JOURNAL OF VIROLOGY, Feb. 1982, p. 449-461
0022-538X/82/020449-13$02.00/0

Vol. 41, No. 2

Initiation and Regulation of Simian Virus 40 Early Transcription In Vitro

PAUL LEBOWITZ* AND PRABHAT K. GHOSH
Department of Internal Medicine, Yale Medical School, New Haven, Connecticut 06510

Received 27 May 1981/Accepted 15 September 1981

We have studied initiation and regulation of early transcription of simian virus (SV40) DNA in vitro by eucaryotic RNA polymerase II, using both a crude HeLa cell extract and a partially purified calf thymus polymerase supplemented with a HeLa cell S100 fraction. Analysis of initiation sites by primer-directed cDNA synthesis and sequencing of cDNA's has revealed that early transcription is initiated at a multiplicity of sites corresponding to the 5′ termini of early viral mRNA's. The pattern of in vitro initiation closely resembles the pattern of 5′ termini of early mRNA's late in the lytic cycle, with principal initiations between residues 5184 to 5194, upstream from the early Hogness-Goldberg (TATA) sequence, and at residue 5123, well downstream from this sequence. In vitro transcription is initiated to a lesser extent at sites between residues 5150 and 5155, the principal 5′ termini of early mRNA's in transformed cells and early in lytic infection, located 21 to 26 nucleotides downstream from the TATA sequence. Initiation occurs at identical sites and with similar efficiencies on form I and linearized DNA templates. There are minor differences in the efficiency of initiation at specific sites by the two transcriptional systems. Studies using a DNA template cleaved just downstream from the TATA sequence and a second template cleaved through a pair of 72-base-pair tandem repeats starting 87 nucleotides upstream from the TATA sequence have revealed that neither the TATA sequence nor the repeats are essential for early transcription in vitro. However, removal of the TATA and upstream sequences shifts initiation of transcription principally to the residue 5123 site. Comparison of the relative efficiencies of transcription on intact wild-type DNA, the two cleaved DNAs, and DNA from a deletion mutant suggests that all or most of the sequences constituting an early promoter lie within the genomic region 60-70 to 140 nucleotides upstream from the principal 5′ termini of the early mRNA's.

The availability of in vitro systems that carry out transcription of eucaryotic DNAs has greatly facilitated inquiries into both initiation and regulation of eucaryotic transcription. To date, two such systems have been described, one making use of a crude HeLa cell extract from which DNA and certain DNA-binding proteins have been removed (24) and the second composed of a partially or fully purified RNA polymerase II supplemented with an S100 cytoplasmic fraction from any one of a number of eucaryotic sources (41). Both systems have been used for transcription of a variety of eucaryotic genes—adenovirus (24, 40) and simian virus 40 (SV40) (16, 25, 33) early and late genes and genes for ovalbumin, conalbumin (40), and globins (23, 28). In the systems so far studied, there is good evidence that initiation of transcription, capping of 5′ termini, and chain elongation are carried out faithfully. On the other hand, chain termination does not take place.

SV40 is an especially favorable system for studying initiation and regulation of eucaryotic transcription since there are two interrelated temporal modes of transcription during lytic infection (early and late transcription) (22, 39); the origin of DNA replication (22, 39), overlapping templates for a multiplicity of 5′ termini of early (6, 8, 15, 19, 29) and late (3, 10, 14, 21, 30) mRNA's, and sequences required for regulation of transcription (2, 13) are clustered in a single control region; and the early protein, large-T antigen, binds in graded fashion to three sites within the origin of DNA replication (38) and in so doing influences the selection of 5′ termini of early mRNA's during the lytic cycle (6) and suppresses early transcription (20, 32, 33, 37).

The present study was initially undertaken in an effort to identify and map sites on the SV40 genome at which early transcription is initiated. We were interested in this problem because of the implications that multiplicity of 5′ termini and the wide expanse of the template for the 5′ termini might have for the mechanism by which

5′ termini are generated and early transcription is controlled. Furthermore, prior in vitro transcriptional studies (16, 33) had suggested early transcription from two sites, but the method used, sizing of transcripts of linearized DNA templates, would not permit detection of transcription from multiple proximate sites, and it was not possible to map the two initiation sites on the map of viral early transcription (Fig. 1) with any accuracy. In addition, a recent study making use principally of S1 nuclease mapping techniques suggested initiation from three sites in the region of the major 5′ termini of transformed-cell mRNA's (Fig. 1), but not from any other major or minor 5′ terminal sites (25).

Using the two aforementioned in vitro transcriptional systems, we have transcribed linearized and form I SV40 DNAs. To map sites of initiation of transcription, we have synthesized DNAs complementary to the 5′ termini of in vitro transcripts and compared them with cDNA's similarly synthesized on the early mRNA's. Our studies have demonstrated a multiplicity of in vitro initiation sites, both upstream and downstream from the early TATA sequence. Furthermore, all of the major and most of the minor in vivo 5′ termini lie at sites of in vitro initiation, suggesting that they arise by transcription initiation.

Having drawn this conclusion, we next conducted studies inquiring into the importance of the early TATA box and a pair of upstream 72-base-pair tandem repeats for initiation of early transcription and also attempting to gain insight into genomic regions regulating SV40 transcription. For these studies, we separately cleaved SV40 DNA slightly downstream from the TATA box and through the repeats and tested the ability of the cleaved DNAs to serve as templates for early transcription. Our main conclusions from these studies are that the TATA box and tandem repeats are not required for early transcription; that in the absence of the TATA box and upstream sequences, initiation is shifted principally to a site approximately 50 nucleotides downstream from the TATA box; and that sequences within the genomic region from 60–70 to 140 nucleotides upstream from the major cap sites of transformed cell and early-lytic-phase early mRNA's constitute the major portion, if not the entirety, of an early transcriptional promoter.

MATERIALS AND METHODS

DNAs from strain 776 wild-type SV40 and the mutant *dl*892 were prepared from infected CV1 cells by Hirt (17) extraction followed by two cycles of banding in CsCl-ethidium bromide gradients. DNA I was digested with the indicated enzymes as specified by the supplier (New England Biolabs, Inc.). Digested DNAs were extracted twice with phenol-chloroform, precipitated twice with ethanol, and taken up in 0.01 M Tris-hydrochloride, pH 8.0, and 1 mM EDTA.

Polyadenylated cytoplasmic RNAs from SV40-transformed line SV80 and CV1 cells infected with wild-type SV40 and deletion mutant *dl*892 were isolated as previously described (10, 29).

A crude extract containing RNA polymerase II activity and an S100 fraction were prepared from HeLa cells as described (24, 42). Partially purified RNA polymerase II was prepared from calf thymus according to Hodo and Blatti (18) with batchwise rather than column DEAE-cellulose chromatography, precipitation of the final DEAE eluate with ammonium sulfate to a final concentration of 75%, solubilization of the pellet in 0.05 M Tris-hydrochloride, pH 7.9, 25% glycerol, 5 mM dithiothreitol, and 1 mM EDTA, and dialysis against storage buffer containing 0.05 M Tris-hydrochloride, pH 7.9, 50% glycerol, 0.1 M ammonium sulfate, 5 mM dithiothreitol, and 0.1 M EDTA. All preparations were stored in small portions in liquid nitrogen.

