MOLECULAR AND CELLULAR BIOLOGY, Apr. 1985, p. 742–750
0270-7306/85/040742-09$02.00/0

Vol. 5, No. 4

# Two Integrated Partial Repeats of Simian Virus 40 Together Code for a Super-T Antigen

ALEXANDRA LEVITT,†* SUZIE CHEN, GEORGE BLANCK,‡ DIANE GEORGE, AND ROBERT E. POLLACK
*Fairchild Life Sciences Center, Department of Biological Sciences, Columbia University, New York, New York 10027*

Received 5 November 1984/Accepted 16 January 1985

**We determined that the coding sequence for a 100-kilodalton super-T antigen found in Simian virus 40 mouse transformants spanned two separate partial repeats of the viral genome. The downstream repeat contained a complete Simian virus 40 large-T-antigen gene, whereas the upstream repeat was a truncated copy of the same gene. When the repeats were separated by subcloning, the capacity to code for the super-T antigen was lost. A small insertion or deletion in the origin-control region which preceded the second repeat could also destroy the ability to code for the 100-kilodalton protein. Our data suggest that differential splicing between parts of two gene copies was responsible for the additional molecular weight of this super-T antigen.**

Simian virus 40 (SV40) DNA usually integrates into mouse chromosomes in the form of tandem partial repeats of the viral genome (2, 3, 13, 19, 20, 35). Some of the repeats are precisely colinear with wild-type DNA, whereas others contain duplications, inversions, or deletions of viral sequences. Quite often these rearrangements of SV40 sequences generate variant SV40 tumor antigens in addition to the virally coded large-T and small-t antigens of 94 and 17 kilodaltons (kDa), respectively (6, 8, 21, 29, 37). Some of the variants are truncated forms of the wild-type large-T antigen (8, 29, 37), whereas others, called super-T antigens, are larger than the wild-type protein (10, 21, 29, 37).

Super-T antigens have been observed with molecular masses which range from 100 to 145 kDa (10, 6, 21, 29, 37). All appear, by immunoprecipitation and tryptic mapping, to be variants of the SV40 large-T antigen. The SV40 T antigen is a multifunctional protein which initiates viral replication in permissive cells and establishes and maintains the transformed state in nonpermissive cells (43). The large-T antigen plays well-characterized biochemical roles in several viral and host processes. These include reduction of early viral transcription by specific binding to SV40 origin sequences (41); stimulation of host DNA synthesis (18) and ribosomal gene transcription (38); binding to the host-coded, 53-kDa tumor antigen (23); and expression of an in vitro ATPase activity (42).

Super-T antigens may retain certain large-T functions but lack others. The coding sequences for two super-T antigens, a 115-kDa protein from a rat transformant (26) and a 145-kDa protein from a mouse (24), have been cloned and transfected into rodent and monkey cells. Both of these super-T antigens are able to transform rodent cells, but neither is able to support lytic replication (12, 27). Like the wild-type large-T antigen, both super-T antigens can interact with the 53-kDa host tumor antigen (12, 27), and both can bind specifically to the SV40 origin of replication (7, 27).

In each case, the additional molecular mass of the super-T antigen is generated by an internal in-phase duplication of coding sequences for the large-T antigen (24, 26). The duplications occur in the region of the gene where point mutations can cause temperature-sensitive growth (22).

We have studied a particular super-T antigen of 100 kDa which is found in fully transformed, anchorage-independent mouse cell lines (10). In contrast to other super-T antigens, the 100-kDa protein appears with extremely high frequency in SV40-transformed mouse cell lines, whether they are generated by viral infection or by DNA transfection with a plasmid carrying a wild-type SV40 genome (9). Only the transformants produced by transfection of an origin-defective SV40 sequence fail to make the 100-kDa protein (9). We wished to determine what sort of mutation or rearrangement could occur in nearly every wild-type mouse transformant to produce the same-size variant. Southern blot experiments on a variety of independently isolated mouse transformants do not reveal any variant viral early region that is common to all lines which produce the 100-kDa antigen (3).

In this paper we present results indicating that the 100-kDa protein originated in a different manner from the other super-T antigens studied to date. Unlike other super-T antigens, the 100-kDa protein was not coded for by a single SV40 insertion containing an internal duplication in the early region. Rather, its coding sequence included two separate partial repeats of the SV40 genome. The downstream repeat contained the complete coding sequence for the SV40 large-T-antigen gene, including the SV40 origin, whereas the upstream repeat was a truncated copy of the same gene which varied in length in different clones.

We believe that this is an unusual example of a coding situation in which the partial amplification of a gene generated a new protein that was coded for by parts of two gene copies.

## MATERIALS AND METHODS

**Cells.** CV1 and BSC-1 monkey cells and SV101 mouse cells were grown in Dulbecco modified Eagles medium (DME) supplemented with 10% fetal calf serum (FCS) in an atmosphere containing 10% $CO_2$. All cells were shown to be free of PPLO (pleuropneumonia-like organisms) by using Hoescht stain.

**Transient-expression assay.** Our transfection procedure was a modified version of the method developed by Rio and Tjian (32). BSC-1 cells were plated at a density of $10^6$ cells per 100-mm dish. At 24 h after plating, each dish was covered with 0.5 to 0.7 ml of DME containing 0.5 mg of

* Corresponding author.
† Present address: Department of Molecular Biology, Albert Einstein College of Medicine, Bronx, NY 10461.
‡ Present address: Department of Biochemistry, Harvard University, Cambridge MA 02138.

