MOLECULAR AND CELLULAR BIOLOGY, May 1988, p. 2021–2033
0270-7306/88/052021-13$02.00/0

Vol. 8, No. 5

Simian Virus 40 Major Late Promoter: a Novel Tripartite Structure That Includes Intragenic Sequences

DONALD E. AYER AND WILLIAM S. DYNAN*

Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215

Received 30 November 1987/Accepted 25 January 1988

Unlike most genes transcribed by RNA polymerase II, the simian virus 40 late transcription unit does not have a TATA box. To determine what sequences are required for initiation at the major late mRNA cap site of simian virus 40, clustered point mutations were constructed and tested for transcriptional activity in vitro and in vivo. Three promoter elements were defined. The first is centered 31 base pairs upstream of the cap site in a position normally reserved for a TATA box. The second is at the cap site. The third occupies a novel position centered 28 base pairs downstream of the cap site within a protein-coding sequence. The ability of RNA polymerase II to recognize this promoter suggests that there is greater variation in promoter architecture than had been believed previously.

Synthesis of eucaryotic mRNAs requires specific DNA sequences that are located close to the transcriptional initiation site. These sequences, termed promoters, bind protein factors that position the transcriptional initiation reaction at the mRNA cap site (for review, see references 17 and 46). The binding of protein factors to promoters has also been implicated in biological regulation, for example, during the induction of transcription by heat shock (69, 74) and phorbol esters (1, 40), and in the tissue-specific expression of several genes (42). For many genes, the level of transcription is also regulated by enhancer sequences that are distinct from promoter sequences. These elements interact with promoters in an orientation- and distance-independent manner.

The sequences that make up eucaryotic promoters have been studied in detail. For protein-coding genes, which are transcribed by RNA polymerase II, the first common sequence to be identified was the sequence TATAA. These so-called TATA boxes are usually located 30 base pairs upstream of the cap site, and significant changes in transcription have been demonstrated when these sequences are mutated (12, 45, 72, 76). Recent work suggests that there may be another class of promoters for protein-coding genes that have no apparent TATA box in their 5′-flanking region. A recent survey identified 14 cellular genes without TATA boxes, most of which have housekeeping functions (20). We wanted to determine what sequences are required for the expression of a gene that does not possess a TATA box, and we chose the simian virus 40 (SV40) late transcription unit as a model system.

SV40 has been a useful system for the study of eucaryotic gene expression in the past. The circular double-stranded 5,243-base-pair genome contains only two transcription units, both of which utilize host RNA polymerase II. The early transcription unit codes for small and large T antigens. The late transcription unit codes for proteins required for virion production and is activated following the onset of DNA replication. The major class of late RNA, 16S, encodes the virion protein VP1 and the agnoprotein. This RNA initiates from nucleotide 325 and accounts for at least 85% of the RNA made late in infection (23, 54). The minor class of late RNA, 19S, encodes the virion proteins VP2 and VP3. The 19S RNA is heterogeneous at its 5′ end, with some 5′ termini mapping to nucleotide 325 and others mapping upstream (24). Because upstream starts represent a small fraction of total late RNA synthesis, we have concentrated on the major start at nucleotide 325 in this study.

Previous studies have indicated that sequences controlling SV40 late gene expression extend over a 350-base-pair region, beginning at the origin of viral replication and extending downstream to, and possibly beyond, the major late cap site at nucleotide 325. One of the major mechanisms controlling the level of late gene expression is *trans*-activation by an early gene product, large T antigen (5, 37). The sequences required for this *trans*-activation lie considerably upstream of the major late start and include sequences within the origin of replication and the 72-base-pair repeats (6, 31, 38). The locations of the upstream control sequences are flexible and, in this respect, the sequences are enhancer-like (38, 52, 64).

In contrast to the upstream regulatory elements, the sequences that position the late transcriptional initiation sites are not well understood. There is no TATA box upstream of the major late initiation site, suggesting that the SV40 late promoter may be fundamentally different from other promoters. However, it seemed likely that sequences near the transcriptional start site would control its selection. Therefore, we have focused our studies on the region surrounding the major late cap site.

Mutations have been previously described in this region, but their phenotypes are often complex (8, 22, 27, 35, 49, 52, 53, 63, 64, 71). Mutants with sequence changes near position 325 or that have deletions immediately downstream of this position do not accumulate mRNAs with 5′ termini mapping to nucleotide 325. Though these mutants are capable of productive infection, they show novel patterns of RNA synthesis; minor upstream cap sites become predominant. The results of these experiments are consistent with an atypical promoter structure requiring sequences both at and downstream of the start site. However, interpretation is complicated by several factors. Because the mRNAs are altered in structure, the observed phenotypes may arise as the result of altered splicing, transport, stability, or coding capacity rather than altered transcriptional initiation. In addition, many of the mutations involve duplications or large deletions, creating changes in the spacing of flanking sequences.

* Corresponding author.

We have reexamined the region around the nucleotide 325 cap site, using new mutants and additional methods of analysis. A series of clustered point mutations was constructed that saturates a 143-base-pair region surrounding the major late cap site at nucleotide 325. The ability of these mutant DNAs to direct transcription has been tested both in vitro and in vivo, permitting transcriptional and posttranscriptional effects to be distinguished. These experiments have defined three distinct sequence elements in the region near the transcriptional start site that are required for efficient synthesis of SV40 late RNA. These elements make up a promoter that is quite different from those that contain a TATA box. This SV40 major late promoter requires sequences upstream of the start site, at the start site, and, interestingly, downstream of the start site in protein-coding sequence. While the three elements appear to act in additive fashion to serve as a functional equivalent of a TATA box, they each have properties uniquely their own.

MATERIALS AND METHODS

Plasmid constructions. Maps of the recombinant plasmids used in these experiments are shown in Fig. 1. The stippled boxes represent SV40 sequences, while the solid lines represent pBR322 sequences. In the pS LS and the pSV LS series of mutants, the darker portion indicates the approximate region where the *Bgl*II linkers have been substituted. The direction of late transcription is indicated on each map. The numbers in parentheses indicate SV40 map positions, and the numbers not in parentheses indicate pBR322 map positions.

The wild-type plasmid used in these experiments was pSVOLO. This construct was made by replacing the SV40 *Hin*dIII fragment of pSVO1O (39) with SV40 sequences between the *Hin*dIII sites at nucleotides 5171 and 1046.

The parental plasmid that provided sequences 3′ to the *Bgl*II linker in the LS mutants was pSVN/B04. This construct was made by inserting a *Sal*I-*Bgl*II linker into the *Pvu*II site (nucleotide 273) of pSVOLO. The *Sal*I-*Bgl*II fragment of pBRN/B (28) was then inserted into this new site. The parental plasmid that provided sequences 5′ to the *Bgl*II linker in the LS mutants was pBRN/BL. This construct was made by replacing the *Hin*dIII-*Bam*HI fragment of pBRN/B (28) with the *Hin*dIII-*Hin*cII (5171-470) of SV40 by means of a *Bam*HI linker at the *Hin*cII site. The 3′- and 5′-deletion parents of the linker-scanning mutants were made by digesting pSVN/B04 and pBRN/BL with *Sal*I and *Bam*HI, respectively, followed by digestion with Bal-31 nuclease. The digested ends were filled by using the Klenow fragment of *Escherichia coli* DNA polymerase I, and the pBR322 spacer DNA was removed by *Nru*I digestion followed by ligation with T4 DNA ligase as described before (28). The exact endpoints of the SV40 sequence in the LS

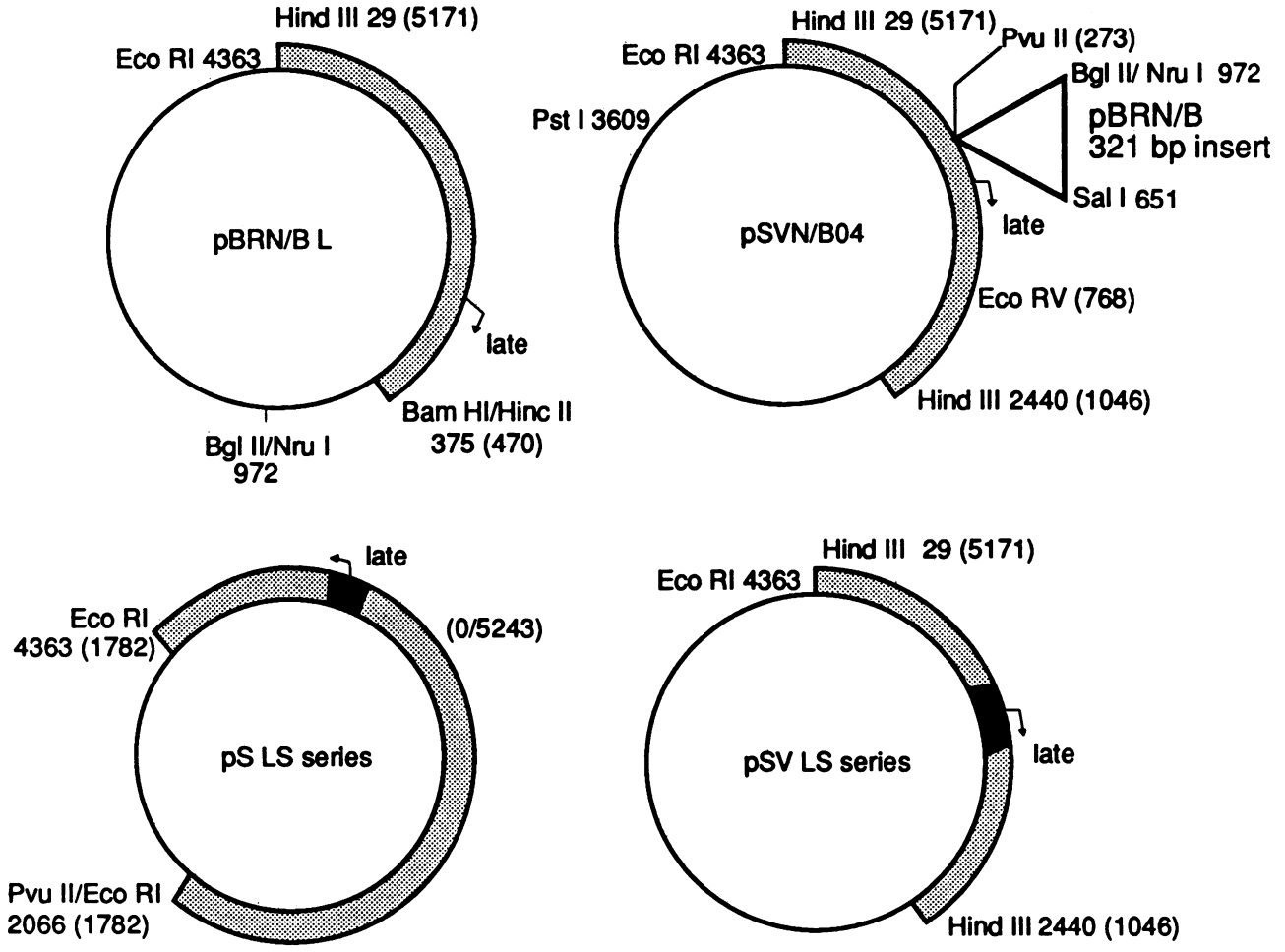

FIG. 1. Maps of recombinant plasmids. See Materials and Methods for details. bp, Base pair.

