Proc. Natl. Acad. Sci. USA
Vol. 77, No. 7, pp. 3898–3902, July 1980
Biochemistry

Expression of early genes of origin-defective mutants of simian virus 40

(deletion mutants/DNA sequence/S1 nuclease analysis/5′ end of early viral mRNA)

YAKOV GLUZMAN, JOSEPH F. SAMBROOK, AND RICHARD J. FRISQUE

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724

Communicated by J. D. Watson, April 9, 1980

ABSTRACT **The nucleotide sequences of eight origin-defective mutants of simian virus 40 have been determined. All of the mutants have suffered deletions, which range in size from 4 to 241 nucleotides. Some of the mutants induce the synthesis of tumor (T) antigen, others do not. Viral mRNA extracted from rat cells transformed by two of the T-antigen-positive mutants has been analyzed by the S1 nuclease technique of Berk and Sharp. Irrespective of the size or the location of the deletions, the 5′ ends of viral mRNAs are located approximately the same distance from the A+T-rich region (A-T-T-T-A-T) rather than at a specific site in the viral genome.**

The genomic organization and developmental cycle of simian virus 40 (SV40) are well understood (for review see ref. 1), and the complete nucleotide sequence of the viral DNA has been determined (2, 3). The region of the SV40 chromosome encompassing the *Bgl* I restriction endonuclease site is of particular interest because various control elements—the origin of viral DNA replication, the 5′ ends of the early mRNAs, and the tumor (T) antigen binding sites—are localized there (Fig. 1). In an attempt to locate these control elements more precisely and to analyze possible interactions between them, we have isolated a series of deletion mutants in this region. The method of isolation of these mutants and their biological properties have been described (10). Here we present their nucleotide sequences and describe the structure of viral mRNAs isolated from cells transformed by the mutants.

MATERIALS AND METHODS

Cells and DNAs. Transformed cells and the preparation of *Bgl* I-resistant SV40 DNA plasmids have been described (10).

DNA Sequence Determination. Wild-type SV40 DNA or mutant DNAs were digested with either *Hin*dIII or *Hin*f (Bethesda Research Laboratories, Rockville, MD). Total digests were dephosphorylated with bacterial alkaline phosphatase (Bethesda Research Laboratories) for 2 hr at 57°C and individual fragments were separated by gel electrophoresis. *Hin*dIII C or *Hin*f A fragments were purified from the gel and labeled at their 5′ termini by using phage T4 polynucleotide kinase (P-L Biochemicals) and [γ-^{32}P]ATP, 2000 Ci/mmol (Amersham) (1 Ci = 3.7×10^{10} becquerels). After digestion with *Hha* I, *Hpa* I, or *Kpn* I (Bethesda Research Laboratories) the appropriate labeled fragments were chemically cleaved and analyzed by the method of Maxam and Gilbert (11).

S1 Nuclease Analysis of the Early mRNAs. Total cytoplasmic RNA from transformed cells was prepared as described (12), and poly(A)-containing mRNAs were selected on an oligo (dT)-cellulose column (13). Unit-length wild-type or mutant 1-11 viral DNA was prepared by digesting the appropriate plasmid with *Bam*HI. The conditions for RNA·DNA hybridization and S1 treatment were described by Berk and Sharp (12).

RESULTS

Properties of Origin-Defective Mutants of SV40. A brief description of eight origin-defective mutants is presented in Table 1 (for further details see ref. 10).

DNA Sequence Determination of Origin-Defective Mutants. Restriction enzyme digestion of all mutant DNAs revealed patterns identical to those of wild-type viral DNA with the exception of those fragments encompassing the *Bgl* I site (10). Mutant DNAs with small deletions (Table 1) were digested with *Hin*dIII, and fragment C was isolated, labeled at its 5′ ends with ^{32}P, and then redigested with *Hha* I. Mutant DNAs with large deletions were digested with endonuclease *Hin*f, and fragment A was isolated, labeled, and recut with *Hpa* I or *Kpn* I. The appropriate fragments were subjected to sequence analysis by the procedure of Maxam and Gilbert (11). Autoradiograms of representative gels of four different mutants and the nucleotide sequences of these mutant DNAs are shown in Fig. 2 *Top*.

