JOURNAL OF VIROLOGY, Jan. 1987, p. 167–176
0022-538X/87/010167-10$02.00/0

Vol. 61, No. 1

Characterization of the Simian Virus 40 Late Promoter: Relative Importance of Sequences within the 72-Base-Pair Repeats Differs before and after Viral DNA Replication

MICHÈLE ERNOULT-LANGE, FRANCIS OMILLI, DAVID R. O'REILLY, AND EVELYNE MAY*

Institut de Recherches Scientifiques sur le Cancer, Laboratoire d'Oncologie Moléculaire, 94802 Villejuif Cedex, France

Received 10 April 1986/Accepted 22 September 1986

We examined sequences involved in the simian virus 40 (SV40) late promoter in vivo, by using quantitative S1 nuclease analysis of a series of deletion mutants within the SV40 regulatory region. These mutants were constructed so as to place the altered promoter region in its normal position relative to the SV40 late genes. The effects of the deletions on late transcriptional activity were analyzed before and after viral DNA replication, by omitting or including SV40 large T antigen. The data show that (i) in the absence of large T antigen, the deletion of the 21-base-pair (bp) repeats results in a fourfold increase in late transcription, and (ii) the sequences within the 72-bp repeats are a component of the SV40 late promoter, acting not only before, but also after viral DNA replication. We identified two domains which contain sequences important for efficient late transcription. Domain I, at the late proximal end of each 72-bp repeat, was found to function before replication and was possibly also involved after replication. The contribution of domain II, at the late distal end of each 72-bp repeat, was much more significant after replication but only of minor importance before replication.

The regulatory sequences of simian virus 40 (SV40) are contained in the 418-base-pair (bp) *Hin*dIII-*Hpa*II fragment (nucleotides [nt] 5171 to 346). This region includes the control elements required for early and late transcription as well as sequences required for viral DNA replication. The importance of control elements for early gene expression has been the subject of a number of investigations. At least three spatially distinct elements compose the SV40 early promoter. The term "promoter" is used here in a broad sense to cover any DNA sequence required for accurate and efficient initiation of transcription. The first element contains the early start sites and an A-T-rich sequence (17 bp) referred to as the TATA homology sequence (TATA box). The TATA box plays a role in fixing the sites used for initiation of transcription (3, 14). The second element, located upstream of the TATA box, is a G-C-rich region comprising two perfect 21-bp direct repeats and a third, degenerate repeat of 22 bp. Each repeat contains two copies of the sequence 5'-CCGCCC-3'. This transcriptional control element is essential for in vivo and in vitro transcription from the SV40 early promoter (13, 14, 19, 38). It may function in a bidirectional manner and is known to contain the binding sites for the cellular transcription factor Sp1 (10). The third element, located upstream of the G-C-rich sequences, comprises two perfect 72-bp direct repeats and 20 bp upstream from these repeats. The deletion of this element, referred to as an enhancer, has a dramatic effect on early gene expression in short-term experiments in vivo (3, 14, 17, 29). In addition, the 72-bp repeat enhancer, acting in *cis*, can stimulate transcription in vivo from heterologous promoters in either orientation (2, 29, 39).

SV40 late transcription occurs in the opposite direction to that of early transcription. Far less is known about the transcriptional control signals of the SV40 late genes. The late mRNAs are known to have heterogeneous 5' ends (15, 18, 24). However, in contrast to observations for the majority of eucaryotic mRNAs (8), there is no recognizable A-T-rich sequence (TATA box) approximately 25 nt upstream from any of these 5' termini.

In abortively infected or in transformed cells, 10 to 20% of the viral transcripts correspond to SV40 late gene expression (11, 25). After replication in lytically infected cells, there is a dramatic increase in late gene expression attributable to template amplification and to a switch in the relative strengths of the early and late promoters, with late mRNAs now accounting for up to 95% of viral transcripts (1). Recently, two groups have proposed that *trans* activation by T antigen is a factor in the stimulation of late promoter activity (4, 5, 22, 23). However, it is possible that this effect is indirect and that other factors are also involved, as suggested by the results of Brady et al. (6) and Tack and Beard (36).

Previously we have studied the structure of the late promoter operational in the absence of the SV40 origin, which is contained within sequences located from nt 332 to 113 (12). This late promoter region comprises two elements. The first element, which we call the +7 to −53 element, is located from nt 332 to 273 and includes the major late cap site (nt 325) and the sequence 5'-GGTACCTAACC-3' (nt 294 to 304), which has been shown to be important for the efficient in vitro initiation of transcription from this site (7). The second element, located from nt 272 to 113, includes the enhancer sequences, which are thus a component of both the early and late promoters in vivo. We have subsequently shown that the principal late promoter element before viral DNA replication is in fact the enhancer element and that this element can function in the absence of T antigen (30). Shaw et al. (34) have also found that enhancer sequences contribute to late promoter activity both in vivo and in vitro in the absence of replication. However, these authors report that T antigen is required for significant late transcription in vivo when replication is prevented.

In this work, we now focus our attention on the sequence elements operating within the promoter region before and after replication and late promoter activation. The data obtained show that sequences within the 72-bp repeats are a

* Corresponding author.

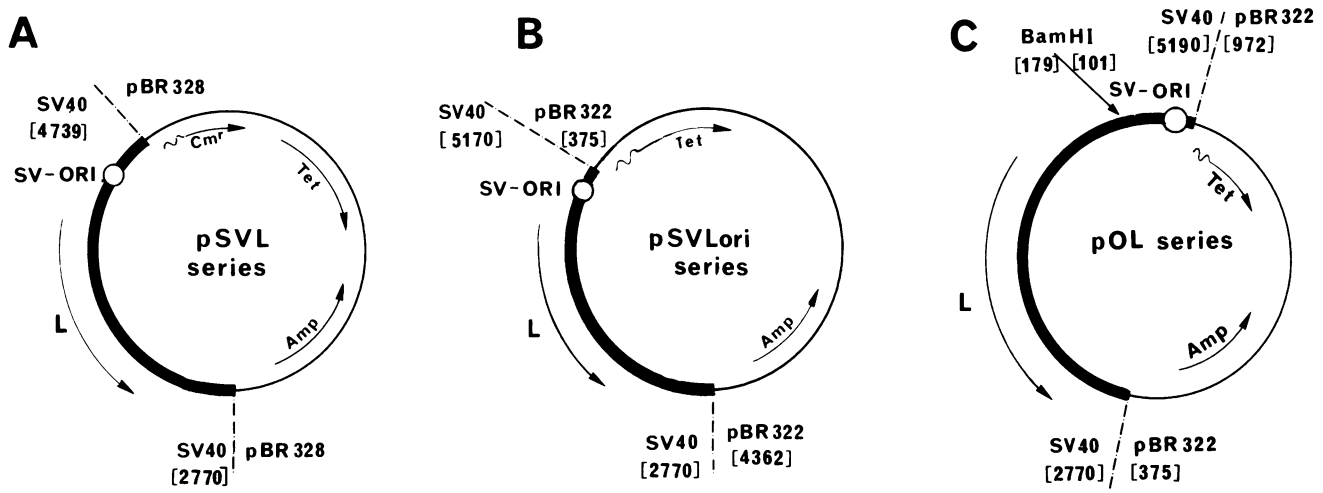

FIG. 1. (A) General structure of the pSVL series plasmids: the thick line represents SV40 sequences (the circle shows the location of the viral origin of replication, the arrow marked L shows the direction of the late transcription). pBR328 sequences are indicated by the thin line. (B) General structure of the pSVLori series plasmids: The thick line represents the SV40 sequences (symbols as in panel A). pBR322 sequences are indicated by the thin line. (C) General structure of pOL series plasmids: The thick line represents the SV40 sequences (symbols as in panel A). The nucleotides at the junction of the 21-bp repeats and the single 72-bp sequences are indicated (*Bam*HI site). The thin line represents pBR322 sequences.

component of the SV40 late promoter, acting not only before but also after viral DNA replication. In addition, we show that the relative importance for late transcription of sequences within the enhancer region changes after replication of the template DNA molecule.