Transcriptions using the crude HeLa cell extract were carried out in reactions containing 0.012 M *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES), pH 7.9, 0.06 M KCl, 7.5 mM $MgCl_2$, 0.06 mM EDTA, 1.2 mM dithiothreitol, 2.5 μg of DNA, and 30 μl of extract per 50-μl reaction volume (24). For analysis of transcripts by gel electrophoresis, 10 to 15 μCi of [α-^{32}P]UTP (Amersham Searle) was included in 50-μl reactions. For analysis of transcripts by synthesis and sequencing of DNAs complementary to 5′ ends, radiolabeled UTP was omitted, and reactions were scaled to 750 μl. Transcriptions with the calf thymus polymerase were carried out in 1-ml reactions containing 0.01 M HEPES, pH 7.9, 0.06 M KCl, 7.5 mM $MgCl_2$, 0.25 mM dithiothreitol, 10% glycerol, 0.60 mM CTP, UTP, GTP, and ATP, 6 μg of DNA, 500 μl of HeLa cell S100 extract, and 2,500 U of RNA polymerase II (4). Incubations in both systems were carried out for 1 h at 30°C. Vanadyl RNase inhibitor (Miles, Inc.) and DNase freed of RNase by treatment with iodoacetic acid (4) were then added to reactions to final concentrations of 10 and 4 μg/ml, respectively, and incubations were continued for another 5 min. Samples were then extracted as described (24). Final RNA pellets were either taken up in 0.01 $MNaPO_4$, pH 6.8–0.01% Sarkosyl, glyoxylated (27) and electrophoresed on 1.7% agarose gels in 0.01 M $NaPO_4$, pH 6.8–0.01% sodium dodecyl sulfate with glyoxylated size markers or taken up in 80% formamide (Eastman Chemical Co.; deionized)–0.4 M NaCl–0.01 M piperazine-*N*,*N*′-bis(2-ethanesulfonic acid (PIPES), pH 6.4 for hybridization to a DNA primer and synthesis of DNAs complementary to 5′ termini.

Methods for synthesis, separation, and sequencing of cDNA's have been described in detail (9). A primer spanning residues 5053 to 5089 and labeled with ^{32}P only at the 5′ end of the early strand (at position 5053) (6) was used for all cDNA syntheses. (Residue numbers used are according to a modification [M. Piatak, K. N. Subramanian, P. Roy, and S. M. Weissman, J. Mol. Biol., in press] of the system of Reddy et al. [31]. Correlation with the BBB [39] numbering system is provided in Fig. 1.) cDNA's were sequenced by either separation on 2-mm-thick slab gels of 8% polyacrylamide–7 M urea followed by Maxam-Gilbert (26) sequencing of individual cDNA's or coelectrophoresis

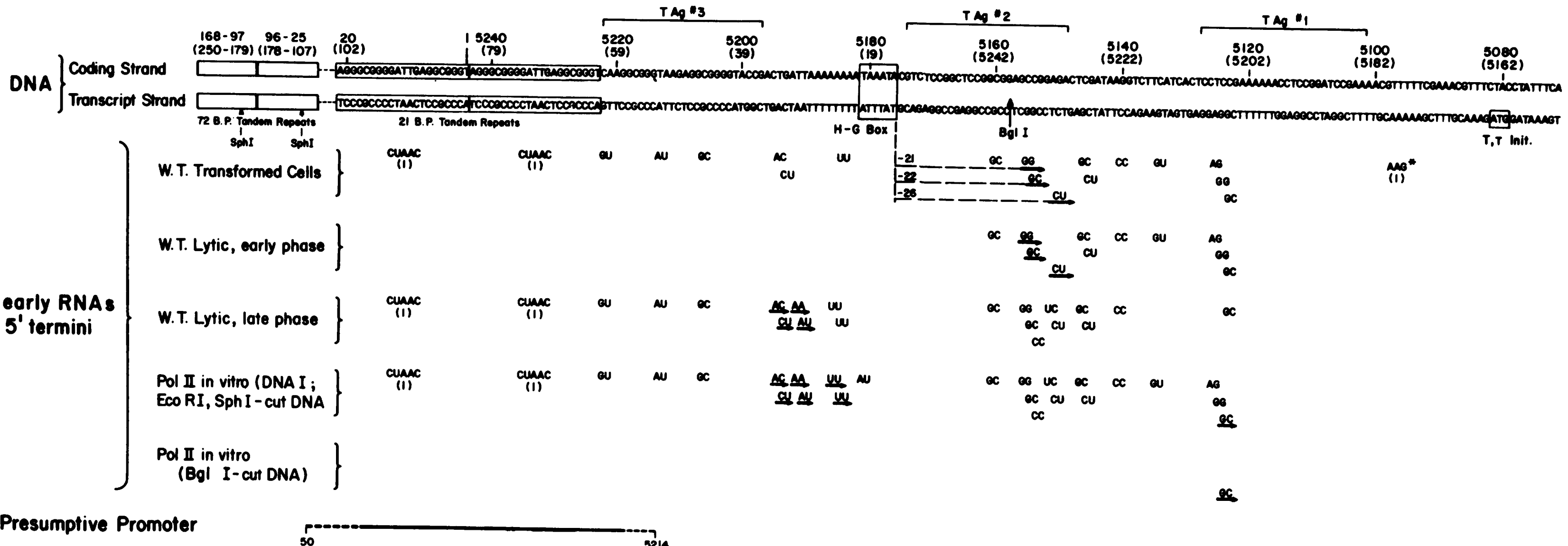


FIG. 1. Transcriptional map of SV40 genome. Upper portion of map demonstrates region of SV40 DNA encoding 5′ termini of early mRNA's and flanking regions. Early transcription takes place on upper DNA strand in a left-to-right direction. Lower portion of map indicates sites (± one or two mucleotides for certain sites) at which early wild-type mRNA's from the indicated sources have 5′ termini (6) and at which in vitro transcription is initiated. Most termini are presented as 5′-terminal transcribed dinucleotides. In certain locations it has not been possible to identify terminal sites precisely; at these loci, the number of termini indicated in parentheses lie within the indicated span of nucleotides. Major 5′ termini and initiation sites are underlined by heavy arrows. Asterisk indicates a G residue in early mRNA at which site an A residue appears on the transcript strand of DNA. DNA residue numbers without parentheses are according to modification (Piatak et al., in press) of the system of Reddy et al. (31); residue numbers in parentheses are according to the BBB system (38).

on 0.3-mm gels of the same composition, with the residue 5053-190 fragment also labeled only at the 5′ end of the early strand and digested according to Maxam and Gilbert. By reference to the DNA sequencing channels, the 3′ termini of cDNA's could be directly determined from the latter gels.

RESULTS

Five-prime termini of early mRNA's. In a recent study (6), we analyzed the 5′ termini of SV40 early mRNA's by synthesis, separation, and sequencing of DNAs complementary to 5′ termini. We found that mRNA's from transformed cells and early and late in lytic infection reproducibly yielded characteristic, but very different, cDNA electrophoretic patterns. As shown in Fig. 2. (lane 7), the cDNA pattern of transformed-cell early mRNA's typically reveals two principal cDNA's (1 and 2), each accounting for about 40 to 50% of total cDNA's, and a large number of additional minor cDNA's (3 to 16). In the early phase of lytic infection, the pattern of cDNA's mimics that of transformed cells for cDNA's 1 to 9, but there are no cDNA's with termini upstream for cDNA 9 (pattern not shown). Late in the lytic cycle, however, the cDNA pattern is radically altered: cDNA's 10.5 and 11 are vastly increased in quantity and almost always become the major cDNAs; cDNA's 1 and 2 are reduced in quantity (reductions always relative, usually absolute); cDNAs 8 and 10 usually show an absolute increase in amounts; and the remaining cDNA's are scarce (Fig. 2, lanes 5 and 6, reveal late lytic cDNA patterns for wild-type SV40 and mutant *dl*892 early mRNA's; although the two patterns differ in certain minor regards, prior studies have documented their similarity [6] and both show typical late pattern features). In some studies, the ratios of cDNA's 10, 10.5, and 11 to 1 and 2 have approached 5 to 10:1 in late lytic infection.

cDNA's 1 to 16 from all three sources have been sequenced individually by degradations adjacent to all four nucleotides according to Maxam and Gilbert (26) and also by coelectrophoresis on a DNA sequencing gel with a Maxam-Gilbert-degraded DNA fragment with the same labeled 5′ end as the primer used for cDNA syntheses (Fig. 2). The specificity of cDNA patterns, the appearance of certain minor wild-type cDNA's as major cDNA's in studies of mutant early mRNA's (7), and the absence of any or the presence of only insignificant quantities of comigrating cDNA's in reverse transcripts of cRNA have led us to conclude that the 3′ termini of all the major and probably most of the minor cDNA's correspond to the 5′ termini of early mRNA's.