The mutants could be divided into three groups, according to the size of the deletions. The first group consists of the mutants 8-4, 8-16 (which is identical to 8-4), 6-1, and 6-17, which have suffered deletions of 4, 6, or 9 nucleotides, respectively. Two identical mutants, 1-11 and 3-20, form the second group and have a larger deletion (58 base pairs). Both the first and second groups of mutants induce T antigen. The third group consists of two identical mutants, 8-11 and 8-12, which have suffered the largest deletion (241 base pairs) and no longer induce detectable T antigen. The structures of all the mutants are presented in Fig. 2 *Bottom*.

The small deletions are all localized at the *Bgl* I site and presumably were produced by nuclease activity of either S1 or DNA polymerase I upon the cohesive ends left by digestion with *Bgl* I. The boundaries of the large deletions generally are positioned far from this restriction site, although 1-11 and 3-20 have one boundary close to this site. A possible mechanism for formation of the large deletions is homologous recombination, and it may therefore be significant that wild-type SV40 DNA contains a direct repeat of 8 nucleotides at positions 5214–5221 and 212–219. Mutant 8-12 has lost one of these repeats and the intervening 233 nucleotides, while the remaining repeat is located at the deletion junction. A similar situation could explain the deletion in mutant 1-11, in which 10 out of 11 nucleotides are shared by the repeats (positions 5186–5196 and 1–11) which are, in this case, separated by 47 nucleotides.

S1 Analysis of Viral mRNA Isolated from Transformed Cells. To compare the locations of the 5′ ends of the mRNAs

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: SV40, simian virus 40; T antigen, tumor antigen; bp, base pair(s).

