MOLECULAR AND CELLULAR BIOLOGY, July 1982, p. 737–751
0270-7306/82/070737-15$02.00/0

Vol. 2, No. 7

DNA Sequences Required for Specific and Efficient Initiation of Transcription at the Polyoma Virus Early Promoter

PARMJIT JAT,[1] ULRIKE NOVAK,[2] ALISON COWIE,[1] CHIARA TYNDALL,[1] AND ROBERT KAMEN[1]*

Transcription[1] and Nucleic Acid Chemistry[2] Laboratories, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, England

Received 24 November 1981/Accepted 3 March 1982

The 5'-flanking DNA sequences involved in the specific and efficient transcription of the polyoma virus early region have been investigated. Sequence requirements for efficient in vivo expression differed from those in vitro. Deletion of DNA located between 200 and 400 base pairs before the principal cap sites severely inhibited in vivo expression as measured by transformation ability, but did not affect in vitro transcription. Viable deletion mutants which lack the principal cap sites and the "TATA" box were very poor templates for in vitro transcription. Analysis of other deletion mutants in vitro demonstrated that no specific sequences more than 46 base pairs before the cap sites were important. Removal of the TATA box reduced in vitro transcriptional efficiency but did not alter the initiation sites. The synthesis of transcripts with abnormal 5' termini did not occur in vitro until sequence between the TATA box and the normal cap sites was also deleted. We further observed a nonspecific requirement for 90 to 100 base pairs of DNA 5' to the cap site for optimal transcription of DNA fragments in vitro.

The early region of the polyoma virus (Py) genome serves as an excellent model for the study of eucaryotic transcription. It expresses three principal mRNAs throughout productive infection and when integrated into the DNA of transformed cell lines (36, 61). These mRNAs share common 5' and 3' termini, but are differentially spliced to encode separately the three viral early gene products, the small, middle, and large T proteins (36, 61, 63). Unlike cloned cellular genes, one can be certain that all of the signals necessary and sufficient for the regulated (12, 35, 61) transcription of the Py early region are present in the viral DNA because it evolved as an independent extrachromosomal element. Transient expression in vivo can be studied by using viral infection, or transfection of recombinant plasmids containing all of the early region into a variety of cell types (68); stable expression can be studied by using the permanent integrants readily selected by the transforming activity of one of the three early gene products, the middle T protein (49, 50, 64). Expression can also be studied in vitro, because the initiation of Py early region transcription occurs faithfully in the HeLa whole-cell extract (34, 41).

The availability of convenient procedures for altering specific DNA sequences and assessing the consequences has permitted extensive characterization of the elements necessary for initiation of transcription from viral and cellular genes (reviewed in references 11 and 56). The principal 5' terminus, located at nucleotides 148 to 153 (nt148–153) in the viral DNA sequence (we used the Py numbering system proposed by Soeda et al. [57], as slightly modified in reference 68), and the minor (at nt300±2) 5' terminus of Py early region transcripts, synthesized both in vivo (R. Kamen, P. Jat, R. Treisman, J. Favaloro, and W. Folk, J. Mol. Biol., in press) and in vitro (34), are preceded by sequences homologous to those present in the 5'-flanking regions of many genes. Both cap site regions are approximately 30 base pairs (bp) from "TATA" boxes. This homology sequence, originally observed by Goldberg and Hogness upstream of histone genes (M. Goldberg, Ph.D. Thesis, Stanford University, Stanford, Calif., 1978), has now been found in more than 60 eucaryotic genes (11) and bears a striking homology to the *Escherichia coli* Pribnow box (53). The exact location of the TATA box varies somewhat, but it is generally located in higher eucaryotes about 30 nucleotides before the cap sites (11). A second region of homology, GGPyCAATCT, 70 to 90 nucleotides before the cap sites, has also been noted in several cellular and viral genes (7, 11, 17, 32, 45). Both the principal and minor Py early cap site regions are preceded by sequences similar to this consensus (Kamen et al., in print). Extensive evidence supports the conclusion that the TATA sequence is a component of the eucaryot-