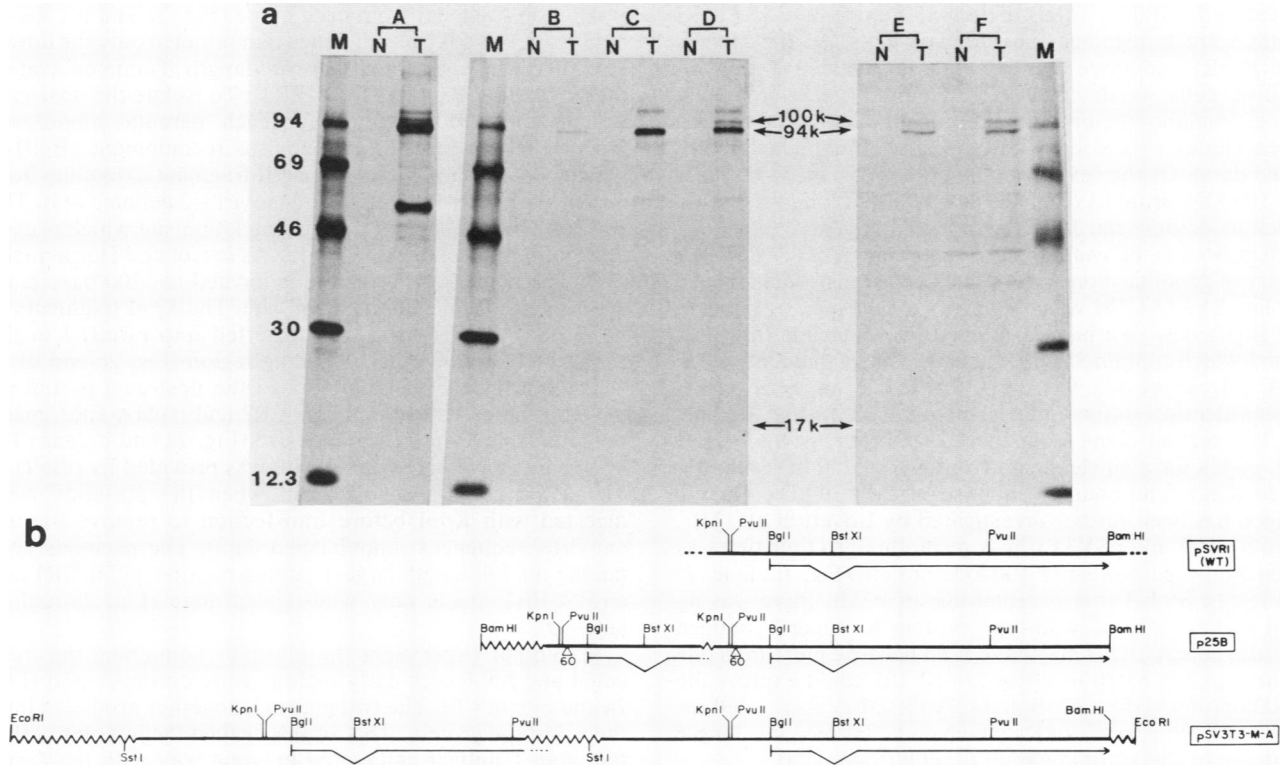


FIG. 1. Identification of coding sequences for an SV40 100-kDa super-T antigen. (a) Transient-expression assay. BSC-1 cells were labeled 48 h after transfection with [$^{35}$S]methionine, extracted, immunoprecipitated, and analyzed on a 10 to 20% sodium dodecyl sulfate-polyacrylamide gradient gel (24). Plasmids transfected were pSVRI (B), pSVOri$^-$ (C), p25B (D), pSV3T3-M-A (E), and p100-D (F). Lane A shows the immunoprecipitation of proteins from the mouse transformant SV101. The 17-kDa small-t antigen is visible in all lanes on a longer exposure of the autoradiogram. T, Tumor serum; N, normal serum. (b) Restriction maps of pSVR1, p25B, and pSV3T3-M-A. Viral DNA is indicated by straight lines, and host DNA is indicated by jagged lines. Arrows indicate the expected position of the large-T transcript. pSVR1 contained a complete wild-type SV40 genome; only its early region is shown.

DEAE dextran per ml (28) and either 5 or 10 μg of uncut plasmid. Plates were incubated at 33°C for 30 min with rocking, washed three times with DME, and fed fresh DME containing 5% FCS and 0.1 mM chloroquine (Sigma Chemical Co.). After 6 to 8 h of incubation at 37°C, the cells were rinsed twice with DME and fed with DME plus FCS. After 48 h, cover slips were removed from each plate and examined for T antigen by using indirect immunofluorescence. Labeling, extraction, and immunoprecipitation were accomplished as previously described (10).

**Rescue of integrated SV40 sequences by cell fusion.** CV1 monkey cells, which are permissive for SV40 replication, were fused with cells from SV101, a mouse transformant which has been extensively studied in our lab (3, 10). Cells ($10^6$) of each type were plated together on a 100-mm dish. On the following day the dish was overlaid with 2 ml of polyethylene glycol 1000 (50% [wt/vol]; Fisher Scientific Co.) for 1 min. The cells were then washed successively with 25, 10, and 5% polyethylene glycol 1000 and fed with DME and 10% FCS. After 24 h, $3 \times 10^5$ to $5 \times 10^5$ fused cells were plated on top of a monolayer of CV1 cells on a 60-mm dish. Once the cells were attached, 4 to 6 h later, 4 ml of 0.9% agar in DME–10% FCS was added to each plate. Plates were fed weekly, and after 21 days a solution of 0.33% Neutral Red (GIBCO Laboratories) was added with the agar. Plaques were picked the next day in DME.

Samples of three of these plaques were passaged on 60-mm plates of CV1 cells. Lysates were collected after 10 days, and DNA was extracted from a portion of each lysate. A few micrograms of DNA from each lysate was tested in the transient-expression assay (see above). DNA from one of these three plaques expressed both the 94- and 100-kDa T antigens (unpublished data). The lysate of this plaque was further amplified by growth on a 100-mm and then on a 150-mm dish of CV1 cells.

**Plasmids.** Plasmids pSV3T3-M-A and pSV3T3-M-C, which contain sequences isolated from the SV3T3 C1M library (13), were provided by P. Rigby. pSVOri$^-$ was a gift from D. Grass and J. Manley; pBglII-J was a gift from R. Jove and J. Manley. pSTER was from D. Lewis and J. Manley, and p6-1 was from Y. Gluzman.

## RESULTS

We identified sequences from two independent SV40 mouse transformants which retain the capacity to code for a 100-kDa super-T antigen when tested in a permissive transient-expression assay. Wild-type SV40 DNA produced 94- and 17-kDa tumor antigens in this assay, but not 100-kDa or any other super-T antigens (Fig. 1a, lane B).

**Identification of coding sequences from SV101.** We recovered integrated SV40 sequences from the mouse transformant SV101 by fusing SV101 cells to permissive monkey cells (4) as described above. Fusion plaques were picked after 21 days. DNA extracted from the lysates made from one out of three of these plaques expressed both 94- and 100-kDa T antigens in the transient-expression assay (unpublished data). A portion of the lysate DNA recovered from this plaque was cut with either *Bam*HI or *Eco*RI restriction enzyme and

ligated into pBR322. Totals of 9 *Bam*HI clones and 15 *Eco*RI clones were transfected into monkey cells. Of the clones, four *Eco*RI and three *Bam*HI clones produced T-antigen-positive cells as determined by indirect immunofluorescence. Immunoprecipitation experiments confirmed that all seven clones expressed wild-type large-T antigen (unpublished data). Of the seven clones, one, p25B, made both 94- and 100-kDa proteins (Fig. 1a, lane D). No clone was found which made only the 100-kDa super-T antigen.