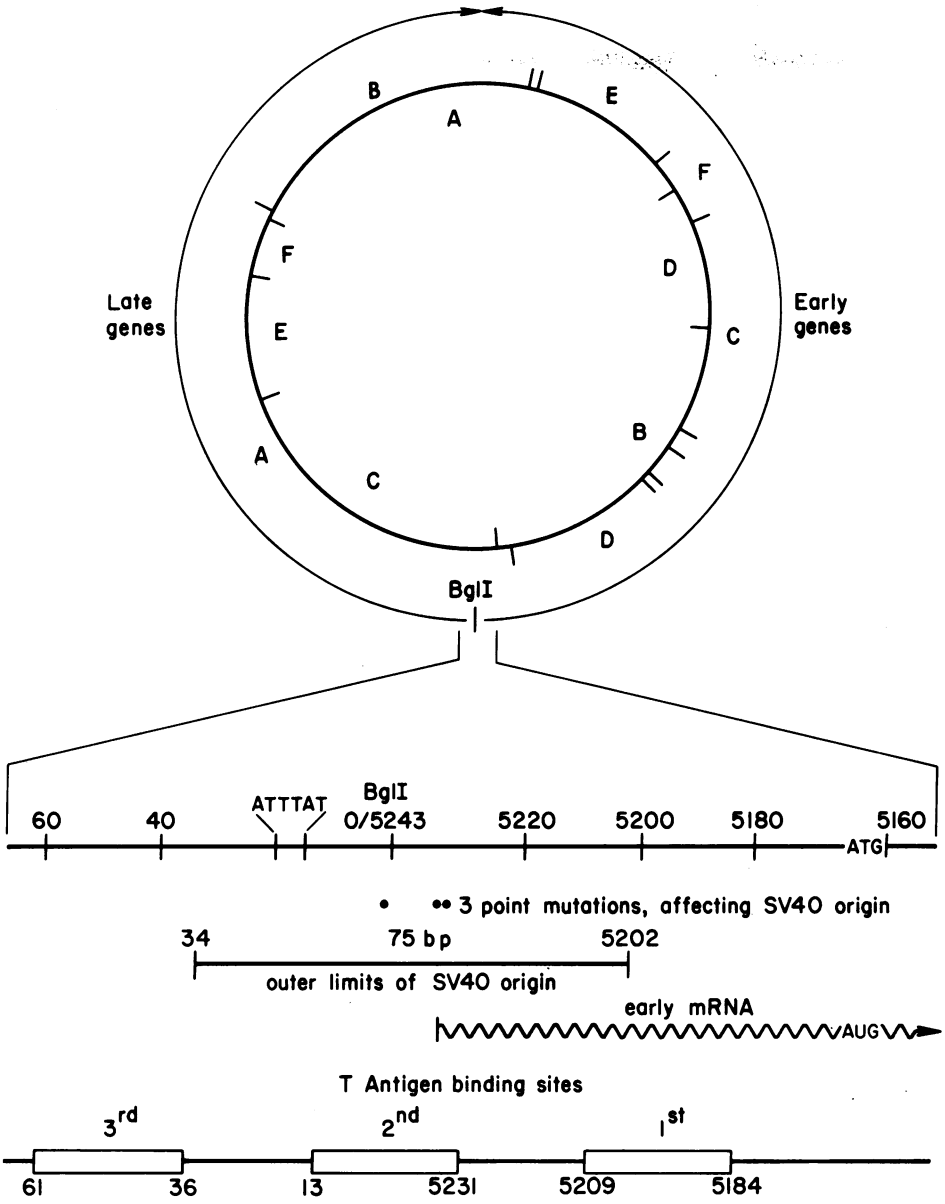

FIG. 1. Genetic organization of the region of viral DNA surrounding the origin of SV40 DNA replication. The sites of restriction endonucleases *Hin*dIII (inside) and *Hinf* (outside) are shown on a circular map of SV40. The circular arrows indicate the direction of early and late transcription and demarcate the junction between the early and late genomic segments. An enlargement of the region of DNA surrounding the *Bgl* I site includes the following features: A+T-rich region A-T-T-T-A-T (for references see ref. 4); the initiating AUG codon for both large and small T antigens (2, 3); the locations of single base pair mutations in three mutants that have altered origins of DNA replication (5); the smallest fragment of SV40 DNA known to contain the origin of replication as deduced from work with viable deletion mutants (6) and evolutionary variants of SV40 (7); the position of the longest stable cytoplasmic early mRNA as determined by reverse transcription of the early messages and sequencing of the cDNA (8); and the boundaries of three T-antigen-binding sites (9). Numbering of the nucleotides is according to the BBB system (1) and the complete nucleotide sequence of this region is presented in Fig. 2.

isolated from cells transformed by mutant (1–11, 6-1) or wild-type plasmids, the mRNAs from the three transformed cell lines were purified and analyzed by the S1 technique (12). These mRNAs were hybridized with wild-type or 1–11 DNA, and the resulting hybrids were treated with S1 nuclease, subjected to electrophoresis on neutral (not shown) or alkaline gels, transferred to nitrocellulose filters (14) and hybridized to ^{32}P-labeled SV40 DNA. An autoradiogram of one such alkaline gel is shown in Fig. 3*A*. The 1900-nucleotide fragments representing the 3′ end of the early transcripts are indistinguishable in size regardless of the source of mRNA or DNA used in the hybridization. However, the sizes of the fragments arising from the 5′ ends of the early messages varied according to the DNA and RNA used (Table 2). The lengths of the two small S1-resistant fragments resulting from the hybridization of wild-type RNA and wild-type DNA were in good agreement with the findings of Berk and Sharp (12) and Reddy *et al.* (8). Similar sized fragments were obtained when wild-type DNA was hybridized to 6-1 RNA. However, significant decreases (50–70 bp) in the size of the fragments were observed after S1 treatment of 1-11 RNA–wild-type DNA hybrids; the magnitude of these reductions correlates well with the size of the deletion in 1-11 DNA. When 1-11 RNA was hybridized with 1-11 DNA, protected fragments were indistinguishable in length from those found when 1-11 RNA was hybridized to wild-type DNA. These fragments were 10 nucleotides shorter than the corresponding fragments produced after hybridization of 1-11 DNA to either 6-1 or wild-type RNA. Because both wild-type and 6-1 RNAs cross the deletion junction of 1-11 DNA, the two pieces of protected DNA must represent the stretch of nucleotides between the downstream boundary of the deletion and the large and small T antigen splice points. Therefore the 5′ ends of 1-11 early mRNA were placed about 10 nucleotides downstream from the 1-11 deletion junction (see Fig. 3*C*).