parental deletion mutants were determined by dideoxy sequencing (59). The linker-scanning mutants were made by selecting 5′- and a 3′-deletion mutant parents which had their deletion endpoints separated by 10 ± 2 base pairs and replacing the *Eco*RI-*Bgl*II fragment of the 3′-deletion parent derived from pSVN/B04 with the *Eco*RI-*Bgl*II fragment of the 5′-deletion parent derived from pBRN/BL to give the structure shown (pSVLS series). The sequence of each LS mutant was confirmed by dideoxy sequencing.

The linker-scanning mutants, pSLS series, used in transfection experiments were made by replacement of the *Sfi*I-*Eco*RV fragment of pS312 (19) with the analogous fragment from pSVOLO or the LS mutants described above. The wild-type plasmid used for transfection was called pSOLO.

The double and triple mutants were constructed from LS17, LS21, and LS24 and were made by using the unique *Asp*718 and *Hpa*II sites at SV40 map positions 294 and 346, respectively. The construction of the triple mutant required a tetramolecular ligation. These constructs were isogenic with the LS parental mutants.

pSVN/B06 was made by replacing the *Bam*HI-*Eco*RI fragment of pBRN/B (28) with the *Xho*I-*Eco*RI fragment, containing the SV40 control region, of pX8 (19), using a *Bam*HI linker at the *Xho*I site. The upstream deletions were made by digesting pSVN/B06 with *Bam*HI followed by digestion with Bal-31 nuclease. The digested ends were filled by *E. coli* Klenow fragment, and the pBR322 spacer DNA was removed by *Nru*I digestion. The unique digestion endpoint was determined by dideoxy sequencing.

The construct, pT7/SV40L, used to generate late cRNA was made by insertion of the *Eco*RI-*Hin*cII fragment of pSVOLO, containing the SV40 control region, into the polylinker of pT7.2 (U.S. Biochemical).

DNA from all of the above clones was isolated by standard methods from *E. coli* HB101 grown in M9 salts supplemented with Casamino Acids and amplified with chloramphenicol. All transcription template DNAs were purified on CsCl-EtBr gradients.

Cell culture and extracts. Whole-cell extracts were prepared by the method of Manley et al. (43) from suspension cultures of HeLa cells. The cells were harvested at a density of 4×10^5 to 6×10^5 cells per ml. The HeLa cell cultures were maintained in Joklik modified minimal essential medium supplemented with 2 g of Na HCO_3, 110 mg of pyruvate, 8.9 mg of L-alanine, 15 mg of L-asparagine · H_2O, 13.3 mg of L-aspartic acid, 14.7 mg of L-glutamic acid, 7.5 mg of glycine, 11.5 mg of L-proline, 10.5 mg of L-serine, and 50 ml of calf serum (M.A. Bioproducts) per liter. Cell density was maintained at 2×10^5 to 5×10^5 cells per ml. BSC-1 cells were grown under 5% CO_2 in 100- or 150-mm dishes in Dulbecco modified Eagle medium supplemented with 7.5% calf serum.

In vitro transcription. Purified supercoiled DNA, 20 μg/ml, was incubated in HeLa whole-cell extracts for 30 min at 30°C in the presence of 250 μM nucleoside triphosphates–25 mM Tris hydrochloride (pH 7.9)–50 mM KCl–6.25 mM $MgCl_2$–0.5 mM dithiothreitol–0.5 mM EDTA–10% glycerol. In a typical 50-μl reaction, 25 μl of extract was used. The amount of extract required for optimal levels of late RNA synthesis was determined empirically for each extract preparation. At 30 min, the reactions were stopped with 2 volumes of 200 mM EDTA–200 mM NaCl–1% sodium dodecyl sulfate–100 μg of carrier tRNA per ml. The reactions were phenol-chloroform-isoamyl alcohol extracted and precipitated with 2.5 volumes of 0.5 M ammonium acetate in absolute ethanol. The nucleic acid pellet was suspended in 400 μl of 0.3 M sodium acetate (pH 7.0) and precipitated with 2.5 volumes of absolute ethanol. The final pellet was washed with 70% ethanol. For primer extension, the RNA was suspended in 8 μl of 50 mM Tris hydrochloride (pH 8.3)–60 mM NaCl–10 mM dithiothreitol containing 100 fmol of ^{32}P-end-labeled primer ([γ-^{32}P]ATP, > 7,000 Ci/mmol; ICN Pharmaceuticals, Inc.). The primer was annealed to the RNA for 5 min at 60°C followed by a 5-min incubation on ice. To the primer-template mix 8 μl of 25 mM Tris hydrochloride (pH 8.3)–30 mM NaCl–15 mM $MgCl_2$–1.25 mM dithiothreitol–1 mM each deoxynucleoside triphosphate–0.6 U of avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc.) was added, and the reaction was allowed to proceed for 45 min at 37°C. The reactions were stopped by addition of an equal volume of formamide containing 0.2% xylene cyanol and 0.2% bromphenol blue, and the cDNAs were resolved on a 0.4-mm-thick 8% polyacrylamide–8.3 M urea sequencing gel. The dried gel was exposed for autoradiography overnight at −70°C with an intensifying screen.

Runoff transcription was performed essentially as described for the primer extension assays except for the following: the template was digested with the appropriate restriction enzyme, the nonradioactive CTP concentration was decreased to 12.5 μM, and 10 μCi of [α-^{32}P]CTP (Du Pont-New England Nuclear Corp.) was added. The transcription products were extracted and precipitated as described above, and the final pellet was suspended in formamide plus dyes. The RNA was analyzed on 1.5-mm-thick 4% polyacrylamide–8.3 M urea gels electrophoresed at 12 V/cm. Autoradiography was as described above.

Transfections and analysis of nucleic acids. A 10-μg portion of purified supercoiled DNA was used to transfect approximately 7×10^6 BSC-1 cells at 80 to 90% confluence in 150-mm plates by the DEAE-dextran procedure (65). Following a 48-h incubation, the cells were harvested by scraping. The cytoplasmic fraction was collected as described by Chikaraishi and Danna (11), and the cytoplasmic RNA was purified by phenol extraction and ethanol precipitation. The nuclei were suspended in 200 μg of proteinase K (Boehringer Mannheim Biochemicals) per ml–0.1% sodium dodecyl sulfate–0.2 M Tris hydrochloride (pH 7.5)–0.3 M NaCl. Following a 90-minute incubation at 37°C and phenol-chloroform-isoamyl alcohol extraction, the nucleic acids were precipitated with 2.5 volumes of absolute ethanol and suspended in 10 mM Tris hydrochloride (pH 7.9)–1 mM EDTA (TE). The nucleic acids were then digested with 50 μg of RNase A per ml for 60 min at 37°C, followed by phenol-chloroform-isoamyl alcohol extraction and precipitation with ethanol. The pellet was suspended in TE, and the concentration of the nucleic acid was determined spectrophotometrically. A 5-μg amount of this nuclear DNA was digested with 25 U of *Eco*RI overnight at 37°C. The digestion products were resolved on a 1% agarose gel and transferred to Magna nylon 66 membrane (Micron Separations Inc.) and hybridized with a SV40 late specific RNA probe as described by Southern (66) with the modifications of Morin and Cech (48). The SV40 late specific RNA probe was synthesized by T7 RNA polymerase (gift of A. Zaug) from *Eco*RI-digested pT7/SV40L in 40 mM Tris hydrochloride (pH 8.0)–20 mM $MgCl_2$–5 mM dithiothreitol–4 mM spermidine–100 μCi of [α-^{32}P]CTP–0.5 mM concentrations of all four nucleoside triphosphates at 37°C for 60 min. Primer extension of the cytoplasmic RNA was carried out as described above, using 5 μg of RNA as the template.

RESULTS

In vitro transcription of wild-type DNA. To determine which sequences make up the SV40 major late promoter, we first established an in vitro transcription system. The *Hin*dIII C fragment spanning the SV40 origin of replication (map position, 5171 to 1046) and containing sequences sufficient for both early and late gene expression was inserted into recombinant plasmid pSVO1O (39) to give recombinant plasmid pSVOLO. This plasmid was digested with *Eco*RV and *Pst*I to provide termination sites for runoff transcription. A RNA runoff pattern generated with this template and a HeLa whole-cell extract is shown in Fig. 2A. The largest SV40-specific transcript, 853 nucleotides in length, was synthesized from the SV40 early promoter. Two other SV40 RNAs, 598 and 443 nucleotides in length, were transcribed in the late direction. The synthesis of all three RNAs was sensitive to low concentrations of the toxin α-amanitin, indicating that they are transcribed by RNA polymerase II. The higher-molecular-weight background RNAs seen in this experiment probably result from end-to-end transcription of the restriction fragments or initiation within the vector sequences.