**Identification of coding sequences from SV3T3 C1M.** A library of genomic clones from the SV40-transformed mouse cell line SV3T3 C1M was previously established by Clayton and Rigby (13). We have examined two plasmids from this library which contain SV40 sequences. One of them, pSV3T3-M-A (13) expresses 94- and 100-kDa T antigens which appear identical to the antigens made by p25B (Fig. 1a, lane E). The second clone from the C1M library, pSV3T3-M-C (13), codes for a single super-T antigen of 130 kDa (unpublished data). The coding sequence of the 130-kDa super-T antigen has been further investigated by Lovett et al. (24).

Both p25B and pSV3T3-M-A made three to four times as much wild-type protein as 100-kDa protein (Fig. 1a, lanes D and E). In SV101, the parental line of p25B, there was an even higher ratio of wild-type protein to super-T antigen (Fig. 1a, lane A). This may have been because more than one of the multiple SV40 insertions of SV101 could express the 94-kDa protein. Both p25B and pSV3T3-M-A also coded for the small-t antigen of 17 kDa which could be seen in a longer exposure of the autoradiogram (unpublished data).

**Structural similarities between pSV3T3-M-A and p25B.** p25B and pSV3T3-M-A were extensively mapped with restriction enzymes (Fig. 1b). They did not share a variant-sized early region which could account for the appearance of the same-size super-T antigen. Instead, each clone contained two tandem partial repeats of the SV40 genome. In both p25B and pSV3T3-M-A, the downstream repeat encompassed a full-length SV40 early region, which included the complete coding sequences for the large-T and small-t antigens, whereas the upstream repeat included a partial duplication of the early-region sequence. Furthermore, the partial duplication in both cases included the 5′ end of the gene, beginning before the origin-control region and proceeding past the donor and acceptor splice sites for the large-T-antigen mRNA.

In p25B the partial duplication proceeded about one-third of the way through the large-T antigen coding sequence and was separated from the second, full-length early region by 50 to 100 base pairs of host DNA. In pSV3T3-M-A the duplication extended considerably further into the large-T coding sequence, containing the *Pvu*II site at base pair 3506. In addition, ca. 600 base pairs of host DNA were present between the upstream duplication and the second coding region.

p25B had a small (ca. 60-base-pair) deletion in both copies of its early-region sequence, just downstream from the *Kpn*I site. This mutation did not affect the origin, which was located near the *Bgl*I site, but involved the deletion of part of the two 72-base-pair repeats which constitute the SV40 enhancer sequence (1, 30). In pSV3T3-M-A, this region was intact.

**Colinearity of long and short coding regions.** When the tandem repeats were separated, the capacity to code for the super-T antigen was lost. To show this, the upstream and downstream repeats were subcloned into separate vectors. In parallel experiments the upstream repeats of p25B and pSV3T3-M-A were removed by the deletion of a *Kpn*I restriction fragment from each clone (Fig. 2c and d). Recircularization produced subclones that retained only the downstream repeat containing the full-length T-antigen coding region (p25B-ER1 and pTMA-ER1). To isolate the upstream repeat, the *Kpn*I fragment of each parental clone was inserted into the pBR322-adenovirus recombinant pBglII-J, which contains a 1.2-kilobase *Bgl*II fragment extending from 60.7 to 64.2 map units on the adenovirus-2 genome (43). The *Bgl*II fragment contains a polyadenylation site which is used efficiently by adenovirus L3 mRNAs in isolated HeLa nuclei (25). The polyadenylation site is located ca. 200 base pairs downstream from a unique *Kpn*I site. The *Kpn*I fragments of p25B and pSV3T3-M-A were inserted into pBglII-J in the same orientation as the L3 coding region (Fig. 2c and d).

The constructs which contained the upstream partial repeats of the early region, p25B-ER2 and pTMA-ER2, made no detectable T antigen of any size (Fig. 2a and b, lanes E), even when a polyadenylation site was provided by pBglII-J. The same result was observed when the plasmids were digested with *Kpn*I before transfection to remove the adenovirus sequences (unpublished data). The plasmids containing the full-length large-T coding regions, p25B-ER1 and pTMA-ER1, made only wild-type protein (Fig. 2a and b, lanes D).

In another experiment the plasmids containing the truncated and full-length early regions were cotransfected (Fig. 2a and b, lanes F). The two plasmids together produced only the wild-type protein. This argues against the possibility that the large-T antigen interacted in some way with the duplicated sequence to produce the 100-kDa protein or that a truncated T antigen made from either of the ER-2 plasmids could act in *trans* to affect the message made from the full-length sequence. As noted above, no such truncated T antigen was seen in the immunoprecipitates. We concluded that the partial and full-length early region must be joined in *cis* to generate the super-T antigen.

**Partial repeats are fragments of wild-type SV40.** One hypothesis to explain the origin of the 100-kDa protein is that the upstream partial repeat of both clones contains an alteration which encodes the additional molecular weight of the super-T antigen. There is a 93-base-pair open reading between the TATA box and the start codon of the large-T-antigen gene (15) which is interrupted by a stop codon at position 5173 (see Fig. 3A). If the stop codon had been mutated or deleted and a transcript had been initiated sufficiently far upstream, an extra 31 amino acids, or about 4 kDa, could have been added to the amino terminal of the large-T antigen. Although early-region transcription usually is initiated just beyond the *Bgl*I site, during the late lytic cycle it switches to a position upstream of the TATA box, and this upstream site is used by a small percentage of transcripts of SV40 transformants (15).

To test the hypothesis above, we constructed the plasmid p25B-J (Fig. 4b). The first exon from the upstream sequence was joined to the second exon from the downstream sequence by deleting a *Bst*XI restriction fragment. *Bst*XI cut once within the large-T intron, so the recircularized plasmid contained a wild-type-sized intron between the two exons. Upon transfection into monkey cells, p25B-J generated only the wild-type 94-kDa tumor antigen (Fig. 4a, lane E). This result suggests that the first exon in the partial duplication was the wild type.

