JOURNAL OF VIROLOGY, Oct. 1984, p. 277–280
0022-538X/84/100277-04$02.00/0

Vol. 52, No. 1

NOTES

Mapping In Vivo Initiation Sites of RNA Transcription and Determining Their Relative Use

MARK KESSLER AND YOSEF ALONI*

Department of Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

Received 12 April 1984/Accepted 13 June 1984

Runoff transcripts were generated on viral transcriptional complexes cleaved with restriction enzymes and incubated in vitro with [α-^{32}P]UTP under pulse-chase conditions. As viral transcriptional complexes in vitro elongated the nascent RNA preinitiated in vivo, size analysis by gel electrophoresis of the runoff transcripts allowed identification of the in vivo initiation sites. Moreover, scanning the intensities of the runoff bands as they appeared in the autoradiogram of the gel allowed determination of the relative use of these sites. A model system in which the initiation sites of simian virus 40 late RNA were identified and their relative use determined is presented.

Mapping of initiation sites of RNA transcripts has been done primarily by S1 nuclease mapping (2) and primer extension (9, 10, 20, 21) and recently by a more simplified approach that uses runoff transcripts (16, 17, 25). Runoffs are generated by transcribing RNA initiated in vitro on a specifically truncated DNA template. By subtracting the length of the RNA, in nucleotides, from the residue number of the 3′ end of the DNA fragment, the residue number of the initiation site can be determined. Such cell-free systems (16, 17, 25) do not reproduce the in vivo activity of different promoters or a promoter which contains multiple initiation sites (4, 11, 15, 18). If RNA transcription is initiated in vitro on the simian virus 40 (SV40) DNA template, the relative use of the early and late initiation sites does not reflect that previously described (7, 10, 15, 20, 26). These discrepancies may be due to inappropriate concentrations or to the complete absence of factors involved in promoter recognition and regulation of transcription initiation (3, 5, 6, 12, 17, 19, 22, 27).

To overcome these problems, we developed a new runoff system in which RNA transcription is initiated in vivo. As a model system viral transcriptional complexes (VTC) isolated from SV40-infected cells were used. VTC are composed of viral DNA, RNA polymerase B, and in vivo preinitiated nascent RNA. The nascent transcripts are elongated and labeled in vitro by the addition of labeled ribonucleotides at a rate of about 20 nucleotides per min (23, 24). In vitro initiation does not occur in the VTC system. Sarkosyl, which is used to extract VTC, solubilizes at least 90% of the complexes which generate viral RNA in intact nuclei (8, 14). A description of how these runoffs are generated and an example of how they can be used to map and determine the relative in vivo use of multiple RNA initiation sites in SV40 are presented.

VTC were prepared from BSC-1 cells infected with SV40 (strain 776). At 48 h postinfection the cells were harvested, and the nuclei were isolated by treatment of the cells with 0.5% Nonidet P-40 (14). Nuclei were suspended in TC buffer [40 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid)–NaOH (pH 8.0), 5 mM KCl, 0.15 M $(NH_4)_2SO_4$] and treated with 0.3% Sarkosyl. Cellular chromatin was pelleted by centrifugation at 30,000 × *g* for 30 min at 2°C, and the supernatant containing VTC was collected. To remove Sarkosyl (an inhibitor of restriction enzymes), we loaded VTC onto 5 to 30% (wt/vol) sucrose gradients containing 50 mM NaCl, 50 mM Tris-hydrochloride (pH 7.9), and 5 mM 2-mercaptoethanol. The gradients were centrifuged in a Beckman SW41 rotor for 5 h at 37,000 rpm and 4°C. Gradients were fractionated, and the ^{3}H-labeled viral DNA peak fractions (21S) (14) were pooled.

The choice of restriction enzymes were based on the location and frequency with which they cut the template. VTC must be cut to completion as rapidly as possible to minimize the loss of endogenous RNA polymerase activity which occurs during digestion. If the template is not cut to completion, the long RNA transcribed from the uncut template can obscure the runoff bands. For digestion, $MgCl_2$ was added to 10 mM, and the NaCl concentration was adjusted to that recommended by the manufacturer (New England Biolabs). Enzyme (50 U) was added to 1 ml of VTC (5 to 10 μg of DNA). Digestion was carried out for 30 min at 28°C instead of 37°C. The endogenous polymerase was found to be more stably maintained at the lower temperature (data not shown).