To confirm the orientation and determine the nucleotide from which the SV40 late transcripts originated, we analyzed the RNA synthesized in vitro by primer extension. A synthetic oligonucleotide complementary to SV40 late RNA was hybridized to the in vitro transcription products and extended with reverse transcriptase to produce two populations of cDNA (Fig. 2B). When these cDNAs are aligned with a dideoxy sequence ladder synthesized from the same oligonucleotide primer and DNA template, it is apparent that the in vitro late RNAs originate from nucleotide 325 and approximately nucleotide 170. The cDNA representing the nucleotide 325 start appears as a doublet, which may be an artifact of reverse transcription. The lower band corresponds to the cap site reported in the literature (24). Our results with the in vitro transcription system are similar to those reported previously by other laboratories (8, 29, 30, 56, 57, 70).

Sequences required in vitro for SV40 late transcription. To determine the 5′ boundary of the sequences required for transcriptional activity from nucleotide 325, a series of mutants that progressively remove greater amounts of upstream sequence was constructed. These mutants were transcribed in vitro with a HeLa whole-cell extract, and the RNA was assayed by primer extension (Fig. 3). PSVN/B06 contains the SV40 control region from the origin of replication, map position 1, to the downstream *Eco*RI site at nucleotide 1782. This plasmid directed transcription from nucleotide 325, although the level of transcription was

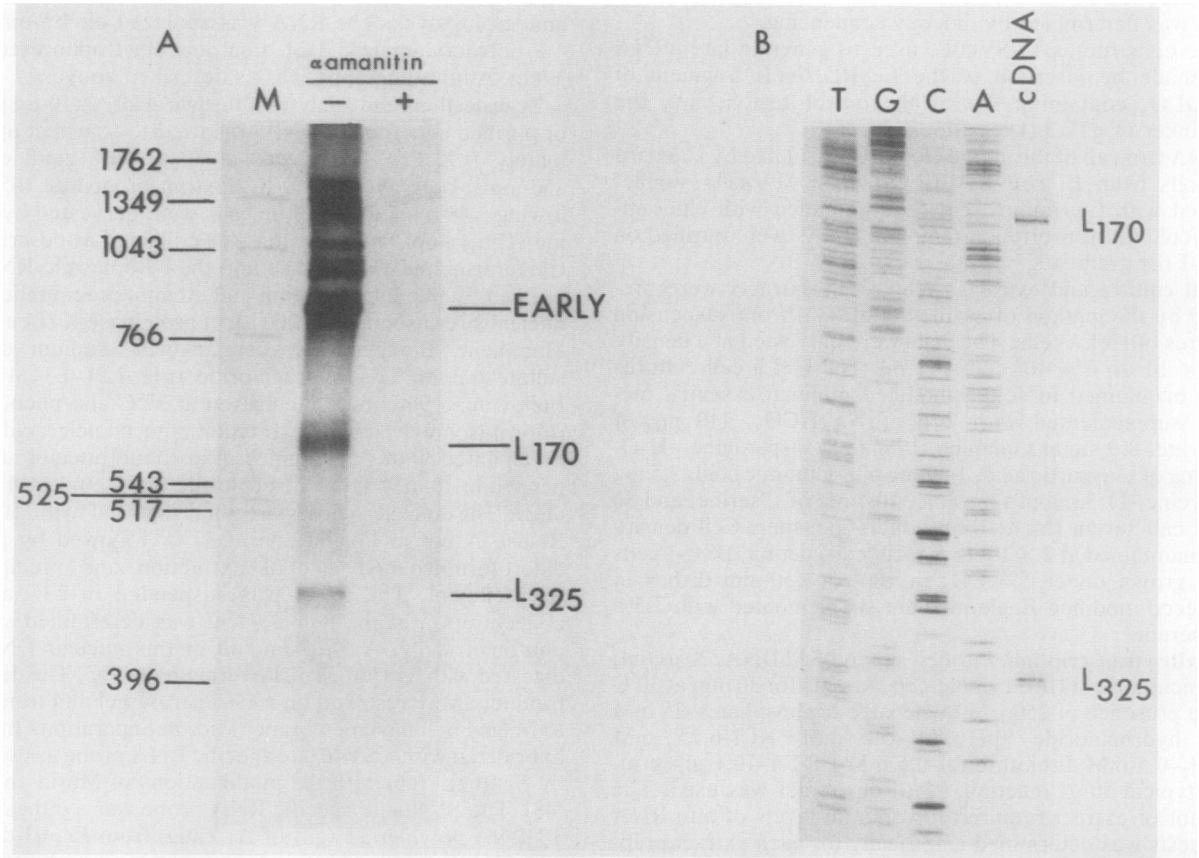

FIG. 2. Analysis of RNAs transcribed in vitro in HeLa whole-cell extracts. (A) Runoff transcription assay, using HeLa whole-cell extracts and pSVOLO digested with *Eco*RV and *Pst*I in the absence (−) and presence (+) of 1 μg of α-amanitin per ml. The three SV40-specific RNAs are denoted Early, L_{170}, and L_{325}. Lane M contains marker DNA with the size indicated in base pairs. (B) Lanes T, G, C, and A are dideoxy sequence ladders of the pSVOLO template. The cDNAs were generated from RNA synthesized from pSVOLO in vitro. The sequence around nucleotide 325 is 5′-AGAGGTTATTTCAGG-3′, and that around nucleotide 170 is 5′-CTGGGACTTTCCACAC-3′. In each case the initiating nucleotide is underlined.

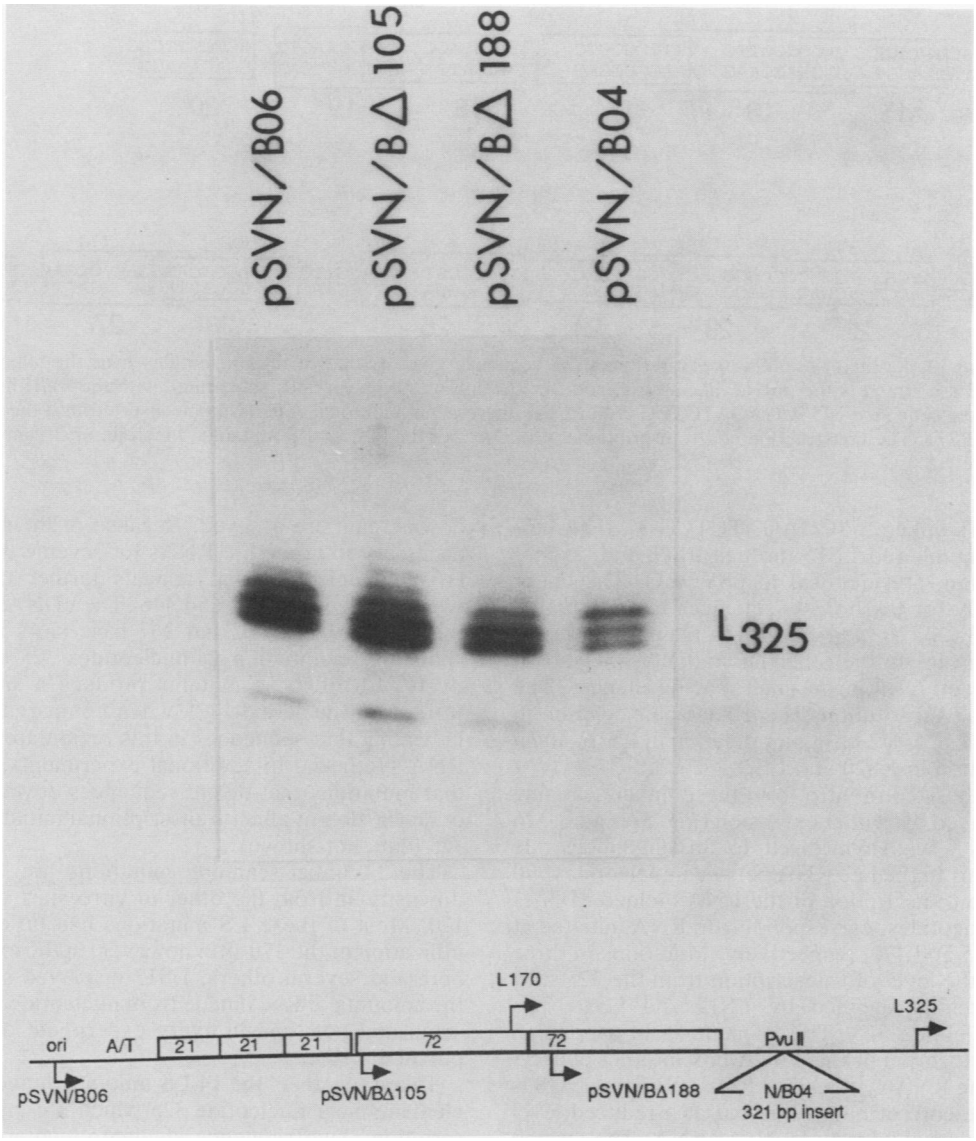

FIG. 3. Primer extension of the late RNA synthesized in vitro from the upstream deletion mutants. The mutant DNAs given at the top were transcribed in vitro with HeLa whole-cell extracts and primed for cDNA synthesis as described in Materials and Methods. L_{325} marks the position of the cDNA that corresponds to RNA initiated at nucleotide 325. The diagram at the bottom is of the SV40 control region. The 5′ boundary of each deletion mutant is denoted by the rightward arrows at the bottom of the diagram. A 321-base-pair (bp) fragment was inserted into the control region at the *Pvu*II site at nucleotide 273 as indicated.

slightly lower than that with pSVOLO (data not shown). PSVN/BΔ105 has sequences deleted between nucleotides 1 and 105 and therefore does not have the majority of the SV40 early promoter, including the TATA box and the GC motifs that bind transcription factor Sp1 (16). PSVN/BΔ188 has an additional 83 nucleotides deleted, removing a portion of the sequences responsible for SV40 enhancer function, including the early-proximal, 72-base-pair repeat and the first 10 nucleotides of the adjacent 72-base-pair repeat. There was no significant decrease in the transcriptional activity from nucleotide 325 with either pSVN/BΔ105 or pSVN/BΔ188. A fourth mutant, pSVN/B04, was constructed by inserting a 321-base-pair fragment of plasmid DNA into the SV40 genome at the *Pvu*II site at nucleotide 273. The amount of RNA synthesized with this mutant was two- to threefold lower than the amount with pSVN/B06. A similar result was obtained when pSVN/B04 was digested within the inserted sequences so that the control region was placed completely in *trans* from the 325 start (data not shown). These data show that the 52 base pairs upstream of the RNA start are sufficient to direct a substantial amount of RNA synthesis. Moreover, in the four clones assayed, nucleotide 325 was the predominant initiating nucleotide; therefore, sequences required for the selection of the initiating nucleotide must be intact.