Figure 1 demonstrates our localizations of the 5′ termini of the early mRNA's. From our cDNA sequence analyses, we have deduced that the early mRNA's of transformed cells and early lytic infection have three principal 5′ termini (corresponding to the termini of cDNA's 1 and 2) located at residues 5150, 5154, and 5155, 21 to 26 nucleotides downstream from the early TATA box. Terminal sequences at the first two sites, CU and GC, correspond to the most abundant caps identified in transformed cell mRNA's by Kahana et al. (19), whereas the GC sequence at position 5155 corresponds to another abundant cap (15, 19). Our analyses have further revealed that there are four principal 5′ termini late in infection (corresponding to the ends of cDNAs 10.5 and 11) located at or within one or two nucleotides of residues 5190 to 5194. Terminal sequences at three of these sites—AU, CU, and AC at positions 5190, 5193, and 5194, respectively—correspond to the three most abundant caps identified late in infection (Y. Groner, personal communication). Furthermore, in the analyses of Groner's group, the AC cap is present in early mRNA's only in late lytic infection, whereas the AU cap is a minor cap in the early phase of infection. Thus, two independent approaches have revealed a shift in the principal early 5′ termini used late in lytic infection. Minor 5′ ends of the early mRNA's extend from as far upstream as residues 9 to 13 to as far downstream as residues 5095 to 5097, a span of 160 nucleotides (Fig. 1). The two most upstream sites identified lie at identical sequences within a pair of 21-base-pair tandem repeats (extending from residues 5223 to 5243 and 1 to 21). With only two exceptions, the minor 5′ termini we have identified also lie at sites with sequences identified in early caps (15, 19).

Early transcription is initiated at multiple sites corresponding to 5′ termini of early RNA's. Figure 3 schematically depicts the region of SV40 DNA coding for 5′ termini of the early and late viral mRNA's, sites at which we have cleaved the genome with restriction endonucleases to obtain in vitro transcriptional templates, and the lengths expected of early and late transcripts of the cleaved templates if transcription is initiated in the regions of the 5′ termini of the in vivo mRNA's and terminates at the cleaved ends of the linear DNA templates.

Initial experiments involved transcriptions of linearized DNA templates with the crude HeLa cell extract (24) and "runoff" gel electrophoretic analysis of transcripts (Fig. 4). The gel electrophoretic pattern of transcripts of *Eco*RI-cleaved DNA reveals one dark band of about 3,500 nucleotides, in the size range expected for early transcripts, and two bands, one dark with a chain length of about 1,600 nucleotides and the other light with an estimated length of 1,450 nucleotides, in the size range expected of late

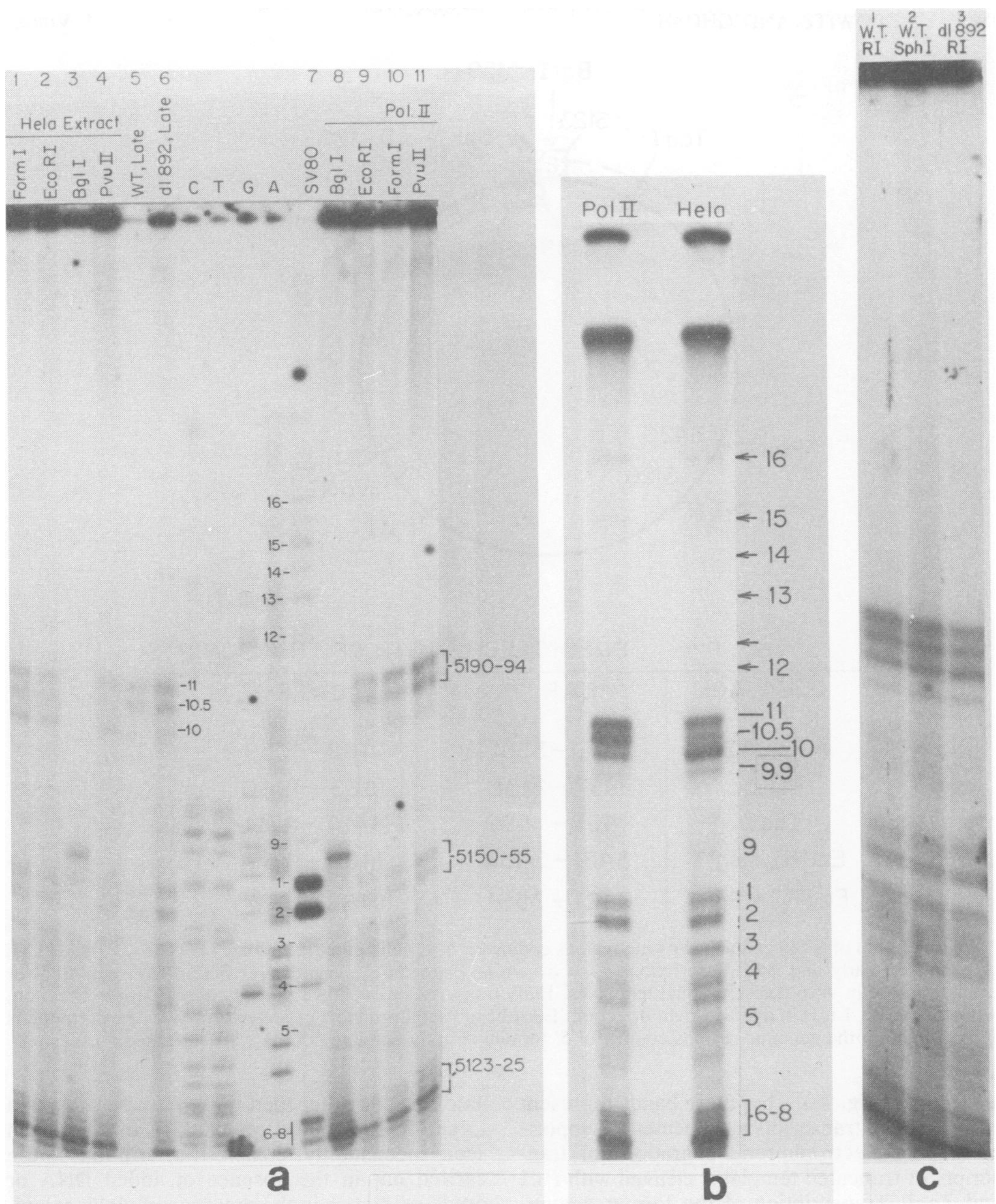
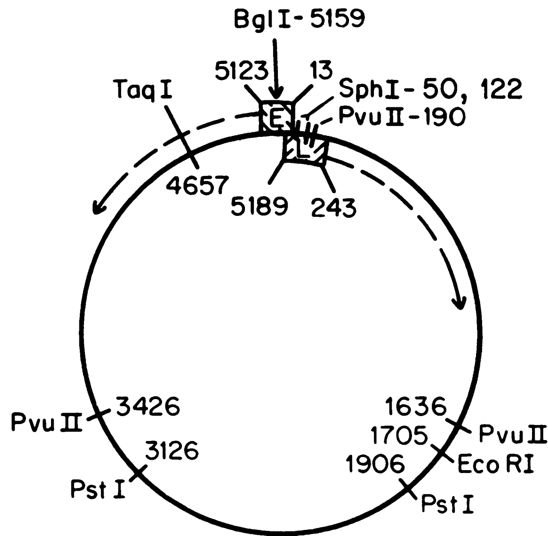