Finally, greater amounts of viral mRNA were observed in cells transformed by the mutants than in cells transformed by the wild-type DNA. To quantitate relative amounts of viral RNA, densitometer tracings of the 1900-bp fragments (Fig. 3*B*, top) were made. The S1 experiment was carried out with excess DNA. In cells transformed by 6-1, a mutant that has lost only the second T-antigen-binding site, there is a 4-fold greater amount of viral mRNA compared to the wild-type counterpart,

Table 1. Properties of origin-defective mutants of SV40

Mutant	Mutagenesis procedure*	T-Ag†	DNA replication	Transformation	Size of deletion, bp
1-11	*Bgl* I	+	−	+	58
3-20	*Bgl* I/S1	+	−	+	58
6-1	*Bgl* I/S1/ligase	+	−	+	6
6-17	*Bgl* I/S1/ligase	+	−	+	9
8-4	*Bgl* I/Pol I/ ligase	+	−	+	4
8-16	*Bgl* I/Pol I/ ligase	+	−	+	4
8-11	*Bgl* I/Pol I/ ligase	−	NT	−	241
8-12	*Bgl* I/Pol I/ ligase	−	NT	−	241

bp, Base pairs; NT, not tested.

* The parental plasmid (pMK16, wild-type SV40) was linearized by digestion with *Bgl* I and either used directly to transform *Escherichia coli* χ1776 or first treated with one or more of the following enzymes: S1 nuclease, T4 ligase, *E. coli* DNA polymerase I.

† CV-1 monkey cells were transfected with mutant DNAs by using the DEAE-dextran technique, and T antigen was detected by indirect immunofluorescence. The mutant DNAs producing T antigen were shown to complement the growth of a temperature-sensitive *A* mutant at the nonpermissive temperature.

whereas 1-11 transformants (1-11 has deleted both the first and second T-antigen-binding sites) contain 7-fold more mRNA than wild type.

Analysis of the S1 digestion products on a neutral gel (data not shown) revealed that large T antigen message was more abundant than small tumor antigen (t) message, as previously reported (12). Stronger hybridization of the probe to the 620-bp fragment (originating from small t mRNA) compared to the 320-bp fragment (large T mRNA) on the alkaline gel (Fig. 3) does not reflect the relative amount of these messages but is due to the inefficient transfer of the smaller fragment to the nitrocellulose filter.

DISCUSSION

The origin of SV40 DNA replication has been located in a segment of viral DNA spanning the *Bgl* I site. Bidirectional DNA replication (15, 16) that is dependent on viral T antigen expression (17) is initiated here, and three T-antigen-binding sites have been assigned to this region (9).

Here we have described the nucleotide sequence of eight origin-defective mutants and the structure of viral mRNAs synthesized in rat cells transformed by two of these mutants. All of these mutants have lost nucleotides around the *Bgl* I site and can be divided into two categories: those that produce functional T antigen (1-11, 3-20, 6-1, 6-17, 8-4, 8-16) and those that do not (8-11, 8-12) (10). The loss of 4-9 nucleotides around the *Bgl* I site in mutants 6-1, 6-17, 8-4, and 8-16 destroys the origin of viral DNA replication (10) and affects the structure of the second T-antigen-binding site (Figs. 1 and 2). Mutants with more extensive deletions in this region (1-11 and 3-20, 58 nucleotides) lack the origin of DNA replication and both the first and the second T-antigen-binding sites. These findings are in good agreement with earlier results (6, 7), which have placed the outer limits of the viral origin of DNA replication in a 75-bp segment spanning the *Bgl* I site (Fig. 1). It is known that point mutations within the *Bgl* I site drastically affect the efficiency of DNA replication (ref. 5; Fig. 1) and a small deletion (<15 nucleotides) at one of two *Bgl* I sites in a viable variant of SV40 with two functional origins completely destroys that origin (18). We have shown that a deletion as small as 4 nucleotides at the *Bgl* I site (Fig. 2) abolishes the origin of DNA replication (10).