To examine the region around nucleotide 325 at higher resolution, we constructed a series of linker-scanning mutants (45). These clustered point mutations extend from nucleotide 263 to 406 and essentially saturate a 143-base-pair region around the start site (Fig. 4). There are a total are 14 mutants encompassing both the 52-base-pair region described above and sequences downstream of the cap site important for transcription from nucleotide 325. The boxes in Fig. 4 represent sequences substituted by a 10-base-pair

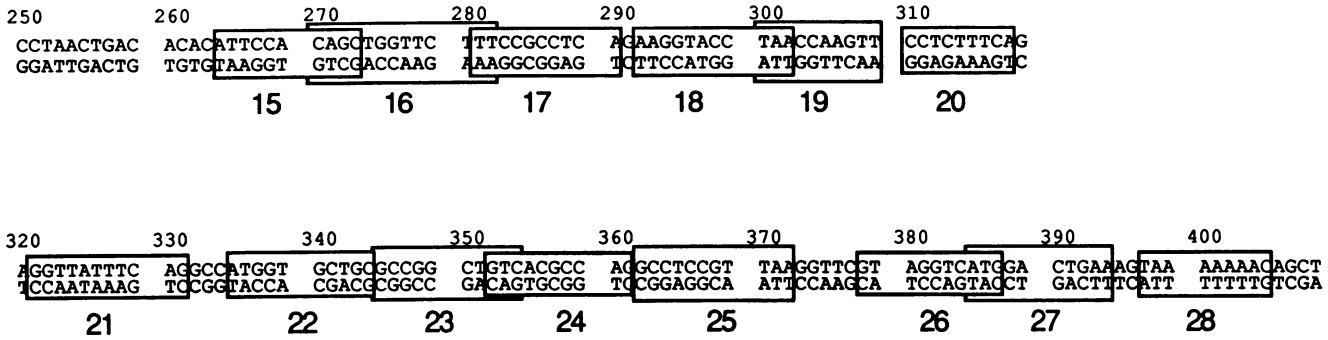

FIG. 4. Locations of the linker-scanning substitutions. The sequence given is that of the nucleotides near the major late start site at nucleotide 325. The top strand is the mRNA-like strand, and the location of the nucleotides substituted with the *Bgl*II linker is denoted by the boxes. The linker sequence is 5′-CGAGATCTCG-3′. In LS16 there is an additional A/T base pair inserted into the sequence between nucleotides 270 and 271. The construction and transcriptional properties of the first 14 LS mutants, which lie upstream of this region, are described elsewhere (15).

palindromic *Bgl*II linker, 5′-CGAGATCTCG-3′. The mutants have been numbered LS15 through LS28 and, except for the substitution, are identical to pSVOLO. The linker substitutes exactly for the boxed sequences in LS mutants 15, 17, 19, 22, 23, 24, 26, and 27; therefore, no changes in spacing are produced. In the other LS mutants, the substitution was not exact, resulting in small spacing changes. The net result of these substitutions is a 1-base-pair deletion in LS18, LS21, and LS25, a 2-base-pair deletion in LS16, and a 1-base-pair insertion in LS20 and LS28.

The RNA synthesized in vitro from these linker-scanning mutants was assayed by primer extension (Fig. 5A and B). In Fig. 5A the RNA was hybridized to an oligonucleotide primer that anneals to the late RNA between nucleotides 370 and 394. Reverse transcription of the RNA yielded cDNAs of 68 and 223 nucleotides, corresponding to RNA initiated at map positions 325 and 170, respectively. Mutations in three regions affected the level of transcription from the 325 start site. The first region, defined by LS17 and LS18, lies upstream of nucleotide 325. This element is in a position analogous to the location of the TATA box in other eucaryotic genes (4). The RNA synthesized from LS17 and LS18 is initiated from the correct nucleotide but at a reduced level when compared with wild-type levels of RNA. The second region, defined by LS21, overlies the start site of transcription. The RNA synthesized from this mutant is reduced in amount, and the initiation sites are shifted slightly from the position corresponding to nucleotide 325 in the wild type. Scanning densitometry suggests that the total amount of RNA synthesized from LS21 was about fourfold less than the total RNA synthesized from the wild-type template. The third region, defined by LS23 and LS24, lies downstream of the start site. Mutations in this region decrease the amount of RNA made from nucleotide 325 without altering the start site.

To determine whether the reduced amount of RNA synthesized from LS23 and LS24 was due to the introduction of sequences that repress transcription, we constructed a deletion that removes sequences between nucleotides 332 and 353. Because of the internal deletion, the cDNA synthesized from this RNA is 20 bases smaller than the others. As shown in the lane marked Δ, this deletion mutant, like LS23 and LS24, produces RNA from the correct start site (data not shown) but at a reduced level. This indicates that LS23 and LS24 act by altering a positive *cis*-acting element rather than by supplying sequences that inhibit RNA synthesis.

In Fig. 5B the linker-scanning mutations that lie farther downstream are assayed. Because of their location, it was necessary to prime the RNAs for reverse transcription with an oligonucleotide that anneals farther downstream, between nucleotides 422 and 446. The cDNAs generated from this primer were 120 and 267 base pairs in length, corresponding to initiation at nucleotides 325 and 170, respectively. All of these mutants produce a similar amount of RNA from nucleotide 325 when compared with wild type, indicating that sequences in this region are not required for RNA synthesis. In additional experiments, we have shown that mutations that delete sequences downstream of nucleotide 406 do not affect transcriptional initiation at nucleotide 325 (data not shown).

The 14 linker-scanning mutations are >100 base pairs downstream from the other in vitro start site at nucleotide 170. Most of these LS mutations had little or no effect on utilization of the 170 site; however, in the experiment shown here and several others, LS17 displayed some decrease in the amounts of RNA made from nucleotide 170. The effect is small and not seen in every experiment and has not been pursued further.

Taken together, the 14 LS mutants define three *cis*-acting elements near nucleotide 325 which are important for transcription. Two sequence elements, centered 31 base pairs upstream and 28 base pairs downstream of the start site, are required for efficient transcription, while sequences at the start site are required for efficient transcription and for determination of the initiating nucleotide.

Sequences required in vivo for SV40 late transcription. To test whether the control of SV40 late transcription requires the same sequences in vivo, we transfected wild-type and mutant DNAs into the permissive monkey cell line BSC-1 and assayed transient SV40 RNA levels. The constructs used in these experiments were made by transferring the wild type and select LS mutants into a vector that has an intact SV40 early region and produces functional T antigen (19). These plasmids contain the splice donor and acceptor sites for the 16S and 19S populations of the SV40 late RNA but are not expected to produce late proteins because the late region is interrupted by vector sequences.

Supercoiled DNA was transfected into BSC-1 cells by the DEAE-dextran procedure (65). After a 48-h incubation, cytoplasmic RNA and nuclear DNA were isolated. To determine the relative efficiencies of transfection, the nuclear DNA from these cells was linearized and analyzed by Southern blotting (Fig. 6A). Only small differences in the amount of transfected DNA were observed; therefore, com-