FIG. 2. Eight percent polyacrylamide–7 M urea gel electrophoreses of DNAs complementary to 5′ termini of SV40 early mRNA's and RNA polymerase II transcripts of SV40 DNA. (a) Coelectrophoresis on a 0.3-mm "thin" gel of cDNA's copied on in vitro transcripts and in vivo early mRNA's. In vitro transcriptions were carried out on form I DNA and DNAs cleaved with the indicated enzymes, using the crude HeLa cell extract (lanes 1 to 4) and calf thymus RNA polymerase II (lanes 8 to 11) systems. Early mRNA's were obtained from wild-type (lane 5)- and mutant *dl*892 (lane 6)-infected cells and SV80-transformed cells (lane 7). cDNA's are numbered as previously (6). In center four lanes is a Maxam-Gilbert (26) sequence analysis of a DNA fragment with the same 5′ end as the primer for cDNA synthesis. Since the cDNA originating at the end of *Bgl*I-cleaved DNA (cleavage at position 5159) comigrates with the G residue at position 5159 of the DNA sequence, comparison of the migration of cDNA's and bands on the DNA fragment sequence analysis allows direct determination of termini of cDNA's. Numbers in margin indicate termini of cDNA's 1 and 2, 6 to 8, and 10.5 and 11. (b) Coelectrophoresis of DNAs complementary to the 5′ termini of in vitro transcripts of *Eco*RI-cleaved wild-type SV40 DNA, using the crude HeLa cell and partially purified RNA polymerase II transcriptional systems. (c) Coelectrophoresis of DNAs complementary to the 5′ termini of in vitro transcripts of the indicated DNAs, using the crude HeLa cell transcriptional system.



PREDICTED LENGTH OF TRANSCRIPTS

CLEAVAGE	EARLY	LATE
EcoRI	3421-3551	1462-1759
PstI	1997-2130	1663-1960
TaqI	466-599	4414-4711
EcoRI + BglI	3421-3454	1462-1759
EcoRI + SphI	3421-3551	1462-1687

FIG. 3. Map of SV40 genome showing regions coding for 5′-termini of early and late mRNAs (E and L); directions of early and late transcription (←---- ----→); and sites of cleavage by restriction enzymes used for preparation of in vitro transcriptional templates. Early transcription is shown in a counterclockwise direction, whereas in Fig. 1 it is in a left-to-right direction. Lengths of runoff products expected if in vitro transcription is initiated within the genomic regions coding for 5′ termini of early and late mRNA's are indicated.

transcripts (Fig. 4a). That these bands represent early and late transcripts as indicated is supported by the electrophoretic migrations of transcripts of truncated templates cleaved with *Pst*I and *Taq*I. Transcription of the former demonstrates a single prominent band of approximately 2,100 nucleotides, as expected for early transcripts, and two bands in the size range expected for late transcripts, the darker one about 1,900 nucleotides and the lighter about 1,750 nucleotides (Fig. 4a). In certain electrophoretic runs, the early transcripts of these DNAs have resolved into two bands separated by about 75 nucleotides (Fig. 4b,d). The transcripts of *Taq*I-cleaved DNA include a single dark band of about 4,600 nucleotides, corresponding to late transcripts, and two light bands with chain lengths of about 460 and 535 nucleotides, in the size range predicted for early transcripts (Fig. 4a,c). Bands other than those noted correspond to end-to-end transcripts of DNA or have been seen inconstantly. When transcriptions were carried out in the absence of added DNA or crude extract or in the presence of alpha-amanitin (0.5 μg/ml), there was no gel electrophoretic evidence of any incorporation of substrates into RNA. The presence of two principal early and two principal late transcripts in our experiments is in accord with the prior findings of others (16, 25, 33).

The early transcripts of *Taq*I-cleaved DNA appeared to provide a good guide to the major sites of initiation of in vitro early transcription since they were well separated from one another and they coelectrophoresed with the accurately sized *Hin*dII/III C, D, and E fragments of SV40 DNA on the linear portion of the log chain length-migration curve of Fig. 4c. Indeed, since

TaqI cleaves the viral DNA at position 4660, the 460 and 535 chain lengths of these transcripts suggested that early transcription in vitro was initiated principally at sites close to residues 5120 and 5185 on the SV40 genome; as noted, the major 5′ ends of the early mRNA's late in lytic infection lie in the residue 5190 to 5194 region, and three moderately abundant 5′ ends of transformed-cell and early lytic mRNA's lie at residues 5123 to 5125 (Fig. 1). In contrast, there appeared to be no runoff transcript arising from residues 5150 to 5155, sites from which the three major 5′ termini of early mRNA's arise in transformed cells and early in the lytic cycle (Fig. 1).

In an effort to determine the precise locations of the 5′ termini of the early in vitro transcripts and to see if there might be additional in vitro initiation sites not detected by runoff gel electrophoresis, we transcribed form I and *EcoRI* and *PvuII*-cleaved DNAs with both crude HeLa cell and calf thymus RNA polymerases, synthesized DNAs complementary to the 5′ termini of the in vitro transcripts, and compared the electrophoretic patterns and terminal sequences of these cDNA's with those of cDNA's prepared by identical means on the early mRNA's. Whereas the runoff gel analyses suggested two early transcripts, the cDNA patterns (Fig. 2a,b) reveal a multiplicity of cDNA's synthesized on the 5′ ends of in vitro transcripts. Furthermore, the patterns of cDNA's copied on transcripts of form I and linearized DNAs are virtually identical, and the patterns obtained on transcripts synthesized with the two polymerases are qualitatively similar. Thus, there appear to be multiple initiation sites, early transcription is initiated at the same sites and with approximately equal efficiencies on supercoiled and linear templates, and the two transcriptional systems recognize the same start signals on SV40 DNA. Most important, Fig. 2a reveals extensive comigration of cDNA's of in vivo- and in vitro-synthesized RNAs, sequencing of cDNAs 1 to 11 of Fig. 2b has revealed the same termini as found for the