ic promoter, but its importance is greater when expression is assayed in vitro (13, 28, 33, 37, 39, 55, 60, 65, 66, 70–72) rather than in vivo (1, 5, 6, 14, 16, 19, 24, 26, 46). The −70 to −90 homology region is necessary for the efficient in vivo expression of certain genes (6, 16, 45, 46); its function in vitro has not been demonstrated (13, 28, 33, 55, 65, 66). Similarly, more remote upstream sequences are known to be critical for transcription of several viral (6, 24, 68) and cellular (19, 27, 59) genes in vivo (11), but such sequences were not found to be necessary in vitro (11, 56).

We have investigated the 5′-flanking sequences of the principal Py early promoter region by the construction of deletion and substitution mutants. In vivo effects of certain mutations were measured by assaying the ability of the DNAs to transform rat fibroblasts. The effects of all mutations on the position and efficiency of in vitro initiation were measured with quantitative assays. The secondary Py early promoter region served as a convenient internal control. Results were in general agreement with those obtained with other genes (11, 56). We confirmed that the sequence requirements for efficient expression in vivo differ in certain ways from in vitro requirements. The novel conclusions of our study were that the TATA box, though important for efficient expression in vitro, was not essential for accurate initiation, and that although sequences upstream of the TATA box may be replaced with other DNA without substantially decreasing the efficiency of transcription in vitro, the DNA template must include 90 to 100 bp of DNA upstream of the cap site for transcription to occur with maximal efficiency.

MATERIALS AND METHODS

Recombinant plasmids. Plasmid pAT153 (67) was the vector used for all constructions. Plasmids were propagated in *E. coli* HB101 (10) or GM113 (42), and DNA was prepared on both large and small scale by the alkaline extraction procedure (9) as modified by D. Ish-Horowitz (personal communication). The structures of most recombinants used are summarized diagrammatically in Fig. 1 and 2.

Construction of substitution mutants. Plasmid p53.A6.6 contains full-length Py DNA inserted at the *Bam*HI site of pA6 (64); plasmid p45.4.A2 contains viral sequences from the *Bcl*I site (nt5022) to the *Eco*RI site (nt1565) inserted between the *Bam*HI and *Eco*RI sites of the vector; plasmid p45.8.A2 contains Py sequence from *Bam*HI (nt4633) to *Eco*RI inserted at the same position; plasmids p33.6.A2 and p33.11.A2 contain viral sequence from the *Pvu*II site at nt5268 to *Eco*RI inserted between the *Sal*I (nt 650) and *Eco*RI sites of the vector. Plasmids pP10 and pP17 were obtained by blunt-ended closure of repaired partial *Pvu*II + *Bam*HI digests of p45.8.A2 and contain viral sequence from the *Pvu*II site at nt5131 (pP10), or from the *Pvu*II site at nt5268 (pP17), to *Eco*RI. Plasmids pB333, pB317, pB38, and pB232 are deleted derivatives of p45.4.A2 obtained by *Bcl*I digestion (after propagation in *dam E. coli* GM113 [42]), nuclease *Bal*31 (Bethesda Research Laboratories, Gaithersburg, Md.; 38) treatment, *Bam*HI digestion, repair of cohesive ends, and reclosure by blunt-ended ligation. Sequence analysis (44) showed that the partially repaired vector *Bam*HI site joined the viral sequence at nt5041 (pB333), nt5044 (pB317), nt5082 (pB38), and nt5105 (pB232), respectively (see Fig. 1).