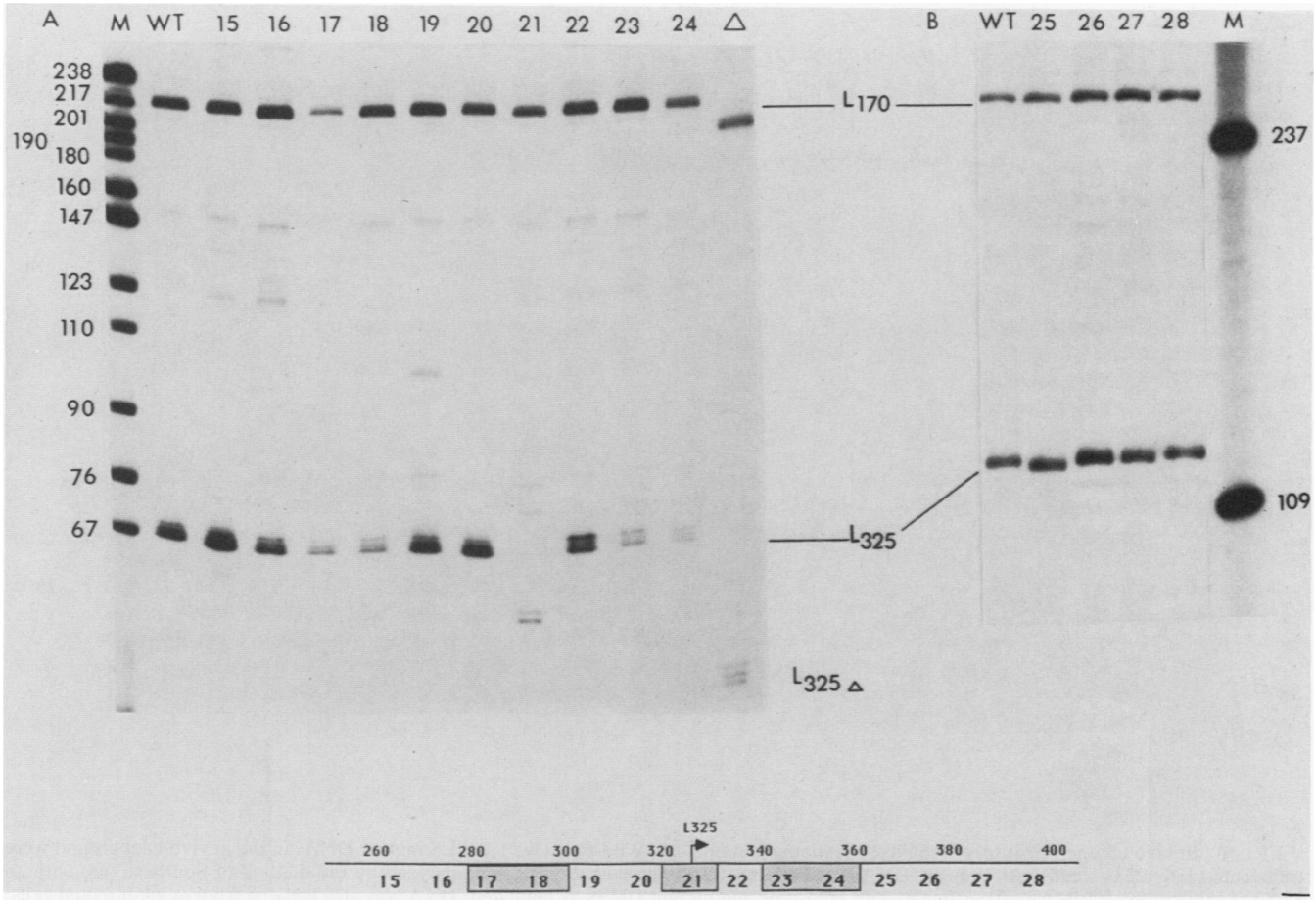

FIG. 5. Primer extension of the RNA synthesized in vitro from the linker-scanning mutants. The LS mutants were transcribed in vitro with HeLa whole-cell extracts, and the resultant RNA was primed for reverse transcription as described in Materials and Methods. The cDNAs shown in panel A were synthesized from a primer that anneals between nucleotides 370 and 394. The cDNAs shown in panel B are synthesized from a primer that anneals to sequences between nucleotides 422 and 446. The numbers at the top of each lane indicate which LS mutant served as the template in the in vitro transcription reaction. WT designates the pSVOLO template. The lane marked Δ contains the cDNA synthesized from a template with a deletion between nucleotides 332 and 353. L_{170} and L_{325} mark the positions of the cDNAs corresponding to RNA initiated at nucleotide 170 and 325, respectively. $L_{325\Delta}$ marks the shortened cDNA synthesized from the RNA produced from the internal deletion mutant. In both panels A and B, lane M contains marker DNA with the size indicated in base pairs. The diagram at the bottom roughly marks the location of each linker-scanning mutant. Stippled boxes indicate which LS mutations had an effect on transcription from nucleotide 325.

parisons of the RNA synthesized from nucleotide 325 with each mutant should primarily reflect transcriptional activity rather than different copy numbers in the cell.

SV40 late RNA in the cytoplasm was assayed with a primer that anneals to the leader sequences of the 16S RNA (Fig. 6B). Only one RNA species was detected. Primer extension in the presence of dideoxynucleoside triphosphates (data not shown) demonstrated that the 5′ terminus mapped to nucleotide 325, the same position as seen previously in vitro. The pattern of expression in vivo is also similar to that seen in vitro. There are three regions, defined by LS17, LS21, and LS24, required for wild-type levels of RNA. The upstream and downstream mutants, LS17 and LS24, produce less RNA than wild type but do so from the correct nucleotide. The start site mutant, LS21, accumulates significantly less RNA than the wild type. Longer exposures of this experiment indicate that the RNA synthesized from LS21 is initiated primarily at the two nearby upstream positions that were evident in Fig. 5A.

Viable deletion mutants of SV40 have previously been reported that are missing either the 325 start or sequences downstream of this nucleotide (22, 26, 27, 52, 53, 64). These mutants compensate for the loss of the 325 start site with initiation at positions upstream of nucleotide 325. We were concerned that such RNAs might have been present in our experiments but were not detected because RNA from upstream starts can be spliced in an alternative fashion to produce molecules that cannot anneal to a 16S leader-specific primer (63). To test this possibility, we performed primer extension on the cytoplasmic RNAs, using oligonucleotide primers complementary to the bodies of the 16 and 19S RNA rather than to the leader region.

The results of the experiment with the 16S body-specific primer are shown in Fig. 6C. Only the one RNA species shown was detected. The 5′ terminus mapped to position 325, and the length of the cDNA corresponds to RNA with a normal 16S splice between nucleotides 526 and 1463. The relative amount of this RNA detected with the different templates was the same as seen with the leader primer. These results suggest that LS17, LS21, and LS24 reduced the overall amount of 16S RNA synthesized and that the

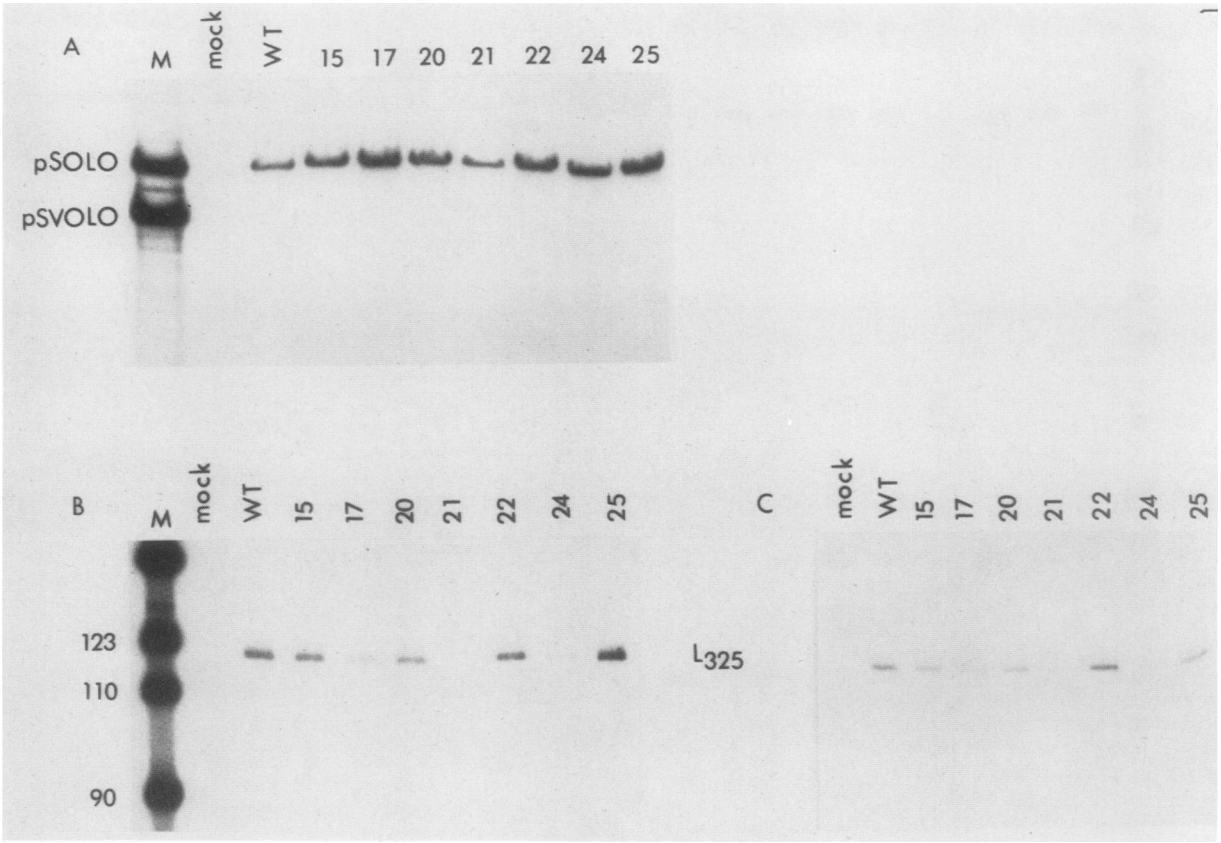

FIG. 6. In vivo characterization of the linker-scanning mutants. (A) Wild-type (WT) and LS mutant DNAs in the in vivo background were transfected into BSC-1 cells. At 48 h the DNA was collected, digested with *Eco*RI, and analyzed by the method of Southern (66) with an SV40-specific probe. Cells were transfected with wild-type and mutant DNAs as indicated. DNA was also isolated from mock-infected cells. The markers in this experiment are pSVOLO and pSOLO digested with *Eco*RI. pSOLO was the wild type used in this experiment and contains sequences from pSVOLO introduced into the in vivo background as described in Materials and Methods. Cytoplasmic RNA was purified from the same cell cultures that yielded the DNA analyzed in panel A. This RNA was primed for reverse transcription with an oligonucleotide that anneals to the leader of the 16S RNA between nucleotides 370 and 394. The resultant cDNAs are shown in panel B. Lane M contains marker DNA with the size indicated in base pairs. The experiment shown in panel C is identical to that in panel B except that the cytoplasmic RNA has been primed for reverse transcription with an oligonucleotide that anneals to the body of the 16S RNA between nucleotides 1477 and 1492. L_{325} marks the location, in both panels B and C, of the cDNA corresponding to RNA initiated at nucleotide 325.

reductions in initiation at the 325 site were not offset by increases in initiation at upstream positions.

In similar experiments with a 19S body-specific primer, no SV40 RNAs were detected. This suggests that in our transfections, as in normal viral infection, 19S mRNA is a minor species. Because of the inadequate sensitivity of the assay, we were unable to draw conclusions about whether initiation of 19S RNA at upstream sites was stimulated in the mutants.

Mechanism of control by downstream sequences. The downstream sequences necessary for late transcription are transcribed into RNA, and it is possible that these sequences are required for the stability or processing of the RNA molecule rather than initiation per se. For example, introduction of mutant sequences into the RNA could render it more sensitive to degradation; however, because the sequences introduced are the same in all mutants, this effect would have to be position dependent. This possibility was tested by examining the stability of the RNAs synthesized in vitro from the wild-type template and two of the downstream LS mutants. RNA was synthesized in vitro from pSVOLO, LS23, and LS24 in duplicate reactions. After 30 min, 1 μg of α-amanitin per ml was added to inhibit RNA polymerase II and one set of three reactions was immediately extracted with phenol and placed on ice. The other set of reactions was allowed to incubate for an additional 30 min at 30°C. Following this incubation, the RNA from both sets of reactions was prepared for primer extension. Figure 7 shows that there was little or no degradation of either mutant or wild-type RNA when the samples were incubated for additional time prior to extraction. This suggests that the mutant and wild-type RNAs are equally stable in vitro, and as such, the downstream element probably does not act by maintaining the integrity of the RNA.