MATERIALS AND METHODS

SV40 nucleotide sequence numbering system. Nucleotides are numbered according to the system of Buchman et al. (9).

Bacteria, plasmids, and cells. The bacterial host strain used was *Escherichia coli* HB101 (27). The plasmids pAS, TB101, and TB208 have been described by Benoist and Chambon (3) and by Moreau et al. (29). DNA sequence analysis has shown that: (i) SV40 nt 5236 to 1 are deleted in pAS; (ii) SV40 nt 162 to 265 are deleted in TB101; and (iii) SV40 nt 137 to 232 are deleted in TB208 (see Fig. 7A). The plasmid pLP, which contains the complete SV40 genome, has been described previously (12).

The eucaryotic cells used were CV-1P and COS-7 monkey cells and HeLa human cells. COS-7 cells are a line of permissive CV-1 cells transformed by origin-defective SV40 DNA, which synthesize a viral A gene product capable of complementing SV40 A gene mutants (16). CV-1P and COS-7 cells were grown in Dulbecco modified Eagle essential medium (GIBCO Laboratories, Grand Island, N.Y.) supplemented with 7% fetal calf serum. HeLa cells were grown in Dulbecco modified Eagle essential medium supplemented with 2.5% fetal and 2.5% newborn calf serum.

Construction of recombinant plasmids: pLP series. The pLP series includes pLAS, pLTB101, pLTB208, pLBK3 and pLBS6.

(i) pLAS, pLTB101, and pLTB208. pLAS, pLTB101, and pLTB208 were constructed from pAS, TB101, and TB208, respectively, by the addition of the late region of SV40 as described previously for pLP (12). These plasmids contain the complete SV40 genome with various deletions in the SV40 regulatory sequences (see Fig. 7A).

(ii) pLBK3. SV40 DNA was cut with *Sph*I and *Kpn*I. The large fragment (from nt 294 to 128) was treated with the Klenow fragment of *E. coli* DNA polymerase I (Klenow fragment DNA pol I), circularized by blunt-end ligation, and recut with *Pst*I. The fragment from nt 1988 to 3204, which includes the origin of replication, was ligated to a partial *Pst*I digest of pLP. The resultant plasmid is the same as pLP but contains a deletion of nt 130 to 300 (see Fig. 7A).

(iii) pLBS6. SV40 DNA was cut with *Sph*I, digested with BAL 31 nuclease, repaired with Klenow fragment DNA pol I, and circularized by blunt-end ligation. The deleted DNA was inserted into pLP by using the same construction as described for pLBK3. This plasmid contains a deletion of nt 66 to 204 (see Fig. 7A).

pSVL series. The pSVL series of plasmids includes wild-type pSVL pSVLAS, pSVL101, pSVL208, pSVLBK3, and pSVLBS6. Plasmids of the pLP series, including pLP, were digested with *Taq*I, and the fragment containing the SV40 late region and replication origin (extending from nt 4739 in SV40 to nt 651 in pBR322) was isolated from each plasmid. These were repaired with Klenow fragment DNA pol I and inserted into the unique *Pvu*II site of pBR328 (Fig. 1A). The deletions within the SV40 regulatory sequences in these plasmids are shown in Fig. 7A.

pSVori and pSVLori series. These series include pSVori, pSVoriΔ21, pSVoriΔ72, pSVLori, pSVLoriΔ21, and pSVLoriΔ72.

(i) pSVori and pSVLori. pSV1 (3), which contains the wild-type SV40 early region (extending from the *Hpa*II site [nt 346] to the *Bam*HI site [nt 2533]), was cut with *Hin*dIII, repaired with Klenow fragment DNA pol I, and recut with *Eco*RI (at the pBR322 [nt 2]-SV40 [nt 346] junction). pBR322 was cut with *Bam*HI (nt 375), repaired by Klenow fragment DNA pol I, and recut with *Eco*RI (nt 2). The pSV1 *Hin*dIII-*Eco*RI fragment, which contains the SV40 regulatory sequences, was then ligated to the large pBR322 *Eco*RI-*Bam*HI fragment, to give pSVori. The SV40 late region (nt 2770 to 294) was isolated from SV40 DNA by *Bcl*I cleavage, Klenow fragment DNA pol I repair, and *Kpn*I cleavage. pSVori was then cut with *Eco*RI, repaired, and recut with *Kpn*I. The large fragment thus obtained was ligated to the SV40 late region to give pSVLori (Fig. 1B).

(ii) pSVoriΔ21 and pSVLoriΔ21. pHS102 (3) was digested with *Bam*HI, repaired with Klenow fragment DNA pol I, recut with *Eco*RI, and the 233-nt fragment of SV40 DNA extending from nt 113 (the unique *Bam*HI site) to nt 346 (the *Eco*RI site at the pBR322-SV40 junction as in pSV1) was isolated. pSVori was cut with *Nco*I (SV40 nt 38), repaired with Klenow fragment DNA pol I, and recut with *Eco*RI (SV40 nt 346). The large pSVori fragment was ligated to the 233-nt fragment isolated from pHS102. The resultant plasmid (pSVoriΔ21) has a deletion from nt 38 to 113 which removes the 21-bp repeats (see Fig. 7A). The SV40 late region was inserted into pSVoriΔ21 by using the same construction as described for pSVLori, to give pSVLoriΔ21 (Fig. 1B).

(iii) pSVoriΔ72 and pSVLoriΔ72. pSVori was cleaved with *Pvu*I and *Pvu*II, and the 704-nt fragment extending from nt 3735 (pBR322) to nt 270 (SV40) was isolated. Separately, a 296-nt fragment extending from nt 128 (SV40) to nt 471 (pBR322) was isolated by cleavage of pSVori with *Sph*I, Klenow fragment DNA pol I repair, and cleavage with *Ban*II. These fragments were ligated to a 3,250-nt fragment isolated from pBR322 via *Pvu*I-*Ban*II cleavage. This recombinant plasmid (pSVoriΔ72) has a deletion from nt 128 to 272 in SV40 which removes the major part of the 72-bp repeats (see Fig. 7A). The SV40 late region was inserted into pSVoriΔ72 as before, to give pSVLoriΔ72 (Fig. 1B).

pOL0 series. The pOL0 series includes pOL0 and pOLΔ3.

(i) pOL0. pA0, which contains only one copy of the 72-bp repeats (nt 107 to 178 are deleted; 42) was digested with *Kpn*I and *Stu*I, and the 275-nt fragment from SV40 nt 294 to 5190 was isolated. The plasmid pL113b, containing the SV40 late region from nt 113 to 2770 (30), was digested with *Kpn*I and *Nru*I, and the large fragment obtained was ligated to the 275-nt fragment described above. The resultant plasmid, pOL0, contains the SV40 late genes and an SV40 regulatory region deleted for one copy of the 72-bp repeats. It has a unique *Bam*HI site between the 72-bp sequence and the 21-bp repeats generated in the construction of pA0 (Fig. 1C; see Fig. 7B).

(ii) pOLΔ3. pOL0 was digested with *Bam*HI, repaired with Klenow fragment DNA pol I, and recut with *Cla*I (pBR322 nt 23). The 3,568-nt fragment obtained was ligated to a 2,921-nt fragment obtained by *Sph*I digestion, Klenow fragment DNA pol I repair, and *Cla*I (pBR322 nt 23) digestion of pOL0 (Fig. 1C). pOLΔ3 is therefore identical to pOL0 except for a deletion from nt 179 to 204 (between the novel *Bam*HI site and the *Sph*I site in the 72-bp sequence), removing the early end of the single 72-bp repeat (see Fig. 7B).