Next, we tested whether the downstream repeat could be replaced by a wild-type SV40 early region without destroying the ability to make the 100-kDa tumor antigen. The *Kpn*I fragment containing the upstream repeat from p25B was

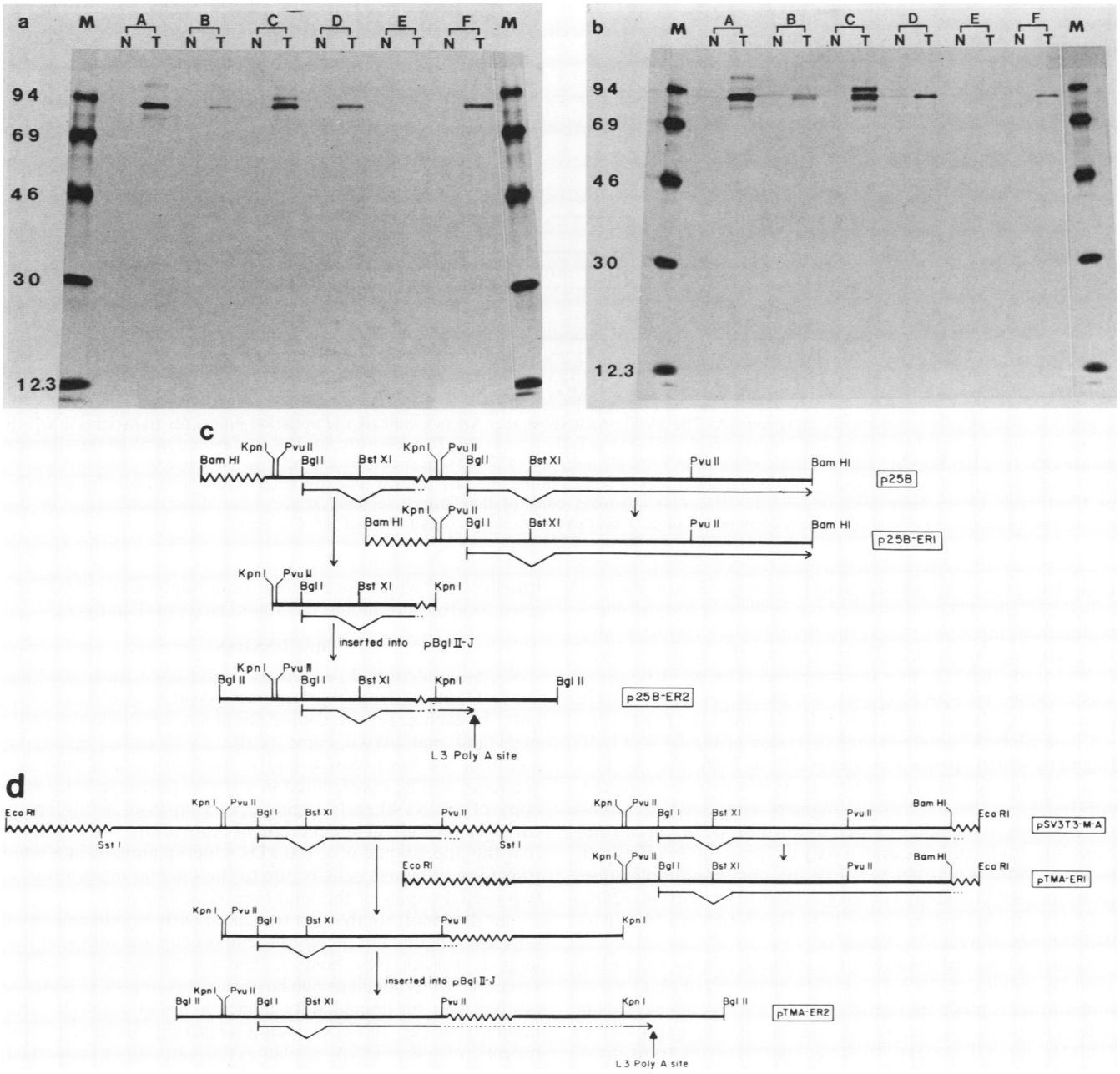


FIG. 2. Separation of the partial repeats from p25B and pSV3T3-M-A. (a) Immunoprecipitation was performed as described in the legend to Fig. 1a. The plasmids transfected were pSVOri$^-$ (A), pSVRI (B), p25B (C), p25B-ER1 (D), p25B-ER2 (E), and p25B-ER1 plus p25B-ER2 (F). T, Tumor Serum; N, normal serum. (b) Immunoprecipitation was performed as described in the legend to Fig. 1a. The plasmids transfected were pSVOri$^-$ (A), pSVRI (B), pSV3T3-M-A (C), pTMA-ER1 (D), pTMA-ER2 (E), and pTMA-ER1 plus pTMA-ER2 (F). T, Tumor serum; N, normal serum. (c) p25B-ER1 was constructed by deleting the *Kpn*I fragment from p25B. p25B-ER2 was constructed by inserting the *Kpn*I fragment from p25B into the *Kpn*I site of pBglII-J so that the truncated SV40 early region was in the same orientation as the adenovirus L3 polyadenylation site contained in the vector. (d) pTMA-ER1 was constructed by deleting the *Kpn*I fragment from pSV3T3-M-A. pTMA-ER2 was constructed by inserting the *Kpn*I fragment from pSV3T3-M-A into the *Kpn*I site of pBglII-J so that the truncated SV40 early region was in the same orientation as the adenovirus L3 polyadenylation site contained in the vector.

inserted into the *Kpn*I site of the wild-type plasmid pSVRI so that both copies of the early region shared the same orientation. The construct containing the wild-type, full-length early region (p25B-R1) retained the capacity to code for the 100-kDa protein (Fig. 4a, lane F).

**Artificial construction of a 100-kDa T-antigen coding sequence.** Taken together these results strongly suggest that no specific mutation or internal rearrangement was needed to code for the 100-kDa protein. All that was required was a full-length gene preceded by a partial duplication of the 5′ end of the gene. To confirm this conclusion, a plasmid was constructed which contained a duplicate copy of the 5′ end of the gene. This construct, called p100-D, mimicked the tandem repeats of p25B and pSV3T3-M-A, but contained no mouse DNA and no possible point mutations or small alterations in sequence. The plasmid was constructed en-