Another mechanistic possibility is that the substitution of mutant sequences downstream either destroys an antitermination function or introduces a position-dependent terminator. By performing in vitro transcription in the presence of [α-^{32}P]CTP, we have attempted to identify small labeled RNAs that would be indicative of premature termination events specific to LS23 and LS24. No short RNAs were detected; therefore, it seems unlikely that the downstream element is an antiterminator or that the introduction of mutant sequences supply an artifactual terminator. On the basis of these results, we suggest that the downstream element influences the rate of transcriptional initiation or steps prior to initiation at nucleotide 325.

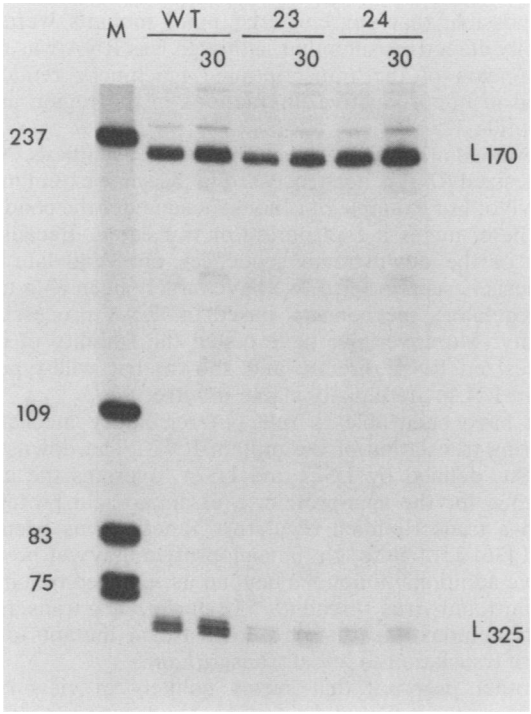

FIG. 7. Stability of wild-type and mutant RNAs synthesized in vitro by a HeLa whole-cell extract. pSVOLO and LS23 and LS24 were transcribed in vitro, and cDNAs were synthesized from an oligonucleotide that anneals to the RNA between nucleotides 370 and 394. The unmarked lanes contain cDNAs synthesized under our standard conditions. The lanes marked 30 contain cDNAs synthesized from RNA that has been incubated in the HeLa whole-cell extract for an additional 30 min as described in the text. L_{170} and L_{325} mark the locations of the cDNAs corresponding to RNA initiated at nucleotides 170 and 325, respectively. Lane M contains marker DNA with the size indicated in base pairs.

Interactions among the three control elements. We have investigated whether the three control elements act independently by constructing three double mutants that combine LS representatives that define each *cis*-acting element (LS17, LS21, and LS24) in each possible pairwise arrangement. The triple mutant from these LS parents was also constructed. If one element required another to perform its function, one would expect that a double mutant would produce the same amount of RNA as either one of its singly mutated parents. By contrast, if the elements function independently, one expects a double mutant to be more severely affected than either of its parents.

The double and triple mutants were transcribed in a HeLa whole-cell extract, and the resultant RNA was assayed by primer extension (Fig. 8). Each of the single mutants behaved as shown before in Fig. 5A and B. The double mutant that combines LS17 and LS24 synthesizes reduced amounts of RNA relative to either of the singly mutated parents. This suggests that the upstream and downstream elements function independently of one another. By contrast, the results with LS21 are less clear-cut. The amount of the minor RNA species synthesized from LS21 appears to be reduced by LS17 but only weakly affected by LS24. As such, it appears that the start site element might interact with the downstream element to facilitate efficient transcription.

There are other cDNAs that correspond to RNA synthesized in the triple mutant, which probably arise from RNA initiated at upstream positions. These cDNAs were only detectable with the triple mutant, and this could be the result of the severity of the mutation. It is possible that in the single and double mutants the transcriptional machinery is sequestered in nonproductive interactions at the 325 initiation site, but when the triple mutant serves as the template, the transcriptional machinery is released and becomes available to utilize upstream initiation sites.

It is clear from longer exposures of the experiment shown in Fig. 8 that the double mutant that combines LS17 and LS24 produces RNA from nucleotide 325, although at much reduced levels. By contrast, none of the mutants with the LS21 mutation produce any RNA from this position within the linker. This suggests that the sequences replaced by the linker in LS21 are both necessary and sufficient for the selection of nucleotide 325 as the initiating nucleotide. However, proof of this will require the demonstration that these sequences can function as a promoter in a heterologous background.

DISCUSSION

Structure of the SV40 late promoter. In the present experiments we have defined three elements that control transcription from the predominant SV40 cap site. These three *cis*-acting elements are different in both sequence and position from the elements of other known promoters. In spite of this dissimilarity, they are able to interact with the cellular

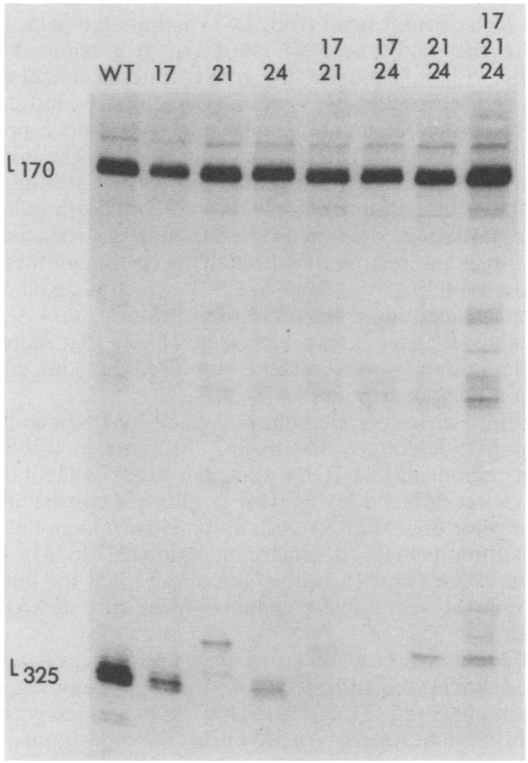

FIG. 8. Three functionally independent promoter elements. The double and triple mutants were transcribed in a HeLa whole-cell extract and the cDNAs were synthesized as described in Materials and Methods. The numbers at the top of each lane indicate which LS mutant or LS mutant combination served as the template for the in vitro transcription. LS17, LS21, and LS24 define the upstream, start site, and downstream elements, respectively. L_{170} and L_{325} mark the locations of the cDNAs which terminate at these nucleotides.

transcriptional machinery to carry out the functions of a promoter, specifying the position of transcriptional initiation and modulating the efficiency of RNA synthesis.

The first of the three elements lies upstream of the start site, in a position analogous to a TATA box. This element, defined by LS17 and LS18, extends from nucleotide 299 upstream to at least nucleotide 290 and it may extend farther upstream to nucleotide 283. It includes the proposed late promoter element affected by point mutations introduced at the *KpnI* site at nucleotide 294 (8, 19). Alterations in this upstream element reduce the amount of late RNA synthesized without altering the position from which transcription initiates. This phenotype contrasts with that seen with mutations in the TATA box region of the SV40 early promoter (3, 44), which lose the ability to initiate from the correct nucleotide without a decrease in the total amount of early RNA synthesis. We conclude that the upstream late element is analogous only in position and not in function to a TATA box. This is consistent with prior work suggesting that there is no measuring function associated with sequences upstream of nucleotide 325 (7, 27, 52, 64).

The second of the three promoter elements is defined by LS21 and spans nucleotides 320 to 332. The present findings are consistent with prior reports suggesting that sequences immediately surrounding the cap site are required for initiation at all of the SV40 late start sites (52, 64). This contrasts with many other promoters transcribed by RNA polymerase II, in which substitutions may be introduced relatively freely at the cap site without impairment of promoter function (21, 25, 45, 60).

The RNA synthesized from LS21 is initiated at positions close to the wild-type RNA start but at a reduced level. Perhaps with this mutant the transcriptional machinery is positioned correctly by accessory factors bound to the flanking control elements, but is unable to find a position within the linker that can be used efficiently for the initiation of transcription. Unable to utilize the correct starting position, the transcription complex is still in position to use cryptic initiation sites nearby, although at reduced efficiency. It is interesting to note that, even though there is a potential initiating A present in LS21, which is displaced by only one nucleotide from its normal position, this A is not used as a start site. It therefore seems likely that more than just the initiating nucleotide is required for the efficient initiation of transcription.

The third promoter element is defined by LS23 and LS24 and spans nucleotides 344 to 362. Mutants in this downstream region initiate RNA synthesis from nucleotide 325 but do so at reduced levels. Our results are consistent with several prior observations with mutants with large deletions or substitutions in this downstream region (22, 26, 53). Based on other work (27, 64), it does not appear that the initiating nucleotide is set a fixed distance from this downstream element.

In previous work, when downstream sequences have been mutated, increased utilization of minor upstream cap sites has been observed (22, 26, 53). This has been seen with both 16S and 19S late RNAs. We are uncertain why stimulation of upstream starts was not observed in our experiments. The explanation may be differences in experimental design, as most of the earlier experiments used cells infected with stocks of viable mutant virus. In contrast, we either transfected with recombinant plasmid or measured RNA synthesis in vitro. It is possible that the upstream shift requires late gene products or high levels of replication or is in some other way dependent on conditions unique to the infected cell. It is also possible that our clustered point mutants were only partially defective, such that, although less RNA was made, components of the transcriptional machinery remain involved in nonproductive interactions in the region around nucleotide 325.

