryotic ρ -independent termination signal (47). These results point to possible similarities between prokaryotic and eukaryotic termination signals. Supporting this notion is the suggestion (43, 44) that transcription termination of some yeast genes may be similar to ρ factor-dependent prokaryotic termination, in which termination sites occur within A-T rich sequences in the absence of an obvious consensus sequence.

The difficulties encountered in identifying discrete termination products, and the apparent lack of consensus sequences, are consistent with the notion that the signals that bring about RNA polymerase II termination are complex. Two types of results add further support to this idea. First, our results have shown that when the position of the SV40 termination-inducing signals relative to the Ad2MLP was altered, as in the plasmid R30, the efficiency of termination was significantly decreased, by roughly a factor of five. Although the basis for this phenomenon is not clear, one possibility is that RNA secondary or tertiary structures can effect the efficiency of transcription termination.

A second reason to believe that RNA polymerase II termination signals are complex derives from experiments suggesting that the mRNA 3' end cleavage/polyadenylation reaction may be an essential step in RNA polymerase II termination in vivo. This possibility was suggested first by experiments with the mouse β -globin gene (11), and more recently, we (S. Connelly and J.L. Manley, manuscript in preparation) have shown that mutations that affect polyadenylation of SV40 early transcripts result in multiple rounds of transcription of circular plasmids as a result of a failure to terminate transcription. These results suggest a model for SV40 early transcription in which a first step is 3' end cleavage of the nascent pre-mRNA, which, in some fashion, signals the polymerase to terminate transcription. The second step would be the actual termination event, and based on the data presented here, we suggest that this occurs in the origin region of the viral genome. The in vitro termination detected here is apparently not dependent on polyadenylation. It may be that such processing is crucial only in vivo, perhaps reflecting the much larger size of the nascent transcript, or, alternatively, that the efficiency and specificity of termination in vitro would be enhanced if the RNAs were cleaved and polyadenylated at an upstream site. We have not yet shown conclusively that the termination products

detected in vitro are produced in vivo, although preliminary S1 nuclease analysis of nuclear RNA isolated from cells transfected with pSVAdE has revealed RNAs with 3' ends located in the origin region (data not shown).

Finally, we wish to point out recent experiments from several labs showing that the sequence required for RNA polymerase I termination is also crucial for transcription initiation (48-50). This sequence, located approximately 170-200 bp upstream of the rRNA transcription start site, terminates transcripts from upstream, tandemly repeated ribosomal RNA genes. Whether this sequence enhances transcription by preventing "read-in" transcription from the upstream gene, or whether the termination event itself enhances transcription, initiation, or whether a protein that binds to this sequence directly enhances transcription, is not yet clear. However, it is interesting to note the possible functional similarity between the tandemly repeated ribosomal RNA genes and the circular SV40 genome. Thus, it may be that the termination products we have detected in the origin-promoter region reflect an important step in controlling transcription initiation from the early promoter. Perhaps the sequences that bring about termination are also important for early transcription, or conversely, one or more of the protein factors that bind to the SV40 promoter (46) or enhancer (51) may also play a role in transcription termination.

ACKNOWLEDGEMENTS

We thank M. Kopcanski for excellent technical assistance and S. Connelly for comments on the manuscript. This work was supported by NIH grant GM28983.

REFERENCES

1. Rosenberg, M. and Court, D. (1979) *Ann. Rev. Genet.* 13, 319-353.
2. Platt, T. (1981) *Cell* 24, 10-23.
3. Holmes, W.M., Platt, T. and Rosenberg, M. (1983). *Cell* 32, 1029-1033.
4. Bogenhagen, D.F. and Brown, D.D. (1981) *Cell* 24, 261-270.
5. Cozzarelli, N.R., Gerrard, S.P., Schlissel, M., Brown, D.D. and Bogenhagen, D.F. (1983) *Cell* 34, 829-835.
6. Bakken, A., Morgan, G., Sollner-Webb, B., Roan, J., Busby, S. and Reeder, R.H. (1982) *Proc. Natl. Acad. Sci. USA* 79, 56-60.
7. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) *Cell* 41, 349-359.
8. Hofer, E. and Darnell Jr., J.E. (1981) *Cell* 23, 585-593.
9. Weintraub, H., Larsen, A. and Groudine, M. (1981) *Cell* 24, 333-344.
10. LeMeur, M.A., Galliot, B. and Gerlinger, P. (1984) *EMBO J.* 3, 2779-2786.
11. Falck-Pederson, E., Logan, J., Shenk, T. and Darnell Jr., J.E. (1985) *Cell* 40, 897-905.
12. Mather, E.L., Nelson, K.J., Haimovich, J. and Perry, R.P. (1984) *Cell* 36, 329-338.
13. Hagenbuchle, O., Wellauer, P.K., Cribbs, D.L. and Schibler, U. (1984) *Cell* 38, 737-744.