Construction of deletion mutations within viral sequence. Construction of deletions within region A (see Fig. 1), the pdl2000 series, has been described (68). Further deletions of the pdl2000 series extending from region A into region B were obtained in the same experiments and have the sequenced endpoints indicated in Fig. 2. Viable Py mutants *dl*1001 and *dl*1004 (40), obtained from G. Magnusson as *Bam*HI linears inserted at the homologous site of plasmid pBR322, were sequenced to determine the precise extents of deletion. Mutants *dl*17, *dl*75, and *dl*2–19 (3, 4) are viable mutants obtained from W. Folk. Mutants of the pdl3000 series were constructed as follows: plasmid p37.3.A2 (*Bam*HI linear viral DNA inserted at the pAT153 *Bam*HI site) was digested with restriction endonuclease *Bgl*I in the presence of ethidium bromide (52) under empirically determined conditions (50 μg of DNA per ml, 15 μg of ethidium bromide per ml, 44 U of *Bgl*I per ml) which maximized the yield of single-cut molecules. These were then treated with excess nuclease Bal31 (38) for 15 to 90 s at 30°C. After addition of *Xho*I linkers, the molecules were recircularized and used to transfect *E. coli* HB101. Plasmid DNA from a bulk preparation grown in ampicillin was digested with *Bgl*I + *Bam*HI. Linear viral DNA which had lost the single *Bgl*I cleavage site at nt93 was purified by agarose gel electrophoresis, eluted, and reinserted into the *Bam*HI site of pAT153. Single ampicillin-resistant colonies produced after transfection of HB101 were isolated, and the plasmids they carried were screened for deletion mutants by restriction enzyme analysis (using *Xho*I + *Ava*I + *Bam*HI for those which proved to have *Xho*I sites, or *Sac*I + *Pvu*II for those which did not). Mutants of interest were propagated on a large scale, and the extent of the deletions was determined by DNA sequence analysis (44).

DNA sequence determination. Recombinant plasmids of the pdl3000 series were restricted with *Dde*I and labeled with [γ-^{32}P]ATP (5000 Ci/mmol, Radiochemical Centre, Amersham, England) using T4 polynucleotide kinase (PL Laboratories, Milwaukee, Wis.). The complementary strands of the DNA fragments were separated on 4% polyacrylamide strand separation gels (44). The appropriate strand of the *Dde*I fragment, extending from the Py DNA site at nt188 to a *Dde*I site in the plasmid (in wild-type DNA this would be either a 2,052- or a 1,294-nucleotide fragment, depending on the orientation in which the Py DNA was inserted into pAT153) was eluted. On 4% strand separation gels, the faster migrating strands of the ca. 2,052-nucleotide fragments and the slower migrating strands of the ca. 1,294-nucleotide fragments were the E DNA strands. These were sequenced by the chemical procedure of Maxam and Gilbert (44),