Construction of the internal marker pSVL7. An *Xho*I linker [d(CCTCGAGG), New England BioLabs] was ligated into pSVL at the *Eco*RI site (nt 1782 in SV40), after cleavage and Klenow fragment DNA pol I repair of this site. The addition of an *Xho*I linker preserved the proper reading frame and permitted the translation of protein.

Quantitative S1 nuclease analysis. HeLa, CV-1P, and COS-7 cells were seeded at 10^6 cells per 10-cm-diameter petri dish. At 24 h after seeding, the cells were transfected with 20 μg of the recombinant plasmid and 10 μg of the marker plasmid pSVL7 by the calcium phosphate coprecipitation technique as previously described (12). At 48 h posttransfection, total RNA was extracted with hot phenol (33) and quantitative S1 nuclease analysis was carried out as follows. The single-stranded probe used was the coding strand of the SV40 *Ava*II fragment (nt 2013 to 1018) of pSVL7, 5′-end labeled with ^{32}P at nt 2013 (see Fig. 2). 5′- end labeling and strand separation were carried out by standard techniques (26). Total RNA (25 μg from HeLa cells, 50 μg from CV-1P cells, or 2.5 μg from COS-7 cells) was hybridized, for 15 h at 42°C, to an excess of 5′-end labeled single-stranded probe in the presence of 0.4 M NaCl–50% formamide–40 mM piperazine-*N*-*N*′-bis(2-ethanesulfonic acid) (PIPES), pH 6.4. Trial hybridization experiments were performed to ensure the presence of excess probe (data not shown). S1 nuclease digestion was performed by ninefold dilution of the hybridization mixture in 0.23 M NaCl, 40 mM $ZnCl_2$, 30 mM sodium acetate (pH 4.5), and 250 U of S1 nuclease (Bethesda Research Laboratories, Inc., Gaithersburg, Md.), followed by 2 h of incubation at 25°C. The digestion products were ethanol precipitated and separated on 5% polyacrylamide–7 M urea gels. The amount of RNA transcribed from the SV40 recombinant plasmids was estimated by scanning the autoradiograms. Each gel was exposed to film, without an intensifying screen, for various lengths of time to avoid saturation of the film. Levels of transcription for each plasmid were normalized relative to transcription of the internal standard (pSVL7). The final figures presented in Fig. 7 represent the average results of at least five independent experiments with different plasmid preparations.

Analysis of DNA replication. COS-7 cells were seeded at 10^6 cells per 10-cm-diameter petri dish. At 24 h after seeding, the cells were transfected with 0.5 μg of recombinant plasmid and 5 μg of marker plasmid pBR327 by the calcium phosphate coprecipitation technique as previously described (12). At 48 h posttransfection, the low-molecular-weight DNA was extracted by the procedure of Hirt (21). This DNA (in the Hirt supernatant) was deproteinized by two extractions with phenol saturated in 10 mM Tris hydrochloride (pH 8)–1 mM EDTA and was then precipitated with ethanol. After linearization of the DNA by *Cla*I (pSVL series) or *Eco*RI (pSVLori series), each DNA sample was fractionated by electrophoresis through a 1% agarose gel in 40 mM Tris hydrochloride (pH 7.8)–50 mM sodium acetate–4 mM EDTA. The gels were immersed in 0.5 M NaOH–1.5 M NaCl for 45 min, and then neutralized in 1 M Tris hydrochloride (pH 7.8)–3 mM NaCl, for 45 min before transfer to a nitrocellulose filter (Schleicher & Schuell, Inc., Keene, N.H.; BA 85) by the method of Southern (35). Transferred DNA was detected by hybridization to a pBR327 probe labeled with ^{32}P by nick translation (31). The relative copy number of each plasmid at 48 h posttransfection was estimated by quantitative scanning of the autoradiograms obtained. Each filter was exposed to film for various periods of time, without intensifying screen, to avoid saturation of the film. The levels of each plasmid present were normalized relative to the amount of the internal marker pBR327 to control for variations in the efficiency of transfection and for extraction of DNA. The relative amount of each plasmid present at 48 h posttransfection was obtained by averaging the results of at least four independent experiments with different plasmid preparations.

RESULTS

Linear relationship between the quantity of transfected gene and the level of transcription. Initially, to ensure that no experiments were carried out under conditions of excess input DNA, we determined the dose-response relationships for the wild-type plasmid pSVL in all three cell lines used. The pSVL plasmid contains the complete SV40 late region and an intact origin of replication. For these experiments, HeLa, CV-1P, or COS-7 cells were cotransfected with

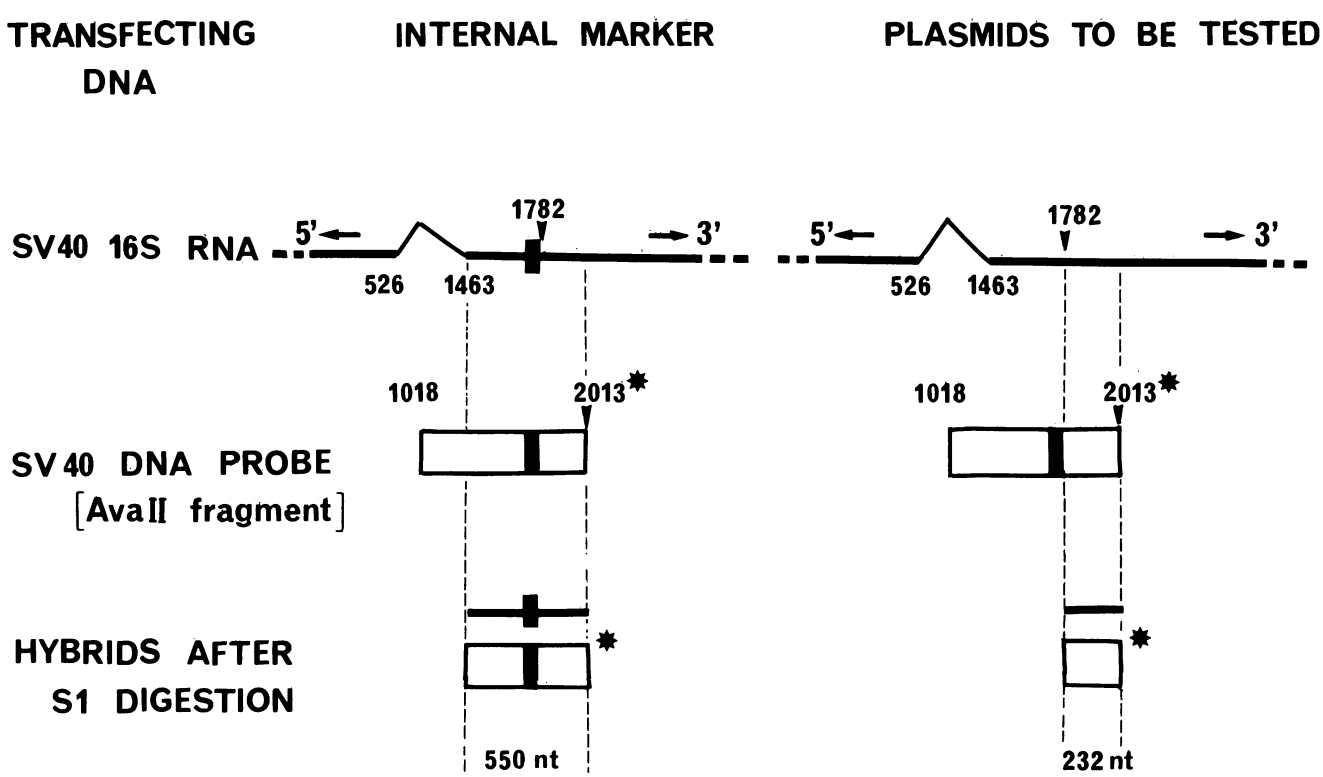

FIG. 2. Illustration of the method used for quantitative S1 nuclease analysis. The 16S mRNA obtained from the marker and test plasmids are shown. Spliced-out sequences are indicated by the thin black line, and the nucleotide numbers at the splice junctions are given. The *Xho*I linker at nt 1782 in the marker plasmid is represented by the black box. Under each RNA, the open rectangle represents the single-stranded DNA probe labeled at its 5' end (*). This probe is prepared from the *Ava*II fragment of pSVL7 (nt 2013 to 1018) which contains the *Xho*I linker. At the bottom of the figure, the hybrids obtained after S1 nuclease digestion are shown. When the probe DNA was hybridized to 16S mRNA transcribed from the marker plasmid, DNA sequence homology continued until the acceptor splice junction at nt 1463 and a 550-nt hybrid fragment was protected from S1 digestion. However, when the probe DNA hybridized with 16S mRNA transcribed from the test plasmid, the sequence homology stopped at the *Xho*I linker at nt 1782 and a 232-nt fragment was protected.