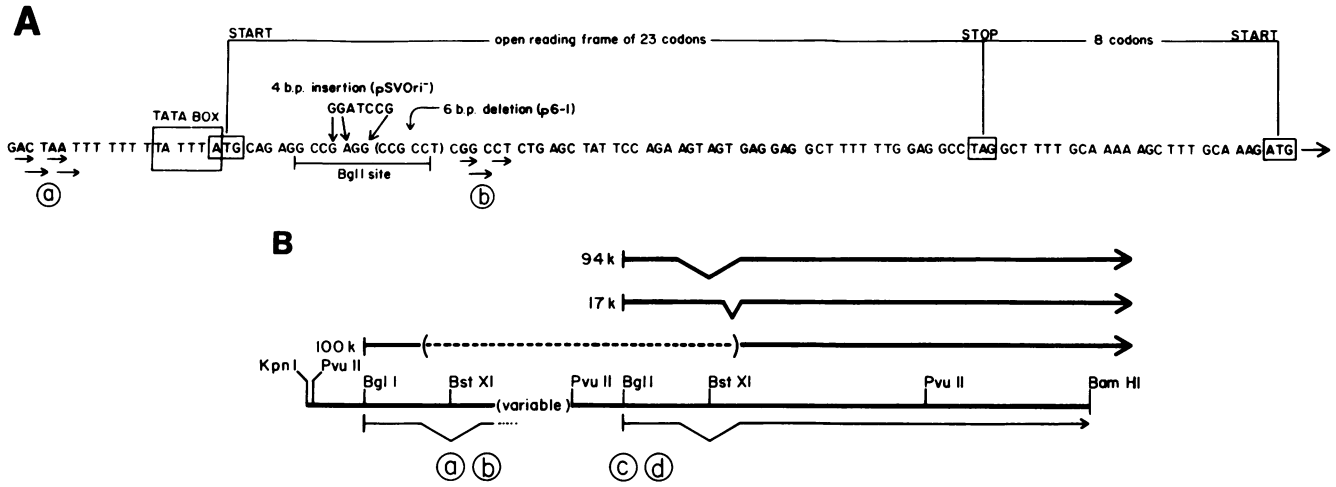


FIG. 3. Transcription in the early direction. (A) The SV40 origin sequence. Arrows indicate the initiation points for transcription which are most commonly used during (a) the late lytic cycle and (b) the early lytic cycle. The downstream sites are favored in SV40-transformed mouse cells (15). (B) Possible mRNAs transcribed from a double insertion of SV40 sequences, which included one complete and one partial (upstream) copy of the early region. The coding sequences which gave the 100-kDa protein its extra length must have come from (a) the upstream copy of the large-T intron; (b) the truncated upstream copy of the large-T second exon; (c) the downstream copy of the origin-control region; (d) the downstream copy of the large-T first exon; or some combination of these sequences.

tirely from the wild-type sequences contained in a plasmid called pSTER, which contained a complete wild-type early region from the *Kpn*I site to the *Bam*HI site. The *Kpn*I site was replaced by an *Eco*RI linker, and the sequence was cloned into pBR322. For our purposes, the pSTER insert was recloned into pAT153 (44), a pBR322 derivative which contains a 622-base-pair deletion that includes the *Pvu*II site.

Clone p100-D contained an extra copy of the SV40 *Pvu*II fragment which spanned positions 270 to 3506 (Fig. 5). The *Pvu*II site at 270 was only 24 base pairs downstream from the *Kpn*I site, so the duplicated fragment contained the complete enhancer sequence as well as a wild-type origin region. Since the duplication extended only as far as the *Pvu*II site at position 3506, it was shorter than the upstream duplication found in pSV3T3-M-A, but longer than the one from p25B. p100-D could indeed make the 100-kDa protein in addition to 94-kDa protein (Fig. 1a, lane F).

**Failure of partial repeats with mutated origins to encode 100-kDa protein.** It has previously been observed that mouse transformants made by transfection of SV40 sequences with defective origins of replication are unable to express the 100-kDa super-T antigen (9). This holds true for transformants made either with the pSVOri$^-$ plasmid (9), which contains a 1-base-pair change plus a 4-base-pair insertion within the *Bgl*I site, or with the p6-1 plasmid (17), which contains a 6-base-pair deletion in the same region (see Fig. 3A). In view of our findings, it seemed possible that the replication-deficient plasmids might be unable to generate the partial-repeat structures necessary to encode the 100-kDa protein.

To test this hypothesis, we constructed tandem repeats containing origin mutations by inserting the *Kpn*I fragment from p25B into the *Kpn*I sites of pSVOri$^-$ and p6-1 (Fig. 4b, plasmids p25B-Ori$^-$ and p25B-6-1). In each case the truncated early region was inserted in the same orientation as the full-length one. Unlike p25B-RI, which had a wild-type origin in front of its full-length repeat, neither p25B-Ori$^-$ or p25B-6-1 expressed 100-kDa super-T antigen (Fig. 4a, lanes G and H). We conclude that it was the alteration in the DNA itself rather than the inability to make a tandem repeat after transfection which prevented origin-defective SV40 sequences from producing the 100-kDa super-T antigen.

## DISCUSSION

We have identified sequences which code for a 100-kDa super-T antigen from two separate SV40 mouse transformants. In each case, the coding sequence contains one complete and one partial copy of the SV40 early region. In addition, we have shown that a 100-kDa-protein template sequence can be constructed in vitro by placing a truncated copy of the SV40 early region upstream from a full-length early region with an intact origin of replication. We conclude that the appearance of a 100-kDa super-T antigen in SV40-transformed mouse cells is due to the tandem integration of two partial copies of the wild-type viral genome.

Our data indicate that the truncated early region can be of different lengths, but includes the first exon and intron of the large-T antigen (see below). The two insertions can be separated by as much as 600 base pairs (as in pSV3T3-M-A), or they can be immediately adjacent (as in p100-D). The finding that the 100-kDa coding sequences include partial repeats of viral DNA explains the previously observed correlation between 100-kDa production and the presence of partial repeats in SV40-transformed cell lines (3).

**Structure of the mRNA for the 100-kDa protein.** The coding regions which give the 100-kDa super-T antigen its additional length must be derived from the sequences which p25B, pSV3T3-M-A, and p100-D have in common. Since p100-D consists entirely of unaltered wild-type sequences, the variant exon or exons must lie within stretches of wild-type viral DNA. The additional coding sequences must come from (i) the upstream copy of the large-T intron; (ii) the small part of the large-T second exon which is retained by the upstream repeat of even p25B; (iii) the origin-control region; or (iv) the downstream copy of the first exon (see Fig. 3B).

Two lines of evidence suggest that one or both of the large-T-antigen splice junctions are necessary for 100-kDa-protein production, whereas the sequences within the intron itself are apparently not required. First, mouse transform-