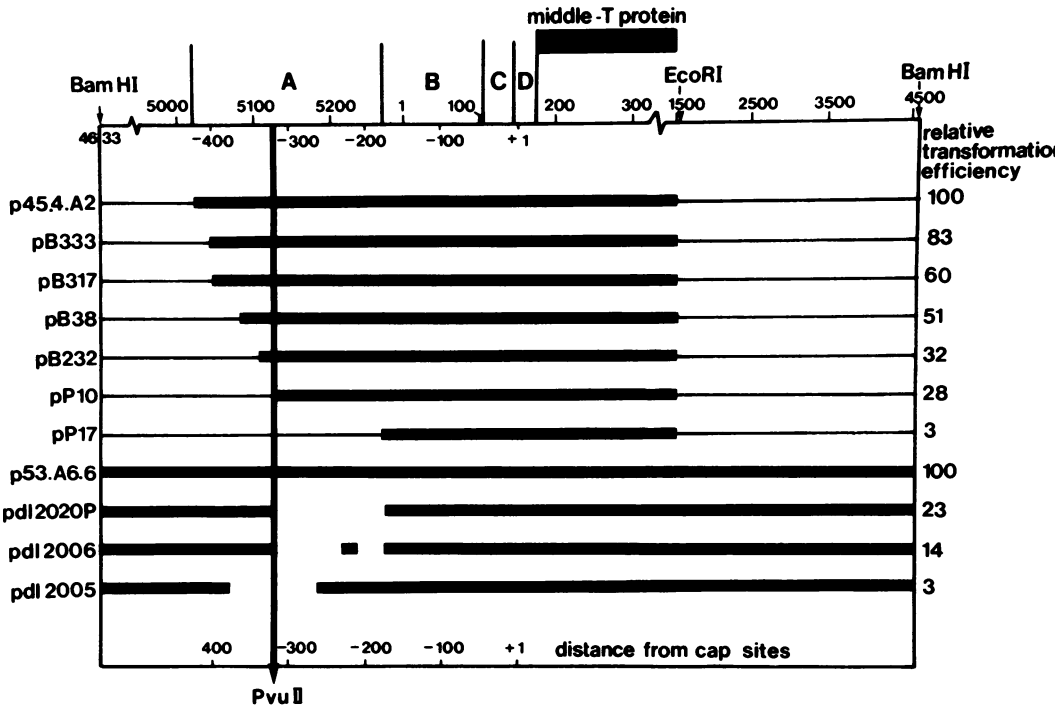

FIG. 1. Structures and transformation activities of recombinant plasmids including Py DNA with region A modifications. The upper coordinate represents the viral genome linearized in the late coding sequences at the unique *Bam*HI site. The numbers above the line are the nucleotide numbering system proposed by Soeda et al. (57; slightly modified by Tyndall et al. [68]). The sequences flanking and including the beginning of the early region are shown on an expanded scale relative to the rest of the sequence. The numbers below the line, and on the coordinate at the bottom of the figure, are approximate distances from the principal cap site region, using nt150 as +1. The subdivisions A through D (see text) and the coding sequences for the middle T protein are indicated. Structures of recombinant plasmids are shown with the viral sequence emboldened. In the first seven constructions, the pAT153 *Bam*HI site is linked to viral sequence at different positions within region A, and the viral insert ends at the *Eco*RI site (see the text). The next four constructions are plasmids containing *Bam*HI linear viral DNA without (p53.A6.6; 64) or with internal deletions in region A (68). Transformation activities (listed to the right) were determined by calcium phosphate-mediated transfection (49, 50, 65, 69) of F2408 Rat1 cells (20); dense foci overgrowing the monolayers were counted 14 to 21 days after transfection (49, 50, 64). Results of two separate experiments are shown. For the first seven plasmids, transformation efficiencies are shown relative to that obtained with plasmid p45.4.A2, which yielded 300 to 400 foci per μg (49, 50) in the same experiment. For the other four constructures, transformation efficiencies were normalized to that of the wild-type clone p53.A6.6 (63), which yielded 4,900 to 5,200 foci per μg of DNA with the procedure (64) used.

using ultrathin polyacrylamide-urea gels (54). The deletions in *dl*1001 and *dl*1004 were determined in the same way, as were those in mutants of the pdl2000 series not described previously.

Preparation of HeLa cell extracts. HeLa cells were grown in suspension culture in Joklik medium (Flow Laboratories, Irvine, Scotland) supplemented with 10% fetal calf serum and glutamine to a density of approximately 5×10^5 cells per ml. Extracts were prepared by the procedure of Manley et al. (41).

In vitro incubations and purification of RNA products. RNA synthesis and purification was as described previously (34), unless otherwise specified in the figure legends.

S1 nuclease gel mapping. For S1 nuclease gel mapping (8, 58, 73), in vitro products were annealed to 5'-^{32}P-labeled single-stranded viral DNA fragments, and the resulting hybrids were treated with S1 nuclease (Kamen et al., in press). S1-resistant DNA products were fractionated on 8 or 12% urea-polyacrylamide sequencing gels (54). The 5' ends of RNAs synthesized in vitro from region A substitution mutants were mapped, with the E-DNA strand of the *Hinf*I fragment (labeled at nt388 and extending to nt5077) or the *Dde*I fragment (labeled at nt188 and extending to nt4389) of Py DNA used as probes. Complementary strands of the DNA probes were separated on 4% polyacrylamide strand separation gels (44); the slower migrating strands were complementary to early mRNAs. The 5' ends of RNAs transcribed from other deletion mutants were mapped by hybridization to the E-strand of the *Dde*I fragment used for sequence determination.