a constant amount of the internal marker pSVL7 (10 μg). At 48 h after transfection, total RNA was extracted and the amount of late transcription directed by pSVL was determined by quantitative S1 nuclease analysis as described in Materials and Methods. A schematic illustration of this methodology is shown in Fig. 2. These results are presented in Fig. 3. As predicted in Fig. 2, bands of 232 and 550 nt were observed (panel A). The 232-nt protected fragment corresponds to 16S and 19S mRNA transcribed from the pSVL plasmid, whereas 16S mRNA transcribed from the internal marker protects a fragment of 550 nt. 19S mRNA synthesized by the internal marker protects the entire probe and so yields a band at 995 nt. As the majority of late mRNA species are 16S molecules (41), this band is much weaker than the 550-nt band. It can be seen in Fig. 3A that the intensity of the 232-nt band increases with input DNA concentration in all three cell types. These results were quantified by scanning suitably exposed autoradiograms from five independent experiments similar to those shown in Fig. 3A. The average levels of expression of pSVL, normalized relative to transcription of the internal marker pSVL7, are plotted versus the amount of DNA transfected in Fig. 3B. In all three cell lines the cotransfection of up to 30 μg of pSVL and 10 μg of pSVL7 did not result in saturation of the system.

For all subsequent experiments, we cotransfected 20 μg of the test plasmid and 10 μg of the internal marker since these levels are within the linear response range of the system.

Effect of deletions within the G-C-rich and 72-bp repeats on late gene expression preceding viral replication. The deletion mutants used, which contain the complete SV40 late region and an intact origin of replication but have varying deletions within the G-C-rich sequence and the 72-bp repeats, are depicted in Fig. 1 and 7. The SV40 early genes are deleted from all these plasmids, and consequently replication from the SV40 origin is blocked. The efficiency of late transcription of each of these mutants was measured after transfection into semipermissive HeLa cells or permissive CV-1P cells. At 48 h after transfection, total RNA was extracted and the amount of late-specific RNA initiated by the various mutants was determined by quantitative S1 nuclease analysis. The results of one such experiment are illustrated in Fig. 4. The efficiency of late transcription directed by each plasmid was estimated by densitometric scanning of suitably exposed autoradiograms (see Materials and Methods) and normalization relative to transcription directed by the internal marker pSVL7. The results were then expressed relative to total transcription from the wild-type recombinants pSVL or pSVLori (taken as 100%). pSVL comprises SV40 sequences from nt 4739 to 2770 in pBR328, whereas pSVLori includes sequences from nt 5170 to 2770 in pBR322. The data presented in Fig. 7A represent the average results from at least five independent quantification experiments.

The effect of deletions within the 21-bp repeats on late gene expression. It can be seen in Fig. 4 and 7A that deletion of the 21-bp repeats (pSVLoriΔ21) resulted in a 3.5- to 4-fold

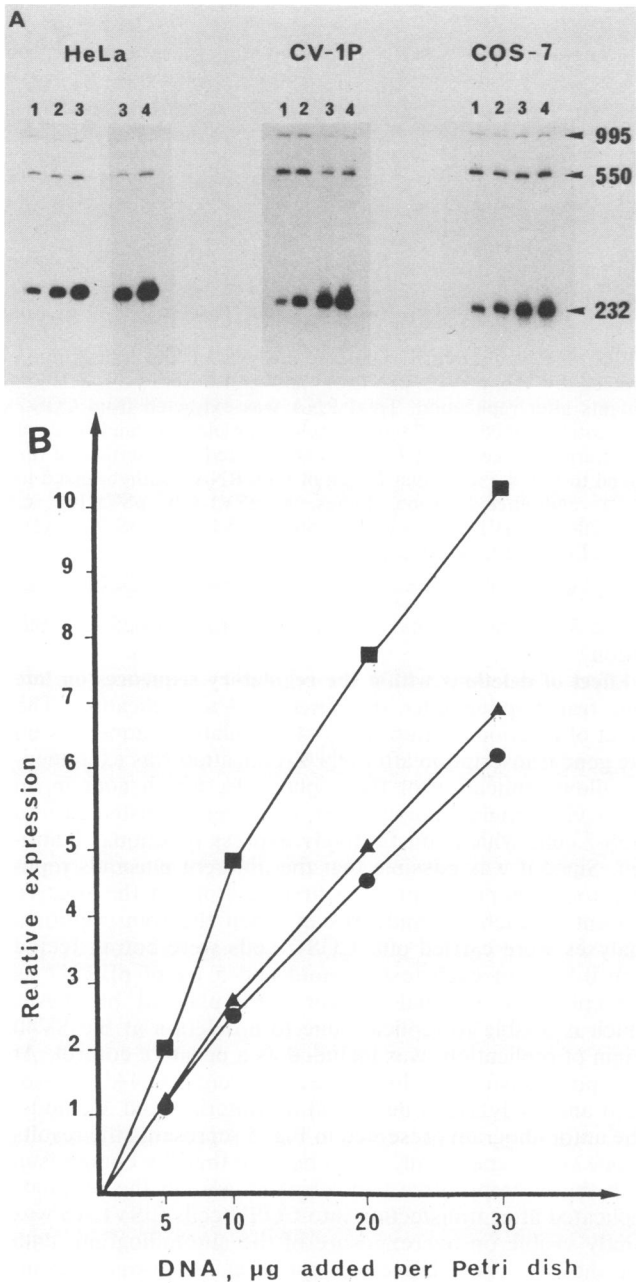

FIG. 3. Dose-response data for pSVL transcription in HeLa, CV-1P, and COS-7 cells. (A) HeLa, CV-1P, and COS-7 cells were cotransfected by calcium phosphate precipitation with increasing concentrations of pSVL DNA and a constant concentration of the internal marker pSVL7 (10 µg). RNA extraction, hybridization, S1 digestion, and electrophoresis were performed as described in Materials and Methods. A total of 5 (lane 1), 10 (lane 2), 20 (lane 3), or 30 µg (lane 4) of pSVL DNA were added per 10-cm-diameter petri dish. (B) The relative expression of pSVL was evaluated by densitometric scanning of suitably exposed autoradiograms similar to those shown in panel A. The transcription of pSVL was normalized relative to transcription from the internal marker pSVL7. Each point in this figure is the average of five independent experiments with different DNA preparations. Symbols: (■), HeLa cells; (▲), CV-1P cells; (●), COS-7 cells.