We have compared the structures of the viral mRNAs synthesized in cells transformed by either wild-type or mutant DNAs and have found that small fragments of 1-11 DNA protected by 1-11 mRNAs are about 10 nucleotides shorter than the corresponding fragments of 1-11 DNAs protected by wild-type or 6-1 mRNAs (Fig. 3, Table 2). It is likely that the splice points remain unchanged in the mutant mRNAs, because the T antigens induced by the mutants in transformed cells are indistinguishable in size from wild-type T antigen and are biologically functional (10). We are therefore able to place the 5′ end of 1-11 mRNAs about 10 nucleotides downstream from the deletion junction (Fig. 3C) or about 50 nucleotides downstream from the 5′ end of wild-type mRNA. However, the 5′ ends of both mutant and wild-type mRNAs are approximately the same distance (25 ± 5 bp) from the A+T-rich region shown in Figs. 1 and 2. To locate the 5′ ends of these transcripts precisely it will be necessary to carry out experiments similar to those of Reddy *et al.* (8), in which the early viral mRNAs of SV40 were transcribed into cDNA and their sequences were determined.

On the basis of the similarities between the nucleotide sequences near the beginning of several different genes it has been suggested that the promoters for RNA polymerase II are localized upstream from the 5′ end of their messages. A common feature is the presence of an A+T-rich region about 30 nucleotides from the 5′ end of the mRNAs (for example see ref. 8). Our findings are consistent with the idea that RNA polymerase II initiates transcription at a fixed distance from the A+T-rich region and that the downstream nucleotide sequences do not play a significant role in this process. Similar conclusions

Table 2. Sizes* of the small protected DNA fragments generated by S1 nuclease digestion of mutant and wild-type (wt) RNA·DNA hybrids

1-11 DNA			wt DNA		
wt RNA	1-11 RNA	6-1 RNA	6-1 RNA	1-11 RNA	wt RNA
560	550	560	610	550	620
285	275	285	325	275	325

* Lengths of the fragments are in bp as calculated from the autoradiogram shown in Fig. 3. All values are ±10 bp.