How the downstream element affects RNA synthesis. Newly synthesized RNA is not processed to the same extent in vitro as in vivo. For example, it is not spliced under the conditions used here, nor is it transported or translated. Because the effect of the downstream sequences on SV40 late RNA synthesis is seen in vitro, we have largely been able to rule out regulatory mechanisms based on RNA processing or stability. Moreover, we have tested the stability of newly synthesized RNA directly and shown that wild-type and mutant RNAs are equally stable in vitro.

We have been able to rule out regulatory mechanisms requiring translation of the mutant RNA. The downstream element, defined by LS23 and LS24, overlaps the coding sequence for the agnoprotein, a 61-amino-acid protein for which a transcriptional regulatory function has been proposed (36, 35). Although the agnoprotein may yet be found to have additional functions beyond its reported role in VP1 transport and virus spread (9, 55), our in vitro transcription results demonstrate that the downstream mutants do not require translation to affect transcription.

Another proposal that seems unlikely in view of the present results is that nucleic acid secondary structure in the downstream region might be important in transcriptional regulation. It has been suggested that stem-and-loop structures in the RNA could provide a processing signal or that analogous structures in the DNA could serve as a recognition signal for transcription factors (23, 27, 36, 52). However, we find no correlation between transcriptional activity in the LS mutants and the potential for secondary structure formation. For example, LS22 alters one-half of an inverted repeat found downstream of the start site, yet synthesizes wild-type levels of RNA. LS23 and LS24 interrupt the other half of the repeat and produce reduced levels of RNA. Furthermore, when assayed in vitro, mutant templates have the same phenotype whether they are linear or supercoiled (data not shown). Linear DNAs are not subject to torsional strain and thus are unlikely to spontaneously form stem-loop structures (47).

In principle, the downstream element might act either by influencing initiation or by changing the rate of elongation through the mutated region. We have attempted to identify short RNAs synthesized from mutant templates, which might be indicative of an effect on elongation. So far, we have failed to find evidence for increased synthesis of short RNAs in the mutants. Therefore, we believe that the most likely mechanism for the downstream element is an effect on the transcriptional initiation process itself.

The RNA polymerase II transcriptional initiation reaction consists of several steps, as evidenced by the ability to uncouple these steps by using the nonionic detergent Sarkosyl (32, 33). The downstream element could affect the rate of formation or the structure of any of the complexes along the pathway to an actively transcribing RNA polymerase, probably by interacting with a protein factor required for the initiation process. We have failed so far to detect such a protein by DNase I protection or gel retention methods, perhaps because it is present in low concentration or the binding is weak. Fractionation of the in vitro transcription activity may help to decide whether a specific protein factor is required for activity of the downstream element.

Downstream elements in other promoters. Our results with

the SV40 late transcription unit suggest the possibility that some other promoters for protein-coding genes may have a similar dependence on downstream sequences. Several reports support this contention. Stenlund et al. (67) have identified sequences downstream of the bovine papillomavirus 1 P1 gene cap site that are required for its expression. In vitro experiments suggest that these sequences act at the transcriptional level and that a cellular factor binds directly to the DNA in the downstream control region. However, this bovine papillomavirus element differs from the downstream element in the SV40 late promoter in both positions relative to the cap site and sequence.

Human immunodeficiency virus also requires downstream sequences for gene expression. Experiments with an SV40 early promoter fusion construct suggest that the position of the human immunodeficiency virus RNA cap site is determined by sequences in that virus downstream of position −17 (58). In addition, *trans*-induction by the viral *tatIII* gene product requires the TAR element, part or all of which lies downstream of the RNA start (58). Although the mechanism of *tatIII trans*-induction has been controversial, recent work suggests that the effect is partly at the RNA level (14, 51). Human immunodeficiency virus, like the SV40 major late promoter, has an inverted repeat downstream of the cap site (50) and a sequence, 5′-CTGGCT-3′, at +41 that is somewhat similar to the SV40 sequence, 5′-CCGGCT-3′, that is altered in LS23.

Other genes that require downstream sequences for their expression include the adenovirus 2 major late transcription unit (41), the herpes simplex virus *tk* gene (13), the α- and β-globin genes when expressed in murine erythroleukemia cells (10, 73), and the *Drosophila* transposible element *copia* (62). Like the downstream element of the SV40 late promoter, these elements may affect transcription, but it is also possible that they affect a posttranscriptional process.

Role of the tripartite late promoter in regulation of SV40 late expression. There is a large body of work directed toward understanding how SV40 late transcription is regulated and how the virus switches from early to late patterns of gene expression. It is not known what role the elements defined in the present work play in these respects.

Late transcription is activated by an early gene product, large T antigen (5, 37). The three elements defined here are required for the selection of the predominant SV40 late cap site, but they have not been examined for their effect on *trans*-activation (6, 31, 38). Furthermore, in two cases (31, 38), the downstream element has been replaced by the reporter gene, chloramphenicol acetyltransferase, at the *Nco*I site at nucleotide 331. Therefore, it remains a formal possibility that either one or more of these elements described here may play a role in T-antigen-mediated *trans*-activation, and these previous studies should be considered with this in mind.

The three promoter elements, although required for efficient initiation at nucleotide 325, are not essential for virus viability. As discussed previously, many workers have reported accumulation of RNA transcribed from upstream sites when initiation at nucleotide 325 is blocked by mutation (8, 22, 26, 27, 52, 53, 64). This upstream RNA is evidently sufficient for production of progeny virus, because SV40 mutants unable to initiate at nucleotide 325 show only minor defects in their ability to grow in cultured cells (2, 68).

This argues that, under conditions that prevail within the infected cell, the sequences that direct late transcription are hierarchical. The start site at nucleotide 325 represents the top of the hierarchy. When the late promoter is intact, nucleotide 325 is used predominantly. However, when nucleotide 325 cannot be recognized by RNA polymerase II, there are starts upstream that possess enough redundant information to function as initiation sites. Under other circumstances, this pattern of expression is not observed. In vitro, or in in vivo experiments with transfected plasmids, initiation at upstream starts was not observed to compensate for the defect in initiation at nucleotide 325. It is interesting to note that the agnoprotein is conserved in the human papovaviruses BK (61, 75) and JC (18). As such, the sequences altered by LS23 and LS24 are also found in BK and JC viruses, and they may play a role in the regulation of late transcription of these viruses as well. These findings raise the possibility that, under other physiological conditions, the sequences directing initiation at nucleotide 325 may be more important than has been apparent previously. The late promoter sequences described here may be required to turn on expression or to repress late gene expression at inappropriate times.

ACKNOWLEDGMENTS

We thank J. Mertz and P. Good for the gift of 16S and 19S body-specific primers, D. Dodds for synthesis of the other oligonucleotide primers used here, and K. Swaggert and J. Schneringer for maintaining the cell cultures. We are grateful to J. Mertz, R. Markowitz, and the members of the lab for helpful discussions and comments on the manuscript.

This work was supported by Public Health Service grants GM 35866 to W.S.D. and RR 07013 to the University of Colorado from the National Institutes of Health. D.E.A. received partial support as a National Institute of Health predoctoral trainee.

LITERATURE CITED

1. **Angel, P., M. Imagawa, R. Chiu, B. Stein, R. J. Imbra, H. J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin.** 1987. Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. Cell **49**:729–739.
2. **Barkan, A., and J. E. Mertz.** 1981. DNA sequence analysis of simian virus 40 mutants with deletions mapping in the leader region of late viral mRNAs: mutants with deletions similar in size and position exhibit varied phenotypes. J. Virol. **37**:730–737.
3. **Benoist, C., and P. Chambon.** 1981. In vivo sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
4. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
5. **Brady, J., J. B. 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.
6. **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.
7. **Brady, J., M. Radonovich, M. Thoren, G. Das, and N. P. Salzman.** 1984. Simian virus 40 major late promoter: an upstream DNA sequence required for efficient in vitro transcription. Mol. Cell. Biol. **4**:133–141.
8. **Brady, J., M. Radonovich, M. Vodkin, V. Natarajan, M. Thoren, G. Das, J. Janik, and N. P. Salzman.** 1982. Site-specific base substitutions and deletions that enhance or suppress transcription of the SV40 major late RNA. Cell **31**:625–633.
9. **Carswell, S., and J. C. Alwine.** 1986. Simian virus 40 agnoprotein facilitates perinuclear-nuclear localization of VP1, the major capsid protein. J. Virol. **60**:1055–1061.
10. **Chamay, P., R. Treisman, P. Mellon, M. Chao, R. Axel, and T. Maniatis.** 1984. Differences in human α- and β-globin expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell **38**:251–263.

