JOURNAL OF VIROLOGY, June 1979, p. 668-673
0022-538X/79/06-0668/06$02.00/0

Vol. 30, No. 3

Nucleotide Sequence Analysis of Viable Deletion Mutants Lacking Segments of the Simian Virus 40 Genome Coding for Small t Antigen

BAYAR THIMMAPPAYA AND THOMAS SHENK*

Department of Microbiology, University of Connecticut Health Center, Farmington, Connecticut 06032

Received for publication 27 September 1978

The deletions in nine viable simian virus 40 mutants have been mapped by direct DNA sequence analysis. The mutant DNAs lack small segments of the early region of the viral chromosome (between 0.535 and 0.600 map unit). The deletions are all located in the region which is removed from the large T antigen transcript by splicing. No one deletion removes this entire region, but no part of this segment is conserved in all of the mutants except for several nucleotides near the splice points of the transcript. Although the deletions do not alter the region coding for the large T polypeptide, they do delete portions of the segment coding for the C-terminal half of the small t polypeptide.

Two transcripts have been identified in simian virus 40 (SV40)-infected cells which map in the early region of the viral genome (Fig. 1). Both transcripts extend (5′ → 3′) from near 0.66 to 0.16 on the SV40 map, and both are spliced (1, 2). One transcript is believed to code for the SV40 small t antigen, a polypeptide of about 17,000 daltons. This polypeptide has been identified in both lytically infected monkey cells and transformed rodent cells (3, 10, 17, 18). The second transcript codes for large T antigen, a polypeptide of about 90,000 daltons which is also present in both lytically infected and transformed cells (10, 13).

The T and t polypeptides are structurally related. Most of the [^{35}S]methionine-labeled tryptic polypeptides of small t are also present in large T antigen (10), and Paucha et al. (9) have shown the N-terminal amino acid sequences of these polypeptides to be identical. This is consistent with the structure of the transcripts coding for the T and t polypeptides. The transcripts share the same 5′-nucleotide sequence, coding for the N-terminal portions of the polypeptides. But, as a result of splicing, the mRNA coding for large T antigen lacks the sequence coding for both the C-terminal portion of small t antigen and a series of translational termination codons (0.534 to 0.600 map unit) (1, 12, 20).

A series of deletion mutants has been isolated with lesions located between 0.535 and 0.600 on the SV40 map (16, 18). This region is not present in the large T transcript, but is contained in the small t antigen transcript (Fig. 1). As predicted, these deletions do not affect the size of large T antigen but do alter small t antigen. The mutants are viable, growing at a slightly slower rate than wild-type virus (16, 18), but are markedly impaired in their ability to transform rodent cells (2, 18). This suggests that at least portions of the C-terminus of small t antigen are dispensable for lytic growth of SV40, but an intact small t antigen is required for efficient viral transformation.

We have determined by direct DNA sequence analysis the nucleotide sequence which nine of these deletion mutants lack. All mutants lack segments of the region between 0.535 and 0.600 on the SV40 chromosome. Although no mutant lacks this entire region, there is no portion of the sequence which is conserved in all of the mutants, except small segments near the splice points of the large T antigen transcript.

MATERIALS AND METHODS

Cells and viruses. The origin and procedures for growth of the CV-1P line of African green monkey cells have been described by Mertz and Berg (8). The wild-type SV40 (designated *wt*830) is a plaque-purified derivative of the SVS strain (19). The deletion mutants *dl*883, *dl*884, *dl*890, and *dl*891 were derived from *wt*830 (16). Mutant *dl*2006 was kindly provided by William Topp (18). Mutants *dl*1410 and *dl*1440 were derived by expansion of the deletion in *dl*884, using the procedure of Shenk (14). Mutant *dl*1410 contains a second deletion at 0.70 to 0.71 map unit (15). The deletion in *dl*891 was expanded to produce *dl*1441 and *dl*1442.