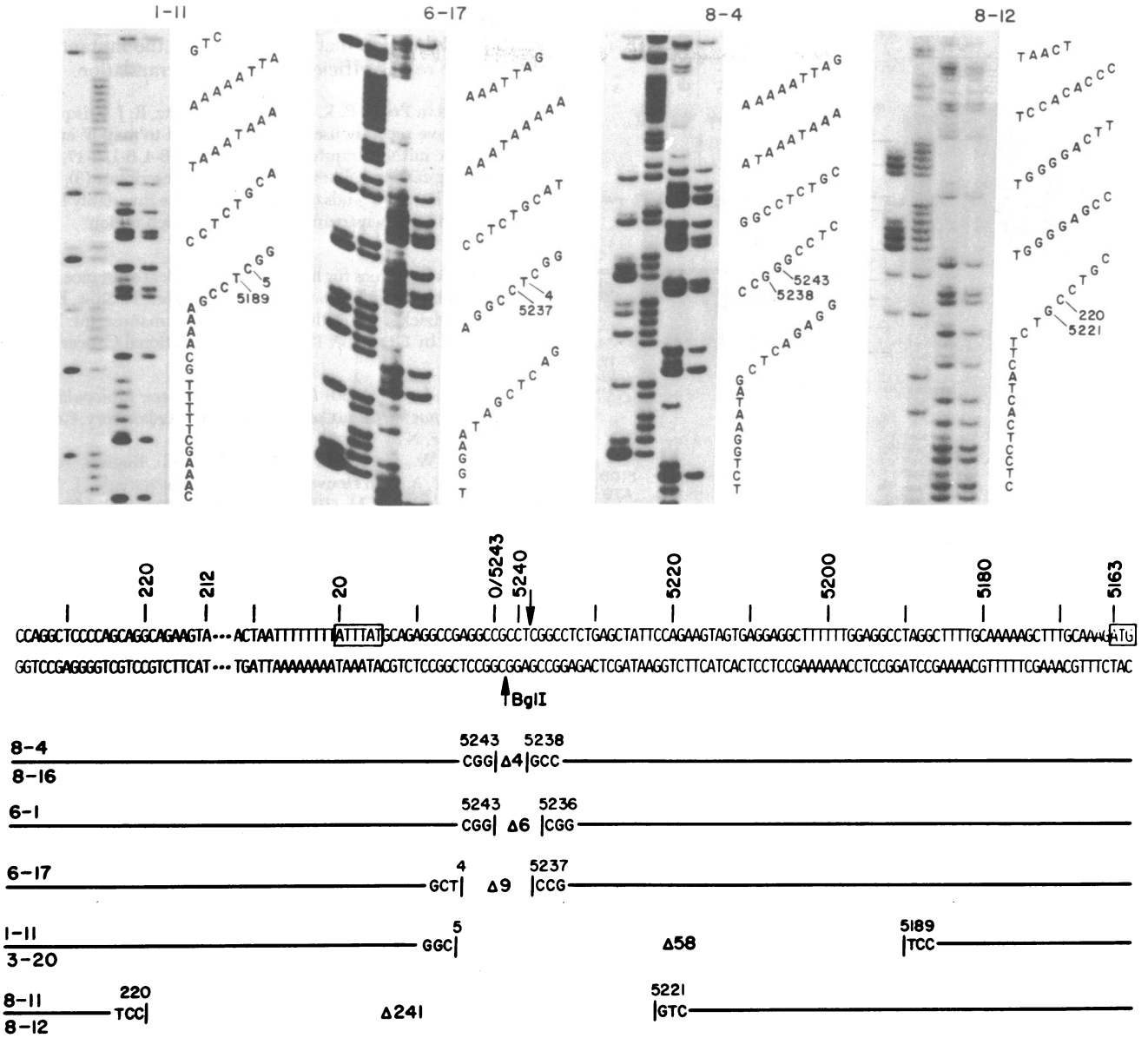

FIG. 2. The nucleotide sequences of origin-defective DNAs. (*Top*) The order of the tracts for the sequencing gels is G, A+G, C+T, and C (left to right). Photographs of the four representative gels were cut so as to place the deletion junctions near the center of the picture. Nucleotides bordering the junctions are labeled with their numbered position on the SV40 map. (*Middle*) The nucleotide sequence of wild-type SV40 DNA from nucleotide 236 to 5161 (excluding 211 to 33) and the ATG initiating codon and A+T-rich region (in boxes) are shown. (*Bottom*) The DNA structures of all of the mutants are shown diagrammatically, with boundaries and sizes of the deletions included.

have been reached for prokaryotic promoters (19, 20), for which it has been shown that nucleotide changes immediately downstream of the promoter do not affect transcription (21). Recently the structure of the promoter for eukaryotic RNA polymerase III has been described. Although it is localized downstream from the 5′ end of the 5S RNA, transcription always starts the same distance (55 bp) from the promoter (22).

One of the mutants, 8-12, is missing a 241-nucleotide segment of DNA that includes the A+T-rich region, the origin of DNA replication, the second and third T-antigen-binding sites, and a stretch of 13 nucleotides (5234–5222) that are the first 13 nucleotides of the wild-type early mRNA (Figs. 1 and 2). This mutant fails to induce T antigen in permissive monkey cells and does not transform rat cells (Table 1), even though 8-12 DNA contains the entire coding sequence for T antigen. The loss of 13 nucleotides, which are transcribed into wild-type mRNA, could not be responsible for this defect because mutant 1-11, which lost 45 nucleotides (5234–5190) from this region still produces T antigen and transforms rat cells. One possible explanation for the phenotype of 8-12 mutant is that the viral early promoter has been deleted and mRNA is not produced. If so, this would define the boundaries of the early promoter between nucleotides 5 and 212.