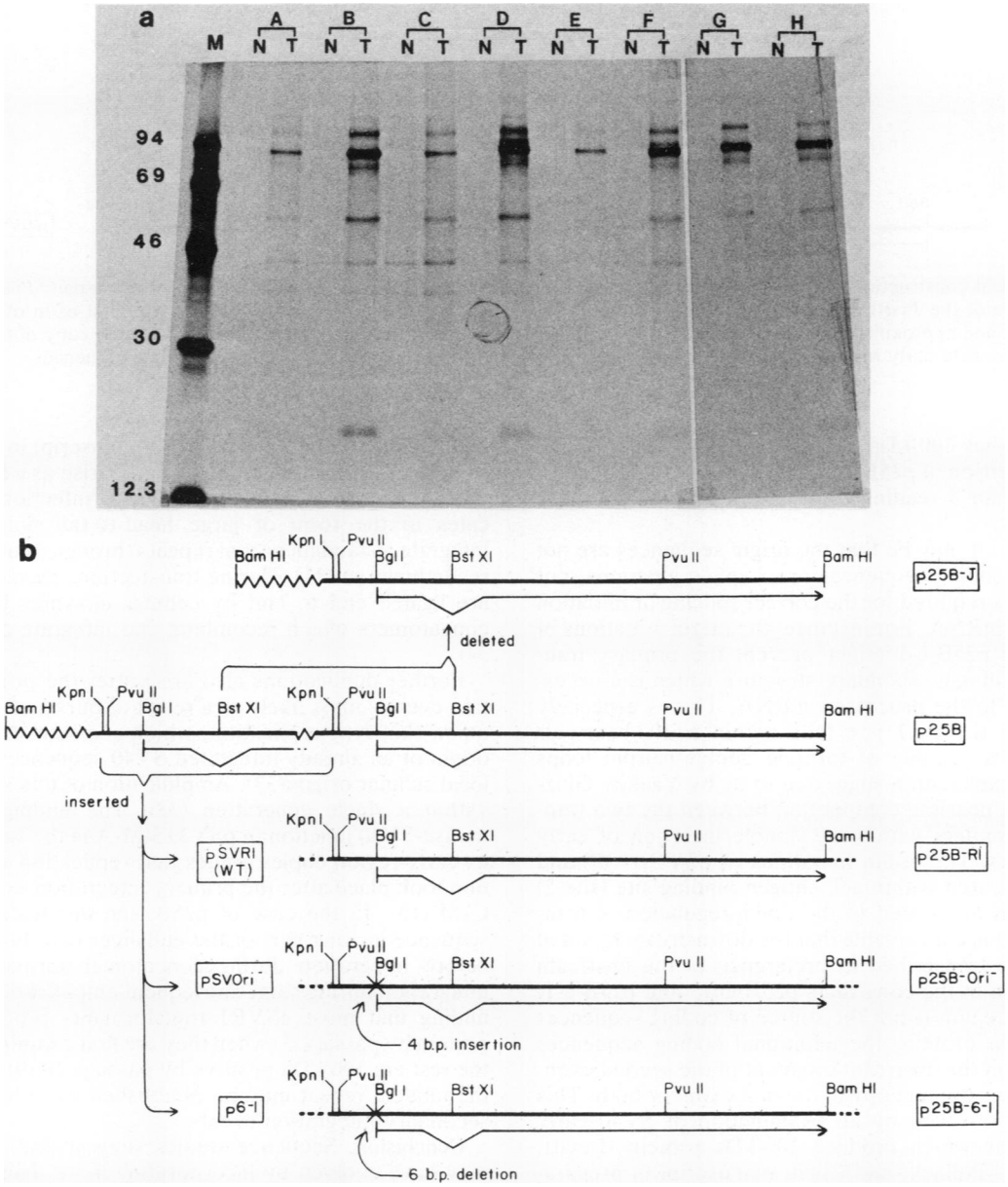


FIG. 4. The 100-kDa coding sequence made of wild-type sequences. (a) Immunoprecipitation was performed as described in the legend to Fig. 1a. The plasmids transfected were pSVRI (A), pSVOri⁻ (B), p6-1 (C), p25B (D), p25B-J (E), p25B-RI (F), p25B-Ori⁻ (G), and p25B-6-1 (H). T, Tumor serum; N, normal serum. (b) p25B-J was constructed by deleting a *Bst*XI fragment from p25B. p25B-RI, p25B-Ori⁻, and p25B-6-1 were constructed by inserting the *Kpn*I fragment from p25B into the *Kpn*I sites of pSVR1, pSVOri⁻, and p6-1, respectively. In each case the truncated early region was inserted in the same orientation as the full-length one.

ants made by transfection of the cDNA for large-T mRNA do not express the 100-kDa protein (Y. Gluzman, personal communication). Second, the mutant virus *dl*884 lacks 245 base pairs within the large-T-antigen intron (including the last 63 codons and donor splice junction of the small-t antigen; 40), and yet the mouse transformants made with this mutant do express 100-kDa protein (10).

The tandem repeat which codes for the 100-kDa super-T antigen also expresses the wild-type protein of 94 and 17 kDa. It seems probable that the transcripts which code for the wild-type proteins are initiated at the downstream origin and processed just as they would be after transcription from any wild-type early region. We propose that the 100-kDa protein is translated from a third transcript which initiates at the upstream origin and continues through the host DNA into the full-length copy of the gene. After this long primary transcript is made, the variable portion of the upstream insert plus part or all of the first exon of the downstream insert must be removed by RNA splicing (see Fig. 3B). However, a short region must be retained to provide the coding sequence for the extra 6 kDa.

Surprisingly, our data indicate that the sequences at the origin of the downstream repeat are crucial for the expression of the 100-kDa protein. The intriguing possibility arises that a cryptic splice junction, the extra coding sequences, or both may come from the origin region itself, which contains a short open reading frame with no known coding function (see Fig. 3A). Since the 6-base-pair, in-phase deletion in

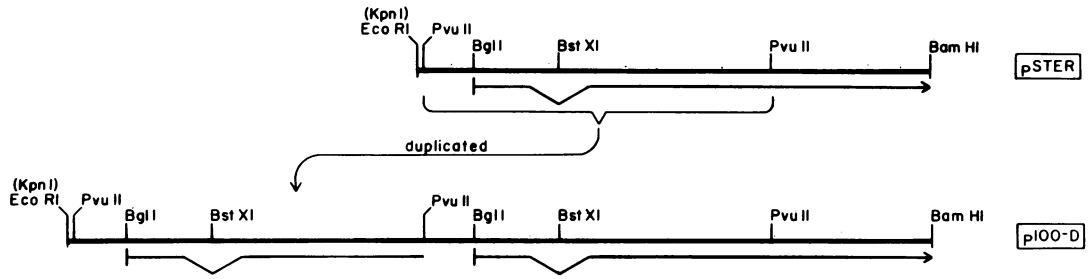


FIG. 5. Artificial construction of a 100-kDa coding sequence. p100-D was constructed by inserting a 2,007-base-pair, *Pvu*II early-region fragment into one of the *Pvu*II sites of pSTER. This fragment contained a complete origin of replication, the first exon of T antigen, the T-antigen intron, and approximately two-thirds of the T-antigen second exon. Thus, p100-D contained a truncated copy of the early region followed by a complete early region in the same orientation. p100-D was composed entirely of wild-type viral sequences.

p25B-6-1 abolishes 100-kDa expression as efficiently as the 4-base-pair insertion of p25B-Ori$^-$, it is possible that a splice signal rather than a reading frame is interrupted by these mutations.

Alternatively, it may be that the origin sequences are not used either as coding sequences or as splice junctions, but are nevertheless required for the correct splicing or initiation of the 100-kDa mRNA. For instance, the origin mutations of p25B-Ori$^-$ and p25B-6-1 might prevent the primary transcript from forming a secondary structure which is a necessary precursor to the processed mRNA. This is especially plausible since the *BglI* site falls within a 100-base-pair sequence that is capable of forming double-hairpin loops (36). Another explanation suggested to us by Yaakov Gluzman involves a possible competition between the two transcriptional promoters within the double insertion of early SV40 sequences. The origin mutations of pSVOri$^-$ (9) and p6-1 (17) are located within a T-antigen binding site (site 2) which has been implicated in the down-regulation of transcription (41). It is conceivable that the downstream mutated origin region is transcribed in preference to the upstream wild-type region in the constructs p25B-Ori$^-$ and p25B-6-1.