DNA and enzymes. SV40 DNA was extracted (5) from CV-1P cells infected at a multiplicity of <0.05 PFU/cell when >90% of the cells showed cytopathic effect. Closed circular viral DNA [SV40(I)] was purified directly from the supernatant by adding CsCl to

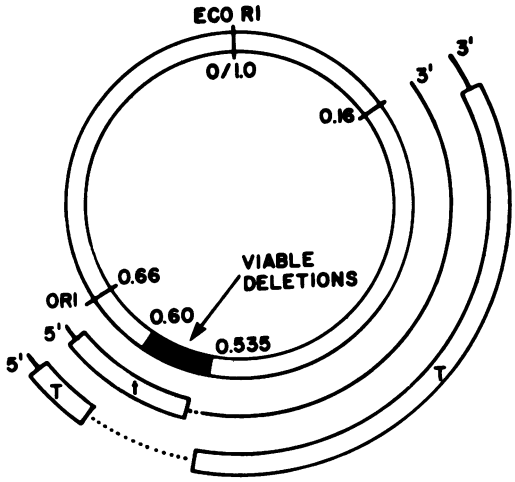

FIG. 1. *Diagram of the SV40 chromosome relating the map position of viable deletion mutants (0.60 to 0.535 map unit) to the location of the two known early transcripts. The portions of the transcripts believed to code for early proteins are indicated by the blocked segments (t = small t antigen; T = large T antigen). Regions removed from transcripts by splicing are represented by dotted lines. ORI = origin of DNA replication.*

1.56 g/cm^3 and ethidium bromide to 200 μg/ml and centrifuging to equilibrium. The band of SV40(I) DNA was collected, and the ethidium bromide was removed by passing the DNA through AG50W-X8 resin (Bio-Rad Laboratories) (11). Restriction endonucleases were either prepared in the laboratory or purchased from New England Biolabs and used according to published protocols. Bacterial alkaline phosphatase was purchased from Worthington Biochemicals Corp., and T4 polynucleotide kinase was from Boehringer Mannheim Corp.

DNA sequence analysis. Restriction endonuclease-generated fragments were end labeled by the procedure of Maniatis et al. (6). Sequence analysis of end-labeled fragments was performed as described by Maxam and Gilbert (7), using 12% polyacrylamide gels (0.7 mm thick, 40 cm long).

RESULTS

To formulate a simple strategy for the nucleotide sequence analysis of alterations present in the deletion mutants, it was first necessary to determine which restriction endonuclease cleavage sites were present and which were missing in each mutant DNA. Accordingly, each DNA was cleaved with several enzymes which cut SV40 DNA between 0.53 and 0.60 map unit. The data from these analyses are summarized in Fig. 2. Mutant *dl*884 DNA, for example, retained the *Hin*f endonuclease cleavage site at 0.533 and the *Hae*III endonuclease cleavage site at 0.595 map unit, but lost all of the cleavage sites assayed between these points (Fig. 2). Mutant *dl*890 DNA retained all of the cleavage sites assayed, but its *Hae*III and *Mbo*I endonuclease A fragments were both shortened, indicating that the deletion was contained between the *Hae*III and *Mbo*I endonuclease cleavage sites at 0.589 and 0.572 map unit, respectively (Fig. 2).

Utilizing the information in Fig. 2, appropriate restriction endonuclease-generated fragments from each mutant were prepared. The fragments were end labeled (6) and finally cleaved with another restriction endonuclease to generate a fragment labeled with ^{32}P at one 5′ end. Mutant *dl*890 DNA, for example, was cleaved with the *Hin*f and *Mbo*I endonucleases, and the 0.572 to 0.642 map unit fragment was end labeled and then cleaved with *Hae*III endonuclease. This produced a fragment with a labeled 5′ end at 0.572 map unit and an unlabeled 5′ end at 0.589 map unit for sequence analysis. The end-labeled fragments prepared from the deletion mutant DNAs are diagramed in Fig. 3. This figure also indicates the sequence which was read (solid lines) on either side of the deletions (dashed lines) after subjecting the end-labeled fragments to nucleotide sequence analysis by the method of Maxam and Gilbert (7).

An example of a sequencing gel (*dl*890 DNA) and the nucleotide sequence determined from it are shown in Fig. 4. The sequence starts from a point near the labeled, *Mbo*I endonuclease-generated 5′ end (complementary to nucleotide 457/4704 in Fig. 5) and is read from bottom to top. The location of the deletion is indicated by an arrow where the sequence jumps ahead 27 nucleotides as compared with the wild-type sequence.