We have also found that the amount of cytoplasmic poly(A)$^+$-selected viral RNA was greater in cells transformed by mutant DNAs than in cells transformed by wild-type DNA. There was a 7-fold increase in the amount of viral RNA in cells transformed by 1-11 DNA (this mutant is missing both the first and second T-antigen-binding sites) and a 4-fold increase in the amount detected in cells transformed by 6-1 DNA (a mutant with an altered nucleotide sequence at the second T-antigen-binding site). It is possible that decreased T antigen binding to the mutant DNAs is responsible for the increased amount of

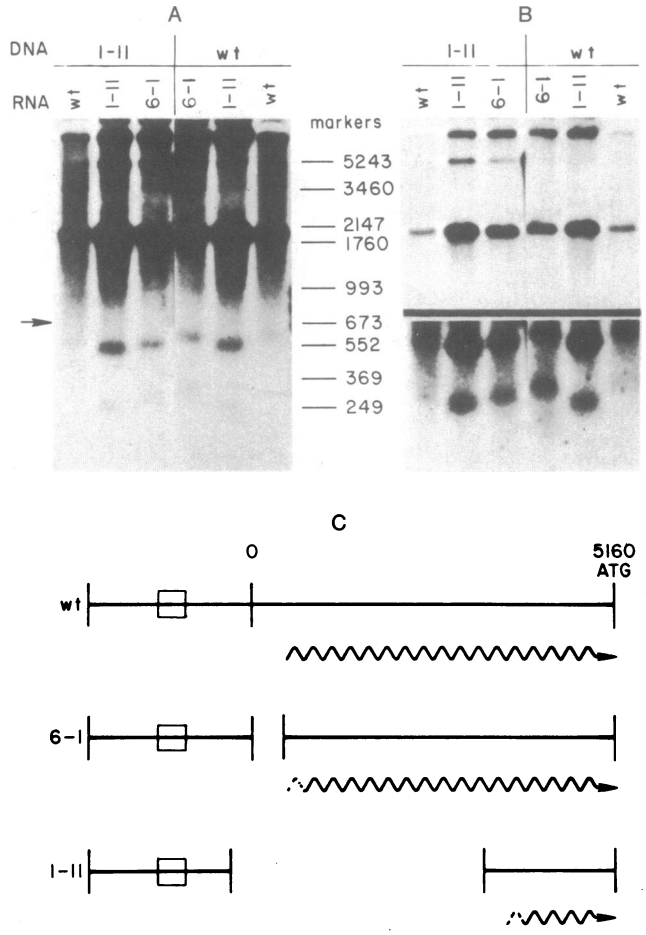

FIG. 3. S1 analysis of the mRNAs extracted from rat cells transformed by wild-type or mutant DNAs. Poly(A)-selected cytoplasmic RNAs (80 μg) were hybridized to unit-length (*Bam*HI cut) wild-type (wt) or 1-11 DNA and digested with S1 (12). DNA fragments resistant to the nuclease were separated by alkaline gel electrophoresis, transferred to nitrocellulose filters, hybridized to ^{32}P-labeled SV40 DNA, and autoradiographed for different lengths of time (intensifier screens used). (*A*) Entire gel exposed for 1 day; (*B*) top half of gel (above arrow) exposed 3 hr, bottom half (below arrow) for 4 days. (*C*) Diagrammatic representation of the structure of wild-type and mutant mRNAs. The solid lines represent the structures of the wild-type and mutant DNAs around the *Bgl* I site and the rectangles depict the A+T-rich sequence (Figs. 1 and 2). The wavy lines represent the stretch of viral mRNA between the 5′ end and the initiation codon.

viral mRNAs. Autoregulation of the synthesis of early mRNAs by T antigen was previously described for temperature-sensitive *A* mutants of SV40 and polyoma virus. Compared to the wild-type viruses, both temperature-sensitive mutants induce a 15-fold increase in the amount of early viral mRNA during a lytic infection at the nonpermissive temperature (23–26). In cells transformed by temperature-sensitive *A* mutants of SV40, only a 2 to 3-fold increase was observed (27, 28). However, other factors, such as changes in the stability of mutant messages, may be responsible for the increased amount of viral RNA in the cells transformed by origin-defective mutants.