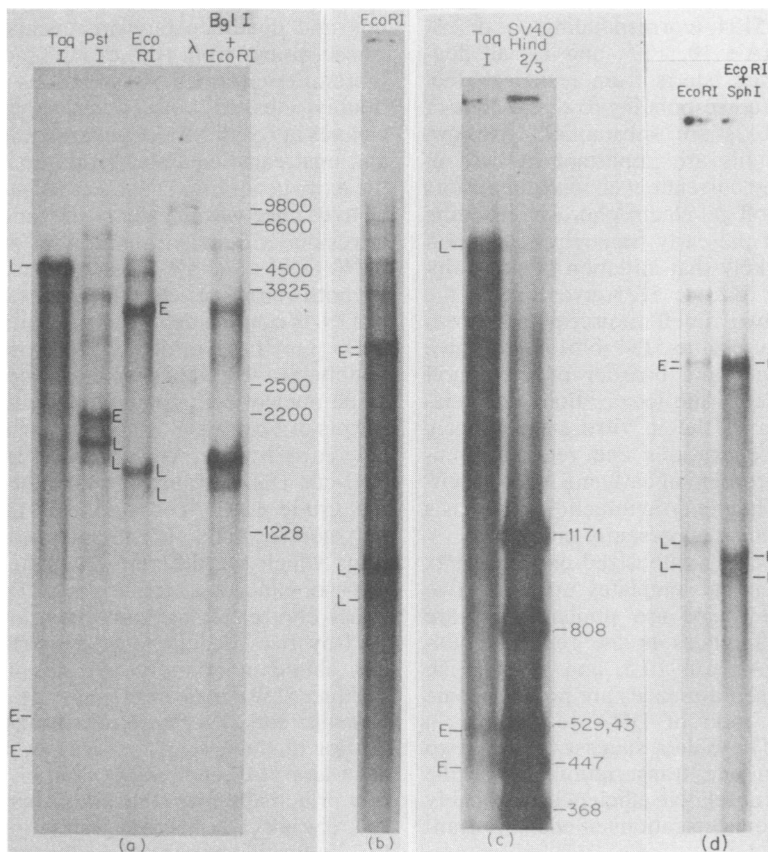


FIG. 4. Agarose (1.7%) gel electrophoresis of glyoxylated early (E) and late (L) RNA polymerase II transcripts of restricted SV40 DNAs. Size markers in (a) were fragments of SV40 mutant *dl809* DNA cleaved with *PstI* (1228 and 3825 nucleotides) and fragments of phage lambda DNA (lane 4) cleaved with *HindIII*; size markers in (c) were fragments of wild-type SV4 DNA derived by cleavage with *HindII/III*.

corresponding cDNA's of the early mRNA's, and analysis of the migration of in vitro cDNA's 12 to 16 of Fig. 2b has suggested that they each terminate at the same sites as the corresponding cDNA's of the early mRNA's (Fig. 2a, lane 8). On the other hand, the termini of two small cDNA's of Fig. 2b migrating between cDNA's 4 and 6 were found not to correspond to the termini of any cDNA's derived from early mRNA's. The 5′ termini of the in vitro transcripts deduced from these studies and the 5′ termini of the early mRNA's are compared in Fig. 1. From these studies, we conclude that most of the in vitro start sites are identical to 5′ termini of early mRNA's, that the 5′ termini of all major and most, if not all, minor early mRNA's arise by transcription initiation, and that initiations occur at sites both upstream and downstream from the early TATA box.

From Fig. 2a and 2b, it is apparent that the most prominent in vitro start sites lie at residue 5123 (corresponding to the 3′ terminus of cDNA 8) and at residues 5184 and 5185, 5190 and 5191, and 5193 and 5194 (corresponding to the 3′ termini of cDNA's 10, 10.5, and 11, all doublets). In contrast, starts from residues 5150, 5154, and 5155 (corresponding to the 3′ ends of cDNA's 1 and 2) are substantially weaker. These observations are important in two regards. First, they correlate well with the results obtained by runoff gel electrophoresis; from the chain lengths of the early transcripts on runoff gels, it seems likely that initiation of transcription at residues 5123 to 5125 gives rise to the shorter of the two runoff transcripts, whereas initiations at the residue 5184 to 5194 sites give rise to the longer and broader of the runoff transcripts. Second, the localizations and relative abundances of the in vitro starts closely resemble the localizations and relative abundances of the 5′ termini of early mRNA's specifically produced late in lytic infection. The basis for this resemblance is presently unknown.

Although cDNA's synthesized on transcripts prepared on identical templates using the two transcriptional systems are similar, there are reproducible differences in the relative abundances of cDNA's 10, 10.5, and 11 (Fig. 2b especially). These differences are not overcome by altering the ratio of DNA to enzyme in transcriptions. These data suggest that the two systems may initiate transcription at specific sites with different relative efficiencies, probably due to different concentrations of ancillary transcription factors.

The TATA box and upstream 72-base-pair repeats are not essential for in vitro early transcription; an early promoter region lies 60–70 to 140 nucleotides upstream from major cap sites. DNA sequences upstream from the 5′ termini of eucaryotic mRNA's are involved in regulation of transcription (for example, see reference 4). For both SV40 early transcription and transcription of the sea urchin histone *H2A* gene, the TATA box or adjacent sequences (or both) play a crucial role in vivo in fixing the positions of major downstream 5′ termini (2, 7, 12). However, there is no absolute requirement for the TATA box in these systems, since SV40 DNA lacking the box is still able to direct the synthesis of early mRNA's upon introduction into permissive cells (1), and the sea urchin gene from which these sequences have been deleted is still transcribed in *Xenopus* oocytes (25). In contrast to the dispensability of the TATA sequence for SV40 early transcription, the presence of at least one copy of a pair of 72-base-pair tandem repeats starting 87 nucleotides upstream the TATA box is essential for viral early transcription in vivo (2, 13).

Having defined multiple sites of initiation of early transcription, it was important to examine the effect of deletion or disruption of the TATA box and upstream tandem repeats on in vitro transcription from these multiple sites. We therefore performed transcriptions on templates doubly cleaved with *Eco*RI and *Bgl*I (*Bgl*I cleaves at residue 5159, severing the TATA box and upstream sequences from the remainder of the early gene; Fig. 1) and *Eco*RI and *Sph*I (*Sph*I cleaves once within each of the tandem repeats, at residues 50 and 122, but at no other sites on SV40 DNA; Fig. 1). Transcripts were analyzed by both runoff gel electrophoresis and analysis of DNAs complementary to 5′ ends.

Despite these critically placed lesions, both doubly cleaved DNAs yield transcripts of about 3,500 nucleotides, comigrating with early transcripts of *Eco*RI-cleaved DNA and also presumably early transcripts (Fig. 4a,d). In the case of *Bgl*I-cut DNA, transcripts obtained with both the crude HeLa cell and calf thymus RNA polymerases yield virtually identical cDNA patterns which are different from the cDNA patterns obtained on transcripts of DNA with an intact control region and which are characterized by two predominant bands (Fig. 2a). The less abundant of the two bands terminates at residue 5159, indicating initiation at the *Bgl*I cleavage site. The more abundant band, representing about 75% of total cDNA's, terminates at residue 5123, indicating initiation of transcription principally from this site. Also seen in Fig. 2a are weak cDNA's comigrating with cDNA's 1 to 3, indicative of persistent albeit very low level initiations from the residue 5150 to 5155 and 5146 sites, and a number of weak cDNA's which do not appear to comigrate with any of the in vivo cDNA's, indicative of weak initiations from a number of new sites. Weak initiations are also

observed further downstream, from residues 5110 to 5112 (results not presented). Thus, as noted recently also by Mathis and Chambon (25), removal of the TATA box and upstream sequences does not abolish early transcription in vitro; however, it does alter the specific sites at which it is initiated.