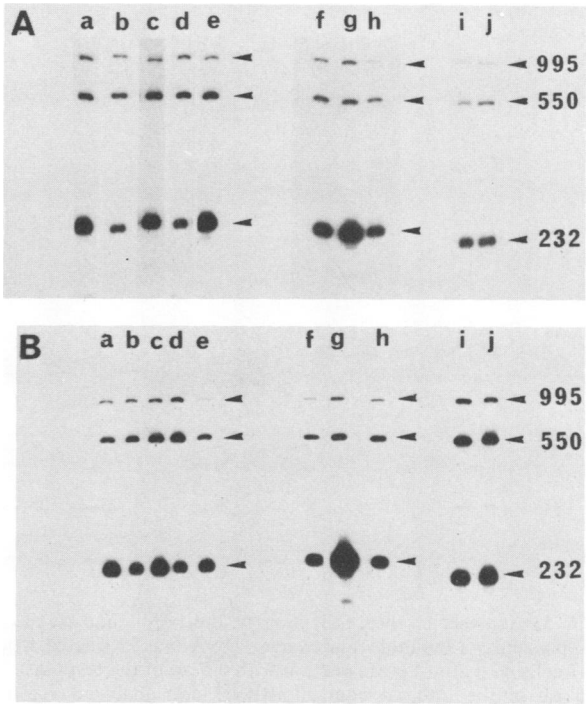

FIG. 4. Quantitative S1 nuclease analysis of late transcription directed by the pSVL, pSVLori, and pOL0 series of deletion mutants before replication. Total RNA was extracted from HeLa or CV-1P cells cotransfected with 20 µg of each of the deletion mutants and 10 µg of the internal marker pSVL7. Total RNA from HeLa cells (25 µg) (A) or 50 µg of total RNA from CV-1P cells (B) was hybridized to an excess of 5'-end-labeled, single-stranded probe (see Fig. 2). After S1 nuclease treatment and separation by electrophoresis through 5% polyacrylamide–7 M urea gels, protected fragments 232 and 550 nt long were obtained, corresponding to RNA transcribed from the test and marker plasmids, respectively. Lanes are identical in panels A and B. Lanes: a, pSVL; b, pSVL101; c, pSVL208; d, pSVLBK3; e, pSVLBS6; f, pSVLori; g, pSVLoriΔ21; h, pSVLoriΔ72; i, pOL0; j, pOLΔ3.

expression. It can be seen in Fig. 4 and 7A that deletion of the 21-bp repeats (pSVLoriΔ21) resulted in a 3.5- to 4-fold increase in late transcription both in HeLa and CV-1P cells. We have recently reported a similar phenomenon with chimeric transposition plasmids in which T antigen was expressed under the control of putative late promoter fragments (30). These results give further support to the hypothesis that before viral DNA replication the presence of the 21-bp repeats between the enhancer and the early start sites favors early transcription. When the 21-bp repeats are absent, transcription is no longer directed preferentially in the early direction, and transcription in the late direction is consequently increased.

This increase in late transcription is not seen with the plasmid pSVLBS6. The deletion in this plasmid extends into the 21-bp repeats as far as nt 66, thus removing four of the six G-C-rich motifs. However, a portion of the 72-bp repeats is also removed, and it may be that an increase in the efficiency of late transcription due to the removal of four of the six G-C repeats is offset by a decrease in promoter strength due to the deletion of important sequences within the enhancer region.

The effect of deletions within the 72-bp repeats on late gene expression. The plasmid pSVL208, which has a deletion of 95 bp within the 72-bp repeats, showed only a moderate reduc-

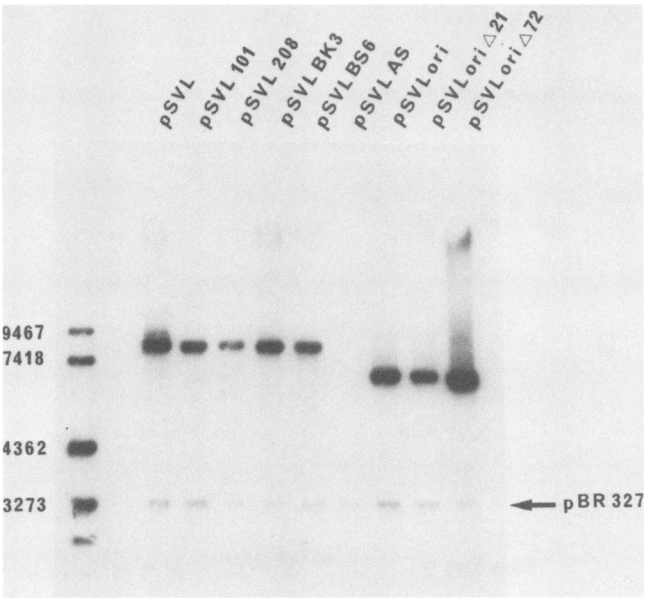

FIG. 5. Analysis of the replication of the pSVL and pSVLori series plasmids. Low-molecular-weight DNA was extracted from COS-7 cells 48 h after cotransfection with 0.5 μg of the test plasmid and 5 μg of the internal control pBR327 and analyzed via the Southern blotting technique (35) as described in Materials and Methods. The relative amount of each test plasmid was determined by scanning of suitably exposed autoradiograms. The position of linear pBR327 is indicated (arrow on the right), and the positions of the molecular-weight markers are shown on the left.

7A). However, the deletion in pSVL101, which extends to nt 265 in the late direction, brought about a 2.5- to 3-fold reduction in late promoter strength (33 to 40% of pSVL). Similarly, pSVLoriΔ72 from which sequences extending from nt 128 to 272 are missing, directed late transcription with an efficiency of 46% of pSVLori in CV-1P cells. Surprisingly, this plasmid appeared to direct late transcription more efficiently in HeLa cells (75% of pSVL). We are unsure of the significance of this observation. These results suggested the existence of a region important for late transcription before viral replication between nt 232 and 265. Extension of the deletion to nt 300 (pSVLBK3; see Fig. 7A) did not further affect the efficiency of transcription (compared with pSVL101) suggesting that, in the absence of a large part of the enhancer (nt 130 to 272), the region from nt

TABLE 1. Relative amounts of replication directed by deletion mutants in COS-7 cells

Mutant[a]	Relative efficiency of replication (%)[b]
pSVL (wild type)	100
pSVL101	90
pSVL208	104
pSVLBK3	150
pSVLBS6	165
pSVLori (wild type)	100
pSVLoriΔ21	60
pSVLoriΔ72	153

[a] The deletion mutants used are described in the legends to Fig. 1 and 7.
[b] The values from four independent experiments with different DNA preparations are expressed as a percentage of the activity determined for undeleted mutants pSVL or pSVLori and are normalized relative to the amount of the internal marker pBR327 as described in Materials and Methods.

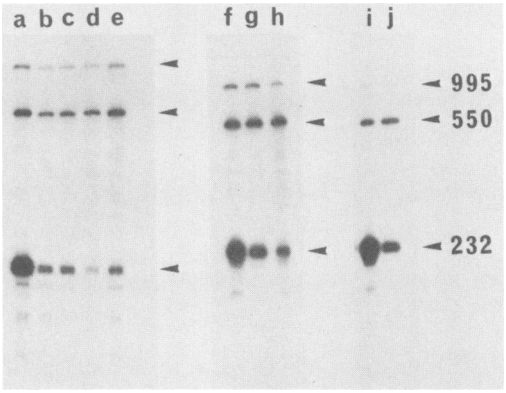

FIG. 6. Quantitative S1 nuclease analysis of late transcription directed by the pSVL, pSVLori, and pOL0 series of deletion mutants after replication. Total RNA was extracted from COS-7 cells cotransfected with 20 μg of each of the plasmids and 10 μg of the internal marker pSVL7. This was analyzed as described in the legend to Fig. 4 except that 2.5 μg of total RNA was hybridized to the 5′ end-labeled probe. Lanes: a, pSVL; b, pSVL101; c, pSVL208; d, pSVLBK3; e, pSVLBS6; f, pSVLori; g, pSVLoriΔ21; h, pSVLoriΔ72; i, pOL0; j, pOLΔ3.

265 to 300 is not critical for late transcription before replication.