If the origin region is not the source of coding sequences for the 100-kDa protein, the additional coding sequences might come from the upstream fragment of the second exon, the second, downstream copy of the first exon, or both. This matter can be resolved by an examination of SV40 early mRNAs in cells which produce 100-kDa protein (Levitt, Lewis, Manley, Pollack, and Chen, manuscript in preparation).

**Function.** In a previous study of a set of cell lines derived from SV101, the amount of 100-kDa protein was correlated with the degree of anchorage-independent growth (10). It would be of interest to determine which functions of the large-T antigen are retained by the 100-kDa variant. However, an experiment which directly compares the transforming and replicating abilities of the 94- and 100-kDa proteins is not possible at the present time since no clone is available which makes only the super-T antigen. An examination of the mRNA for the 100-kDa protein may reveal whether any of the functional domains of the wild-type T antigen are duplicated or altered in the super-T antigen.

**Frequency of occurrence.** The discovery that a tandem repeat rather than a specific mutation or alteration generates 100-kDa protein provides an explanation for the high frequency with which the super-T antigen appears in mouse transformants. Tandem duplications of the early region are ubiquitous in SV40-transformed mouse cells, whether they are produced by viral infection (3, 13) or by plasmid transfection (Blanck, Pollack, and Chen, manuscript in preparation). In both cases the repeats apparently arise as a by-product of the mechanism of integration. During infection, SV40 replicates in the form of large head-to-tail polymers which integrate as tandem partial repeats through double-crossover recombination (11). During transfection, plasmid sequences are ligated end to end by cellular enzymes to form large concatomers which recombine and integrate en masse (31, 34).

Further duplications also arise after the primary integration events, most likely as a result of bursts of out-of-phase, onion-skin-type replication, which can start either from the origin of an already integrated SV40 sequence (5) or from a local cellular origin (33). Amplification of this sort may arise within a single generation (33). The finding that the 5′ mouse-SV40 junction in pSV3T3-M-A is the same in both of its early-region copies argues that replication and reintegration took place after the primary integration event in SV3T3 C1M (13). In the case of p25B, the duplication of a viral sequence lacking part of the enhancer could have occurred before integration during concatomer formation or after integration during some subsequent amplification event. The finding that most pSVR1 transformants produce 100-kDa protein by passage 3 (when they are first examined), whereas the rest are 100 kDa positive by passage 10 (9) suggests that the tandem repeat may be established by either primary or secondary integration events.

**Conclusion.** Sequence studies suggest that certain genes may have evolved by incorporating more than one copy of an amplified coding unit into a single transcript. For example, the chicken conalbumin protein apparently arose from the doubling of an ancestral gene with seven or eight exons (14), whereas the chicken ovomucoid protein was probably created by the triplication of a gene with a single intron (39). Furthermore, the collagen gene family may have evolved from the extensive amplification of a small coding unit of 54 base pairs (45).

The 100-kDa super-T antigen found in SV40-transformed mouse cells is produced as the consequence of a partial gene duplication that occurs during or soon after the integration of viral DNA into mouse chromosomes. Although the duplication includes the first exon, the intron, and part of the second exon of the large-T-antigen gene, it has not yet been established that two copies of the same sequence are expressed in the protein. Our data suggest that the additional coding sequences may come, instead, from the duplicated control region which precedes the second copy of the gene. In either case, it seems probable that the variable portion of the transcribed duplication is removed by splicing and that

coding sequences which are separated by different lengths of DNA are joined together to produce the same—or a very similar—protein. This observation implies a striking degree of flexibility in the process by which genes in pieces can be assembled to form new proteins (16).

## ACKNOWLEDGMENTS

We thank Peter Rigby, Yaakov Gluzman, David Grass, Richard Jove, E. Diann Lewis, and James Manley for plasmids used in this study. We thank Steven Kadish and Emil Lev for excellent technical assistance and Ruth Montague for expert preparation of figures. Lastly, we thank Carol Prives, Diane Robins, and Richard Jove for helpful discussions.

This work was supported by Public Health Service grant CA 336207, Public Health Service New Investigator Research Award CA 36319 (S.C.), and Public Health Service training grant GM 07216 (A.L.), from the National Institutes of Health.