In the case of *Sph*I-cleaved DNA, retention of early transcription cannot be ascribed to religation of *Sph*I-cut ends, since the late transcriptional pattern observed differs from that obtained with DNA cut with *Eco*RI alone. In addition, the cDNA pattern of transcripts of *Sph*I-cut DNA is identical to that of transcripts of DNA with intact tandem repeats (Fig. 2c). Thus, as also found by Mathis and Chambon (25), intact repeats are not required for early transcription in vitro as they are in vivo; furthermore, cleavage through the 72-base-pair repeats does not alter the specific sites on SV40 DNA at which early transcription is initiated in vitro.

Cleavage of SV40 DNA with *Bgl*I reproducibly reduces the level of early transcription by about two-thirds (Fig. 4a). In contrast, cleavage upstream with *Sph*I does not reduce early transcription and may actually lead to some enhancement (Fig. 4d). The former suggests that sequences upstream from the *Bgl*I cleavage site are essential for high-level early transcription, whereas the latter suggests that sequences downstream from the *Sph*I site suffice for maximal transcription. Taken together, these observations suggest that the region from about residues 50 to 5160 contains all of the sequences which are necessary for high-level early transcription and hence which constitute all or most of an early transcriptional promoter. Recently it has been shown that SV40 mutants with deletions extending from downstream sites to residues 5204 to 5225 express early functions at near wild-type levels whereas mutants with deletions extending further upstream suffer sharp reductions in early gene expression (1, 2). Since these results suggest that the downstream end of an early promoter might lie in the region of residues 5204 to 5225, we examined early transcription on DNA from an SV40 mutant, *dl*892, which lacks 19 nucleotides extending from residues 5195 to 5213. As shown by analysis of cDNA's, the early starts on the mutant DNA are qualitatively and quantitatively identical to those on wild-type DNA. Thus, the span from residue 50 to residues 5214 to 5225 appears to contain sequences which specify most, if not all, of an early in vitro transcriptional promoter (Fig. 1).

DISCUSSION

The early and late genes of the papovaviruses are prime examples of eucaryotic genes giving rise to mRNA's with heterogeneous 5′ termini. For the late mRNA's of wild-type SV40, there are five or six major and a number of minor 5′ ends (10, 30). The early mRNA's appear to be even more complex, with three major 5′ termini in early lytic infection and in transformed cells, four different major 5′ ends during late infection, and a multiplicity of minor 5′ ends as detailed in Fig. 1 (6). Our present studies have shown that the 5′ termini of the principal as well as most, if not all, minor SV40 early mRNA's in transformed cells and early and late in infection arise by transcription initiation. This conclusion is based on identity of most cDNA's copied from the 5′ termini of in vivo mRNA's and in vitro transcripts, and also the strong likelihood that the 5′ termini of the in vitro transcripts arise by initiation of transcription and not post-transcriptional processing, even though the presently available transcriptional systems are crude. The latter point is strongly supported by the recent findings that m^7GpppCmpU is the most abundant SV40 early cap (19) and that [β-^{32}P]CTP enters caps with this sequence in SV40 early mRNA's synthesized by isolated nuclei from virus-infected cells (11).

The early transcription initiation sites we have identified span a minimum of about 130 nucleotides (from residue 5123 downstream to residues 9 to 13 upstream). Thus, any mechanism of transcription initiation in SV40 must be able to explain initiation at a multiplicity of sites spanning this distance.

Not only do the sites of in vitro initiation we have identified correspond to the 5′ termini of the early mRNA's, they resemble to a considerable extent the pattern of 5′ termini of mRNA's specifically present late in the lytic cycle (Fig. 1). That is, both in vitro transcripts and early mRNA's late in infection demonstrate prominent starts from residues 5184 to 5194 and 5123 and weaker starts in the region of residues 5150 to 5155. In contrast, initiations at residues 5150 to 5155 represent 80 to 90% of total initiations early in infection and in transformed cells, whereas initiations at upstream sites are either absent or scarce. On the one hand, it is possible that the in vitro pattern of initiation simply reflects the fact that the genomic region spanning residues 5184 to 5194 and residue 5123 do not appear to enter into extensive base pairing, whereas the region from residues 5130 to 5175, owing to the presence of two long sequences with dual axes of symmetry, may form two long hairpin loops (22, 35). The former sites may thus be available for initiation of transcription, whereas the 5150 to 5155 sites and other nearby sites may be unavailable to RNA polymerase. It seems more likely to us, however, that the similarity of in vitro and late lytic transcriptional

patterns has more biological significance. The SV40 transcriptional template late in infection contains significant quantities of bound large-T antigen (33, 37), the shift of early 5′ termini during late phase from residues 5150 to 5155 to residues 5190 to 5194 is mediated by T antigen (6), and we have postulated that binding of T to sites downstream from the TATA box may be involved in this upstream shift in late phase (6). Clearly, the in vitro transcriptional template does not contain bound large-T antigen. However, it seems possible that a cellular DNA-binding protein in the crude transcriptional systems may bind to SV40 DNA in vitro in a manner similar to in vivo binding by large T, thereby imparting a structure to the DNA similar to that in late lytic infection. Alternatively, an ancillary factor removed in the preparation of extracts or some specific configuration of the DNA, specifically a chromatin structure, not attained in vitro may be required for transcription predominantly from the downstream sites as in transformed cells and early lytic infection.

Only one other group, that of Chambon and collaborators, has performed comprehensive studies to map the 5′ termini of the early mRNA's and in vitro transcripts (2, 25). Their studies have been carried out principally by analyses of S1 nuclease-digested, end-labeled DNAs complementary to the early RNAs and have yielded results which are in general agreement with ours on one important point but at odds with ours on several additional points. First, we agree that transformed-cell early mRNA's have three principal 5′ termini in the residue 5148 to 5155 region (we locate termini at residues 5150, 5154, and 5155, whereas they place termini at residue 5148 and the latter two sites). Whereas we find the same principal 5′ termini early in lytic infection, they have not studied this question. With respect to points of disagreement, we have identified a shift in 5′ termini late in lytic infection to four principal sites between residues 5190 and 5194 and in certain experiments to two additional sites at residues 5184 and 5185; in contrast, they have found the same 5′ termini as in transformed-cell early mRNA's between residues 5148 and 5155. In addition, we have identified a number of minor 5′ termini of early mRNA's and corresponding minor initiation sites both upstream from residues 5190 to 5194 and downstream from residues 5150 to 5155, whereas they do not deal with this point. Finally, they have identified major initiations in vitro in the residue 5148 to 5155 region and at a site about 80 nucleotides downstream (this site lies five nucleotides downstream from the nearest potential in vivo 5′ terminus and downstream from the large- and small-T initiation codon), whereas we have localized principal initiations to residues 5184 to 5194 and residue 5123, as previously noted.

It is difficult to reconcile the differences between our results and those of the French group. Differences in in vitro transcriptional conditions may account for the different patterns of initiation in vitro, and different methodological approaches probably account for the differences in the number of 5′ termini identified and their localizations. Whereas "short stops" may arise in cDNA synthetic analyses due to premature termination by reverse transcriptase, giving rise to erroneous 5′ ends downstream from true ends, artifacts may arise in S1 nuclease analyses from overdigestion with the enzyme, resulting in shortening of cDNA's or even degradation of cDNA's. In addition, cDNA analyses as we have applied them may be more sensitive for detecting minor 5′ termini than methods involving S1 nuclease digestion of cDNA's. In our prior studies (6), we attempted to control for procedural artifacts and presented arguments supporting our identifications and localizations of 5′ termini. The most important points supporting our findings are as follows: identification of six separate capped dinucleotides by Kahana et al. (19) and four by Haegeman and Fiers (15), including one not detected by the former workers, in mRNA's from transformed cells and early lytic infection indicate a minimum of six to seven 5′ termini; the most abundant cap in transformed cells and early in lytic infection terminates with the transcribed sequence CU (19), in agreement with our placement of a principal 5′ terminus under these conditions at residue 5150 (Fig. 1); the most abundant caps late in infection terminate in CU, AC, and AU (Y. Groner, personal communication), in agreement with our localizations of major 5′ termini late in infection at residues 5190 to 5194 (Fig. 1); initiations as far upstream as residues 9 to 13 must exist, since there is no way upstream sequences could appear in cDNA's without being present in template mRNA's; and the termini of early mRNA's in a series of origin-defective mutants and correlation with known cap sequences suggest the validity of our localizations of minor 5′ termini downstream from residue 5150. We are presently experimenting with altered transcriptional conditions in an effort to see if we can influence the selection of in vitro initiation sites.