Effect of deletions within the regulatory sequences on late gene transcription after the onset of viral replication. The effect of deletions within the SV40 regulatory sequences on late gene transcription after DNA replication was examined. To allow replication of these plasmids, which contain an intact viral origin of replication, they were transfected into COS-7 cells which constitutively express functional T antigen. Since it was possible that the different plasmids replicate to different extents, we first determined the relative amount of each plasmid present when the transcriptional analyses were carried out. COS-7 cells were cotransfected with 0.5 μg of each test plasmid and 5 μg of pBR327 as nonreplicative internal marker. The plasmid pSVLAS, which is unable to replicate due to a deletion at the SV40 origin of replication, was included as a negative control. At 48 h posttransfection, low-molecular-weight DNA was isolated and analyzed as described in Materials and Methods. The autoradiogram presented in Fig. 5 represents the results of one such experiment. It can be seen that, by comparison with the nonreplicative control pSVLAS, all the plasmids replicated after transfection into COS-7 cells (pSVLAS was faintly visible on overexposure of the autoradiogram; data not shown). The relative amount of each plasmid present was then estimated by scanning suitably exposed autoradiograms (see Materials and Methods) and normalizing relative to the amount of the internal marker pBR327. The results (Table 1) of at least four such experiments were then averaged to provide an estimation of the relative copy number of each plasmid 48 h posttransfection of COS-7 cells.

The relative levels of transcription were determined in a series of quantitative S1 nuclease analysis experiments by using the same internal marker and probe as in the studies carried out before replication. The results of one such experiment are presented in Fig. 6. The variations in the amount of late RNA synthesized by each test plasmid were quantified by densitometric scanning of suitably exposed autoradiograms as before. A summary of the results of this series of experiments is presented in Fig. 7A. All results are again expressed as a percentage of the appropriate wild-type

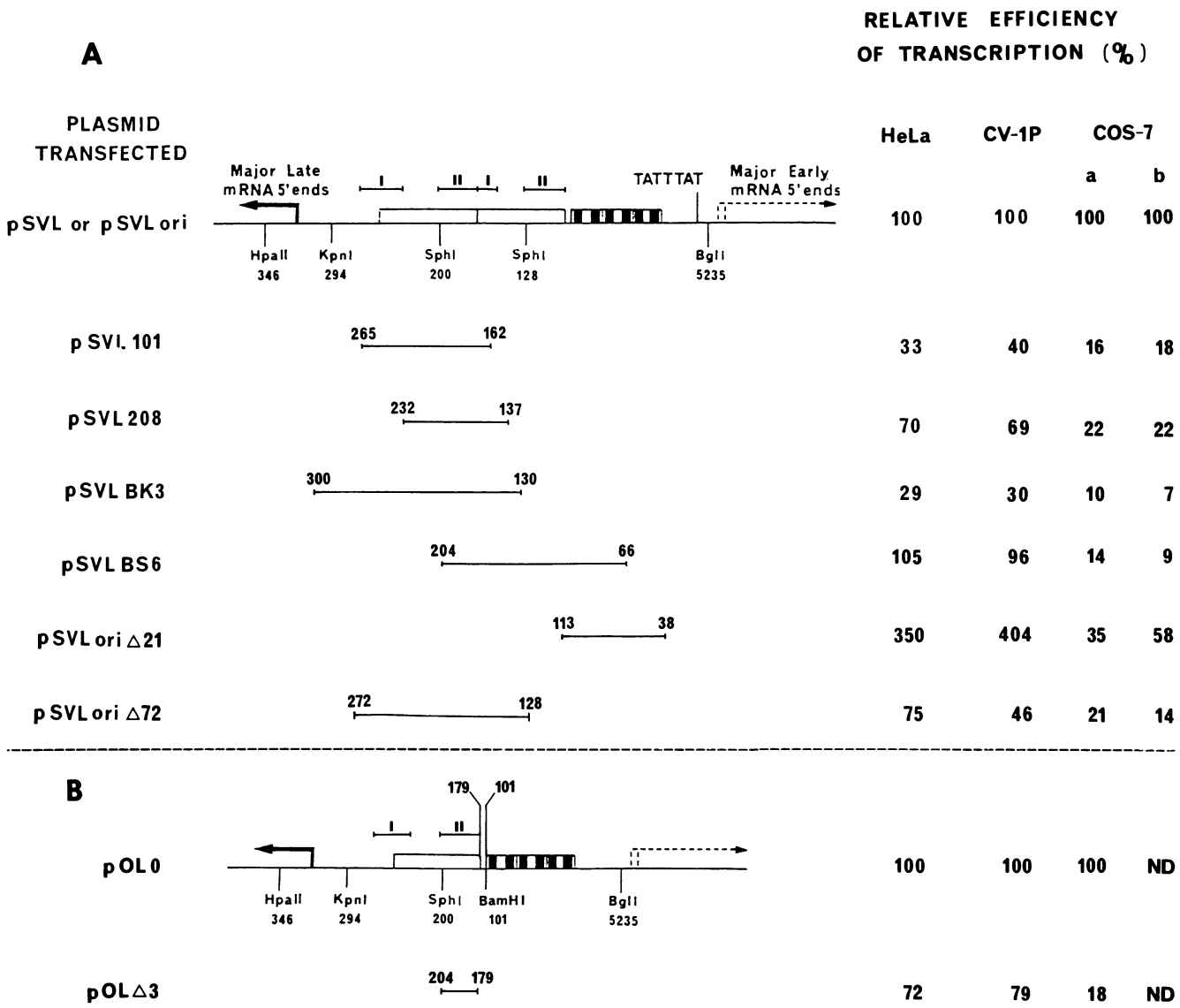

FIG. 7. Summary of levels of late transcription directed by the pSVL, pSVLori, and pOL0 series of deletion mutants in HeLa, CV-1P, and COS-7 cells. After transfection of cells with each of the deletion mutants and the internal marker pSVL7, the amount of RNA transcribed was measured by quantitative S1 nuclease analysis (see Fig. 4 and 6), followed by scanning of suitably exposed gels. The results were normalized relative to transcription of the internal marker pSVL7 and are expressed as a percentage of transcription of the undeleted plasmid pSVL, pSVLori, and pOL0. They are corrected (column b) or not (column a) to account for variations in the copy number of the different plasmids in COS-7 cells. The values presented are the averages of five independent experiments performed with different DNA preparations. (A) Levels of late transcription directed by the pSVL and pSVLori series. A schematic representation of SV40 regulatory sequences and the locations of the deletions in mutant plasmids are presented. The top of the figure shows, from right to left: the 5' ends of early mRNA sites (arrow shows the direction of early transcription); the *BglI* site (the minimal origin of replication is located around the *BglI* site from nt 5208 to 30); the TATTTAT sequence, (the TATA box of the early promoter); the G-C-rich region (one 22-bp degenerate repeat and two 21-bp directed repeats) in which the six 5'-CCGCCC-3' repeated sequences are indicated by six black blocks; the two 72-bp direct repeats (open boxes) with the positions of domains I and II indicated above; and the major late transcription initiation site (arrow shows the direction of late transcription). The deletion mutants used are depicted underneath. The nucleotides at the boundaries of each deletion (inclusive) are given. (B) Levels of late transcription directed by the pOL0 series. A schematic representation of the SV40 regulatory sequences included in pOL0 is presented. The symbols used are as in panel A. The nucleotides at the junction of the 21-bp repeats and the single 72-bp sequence are indicated, as are the nucleotides at the boundaries of the pOLΔ3 deletion. ND, Not determined.

plasmid (pSVL or pSVLori), both of which can replicate in COS-7 cells. The variations in the relative amount of replication of the different plasmids did not significantly affect the efficiency of late transcription. In general, the various deletions had a more pronounced effect on the relative efficiency of late transcription after replication. pSVL101, pSVL208, pSVLBK3, and pSVLoriΔ72 all showed relative promoter strengths two- to threefold less than before replication. After correction for activity of replication, the plasmid pSVLoriΔ21 showed an efficiency of transcription 58% that of pSVLori. This is in contrast to the situation before replication, in which the removal of the 21-bp repeats resulted in a stimulation of late transcription. The fact that this effect was no longer seen suggests that, after replication,

the 21-bp repeats favor early transcription to a much lesser extent.