Even though mRNA is significantly overproduced in mutant transformants, we have not observed a proportional increase in T antigen production (10). One possible explanation for this finding is that the mutant RNAs are translated less efficiently than their wild-type counterparts. It may therefore be that the untranslated region that is missing from the mutant mRNAs plays some role in efficient initiation of translation.

Note Added in Proof. P. K. Ghosh, P. Lebowitz, R. J. Frisque, and Y. Gluzman have recently used another method to map 5′ ends of the virus-specific mRNAs transformed by mutants 8-4, 6-1, 6-17, and 1-11. By analyzing cDNAs synthesized by reverse transcriptase (8), they have confirmed that the 5′ ends of these mRNAs are invariably located 25 ± 2 nucleotides downstream from the A+T-rich region.

We thank B. Ahrens for her excellent technical assistance, Shiu-Lok Hu for helping us with S1 analysis, and S. Hughes, N. Stow, B. Ozanne, and T. Grodzicker for critical reading of the manuscript. This work was funded by Grant CA 13106 from the National Cancer Institute.

1. Tooze, J., ed. (1980) *DNA Tumor Viruses: Molecular Biology of Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Part 2.
2. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. & Ysebaert, M. (1978) *Nature (London)* **273**, 113–120.
3. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. & Weissman, S. M. (1978) *Science* **200**, 494–502.
4. Baker, C. C., Herisse, J., Courtois, G., Galibert, F. & Ziff, E. (1979) *Cell* **18**, 569–580.
5. Shortle, D. & Nathans, D. (1979) *J. Mol. Biol.* **131**, 801–817.
6. Subramanian, K. N. & Shenk, T. (1978) *Nucleic Acids Res.* **5**, 3635–3642.
7. Gutai, M. W. & Nathans, D. (1978) *J. Mol. Biol.* **126**, 259–274.
8. Reddy, V. B., Ghosh, P. K., Lebowitz, P., Piatak, M. & Weissman, S. M. (1979) *J. Virol.* **30**, 279–296.
9. Tjian, R. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 655–662.
10. Gluzman, Y., Frisque, R. J. & Sambrook, J. (1980) *Cold Spring Harbor Symp. Quant. Biol.* **44**, 293–299.
11. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
12. Berk, A. J. & Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1274–1278.
13. Dunn, A. R. & Hassell, J. A. (1977) *Cell* **12**, 23–26.
14. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
15. Danna, K. J. & Nathans, D. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 3097–3100.
16. Fareed, G. C., Garon, C. F. & Salzman, N. P. (1972) *J. Virol.* **10**, 484–491.
17. Tegtmeyer, P. (1972) *J. Virol.* **10**, 591–598.
18. Shenk, T. (1978) *Cell* **13**, 791–798.
19. Johnston, H. M., Barnes, W. M., Chumley, F. G., Bossi, L. & Roth, J. R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 508–512.
20. Bennet, G. N., Brown, K. D. & Yanofsky, C. (1978) *J. Mol. Biol.* **121**, 139–152.
21. Brown, K. D., Bennet, G. N., Lee, F., Schweingruber, M. E. & Yanofsky, C. (1978) *J. Mol. Biol.* **121**, 153–177.
22. Sakonju, S., Bogenhagen, D. F. & Brown, D. D. (1980) *Cell* **19**, 13–25.
23. Tegtmeyer, P., Schwartz, M., Collins, J. K. & Rundell, K. (1975) *J. Virol.* **16**, 168–178.
24. Reed, S. I., Stark, G. R. & Alwine, J. C. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3083–3087.
25. Khoury, G. & May, E. (1977) *J. Virol.* **23**, 167–176.
26. Cogen, B. (1978) *Virology* **85**, 222–230.
27. Alwine, J. C., Reed, S. I. & Stark, G. R. (1977) *J. Virol.* **24**, 22–27.
28. Edwards, C. A. F., Khoury, G. & Martin, R. G. (1979) *J. Virol.* **29**, 753–762.