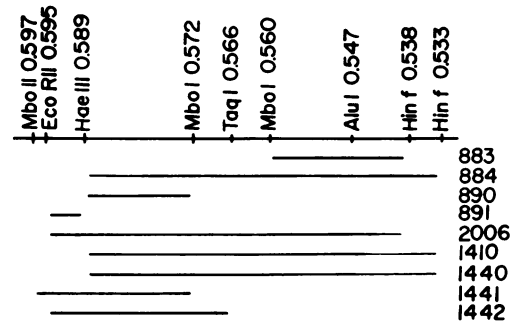

FIG. 2. *Location of viable deletion mutations determined by restriction endonuclease cleavage of mutant DNAs. The map positions at which the enzymes cleave were determined by locating the known recognition sequences of the enzymes in the SV40 nucleotide sequence of Reddy et al. (12). The solid bars are bounded by restriction endonuclease cleavage sites present in mutant DNAs and extend through cleavage sites which the mutants lack.*

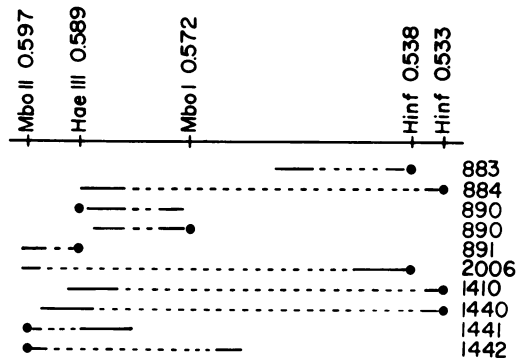

FIG. 3. *Strategy for nucleotide sequence analysis of mutant DNAs. The solid circles represent ^{32}P-labeled ends; the solid lines represent regions in which the sequence was read; the dashed lines indicate the segments deleted in mutant DNAs.*

The nucleotide sequence of SV40 in the region coding for small t antigen is reproduced in Fig. 5 (taken from the data of Fiers et al. [4] and Reddy et al. [12]). The segments which various deletion mutants lack are indicated by brackets. The points at which Reddy et al. (12) have determined splices to occur in the large T and small t transcripts are also indicated. All of the mutants lack portions of the nucleotide segment which are not present in the spliced large T transcript.

DISCUSSION

It was relatively easy to determine the location of these deletions in the SV40 genome since the nucleotide sequence of the viral DNA is known (4, 12). In general, the deletions could be unambiguously fixed in the nucleotide sequence by analyzing only one strand of the viral DNA in the region of interest. Volckaert, Fiers, and Berg (personal communication) have also determined the nucleotide sequence alterations in *dl*883, *dl*884, *dl*890, and *dl*891. Their results are in complete agreement with ours.

The deletions range in size from 25 to 253 base pairs. They are all located in the region removed from the large T antigen transcript by splicing (Fig. 5). None removes this entire region (346 nucleotides), but no part of this segment is retained in all of the mutants except for several nucleotides near the splice points. Although the deletions do not alter the region coding for large T antigen, they do delete portions of the segment coding for the small t polypeptide. This is consistent with reports that the mutants synthesize normal large T but altered small t polypeptides (3, 16, 18; G. Khoury, personal communication).

The alterations predicted by the nucleotide sequence agree well with the size of the mutant