The most interesting result, however, is that obtained with the plasmid pSVLBS6. Before replication, this plasmid showed an unimpaired efficiency of late transcription, whereas afterward it showed levels of late transcription only 9% those of pSVL. Thus, the deletion in pSVLBS6 removed sequences critical for late transcription after replication, and transcription prereplication remained unaffected.

Significance of the late distal moiety of the 72-bp repeats for late gene expression. The results described above suggested that sequences important for late transcription both before and after replication are found between nt 130 and 300. Furthermore, the results obtained with pSVLBS6 show that the deletion in this plasmid clearly affects late transcription more severely after replication. We noticed that this plasmid is the only one in which the sequences downstream from the *Sph*I site of both 72-bp repeats were deleted (Fig. 7A). These results suggested that sequences required only after replication are located in these regions.

However, since the deletion of four of the six G-C-rich motifs in pSVLBS6 can have a stimulatory effect on late transcription, it is possible that the deletion in the 72-bp repeats does indeed remove late promoter sequences active before replication. To clarify this point, we constructed the plasmid pOLΔ3 in which the sequences downstream from the *Sph*I site (nt 179 to 204) are deleted from the single 72-bp sequence of the plasmid pOL0. This plasmid retained the G-C-rich motifs. pOLΔ3 or pOL0 were cotransfected into HeLa, CV-1P, or COS-7 cells with the internal marker pSVL7. Relative levels of transcription were determined in a series of quantitative S1 nuclease analysis experiments as described above. The results of these analyses are presented in Fig. 4 and 6 and summarized in Fig. 7B. We found that, whereas before replication pOLΔ3 directed late transcription with an efficiency slightly less than wild type (70 to 80% of pOL0), after replication it was more inhibited, with a promoter strength only 18% that of pOL0. This result therefore confirms that the late distal moiety of each 72-bp repeat is critically important after replication and of lesser importance before replication.

DISCUSSION

In this study, we analyzed and compared the structure of the SV40 late promoter before and after the onset of DNA replication. To this end, we constructed an extensive series of mutants deleted within the late promoter region. These mutants all contain the altered promoter in its normal position relative to the SV40 late genes. The deletions encompass both the G-C-rich repeats and the enhancer region so that the contribution of both these regions can be assessed. An important observation that can be made concerning the operation of the late promoter before replication is that this promoter apparently functions in a similar manner in HeLa and CV-1P cells. This confirms the significance of previous results dealing with the functioning of this promoter in HeLa cells (12, 30).

The analysis of the plasmid pSVLoriΔ21 indicates that, whereas before replication removal of the G-C-rich region stimulates late transcription approximately fourfold, after replication the removal of these sequences results in a moderate drop in late promoter strength (Fig. 7A). The fact that the removal of the G-C-rich region did not entail an increase in late transcription postreplication suggests that the G-C-rich repeats do not favor early transcription to the same extent after replication as before. The slight decrease in the late promoter strength of pSVLoriΔ21 after replication may be due to a reduction in transcription from the late initiation sites closest to the G-C-rich repeats in the 72-bp repeats. These sites are preferentially used in vitro (28, 32, 38) or in vivo after viral DNA replication (11, 15).

The study of mutants within the enhancer element allowed us to identify two distinct domains in this region which appeared to be particularly important for late promoter activity. One important domain, which we call domain I, was shown to be between nt 232 and 265 (compare pSVL101 and pSVL208, Fig. 7A) comprising the late proximal end of the 72-bp repeats and 15 bp to the late side of these repeats. Our results show that domain I is active before viral replication. It is interesting that this domain includes sequences which other workers have shown are important for early gene transcription. Weiher et al. (40) used a series of multiple-point mutations within the 72-bp repeats to define a core sequence from nt 239 to 246 (5′-TGGAAAGT-3′) which is essential for the enhancer activity of this region in CV-1 cells. Shaw et al. (34) have subsequently found that certain of the mutants studied by Weiher et al. (40) also display reduced late transcription in vivo before replication. More recently, Zenke et al. (42) have used an extensive series of point and deletion mutations to examine the relative importance of all sequences within the enhancer region. By analysis of an SV40 early promoter with only one 72-bp sequence, they identified two broad domains, A and B, corresponding to the early and late moieties, respectively, of the enhancer element used. Our results now show that some of the sequences within domain B (nt 226 to 278) are also required for late transcription. Notice, however, that we find that deletion of this domian is less deleterious to late transcription than to early transcription. Moreau et al. (29) reported that the deletion of TB101 causes a drop of greater than 100-fold in the activity of the early promoter, whereas we find here that the same deletion reduces late promoter activity three- to fivefold. It might be that the differential effect of this deletion reflects the possible ability of the 21-bp repeats to function as a substitute late promoter in the absence of the enhancer sequences, as suggested by Fromm and Berg (14).

The second domain of importance (domain II, nt 179 to 204) functioned chiefly after the onset of replication. This is clearly demonstrated by the results obtained with the plasmid pOLΔ3 (Fig. 7B). Again we notice that the domain we identify here includes sequences that Zenke et al. (42) have shown are important for the early enhancer effect. The deletion in pOLΔ3 removes sequences belonging to the domain A described by these workers.

The finding that the enhancer element is required for late transcription both before and after replication would appear to be in disagreement with the results of Shaw et al. (34), who were unable to demonstrate that enhancer mutations have a deleterious effect on late transcription postreplication. We do not know the basis for this apparent contradiction. However, we note that the results of Hartzell et al. (20) also indicate that sequences in the 72-bp repeats are important for late gene expression in vivo postreplication. Indeed, these authors found that the late distal 22-bp portion of the 72-bp repeat, corresponding in position to domain II identified here, is sufficient for induction of late promoter activity when the minimal origin of replication is present. Other authors have also provided evidence supporting the importance of the domains identified. Both Brady and Khoury (5) and Keller and Alwine (23) found that sequences in the

enhancer region are required for late gene expression in the presence of T antigen. The latter authors further subdivided the sequences of importance into two domains, extending from nt 200 to 270 and nt 168 to 200, which are in good agreement with the domains described here. However, our results show that the presence of T antigen is not required for the activation of domain I (nt 232 to 265).

The existence of elements of the late promoter which are more critical after replication can account for the fact that all the enhancer deletion mutants examined are more deficient postreplication, since all these mutants lack at least one copy of this domain. Conversely, in none of the mutants has late transcription been completely abolished. In this respect, it is important to note that all mutants retain at least one copy of either domain. It is possible that each copy of each domain can function, to a certain extent, independently of the other important sequences (as Zenke et al. [42] have demonstrated for the two domains of the early enhancer). Note that if this is so, then the results obtained with pSVLBS6 and pOLΔ3 suggest that the domain I (in the presence of the +7 to −53 element) continues to function after replication. These results are confirmed by a study of point mutations in this domain (manuscript in preparation).

The increased importance of the domain II after replication suggests that there is a change in the operation of the late promoter concomitant with replication. We have previously observed that the 5′ termini of late mRNAs are more heterogeneous after viral replication than before (11), which also suggests the occurrence of an alteration in the late promoter. It seems highly likely that this change reflects the differential interaction of a transcriptional factor(s) with the promoter region. Indeed the domain II (nt 179 to 204) virtually corresponds to the sequence identified by Keller and Alwine (23) from nt 160 to 200, which appeared to be responsible for T-antigen-mediated *trans* activation.