## LITERATURE CITED

1. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a beta-globin gene is enhanced by remote SV40 sequences. Cell **27**:229–308.
2. **Bender, M., and W. Brockman.** 1981. Rearrangement of integrated viral DNA sequences in mouse cells transformed by simian virus 40. J. Virol. **38**:872–879.
3. **Blanck, G., S. Chen, and R. Pollack.** 1982. Integration, loss, and reacquisition of defective viral DNA in SV40-transformed mouse cell lines. Virology **126**:413–428.
4. **Botchan, M., J. Stringer, T. Mitchison, and J. Sambrook.** 1980. Integration and excision of SV40 DNA from the chromosome of a transformed cell. Cell **20**:143–152.
5. **Botchan, M., W. Topp, and J. Sambrook.** 1978. Studies on SV40 excision. Cold Spring Harbor Symp. Quant. Biol. **43**:708–720.
6. **Chang, C., D. T. Simmons, M. A. Martin, and P. T. Mora.** 1979. Identification and partial characterization of new antigens from simian virus 40-transformed mouse cells. J. Virol. **31**:463–471.
7. **Chaudry, F., G. J. Belsham, and A. E. Smith.** 1983. Biochemical properties of the 145,000-dalton super-T antigen from simian virus 40-transformed BALB/c 3T3 clone 20 cells. J. Virol. **45**:1098–1106.
8. **Chaudry, F., R. Harvey, and A. E. Smith.** 1982. Structure and biochemical functions of four simian virus 40 truncated large-T antigens. J. Virol. **44**:54–66.
9. **Chen, S., D. S. Grass, G. Blanck, N. Hoganson, J. L. Manley, and R. E. Pollack.** 1983. A functional simian virus 40 origin of replication is required for the generation of a super T antigen with a molecular weight of 100,000 in transformed mouse cells. J. Virol. **48**:492–502.
10. **Chen, S., M. Verderame, A. Lo, and R. Pollack.** 1981. Nonlytic simian virus 40-specific 100K phosphoprotein is associated with anchorage-independent growth in simian virus 40-transformed and revertant mouse cell lines. Mol. Cell. Biol. **1**:994–1006.
11. **Chia, W., and P. W. J. Rigby.** 1981. Fate of viral DNA in nonpermissive cells infected with simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **78**:6638–6642.
12. **Clayton, C. E., M. Lovett, and P. W. J. Rigby.** 1982. Functional analysis of a simian virus 40 super-T antigen. J. Virol. **44**: 974–982.
13. **Clayton, C. E., and P. W. J. Rigby.** 1981. Cloning and characterization of the integrated viral DNA from three lines of SV40-transformed mouse cells. Cell **25**:547–556.
14. **Cochet, M., F. Gannon, R. Hen, L. Maroteaux, F. Perrin, and P. Chambon.** 1980. Organization and sequence studies of the 17 piece chicken conalbumin gene. Nature (London) **282**:567–574.
15. **Ghosh, P. K., and P. Lebowitz.** 1981. Simian virus 40 early mRNA's contain multiple 5′ termini upstream and downstream from a Hogness-Goldberg sequence; a shift in 5′ termini during the lytic cycle is mediated by large T antigen. J. Virol. **40**:224–240.
16. **Gilbert, W.** 1978. Genes in pieces. Nature (London) **271**:501.
17. **Gluzman, Y., J. Sambrook, and R. Frisque.** 1980. Expression of early genes of origin-defective mutants of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **77**:3898–3902.
18. **Henry, P., P. H. Black, M. N. Oxman, and S. Weissman.** 1966. Stimulation of DNA synthesis in mouse cell line 3T3 by simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **56**:1170–1176.
19. **Hiscott, J., D. Murphy, and V. Defendi.** 1980. Amplification and rearrangement of integrated SV40 DNA sequences accompany the selection of anchorage independent transformed mouse cells. Cell **22**:534–543.
20. **Ketner, G., and T. Kelly.** 1980. Structure of integrated simian virus 40 in transformed mouse cells. J. Mol. Biol. **144**:163–182.
21. **Kress, M., E. May, R. Cassingena, and P. May.** 1979. Simian virus 40-transformed cells express new species of proteins precipitable by anti-simian virus 40 tumor serum. J. Virol. **31**:472–483.
22. **Lai, C. J., and D. Nathans.** 1975. A map of temperature-sensitive mutants of simian virus 40. Virology **66**:70–81.
23. **Lane, D. P., and L. Crawford.** 1979. T antigen is bound to a host protein in SV40 transformed cells. Nature (London) **278**:261–263.
24. **Lovett, M., C. E. Clayton, D. Murphy, P. W. J. Rigby, A. E. Smith, and F. Chaudry.** 1982. Structure and synthesis of a simian virus 40 super T-antigen. J. Virol. **44**:963–973.
25. **Manley, J. L., P. A. Sharp, and M. L. Gefter.** 1982. RNA synthesis in isolated nuclei: processing of adenovirus serotype 2 late messenger RNA precursers. J. Mol. Biol. **159**:581–599.
26. **May, E., J.-M. Jeltsch, and F. Gannon.** 1981. Characterization of a gene encoding a 115k super-T antigen expressed by an SV40-transformed rat cell line. Nucleic Acids Res. **9**:4111–4128.
27. **May, E., C. Lasne, C. Prives, J. Borde, and P. May.** 1983. Study of the functional activities concomitantly retained by the 115,000 $M_r$ super T antigen, an evolutionary variant of the simian virus 40 large T antigen expressed in transformed rat cells. J. Virol. **45**:901–913.
28. **McCutchan, J. H., and J. S. Pagano.** 1968. Enhancement of the infectivity of simian virus 40 deoxyribonucleic acid with diethylaminoethyl-dextran. J. Nat. Cancer Inst. **41**:351–357.
29. **McCormick, F., F. Chaudry, R. Harvey, R. Smith, P. W. J. Rigby, E. Paucha, and A. E. Smith.** 1980. T antigens of SV40-transformed cells. Cold Spring Harbor Symp. Quant. Biol. **44**:171–178.
30. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gau, and P. Chambon.** 1981. The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. **9**:6047–6068.
31. **Perucho, M., D. Hanahan, and M. Wigler.** 1980. Genetic and physical linkage of exogenous sequences in transformed cells. Cell **22**:309–317.
32. **Rio, D. C., and R. Tjian.** 1983. SV40 T antigen binding site mutations that affect autoregulation. Cell **32**:1227–1240.
33. **Roberts, R. M., L. Buck, and R. Axel.** 1983. A structure for amplified DNA. Cell **33**:53–63.
34. **Robins, D. M., S. Ripley, A. S. Henderson, and R. Axel.** 1981. Transforming DNA integrates into the host chromosome. Cell **23**:29–39.
35. **Sager, R., A. Anisowicz, and N. Howell.** 1981. Genomic rearrangement in a mouse cell line containing integrated SV40 DNA. Cell **23**:41–50.
36. **Shen, C.-K., and J. E. Hearst.** 1977. Mapping of sequences with two-fold symmetry on the simian virus genome: a photochemical crosslinking approach. Proc. Natl. Acad. Sci. U.S.A. **74**: 1363–1367.
37. **Smith, A. E., R. Smith, and E. Paucha.** 1979. Characterization of different tumor antigens present in cells transformed by simian virus 40. Cell **18**:335–346.
38. **Soprano, K. J., V. Dev, C. M. Croce, and R. Baserga.** 1980. Reactivation of silent rRNA genes by simian virus 40 in human-mouse hybrid cells. Proc. Natl. Acad. Sci. U.S.A. **76**:3885–3889.
39. **Stein, J. P., J. F. Catterall, P. Kristo, A. R. Means, and B. W. O'Malley.** 1980. Ovomucoid intervening sequences specify functional domains and generate protein polymorphism. Cell **21**:681–687.
40. **Thimmappaya, B., and T. Shenk.** 1979. Nucleotide sequence analysis of viable deletion mutants lacking segments of the simian virus 40 genome coding for small t antigen. J. Virol.

30:668–673.
41. **Tjian, R.** 1978. The binding site on SV40 DNA for a T antigen-related protein. Cell **13**:165–179.
42. **Tjian, R., and A. Robbins.** 1979. Enzymatic activities associated with a purified SV40 T antigen-related protein. Proc. Natl. Acad. Sci. U.S.A. **76**:610–614.
43. **Tooze, J. (ed.).** 1980. DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
44. **Twigg, A. J., and D. Sherratt.** 1980. Trans-complementable copy-number mutants of plasmid ColE1. Nature (London) **283**:216–218.
45. **Yamada, Y., V. E. Avvedimento, M. Mudryj, H. Ohkubo, G. Vogeli, M. Irani, I. Pastan, and B. de Combrugge.** 1980. The collagen gene: evidence for its evolutionary assembly by amplification of a DNA sequence containing an exon of 54 base pairs. Cell **22**:887–892.