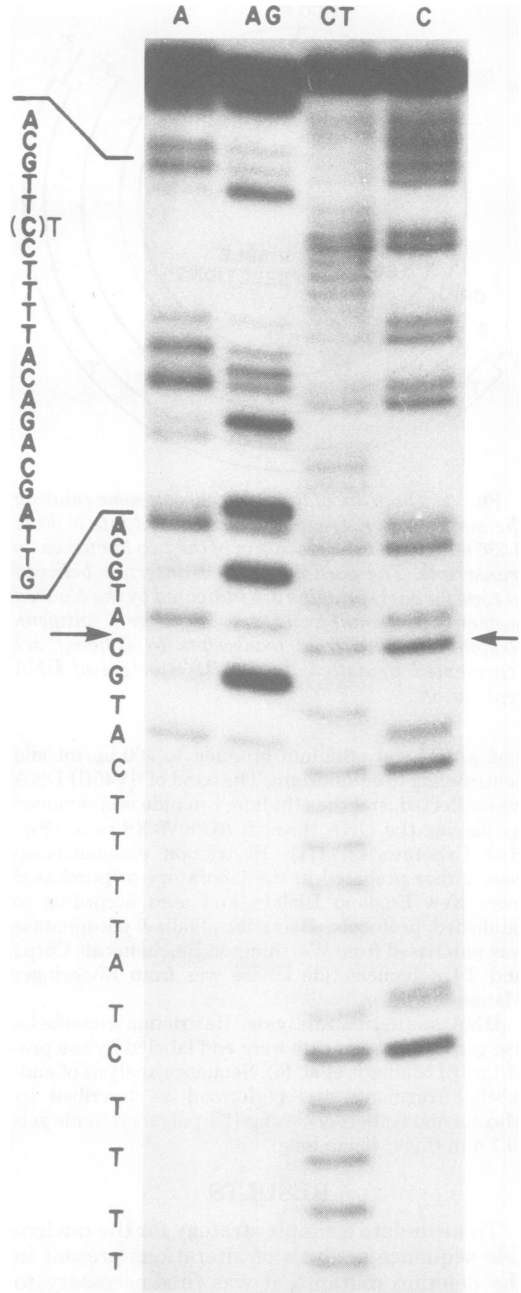

FIG. 4. *Autoradiograph of a sequencing gel containing dl890 DNA. The sequence, which is read from bottom to top, begins at the residue complementary to position 457/4704 (Fig. 5) and reads through the deleted segment (indicated by arrows). The C residue, which is six residues from the top of the sequence listed beside the gel, is artifactual. It should be a T. We find an A residue when the opposite strand is analyzed as predicted by the data of Fiers et al. (4) and Reddy et al. (12).*

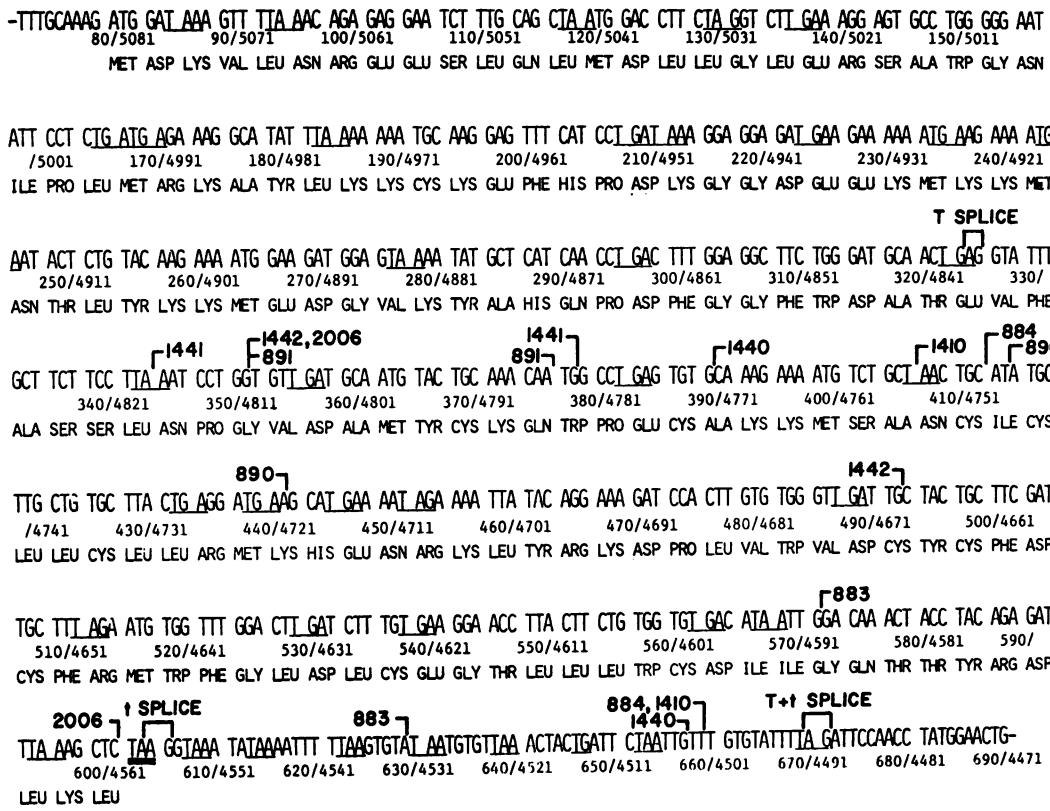

FIG. 5. *Nucleotide sequence of the portion of the SV40 genome coding for small t antigen. The region extends from about 0.65 to 0.53 map unit. The sequence is from Fiers et al. (4) and Reddy et al. (12), and the numbers correspond to those assigned by Fiers et al./Reddy et al. The splice points present in the large T and small t transcripts are from Reddy et al. (12). The segments which various deletion mutants lack are indicated by brackets.*