14. Sheffrey, M., Marks, P.A. and Rifkind, R.A. (1984) *J. Mol. Biol.* 172, 417-436.
15. Frayne, E.G., Leys, E.J., Crouse, G.F., Hook, A.G. and Kellems, R.E. (1984) *Mol. Cell. Biol.* 4, 2921-2924.
16. Amara, S.G., Evans, R.M. and Rosenfeld, M.G. (1984) *Mol. Cell. Biol.* 4, 2151-2160.
17. Sato, K., Ito, R., Baek, K.-H. and Agarwal, K. (1986) *Mol. Cell. Biol.* 6, 1032-1043.
18. Johnson, M.R., Norman, C., Reeve, M.S., Scully, J. and Proudfoot, N. (1986) *Mol. Cell. Biol.* 6, 4008-4018.
19. Whitelaw, E. and Proudfoot, N. (1986) *EMBO J.* 5, 2915-2922.
20. Hay, N., Skolnik-David, H. and Aloni, Y. (1982) *Cell* 29, 183-193.
21. Maderious, A. and Chen-Kiang, S. (1984) *Proc. Natl. Acad. Sci. USA* 81, 5931-5935.
22. Mok, M., Maderious, A. and Chen-Kiang, S. (1984) *Mol. Cell. Biol.* 4, 2031-2040.
23. Tooze, J. (1981) *DNA tumor viruses*. 2nd edition, revised (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
24. Lebowitz, P. and Weissman, S.M. (1979) *Curr. Topics in Microbiol.* 87, 44-172.
25. Fried, M. and Prives, C. (1986) In M. Botchan, T. Grodzicker, P.A. Sharp (eds.) *Cancer Cells* 4, 1-16.
26. Acheson, N.H., Buetti, E., Sherrer, K. and Weil, R. (1971) *Proc. Natl. Acad. Sci. USA* 68, 2231-2235.
27. Acheson, N.H. and Mievillie, F. (1978) *J. Virol.* 28, 885-899.
28. Grass, D.S. and Manley, J.L. (1986) *J. Virol.* 57, 129-137.
29. Lewis, E.D. and Manley, J.L. (1985) *Mol. Cell. Biol.* 5, 2433-2442.
30. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucl. Acids Res.* 12, 7035-7056.
31. Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gefter, M.L. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3855-3859.
32. Jove, R. and Manley, J.L. (1982) *Proc. Natl. Acad. Sci. USA* 79, 5842-5846.
33. McMaster, G.K. and Carmichael, G.C. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4835-4838.
34. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 449-561.
35. Berk, A.J. and Sharp, P.A. (1978) *Proc. Natl. Acad. Sci. USA* 75, 1274-1278.
36. Hansen, U. and Sharp, P.A. (1983) *EMBO J.* 2, 2293-2304.
37. Rio, D., Robbins, A., Myers, R., and Tjian, R. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5706-5710.
38. Handa, H., Kaufman, R., Manley, J., Gefter, M. and Sharp, P. (1981) *J. Biol. Chem.* 256, 478-482.
39. Brady, J., Radonovich, M., Thoren, M., Das, G. and Salzman, N.P. (1984) *Mol. Cell. Biol.* 4, 133-141.
40. Manley, J.L. (1983) *Prog. Nucleic Acid Res. Mol. Biol.* 30, 195-244.
41. Hu, S.-L., and Manley, J.L. (1981) *Proc. Natl. Acad. Sci. USA* 78, 820-824.
42. Jove, R. and Manley, J.L. (1984) *J. Biol. Chem.* 259, 8513-8521.
43. Henikoff, S. and Cohen, E.H. (1984) *Mol. Cell. Biol.* 4, 1515-1520.
44. Henikoff, S., Kelly, J.D. and Cohen, E.H. (1983) *Cell* 33, 607-614.
45. Baek, K.-H., Sato, K., Ito, R. and Agarwal, K. (1986) *Proc. Natl. Acad. Sci. USA* 83, 7623-7627.
46. Gidoni, D., Kadonaga, J., Barrera-Saldana, H., Takahashi, K., Chambon, P., and Tjian, R. (1985) *Science* 230, 511-516.
47. Hatfield, G.W., Sharp, J.A. and Rosenberg, M. (1983) *Mol. Cell. Biol.* 3, 1687-1693.

48. Henderson, S. and Sollner-Webb, B. (1986) Cell 47, 891-900.
49. Grummt, I., Kuhn, A., Bartsch, I. and Rosenbauer, H. (1986) Cell 47, 901-911.
50. McStay, B. and Reeder, R.H. (1986) Cell 47, 913-920.
51. Sassone-Corsi, P., Wildeman, A. and Chambon, P. (1985) Nature 313, 458-463.