A striking observation made in this study was of the potential identity of elements within the enhancer required for the stimulation of early and late transcription. In light of our results concerning the role of the G-C-rich region, we notice that this raises an interesting possibility concerning the change in strength of the later promoter pre- and postreplication. We have already shown that, before replication, the location of the G-C-rich repeats between the enhancer and the early start sites causes transcription to proceed more efficiently in the early direction (30). Results presented here suggest that the G-C-rich region does not favor early transcription to the same extent after replication. Simultaneously, domain II (which corresponds in position to domain A of Zenke et al. [42]) becomes more critical for late transcription. Takahashi et al. (37) have provided evidence which strongly suggests that protein-protein contacts exist between factors bound to the G-C-rich region and to domain A of the enhancer. Thus, it is possible that, whereas before replication the G-C-rich region directs transcription chiefly in the early direction, after replication this effect is reduced and domain II is now free to stimulate transcription in the late direction.

ACKNOWLEDGMENTS

We thank P. May for helpful discussions. We thank C. Benoist, P. Moreau, M. Zenke, and P. Chambon for the generous gifts of the plasmids pSV1, HS102, TB101, TB208, and pA0. We also thank J. Borde and C. Breugnot for their expert technical assistance. We are grateful to M. Maillot for the preparation of this manuscript.

This work was supported by grant ATP 003001 from the Centre National de la Recherche Scientifique.

LITERATURE CITED

1. **Acheson, H. N.** 1981. Lytic cycle of SV40 and polyoma virus, p. 125–204. *In* J. Tooze (ed.), DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
2. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
3. **Benoist, C., and P. Chambon.** 1981. In vivo sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
4. **Brady, J., B. J. Bolen, M. Radonovich, N. Salzman, and G. Khoury.** 1984. Stimulation of simian virus 40 late gene expression by simian virus 40 tumor antigen. Proc. Natl. Acad. Sci. USA **81**:2040–2044.
5. **Brady, J., and G. Khoury.** 1985. *trans* activation of the simian virus 40 late transcription unit by T-antigen. Mol. Cell. Biol. **5**:1391–1399.
6. **Brady, J., M. R. Loeken, and G. Khoury.** 1985. Interaction between two transcriptional control sequences required for tumor-antigen-mediated simian virus 40 late gene expression. Proc. Natl. Acad. Sci. USA **82**:7299–7303.
7. **Brady, J., M. Radonovich, M. Vodkin, V. Natarajan, M. Thoren, G. Das, J. Janik, and N. P. Salzman.** 1982. Site-specific base substitution and deletion mutations that enhance or suppress transcription of the SV40 major late RNA. Cell **31**:624–633.
8. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
9. **Buchman, A. R., L. Burnett, and P. Berg.** 1981. The SV40 nucleotide sequence, p. 799–829. *In* J. Tooze (ed.), DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–87.
11. **Ernoult-Lange, M., and E. May.** 1983. Evidence of transcription from the late region of the integrated simian virus 40 genome in transformed cells: location of the 5′ ends of late transcripts in cells abortively infected and in cells transformed by simian virus 40. J. Virol. **46**:756–767.
12. **Ernoult-Lange, M., P. May, P. Moreau, and E. May.** 1984. Simian virus 40 late promoter region able to initiate simian virus 40 early gene transcription in the absence of the simian virus 40 origin sequence. J. Virol. **50**:163–173.
13. **Everett, R. D., D. Baty, and P. Chambon.** 1983. The repeated G-C motif upstream from the TATA box are important elements of the SV40 early promoter. Nucleic Acids Res. **11**:2447–2464.
14. **Fromm, M., and P. Berg.** 1982. Deletion mapping of DNA regions required for SV40 early region promoter function in vivo. J. Mol. Appl. Genet. **1**:457–481.
15. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. W. Weissman.** 1978. Heterogeneity and 5′ terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
16. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell **23**:175–182.
17. **Gruss, P., R. Dhar, and G. Khoury.** 1981. Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. USA **78**:943–947.
18. **Haegeman, G., H. Van Heuverswyn, D. Gheysen, and W. Fiers.** 1979. Heterogeneity of the 5′ terminus of late mRNA induced by a viable simian virus 40 deletion mutant. J. Virol. **31**:484–493.
19. **Hansen, U., and P. A. Sharp.** 1983. Sequences controlling in vitro transcription of SV40 promoters. EMBO J. **2**:2293–2303.
20. **Hartzell, S. W., B. J. Byrne, and K. N. Subramanian.** 1984. The simian virus 40 minimal origin and the 72 base pair repeat are required simultaneously for efficient induction of late gene expression with large tumor antigen. Proc. Natl. Acad. Sci. USA **81**:6335–6339.
21. **Hirt, B.** 1967. Selective extraction of polyoma from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.

22. **Keller, J. M., and J. C. Alwine.** 1984. Activation of the SV40 late promoter: direct effects of T antigen in the absence of viral DNA replication. Cell **36**:381–389.
23. **Keller, J. M., and J. C. Alwine.** 1985. Analysis of an activatable promoter: sequences in the simian virus 40 late promoter required for T-antigen-mediated *trans* activation. Mol. Cell. Biol. **5**:1859–1869.
24. **Lai, C., R. Dhar, and G. Khoury.** 1978. Mapping of spliced and unspliced late lytic SV40 RNAs. Cell **14**:971–982.
25. **Lange, M., E. May, and P. May.** 1981. Ability of nonpermissive mouse cells to express a simian virus 40 late function(s). J. Virol. **38**:940–951.
26. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
27. **Messing, J., R. Crea, and P H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
28. **Mishoe, H., J. N. Brady, M. Radonovich, and N. P. Salzman.** 1984. Simian virus 40 guanine-cytosine-rich sequences function as independent transcriptional control elements in vitro. Mol. Cell. Biol. **4**:2911–2920.
29. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, 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**:6047–6067.
30. **Omilli, F., M. Ernoult-Lange, J. Borde, and E. May.** 1986. Sequences involved in the initiation of simian virus 40 late transcription in the absence of T antigen. Mol. Cell. Biol. **6**:1875–1885.
31. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase. J. Mol. Biol. **113**:237–251.
32. **Rio, D. C., and R. Tjian.** 1984. Multiple control elements involved in the initiation of SV40 late transcription. J. Mol. Appl. Genet. **2**:423–435.
33. **Scherrer, K.** 1969. Isolation and sucrose gradient analysis of RNA, p. 413–432. *In* K. Habel and N. P. Salzman (ed.), Fundamental techniques in virology. Academic Press, Inc., New York.
34. **Shaw, P. E., D. Bohmann, and A. Sergeant.** 1985. The SV40 enhancer influences viral late transcription in vitro and in vivo but not on replicating templates. EMBO J. **4**:3247–3252.
35. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–518.
36. **Tack, L. C., and P. Beard.** 1985. Both *trans*-acting factors and chromatin structure are involved in the regulation of transcription from the early and late promoters in simian virus 40 chromosomes. J. Virol. **54**:207–217.
37. **Takahashi, K., M. Vigneron, H. Matthes, A. Wildeman, M. Zenke, and P. Chambon.** 1986. Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. Nature (London) **319**:121–126.
38. **Vigneron, M., H. A. Barrera-Saldana, D. Baty, R. E. Everett, and P. Chambon.** 1984. Effect of the 21-bp repeat upstream element on in vitro transcription from the early and late SV40 promoters. EMBO J. **3**:2373–2382.
39. **Wasylyk, B., C. Wasylyk, P. Augereau, and P. Chambon.** 1983. The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell **32**:503–514.
40. **Weiher, H., M. Konig, and P. Gruss.** 1983. Multiple point mutations affecting the simian virus 40 enhancer. Science **219**:626–631.
41. **Weinberg, R. A., Z. Ben-Ishai, and J. E. Newbold.** 1974. Simian virus 40 transcription in productively infected and transformed cells. J. Virol. **13**:1263–1273.
42. **Zenke, M., T. Grundstrom, H. Matthes, M. Wintzerith, C. Schatz, A. Wildeman, and P. Chambon.** 1986. Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. **5**:387–397.