t proteins which have been observed. Mutant *dl*884 lacks a DNA segment coding for 63 of the 174 amino acids which the wild-type sequence predicts comprise the little t polypeptide (Fig. 5). Three amino acids not found in the wild-type protein should be added to the C-terminal end of the *dl*884 polypeptide before reaching a termination codon (UAG) at position 669/4492 (Fig. 5). This predicts that the *dl*884 genome should code for a 114-amino acid little t polypeptide. Sleigh et al. (18) and Khoury (personal communication) have observed a 12,000-dalton polypeptide in *dl*884-infected cells. Mutant *dl*890, whose DNA lacks 27 base pairs, should synthesize a 165-amino acid small t protein. Khoury (personal communication) has found a 16,000-dalton polypeptide in cells infected with this mutant. Finally, the relatively small deletion (25 base pairs) in *dl*891 should alter the translational reading frame (the deletion is not a multiple of 3 base pairs) to produce a dramatically shortened polypeptide. Synthesis of this mutant polypeptide should terminate at a UGA codon (Fig. 5, position 432/4729) after adding 19 amino acids in a new reading frame. Thus, the sequence predicts a 109-amino acid polypeptide. Khoury (personal communication) observes an 11,000-dalton polypeptide.

Mutant *dl*884, among others, removes one splice point utilized for production of the small t antigen transcript (12; Fig. 5, position 604/4557). It follows that this mutant cannot synthesize a small t antigen transcript which is spliced at the normal position; yet *dl*884-infected cells contain a 12,000-dalton small t polypeptide (18; Khoury, personal communication). Possibly, the mutant polypeptide is produced utilizing an unspliced transcript. Alternatively, a suitable transcript may be produced which is spliced at a new location. Analysis of the early viral transcripts present in the cytoplasm of *dl*884-infected cells should distinguish between these possibilities.

Several of the deletions described in this re-

port extend to positions very near splice points in both the large T and small t transcripts. Using such mutants, we can ask how many nucleotides, if any, adjacent to a splice point must remain intact if a mature, spliced transcript is to be generated. The *dl*2006 deletion extends to within 2 base pairs of the splice point for the small t transcript at position 604/4557 (Fig. 5). If this mutant can produce a bona fide small t transcript, then no more than two nucleotides on one side of the splice point are required by the splicing machinery, unless, of course, the nucleotides at the far end of the deletion which are fused to the region of interest can fortuitously perform the required function. If *dl*2006-infected cells do not contain a small t transcript with a normal splice point, it should be safe to conclude that a critical segment has been deleted.

The *dl*884 deletion extends through position 659/4502 (Fig. 5) and is only 10 base pairs from the splice point (position 670/4991) utilized in both the large T and small t transcripts. The mutant makes a normal large T antigen (3, 16, 18) and, therefore, is presumed to synthesize a properly spliced large T transcript. The *dl*884 deletion was expanded by the procedure of Shenk (14) to determine whether any of the remaining 10 base pairs between the deletion and the splice point could be removed while maintaining viability (and, thus, the ability to produce a functional large T transcript). Two expanded mutants were produced, *dl*1410 and *dl*1440 (Fig. 5). Strikingly, neither deletion extended closer to the splice point, but the *dl*1410 and *dl*1440 deletions extended 6 and 23 base pairs in the opposite direction, respectively. Since the deletions did not move further toward the splice point, it is tempting to speculate that all or most of the 10 nucleotides between the ends of the deletions and the splice point are required to produce a functional (i.e., spliced) large T transcript. We are presently isolating additional expanded deletion mutants, both viable and defective, to test this notion. Mutant *dl*1440 has acquired two T/A base pairs at positions 658–659/4502–4503 (Fig. 5) which are not present in its parent (*dl*884). We cannot offer an explanation for this observation. Nucleotide sequence analysis of additional expanded deletion mutants will tell whether this is a common event.