VTC were pulse-labeled for 5 min at 28°C in the presence of 150 mM NaCl–2 mM $MnCl_2$–1.0 mM dithiothreitol–a 400 μM concentration each of ATP, CTP, and GTP–0.25 μM concentration of [α-^{32}P]UTP (400 Ci/mmol; Amersham Corp.). Unlabeled UTP was added to 400 μM for the chase, and transcription was allowed to proceed for another 60 min. Under such conditions, incorporation of labeled ribonucleotide is not dependent on the length of the RNA transcript. The RNA was extracted with phenol-chloroform at room temperature and collected by ethanol precipitation. The precipitate was suspended in TKM (50 mM Tris-hydrochloride [pH 6.7], 25 mM KCl, 2.5 mM $MgCl_2$) and digested with DNase (13). The digest was reextracted with phenol-chloroform and passed through Sephadex G-25. The RNA excluded in the void volume was precipitated twice with ethanol. It was then denatured in 90% formamide in TBE (90 mM Tris-borate [pH 8.3], 90 mM boric acid, 2 mM EDTA) for 1 min at 90°C and subjected to electrophoresis on a 6% polyacryl-

* Corresponding author.

amide (bis-acrylamide 1:19) gel containing 7 M urea in TBE.

The criteria for determining if a band observed on the gel is a runoff transcript are that it must not be present in the lane of the uncut template and that the length of the transcript must change equally with each change in length of the truncated DNA template. Three runoff transcripts were generated when *Eco*RV and *Hae*II were used to cleave VTC, and two were observed when *Hin*dIII was used (Fig. 1). The third band is only barely recognizable because of the reduced amount of radioactivity loaded onto that slot and the poor resolution of bands within that region of the gel. None of these bands was observed in the lane of the uncut template. The length of the RNA in each band was determined by plotting the distance of migration versus the length, in nucleotides, of the RNA (Fig. 2). It is evident from Fig. 1 and 2 that the shift in the position of the bands from one runoff reaction to the next corresponds directly to the change in location of the restriction enzyme site. Three reproducible initiation sites were located at residues 266 ± 5, 325 ± 5, and 346 ± 5 (Fig. 2, table insert).

The relative use of the three initiation sites was determined by scanning the gel (Fig. 3). As expected, the RNA generated from the uncut template did not contain distinct bands in the region scanned (Fig. 3A). The three peaks (Figs. 3B and C) correspond to the three bands observed in the lanes of the *Eco*RV and *Hae*II runoff reactions (Fig. 1). The peaks contain, from left to right, 24 ± 3, 66 ± 5, and 10 ± 2% of the total radioactivity. These percentages reflect the relative use of the initiation sites located at residues 266 ± 5, 325 ± 5, and 346 ± 5, respectively. The intensity of the band