Although our pattern of in vitro initiations resembles that of the 5′ termini of mRNA's in late infection, the resemblance is still only partial. Stated otherwise, the in vitro transcriptional systems appear to faithfully initiate early transcription in a qualitative sense, but the relative quantities of initiation from the various sites only partially reflect the relative quantities of in

vivo 5′ termini from the same sites. Two explanations may be offered. On the one hand, the relative quantities of 5′ ends may reflect differential post-transcriptional processing. On the other hand, the presently available in vitro transcriptional systems may not accurately duplicate in vivo transcription in quantitative terms. Indeed, the current transcriptional systems are only approximations of transcriptional systems in vivo. This seems especially true in that DNA templates used in vitro are naked, whereas in vivo templates have a chromatin structure. In addition, a number of factors influence the efficiency and specificity of transcription initiation from specific sites (34; R. Roeder, personal communication) and, as suggested above, the relative concentrations of active ancillary factors in the transcriptional systems used are almost certainly not the same as in vivo. Accurate duplication of in vivo transcription will probably require full purification of all the factors involved in transcription, recombination of these factors and RNA polymerase II in very precise amounts, and use of native chromatin as transcriptional templates.

Although the 72-base-pair tandem repeats are essential for early transcription in vivo, two lines of evidence suggest that they do not constitute the major portion of an early transcriptional promoter: as shown here and by Mathis and Chambon (25), they are not required for in vitro transcription, and they can be translocated to distant genomic sites without impairing early transcription (P. Berg, personal communication). Although the function of these repeats remains unknown, they appear to establish a permissive state for early transcription in vivo, perhaps by imposing a specific chromatin structure on the viral DNA.

We have presented data which indicate that the genomic span from residues 50 to residues 5214 to 5225 contains sequences which are essential for high-level early transcription and probably constitute all or most of an early in vitro transcriptional promotor. This 70- to 80-nucleotide span lies 60–70 to 140 nucleotides upstream from the major early cap sites at residues 5150 to 5155 and contains sequences analogous to the CCAAT consensus sequence present 75 to 95 nucleotides upstream from many RNA polymerase II initiation sites (5), a pair of 21-base-pair tandem repeats (from residues 5223 to 5243 and 1 to 21; Fig. 1) and the downstream third of the downstream copy of the 72-base-pair repeats (residues 22 to 50; Fig. 1). It also lies adjacent to an additional copy of the 21-base-pair repeats, imperfect and extending from residues 5201 to 5220 (Fig. 1). It has recently been shown that deletion of one or two of the three 21-base-pair repeats has no significant effect on early transcription in vitro, but that removal of all three copies of the repeats drastically reduces early transcription (P. K. Ghosh, K. N. Subramanian, and P. Lebowitz, manuscript in preparation). Thus, although the three copies of the 21-base-pair repeats likely function in the same manner, only one copy appears to be essential for promotion of early transcription.

The proposed early promoter region also includes the most upstream of the known early transcription initiation sites. In view of the significance of the 21-base-pair repeats and the fact that the two most upstream early start sites (between residues 9 to 13 and 5232 to 5236; Fig. 1) lie at identical positions within each of the perfect repeats, it seems reasonable to postulate that early transcription involves an initial interaction, likely binding, of RNA polymerase II with the 21-base-pair repeats, initiations within the repeats, and then some form of movement of enzyme downstream with initiation at a multiplicity of sites determined by local nucleotide sequences, DNA secondary structure, or both. Initiations occur at pyrimidines as well as purines (Fig. 1) (6, 11, 19) and are directed toward specific downstream sites by the TATA sequence (2, 3, 7, 12).

Binding sites with high and moderate affinity for T antigen (1 and 2 in Fig. 1) lie downstream from the proposed promoter, whereas a low-affinity site (site 3) either overlaps or lies close to the promoter. Thus, if T antigen suppresses early transcription at the level of promotion, it must do so by binding to the third site. Such an interaction would be expected to suppress transcription from all initiation sites about equally. Alternatively, T antigen could inhibit early transcription, not by an interaction with the promoter, but by sequential binding to its three binding sites and sequential blockage of transcription initiation from loci within each of the binding sites. Initiations from loci outside binding sites would not be blocked. Favoring the latter mechanism is the aforementioned finding that early transcription from residues 5150 to 5155 and certain other downstream sites which fall within binding sites 1 and 2 is inhibited late in infection and that initiation is shifted principally to sites between residues 5190 and 5194 (6) which lie outside the binding sites (Fig. 1). Indeed, since this shift is mediated by T antigen, we have proposed that it might be the result of binding of T antigen to the higher-affinity binding sites with diversion of RNA polymerase molecules to the upstream sites which are unencumbered by T (6). We have also suggested that the shift in initiation to residues 5190 to 5194 may be the means by which initiation of late transcription from the adjacent residue 5189 site is accomplished (6).

Maintenance of early transcription in the absence of the TATA box and all upstream sequences was initially a surprising finding but suggests the existence of sequences downstream from the box which can bind RNA polymerase and function as a secondary or alternative promoter. The fact that residue 5123 serves as the principal early initiation site upon removal of the TATA box and all upstream sequences indicates that the signal which leads to strong initiation at this site is active whether a primary promoter upstream from the TATA box or a secondary promoter downstream is used for early transcription. Using mutants with deletions which start 30 nucleotides downstream from the TATA box (at residue 5146), take out the residue 5150 to 5155 start sites and the TATA box, and extend varying distances upstream, Chambon and colleagues (2, 25) have found both in vivo and in vitro that initiation of early transcription is extremely heterogeneous, with most starts falling in the region from approximately residues 5120 to 5090 (Fig. 1). It seems possible that the greater heterogeneity of start sites in their mutants and the fact that a principal start site at residue 5123 is not observed may relate to the deletions in their mutants extending 13 nucleotides further downstream (i.e., to within 20 nucleotides of the 5123 site) than the deletion in our in vitro template. If initiation cannot take place at residue 5123, the virus may compensate by using a multiplicity of start sites further downstream.

ACKNOWLEDGMENTS

We thank Jim Manley for teaching us to perform in vitro transcriptions with his system, Sherman Weissman for helpful suggestions and reviewing cDNA sequencing data with us, and Charlene Ivory for excellent technical assistance.

This work was supported by Public Health Service grant 16038 from the National Cancer Institute and by grants from the Leonard Eckstein Fund for Leukemia Research and the Swebilius Trust of the Yale Comprehensive Cancer Center.

ADDENDUM IN PROOF

After submission of this paper, Myers et al. (Cell 25:373–384, 1981) reported that an SV40 early promoter lies approximately 70 to 155 nucleotides upstream from the major early cap sites, in agreement with the results reported here. In addition, Hansen et al. (Cell, in press) have confirmed the findings discussed here and previously (6), that the principal early starts late in lytic infection are shifted to a series of sites 10 to 15 nucleotides upstream from the early TATA box and that the shift is mediated by large T antigen.