Mutant *dl*891 lacks a segment extending through position 352/4809, 26 base pairs from the second large T transcript splice point (Fig. 5, position 325/4836). This deletion was expanded to generate *dl*1441 and *dl*1442. The *dl*1442 deletion has not extended further toward the splice point, but the *dl*1441 deletion has moved an additional 8 base pairs in this direction (Fig. 5). Thus, we set the maximum number of base pairs required on the deletable side of this splice point at 18. Further experiments are in progress to refine this estimate.

ACKNOWLEDGMENTS

We acknowledge the competent technical assistance of Bonnie Massey.

This work was supported by Public Health Service research grant CA-19151 from the National Cancer Institute and was carried out during the tenure of an Established Investigatorship of the American Heart Association to T.S.

We thank P. Berg, W. Fiers, G. Khoury, and G. Volckaert for communicating their unpublished data to us.

LITERATURE CITED

1. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of SV40. Proc. Natl. Acad. Sci. U.S.A. **75**:1274–1278.
2. **Bouck, N., N. Beales, T. Shenk, P. Berg, and G. diMayorca.** 1978. New region of the SV40 genome required for efficient viral transformation. Proc. Natl. Acad. Sci. U.S.A. **75**:2473–2477.
3. **Crawford, L. V., C. N. Cole, A. E. Smith, E. Pancha, P. Tegtmeyer, K. Rundell, and P. Berg.** 1978. Organization and expression of early genes of SV40. Proc. Natl. Acad. Sci. U.S.A. **75**:117–121.
4. **Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. Vande Voorde, H. Van Heuverswyn, J. Van Herreweghe, G. Volckaert, and M. Ysebaert.** 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) **273**:113–120.
5. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
6. **Maniatis, T., A. Jeffrey, and D. Kleid.** 1975. Nucleotide sequence of the rightward operator of phage λ. Proc. Natl. Acad. Sci. U.S.A. **72**:1184–1188.
7. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
8. **Mertz, J. E., and P. Berg.** 1974. Defective SV40 genomes: isolation and growth of individual clones. Virology **62**: 112–124.
9. **Paucha, E., A. Mellor, R. Harvey, A. E. Smith, R. M. Hewick, and M. D. Waterfield.** 1978. Large and small tumor antigens from SV40 have identical amino termini mapping at 0.65 map units. Proc. Natl. Acad. Sci. U.S.A. **75**:2165–2169.
10. **Prives, C., E. Gilboa, M. Revel, and E. Winocour.** 1977. Cell-free translation of SV40 early messenger RNA coding for viral T-antigen. Proc. Natl. Acad. Sci. U.S.A. **74**:457–461.
11. **Radloff, R., W. Bauer, and J. Vinograd.** 1967. A dye-buoyant density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. Proc. Natl. Acad. Sci. U.S.A. **57**:1514–1521.
12. **Reddy, V. B., B. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, P. K. Ghosh, M. L. Celma, and S. M. Weissman.** 1978. The genome of SV40. Science **200**:494–502.
13. **Rundell, K., J. K. Collins, P. Tegtmeyer, H. L. Ozer, C.-J. Lai, and D. Nathans.** 1977. Identification of simian virus 40 protein A. J. Virol. **21**:636–646.
14. **Shenk, T.** 1977. A biochemical method for increasing the size of deletion mutations in SV40 DNA. J. Mol. Biol. **113**:503–515.
15. **Shenk, T.** 1978. Construction of a viable SV40 variant containing two functional origins of DNA replication. Cell **13**:791–798.

16. **Shenk, T. E., J. Carbon, and P. Berg.** 1975. Construction and analysis of viable deletion mutants of simian virus 40. J. Virol. **18**:664–671.
17. **Simmons, D. T., and M. A. Martin.** 1978. Common methionine-tryptic peptides near the amino terminal end of primate papovavirus tumor antigens. Proc. Natl. Acad. Sci. U.S.A. **75**:1131–1135.
18. **Sleigh, M. J., W. C. Topp, R. Hanich, and J. F. Sambrook.** 1978. Mutants of SV40 with an altered small t protein are reduced in their ability to transform cells. Cell **14**:79–88.
19. **Takemoto, K. K., R. L. Kirchstein, and K. Habel.** 1966. Mutants of simian virus 40 differing in plaque size, oncogenicity, and heat sensitivity. J. Bacteriol. **92**:990–994.
20. **Thimmappaya, B., and S. M. Weissman.** 1977. The early region of SV40 DNA may have more than one gene. Cell **11**:837–843.