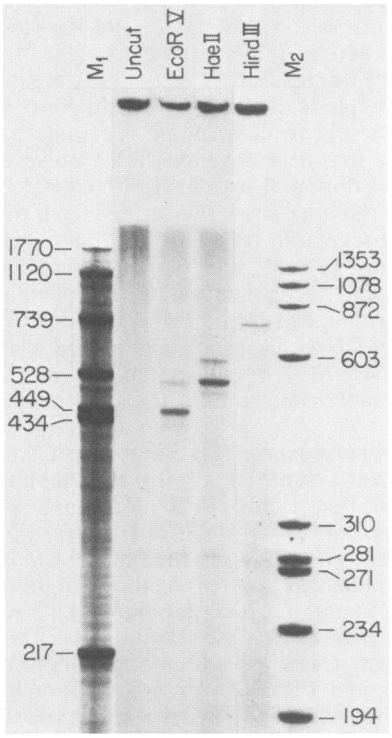
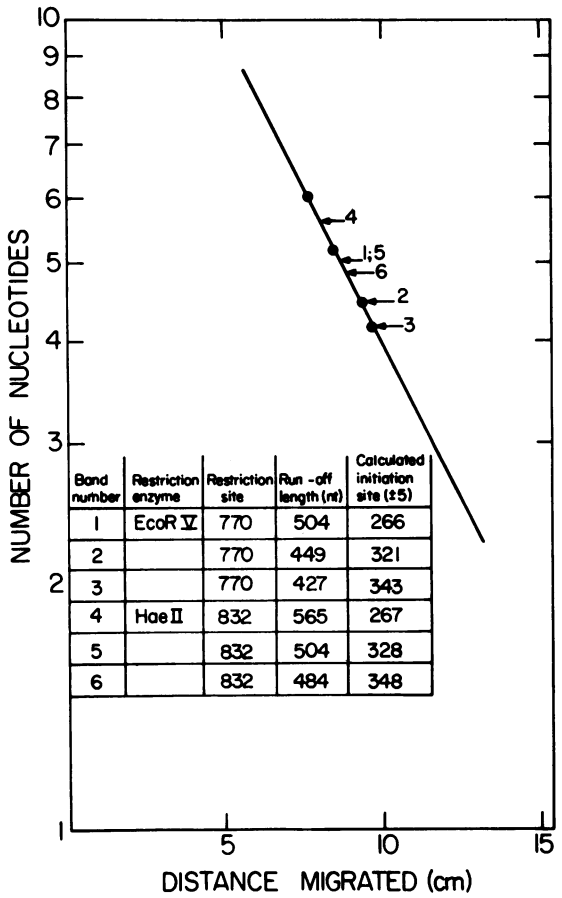


FIG. 1. Runoff transcripts synthesized on truncated VTC. VTC were extracted from SV40-infected cells by the Sarkosyl method (8, 14). The purified VTC were subjected to restriction enzyme digestion as indicated, followed by in vitro transcription. RNAs produced were purified and analyzed by polyacrylamide gel electrophoresis. The labeled RNAs were detected by autoradiography, M1 and M2 are labeled DNA markers.



Band number	Restriction enzyme	Restriction site	Run-off length (nt)	Calculated initiation site (±5)
1	EcoR V	770	504	266
2		770	449	321
3		770	427	343
4	Hae II	832	565	267
5		832	504	328
6		832	484	348

FIG. 2. Semilogarithmic plot relating the length in nucleotides of the runoff transcripts to their migration in the gel shown in Fig. 1. Markers of known nucleotide length are indicated by closed circles; runoff transcripts are indicated by arrows. The numbers adjacent to the arrows indicate the runoff transcript to which they refer (see table insert). The table insert presents the locations of the initiation sites calculated by subtracting the length, in nucleotides (nt) of the runoff transcript from the residue number of the restriction site on the DNA template strand.

corresponding to RNA initiated at residue 266 ± 5 in all of the runoff reactions was about 35% that of the band of the major initiation site.

The locations of the major initiation site at nucleotide 325 and the minor site at nucleotide 264 have been previously mapped by primer extension and S1 nuclease mapping (9, 10, 20, 21, 26). Variability with respect to other initiation sites and their use has been reported (9, 21). This variability may be due to several reasons. These include the time after infection that the RNA is extracted and the techniques used for RNA analysis. Both primer extension and S1 nuclease mapping utilize cytoplasmic and nuclear RNAs extracted from infected cells. These RNAs probably represent a population which is stably maintained within the cell and not one which is initially transcribed. This may explain why we observed, in our system, an RNA species that is initiated at nucleotide 346 ± 5 and comprises 10% of the total late RNA transcribed and that has not been previously identified.