LITERATURE CITED

1. **Benoist, C., and P. Chambon.** 1980. Deletions covering the putative promoter region of early mRNAs of simian virus 40 do not abolish T antigen expression. Proc. Natl. Acad. Sci. U.S.A. 77:3865–3869.
2. **Benoist, C., and P. Chambon.** 1981. In vivo sequence requirements of the SV40 early promoter region. Nature (London) 290:304–310.
3. **Canaani, D., C. Kahana, A. Mukamel, and Y. Groner.** 1979. Sequence heterogeneity at the 5′-termini of late simian virus 40 19S and 16S mRNAs. Proc. Natl. Acad. Sci. U.S.A. 76:3078–3082.
4. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science 209:1406–1414.
5. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. C. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell 21:653–668.
6. **Ghosh, P. K., and P. Lebowitz.** 1981. Simian virus 40 early mRNA's contain multiple 5′ termini upstream and downstream from a Hogness-Goldberg sequence; a shift in 5′ termini during the lytic cycle is mediated by large T antigen. J. Virol. 40:224–240.
7. **Ghosh, P. K., P. Lebowitz, 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. U.S.A. 78:100–104.
8. **Ghosh, P. K., M. Piatak, V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1980. Transcription of the simian virus 40 genome in virus transformed cells and early lytic infection. Cold Spring Harbor Symp. Quant. Biol. 44:31–39.
9. **Ghosh, P. K., V. B. Reddy, M. Piatak, P. Lebowitz, and S. M. Weissman.** 1980. Determination of messenger RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. Methods Enzymol. 65:580–594.
10. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. The heterogeneity and 5′-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. 126:813–846.
11. **Gidoni, D., C. Kahana, D. Canaani, and Y. Groner.** 1981. Specific in vitro initiation of transcription of simian virus 40 early and late genes occurs at the various cap nucleotides including cytidine. Proc. Natl. Acad. Sci. U.S.A. 78:2174–2178.
12. **Grosschedl, R., and M. L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci. U.S.A. 77:1432–1436.
13. **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. U.S.A. 78:943–947.
14. **Haegeman, G., and W. Fiers.** 1978. Characterization of the 5′-terminal capped structures of late simian virus 40-specific mRNA. J. Virol. 25:824–830.
15. **Haegeman, G., and W. Fiers.** 1980. Characterization of the 5′-terminal cap structures of early simian virus 40 mRNA. J. Virol. 35:955–961.
16. **Handa, H., R. J. Kaufman, J. Manley, M. Gefter, and P. A. Sharp.** 1981. Transcription of simian virus 40 DNA in a HeLa whole cell extract. J. Biol. Chem. 256:478–482.
17. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. 26:365–369.
18. **Hodo, H. G., and S. P. Blatti.** 1977. Purification using polyethyleneimine precipitation and low molecular weight subunit analyses of calf thymus and wheat germ DNA-dependent RNA polymerase II. Biochemistry 16:2334–2343.
19. **Kahana, C. D. Gidoni, D. Canaani, and Y. Groner.** 1981. Simian virus 40 early mRNA's in lytically infected and transformed cells contain six 5′-terminal caps. J. Virol. 37:7–16.
20. **Khoury, G., and E. May.** 1977. Regulation of early and late simian virus 40 transcription: overproduction of early viral RNA in the absence of a functional T antigen. J. Virol. 23:167–176.
21. **Lai, C. J., R. Dhar, and G. Khoury.** 1978. Mapping the

spliced and unspliced late lytic SV40 RNAs. Cell **14**:971–982.

22. **Lebowitz, P., and S. M. Weissman.** 1979. Organization and transcription of the simian virus 40 genome. Curr. Top. Microbiol. Immunol. **87**:43–172.
23. **Luse, D. S., and R. G. Roeder.** 1980. Accurate transcription initiation on a purified mouse β-globin DNA fragment in a cell-free system. Cell **20**:691–699.
24. **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. U.S.A. **77**:3855–3859.
25. **Mathis, D., and P. Chambon.** 1981. The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature (London) **290**:310–315.
26. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:498–560.
27. **McMaster, G., and G. C. Carmichael.** 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. U.S.A. **74**:4835–4838.
28. **Proudfoot, N. J., M. H. M. Shander, J. L. Manley, M. L. Gefter, and T. Maniatis.** 1980. Structure and in vitro transcription of human globin genes. Science **209**:1329–1336.
29. **Reddy, V. B., P. K. Ghosh, P. Lebowitz, M. Piatak, and S. M. Weissman.** 1979. Simian virus 40 early mRNA's. I. Genomic localization of 3' and 5' termini and two major splices in mRNA from transformed and lytically infected cells. J. Virol. **30**:279–296.
30. **Reddy, V. B., P. K. Ghosh, P. Lebowitz, and S. M. Weissman.** 1978. Gaps and duplicated sequences of the leaders of SV40 16S RNA. Nucleic Acids Res. **5**:4195–4215.
31. **Reddy, V. B., B. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, M. L. Celma, P. K. Ghosh, and S. M. Weissman.** 1978. The genome of SV40. Science **200**:494–502.
32. **Reed, S. I., G. R. Stark, and J. C. Alwine.** 1976. Autoregulation of simian virus 40 gene A by T angigen. Proc. Natl. Acad. Sci. U.S.A. **73**:3083–3087.
33. **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. U.S.A. **77**:5706–5710.
34. **Sekimizu, K., N. Koboyashi, D. Mizuno, and S. Natori.** 1976. Purification of a factor from Ehrlich ascites tumor cells specifically stimulating RNA polymerase II. Biochemistry **15**:5064–5070.
35. **Subramanian, K. N., R. Dhar, and S. M. Weissman.** 1977. Nucleotide sequence of a fragment of SV40 DNA that contains the origin of DNA replication and specifies the 5' ends of "early" and "late" viral RNA. III. Construction of the total sequence of the EcoRII-G fragment of SV40 DNA. J. Biol. Chem. **252**:355–367.
36. **Subramanian, K. N., and T. She .** 1978. Definition of the boundaries of the origin of DI A replication in simian virus 40. Nucleic Acids Res. **5**:3635–3642.
37. **Tegtmeyer, P., M. Schwartz, J. K. Collins, and K. Rundell.** 1975. Regulation of tumor antigen synthesis by simian virus 40 gene *A*. J. Virol. **16**:168–178.
38. **Tjian, R.** 1978. The binding site on SV40 DNA for a T antigen related protein. Cell **13**:165–180.
39. **Tooze, J. (ed.).** 1980. DNA tumor viruses. Molecular biology of tumor viruses, part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
40. **Wasylyk, B., C. Kedinger, J. Corden, O. Brison, and P. Chamon.** 1980. Specific in vitro initiation of transcription on conalbumin and ovalbumin genes and comparison with adenovirus-2 early and late genes. Nature (London) **285**:367–373.
41. **Weil, P. A., D. S. Luse, J. Segall, and R. G. Roeder.** 1979. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell **18**:469–484.
42. **Wu, G. J.** 1978. Adenovirus DNA-directed transcription of 5.5S RNA in vitro. Proc. Natl. Acad. Sci. U.S.A. **75**:2175–2179.
43. **Zimmerman, S. B., and G. Sandeen.** 1966. The ribonuclease activity of crystallized pancreatic deoxyribonuclease. Anal. Biochem. **14**:269–277.