11. **Chikaraishi, D. M., and K. J. Danna.** 1981. Simian virus 40 transcriptional complexes incorporate mercurated nucleotides into RNA in vitro. J. Virol. **37**:511–517.
12. **Cocino, M., R. A. Goldman, M. H. Caruthers, and R. Weinman.** 1983. Point mutations of the adenovirus major late promoter with different transcriptional efficiencies in vitro. J. Biol. Chem. **258**:8493–8496.
13. **Coen, D. M., S. P. Weinheimer, and S. L. McKnight.** 1986. A genetic approach to promoter recognition during trans-induction of viral gene expression. Science **234**:53–59.
14. **Cullen, B. R.** 1986. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. Cell **46**:973–982.
15. **Dynan, W. S., and D. Ayer.** 1987. Site-specific DNA binding proteins required for the transcription of the simian virus 40 early and late promoters, p. 303–311. *In* W. S. Reznikoff, R. R. Burgess, J. E. Dahlberg, C. A. Gross, T. M. Record, Jr., and M. P. Wickens (ed.), RNA polymerase and the regulation of transcription: a Steenbock Symposium. Elsevier Science Publishing Co., Inc., New York.
16. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–87.
17. **Dynan, W. S., and R. Tjian.** 1985. Control of eukaryotic messenger RNA synthesis by sequence specific DNA binding proteins. Nature (London) **316**:774–778.
18. **Frisque, R., G. L. Bream, and M. T. Cannella.** 1984. Human polyomavirus JC virus genome. J. Virol. **51**:458–469.
19. **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.
20. **Gardiner-Garden, M., and M. Frommer.** 1987. CpG islands in vertebrate genomes. J. Mol. Biol. **196**:261–282.
21. **Ghosh, P. K., P. Lebowitz, R. J. Frisque, and Y. Gluzman.** 1981. Identification of a promoter component involved in positioning the 5′ termini of simian virus 40 early mRNAs. Proc. Natl. Acad. Sci. USA **78**:100–104.
22. **Ghosh, P. K., M. Piatak, J. E. Mertz, S. M. Weisman, and P. Lebowitz.** 1982. Altered utilization of splice sites and 5′ termini in late RNAs produced by leader region mutants of simian virus 40. J. Virol. **44**:610–624.
23. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. V. Choudary, P. Lebowitz, and S. M. Weissman.** 1978. The 5′-terminal leader sequence of 16S mRNA from cells infected with simian virus 40. J. Biol. Chem. **253**:3643–3647.
24. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5′-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
25. **Gluzman, Y., J. F. Sambrook, and R. J. Frisque.** 1980. Expression of early genes in origin-defective mutants of simian virus 40. Proc. Natl. Acad. Sci. USA **77**:3898–3902.
26. **Gruss, P., A. Efstratiadis, S. Karathanasis, M. Konig, and G. Khoury.** 1981. Synthesis of stable unspliced mRNA from an intronless simian virus 40-rat preproinsulin gene recombinant. Proc. Natl. Acad. Sci. USA **78**:6091–6095.
27. **Haegeman, G., D. Iserentant, D. Gheysen, and W. Fiers.** 1979. Characterization of the altered leader sequence of late mRNA induced by SV40 deletion mutant dl-1811. Nucleic Acids Res. **7**:1799–1814.
28. **Haltiner, M., T. Kempe, and R. Tjian.** 1985. A novel strategy for constructing clustered point mutations. Nucleic Acids Res. **13**:1015–1025.
29. **Handa, H., R. Kaufman, J. Manley, M. Gefter, and P. Sharp.** 1981. Transcription of simian virus 40 DNA in a HeLa whole cell extract. J. Biol. Chem. **256**:478–482.
30. **Hansen, U., and P. A. Sharp.** 1983. Sequences controlling in vitro transcription of SV40 promoters. EMBO J. **2**:2293–2303.
31. **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.
32. **Hawley, D. K., and R. G. Roeder.** 1985. Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. J. Biol. Chem. **260**:8163–8172.
33. **Hawley, D. K., and R. G. Roeder.** 1987. Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. J. Biol. Chem. **262**:3452–3461.
34. **Hay, N., and Y. Aloni.** 1985. Attenuation of simian virus 40 mRNA synthesis is enhanced by the agnoprotein and is temporally regulated in isolated nuclear systems. Mol. Cell. Biol. **5**: 1327–1334.
35. **Hay, N., M. Kessler, and Y. Aloni.** 1984. SV40 deletion mutant (d1861) with the agnoprotein shortened by four amino acids. Virology **137**:160–170.
36. **Hay, N., H. Skolnik-David, and Y. Aloni.** 1982. Attenuation in the control of SV40 gene expression. Cell **29**:183–193.
37. **Keller, J. M., and J. C. Alwine.** 1984. Activation of the SV40 late promoter: direct effects of T antigen in the absence of viral replication. Cell **36**:381–389.
38. **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.
39. **Learned, R. M., R. M. Meyers, and R. Tjian.** 1981. Replication in monkey cells of plasmid DNA containing the minimal SV40 origin. ICN-UCLA Symp. Mol. Cell. Biol. **21**:555–566.
40. **Lee, W., P. Mitchell, and R. Tjian.** 1987. Purified transcription factor AP-1 interacts with TPA inducible enhancer elements. Cell **49**:741–752.
41. **Lewis, E. D., and J. L. Manley.** 1985. Control of adenovirus late promoter expression in two human cell lines. Mol. Cell. Biol. **5**:2433–2442.
42. **Maniatis, T., S. Goodbourn, and J. A. Fisher.** 1987. Regulation of inducible and tissue-specific gene expression. Science **236**: 1237–1245.
43. **Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter.** 1980. DNA-dependent transcription of adenovirus genes in a soluble whole cell extract. Proc. Natl. Acad. Sci. USA **77**:3855–3859.
44. **Mathis, D. J., and P. Chambon.** 1981. The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature (London) **290**:310–315.
45. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science **217**: 316–324.
46. **McKnight, S. L., and R. Tjian.** 1986. Transcriptional selectivity of viral genes in mammalian cells. Cell **46**:795–805.
47. **Mizuuchi, K., M. Mizuuchi, and M. Gellert.** 1982. Cruciform structures in palindromic DNA are favored by DNA supercoiling. J. Mol. Biol. **156**:229–243.
48. **Morin, G. B., and T. R. Cech.** 1986. The telomeres of the linear mitochondrial DNA of Tetrahymena thermophila consist of 53 bp tandem repeats. Cell **46**:873–883.
49. **Nandi, A., G. Das, and N. P. Salzman.** 1985. Characterization of a surrogate TATA box promoter that regulates in vitro transcription of the simian virus 40 major late gene. Mol. Cell. Biol. **5**:591–594.
50. **Okamoto, T., and F. Wong-Staal.** 1986. Demonstration of virus-specific transcriptional activator(s) in cells infected with HTLV-III by an in vitro cell-free system. Cell **47**:29–35.
51. **Peterlin, B. M., P. A. Leciw, P. J. Bar, and M. D. Walker.** 1986. Elevated levels of mRNA can account for the trans-activation of human immunodeficiency virus. Proc. Natl. Acad. Sci. USA **83**:9743–9738.
52. **Piatak, M., P. K. Ghosh, L. C. Norkin, and S. M. Weissman.** 1983. Sequences locating the 5′ ends of the major simian virus 40 mRNA late forms. J. Virol. **48**:503–520.
53. **Piatak, M., K. N. Subramanian, P. Roy, and S. M. Weissman.** 1981. Late messenger RNA production by viable simian virus 40 mutants with deletions in the leader region. J. Mol. Biol. **153**: 589–618.
54. **Reddy, V. B., P. K. Ghosh, P. Lebowitz, and S. M. Weissman.** 1978. Gaps and duplicated sequences in the leaders of SV40 16S mRNA. Nucleic Acids Res. **5**:4195–4213.

55. **Resnick, J., and T. Shenk.** 1986. Simian virus 40 agnoprotein facilitates normal nuclear location of the major capsid polypeptide and cell-to-cell virus spread. J. Virol. **60**:1098–1106.
56. **Rio, D., A. Robbins, R. Myers, and R. Tjian.** 1980. Regulation of simian virus 40 early transcription in vitro by a purified tumor antigen. Proc. Natl. Acad. Sci. USA **77**:5706–5710.
57. **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.
58. **Rosen, C. A., J. G. Sodroski, and W. A. Haseltine.** 1985. The location of cis-acting regulatory sequences in the human T cell lymphotrophic virus type III (HTLV-III/LAV) long terminal repeat. Cell **41**:813–823.
59. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
60. **Sassone-Corsi, P., J. Corden, C. Kedinger, and P. Chambon.** 1981. Promotion of specific in vitro transcription by excised "TATA" box sequences inserted in a foreign nucleotide environment. Nucleic Acids Res. **9**:3941–3958.
61. **Seif, I., G. Khoury, and R. Dahr.** 1979. The genome of the human papova virus BKV. Cell **18**:963–977.
62. **Sinclair, J. H., J. F. Burke, D. Ish-Horowitz, and J. H. Sang.** 1986. Functional analysis of the transcriptional control regions of the copia transposable element. EMBO J. **5**:2349–2354.
63. **Somasekhar, M. B., and J. E. Mertz.** 1985. Exon mutations that affect the choice of splice sites used in processing of the SV40 late transcripts. Nucleic Acids Res. **13**:5591–5609.
64. **Somasekhar, M. B., and J. E. Mertz.** 1985. Sequences involved in determining the locations of the 5′ ends of the late RNAs of simian virus 40. J. Virol. **56**:1002–1013.
65. **Sompayrac, L. M., and K. J. Danna.** 1981. Efficient infection of monkey cells with DNA of simian virus 40. Proc. Natl. Acad. Sci. USA **78**:7575–7578.
66. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**: 503–517.
67. **Stenlund, A., G. Bream, and M. R. Botchan.** 1987. A promoter with a internal regulatory domain is part of the origin of replication in BPV-1. Science **236**:666–671.
68. **Subramanian, K. N.** 1979. Segments of simian virus 40 DNA spanning most of the leader sequence of the major late viral messenger RNA are dispensable. Proc. Natl. Acad. Sci. USA **76**:2556–2560.
69. **Topal, J., D. M. Ruden, and C. S. Parker.** 1985. Sequences required for in vitro transcriptional activation of a Drosophila hsp 70 gene. Cell **42**:527–537.
70. **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.
71. **Villarreal, L. P., R. T. White, and P. Berg.** 1979. Mutational alterations within the simian virus 40 leader segment generate altered 16S and 19S mRNAs. J. Virol. **29**:209–219.
72. **Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, D. Roget, R. Teoule, and P. Chambon.** 1980. Specific in vitro transcription of the conalbumin gene is drastically decreased by a single-point mutation in the T-A-T-A box homology sequence. Proc. Natl. Acad. Sci. USA **77**:7024–7028.
73. **Wright, S., A. Rosenthal, R. Flavell, and F. Grosveld.** 1984. DNA sequences required for expression of β-globin genes in murine erythroleukemia cells. Cell **38**:265–273.
74. **Wu, C.** 1984. Activating protein factor binds in vitro to upstream control sequences in heat shock gene chromatin. Nature (London) **311**:81–84.
75. **Yang, R. C. A., and R. Wu.** 1979. BK virus DNA: complete nucleotide sequence of a human tumor virus. Science **206**:456–462.
76. **Zarucki-Schultz, T., S. Y. Tsai, K. Itakura, X. Soberon, R. B. Wallace, M. Tsai, S. L. C. Woo, and B. W. O'Malley.** 1982. Point mutagenesis of the ovalbumin gene promoter sequence and its effect on in vitro transcription. J. Biol. Chem. **257**:11070–11077.