This runoff system has several other advantages over the S1 nuclease mapping and primer extension techniques: both systems are more complex than the runoff system, both utilize indirect instead of direct means to map initiation sites,

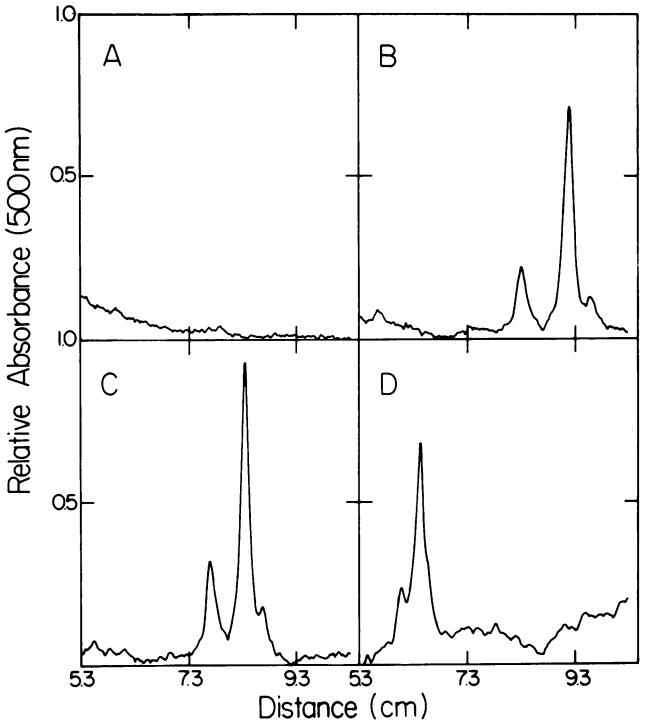


FIG. 3. Phototracing of the gel shown in Fig. 1. (A) Uncut template; (B) *Eco*RV-digested template; (C) *Hae*II-digested template; (D) *Hin*dIII-digested template. Phototracing was carried out on a Beckman DU-8 scanner.

and both are more prone to misidentification of initiation sites through artifacts. It has been previously noted that such artifacts can result from premature termination of reverse transcriptase during synthesis of cDNA, synthesis of cDNA from degraded RNA, or nonspecific cleavage of the DNA probe by S1 nuclease at points of instability along the DNA-RNA hybrid (such as at deoxyribosyladenine-ribouridylate bonds) (11, 20).

It is possible that the locations of the initiation sites determined by the runoff system presented here can be incorrectly mapped if large runoff transcripts are analyzed. It is also possible that the polymerase activity which is lost during extraction, purification, and digestion of the VTC by restriction enzymes is not random. The relative use of the different initiation sites reflected by the abundance of the different RNA species would therefore not be accurate, but we have no evidence to support this possibility.

The generation of runoffs by transcriptional complexes can be done with any extrachromosomal DNA system in which it is possible to extract the DNA template with active RNA polymerase. We recently used this technique to map the transcription initiation sites of the autonomous parvovirus minute virus of mice (1); it has also been used to map the transcription initiation sites of SV40 mutants (N. Hay, M. Kessler, and Y. Aloni, Virology, in press). This method thus provides a relatively simple means of screening active genes for their in vivo initiation sites and for determining their relative use.

We thank Ruchama Leiserowitz for technical assistance and Orna Resnekov for comments on the manuscript.

This research was supported by U.S. Public Health Service research grant CA 14995 and in part by the Leo and Julia Forcheimer Center for Molecular Genetics.

LITERATURE CITED

1. **Ben-Asher, E., and Y. Aloni.** 1984. Transcription of minute virus of mice, an autonomous parvovirus, may be regulated by attenuation. J. Virol. **52**:266–276.
2. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **75**:1274–1278.
3. **Brady, J., M. Radonovich, M. Thoren, G. Das, and N. P. Sulzman.** 1984. Simian virus 40 major late promoter: an upstream DNA sequence required for efficient in vitro transcription. Mol. Cell. Biol. **4**:133–141.
4. **Cepko, C. C., U. Hansen, H. Handa, and P. A. Sharp.** 1981. Sequential transcription-translation of simian virus 40 by using mammalian cell extracts. Mol. Cell. Biol. **1**:919–931.
5. **Davidson, B. L., J. M. Egly, E. R. Mulvihill, and P. Chambon.** 1983. Formation of stable preinitiation complexes between eukaryotic class transcription factors and promoter sequences. Nature (London) **301**:680–686.
6. **Dynan, W. S., and R. Tijin.** 1983. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell **32**:669–680.
7. **Fromm, M., and P. Berg.** 1982. Deletion mapping of DNA regions required for simian virus 40 early region promoter function *in vivo*. J. Mol. Appl. Genet. **1**:457–481.
8. **Gariglio, P., and S. Mousset.** 1975. Isolation and partial characterization of a nuclear RNA polymerase simian virus 40 DNA complex. FEBS Lett. **56**:149–155.
9. **Ghosh, P. K., M. Piatak, J. E. Mertz, S. M. Weissman, and P. Lebowitz.** 1982. Altered utilization of splice sites and 5′ termini in late RNAs produced by leader region mutants of simian virus 40. J. Virol. **44**:610–624.
10. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5′ terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
11. **Handa, H., R. J. Kaufman, J. L. Manley, M. Gefter, and P. A. Sharp.** 1981. Transcription of simian virus 40 DNA in a HeLa whole cell extract. J. Biol. Chem. **256**:478–482.
12. **Hansen, V., D. G. Tenen, D. M. Livingston, and P. A. Sharp.** 1981. T antigen repression of simian virus 40 early transcription from two promoters. Cell **27**:603–612.
13. **Hay, N., H. Skolnik-David, and Y. Aloni.** 1982. Attenuation in the control of simian virus 40 gene expression. Cell **29**:183–193.
14. **Laub, O., and Y. Aloni.** 1976. Transcription of simian virus 40. VI. Simian virus 40 RNA-DNA polymerase complex isolated from productively infected cells transcribed *in vitro*. Virology **75**:346–354.
15. **Lebowitz, P., and P. K. Ghosh.** 1982. Initiation and regulation of simian virus 40 early transcription in vitro. J. Virol. **41**:449–461.
16. **Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter.** 1980. DNA dependent transcription of adenovirus genes in a soluble whole cell extract. Proc. Natl. Acad. Sci. U.S.A. **77**:3855–3859.
17. **Matsui, T., J. Segall, P. A. Weil, and R. G. Roeder.** 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. **255**:11992–11996.
18. **Natarajan, V., M. J. Madden, and N. P. Salzman.** 1983. Preferential stimulation of transcription from simian virus 40 late and adeno virus IVa2 promoters in HeLa cell extract. J. Biol. Chem. **258**:14652–14655.
19. **Parker, C. S., and J. Topol.** 1984. A Drosophila RNA polymerase II transcription factor contains a promoter region-specific DNA binding activity. Cell **36**:357–369.
20. **Piatak, M., P. K. Ghosh, L. C. Norkin, and S. M. Weissman.** 1983. Sequences locating the 5′ ends of the major simian virus 40 late mRNA forms. J. Virol. **48**:503–520.
21. **Piatak, M., K. N. Subramanian, P. Roy, and S. M. Weissman.** 1981. Late messenger RNA production by viable simian virus 40 mutants with deletions in the leader regions. J. Mol. Biol. **153**:589–618.
22. **Samuels, M., A. Fire, and P. A. Sharp.** 1982. Separation and characterization of factors mediating accurate transcription by RNA polymerase II. J. Biol. Chem. **257**:14419–14427.

23. **Shani, M., E. Birkenmeier, E. May, and N. Salzman.** 1977. Properties of simian virus 40 transcriptional intermediates isolated from nuclei of permissive cells. J. Virol. **23**:20–28.
24. **Skolnik-David, H., and Y. Aloni.** 1983. Pausing of RNA polymerase molecules during *in vivo* transcription of the SV40 leader region. EMBO J. **2**:179–184.
25. **Weil, P. A., D. S. Luse, D. Segall, and R. G. Roeder.** 1979. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on RNA polymerase II and DNA. Cell **18**:469–484.
26. **Weissman, S. M.** 1980. Relationship between the simian virus 40 genome and viral mRNA. Mol. Cell. Biochem. **25**:29–38.
27. **Wesylyk, B., C. Wesylyk, H. Matthes, M. Wintzerith, and P. Chambon.** 1983. Transcription from the simian virus 40 early-early and late-early overlapping promoters in the absence of DNA replication. EMBO J. **2**:1605–1611.