Reverse transcriptase primer extension. The modifications of the primer extension procedure (23) recently described (62; Kamen et al., in press) were used. A single-stranded 63-nucleotide-long primer (a *Dde*I

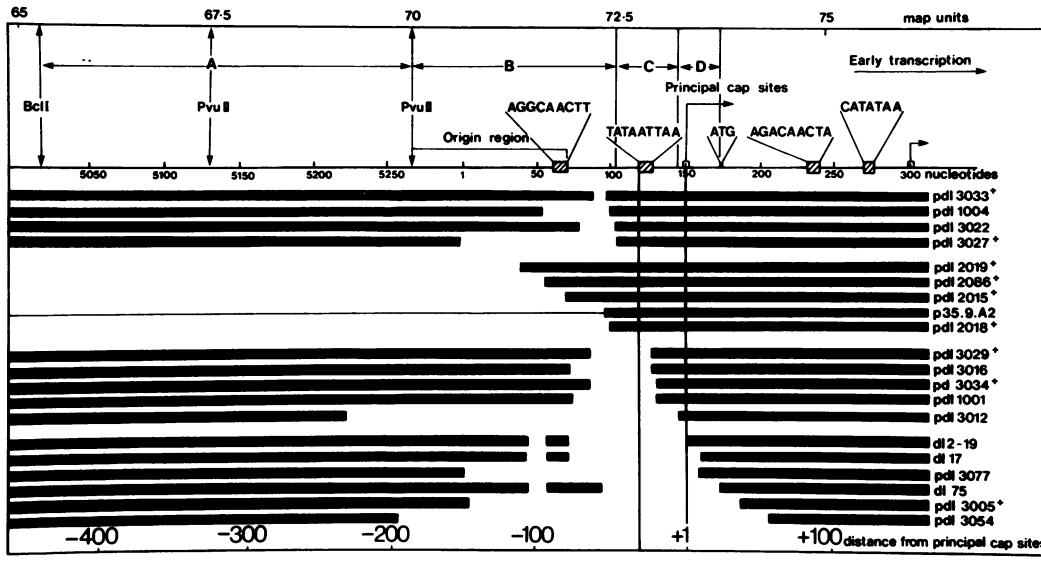

FIG. 2. Structures of recombinant plasmids containing Py DNA with modifications in regions A, B, C, and D. The portion of the Py DNA sequence spanning the sequences flanking and including the beginning of the early region is shown. The subdivisions A through D (see text) are indicated. Coordinates are given in map units (top line, 25), nucleotide number (middle line; 57, 68), and distance from the principal cap sites (bottom of the figure). The direction of early transcription is indicated. Locations of the mRNA cap sites (Kamen et al., in press) TATA boxes (Goldberg, thesis), CAATC boxes (11), and the translational initiation codon (57) are shown along the nucleotide number coordinate. Structures of the recombinant plasmids are shown with the viral sequence emboldened. All plasmids except p35.9.A2 contain *Bam*HI linear viral DNA with internal deletions; plasmid p35.9.A2 contains the viral *Hae*II-*Eco*RI fragment (nt95–1565) inserted between the *Hae*II site at nt547 and the *Eco*RI site of pAT153 (67). Plasmids denoted "+" contain an *Xho*I linker inserted at the site of the deletion. Mutants *dl*17, *dl*75, and *dl*2–19 (3, 4) are viable mutants obtained from W. Folk. Note that these were derived from the Py A3 strain, which differs from our wild-type strain A2 by an 11-bp deletion (57); this deletion has no effect on RNA synthesis in vivo (Kamen et al., in press).

fragment labeled at nt443 and extending to nt381) was used. The complementary strands of the primer were separated on 8% polyacrylamide strand separation gels (44); the faster migrating strand was complementary to early region mRNAs.

RESULTS

The structures of the substitution and deletion mutants used in this study are shown in Fig. 1 and 2, relative to positions of sequence elements potentially important in the specification of Py early region transcription (Fig. 2). Some of the mutants were constructed in recombinant plasmids expressly for the experiments discussed here. Others are viable Py deletion mutants isolated in mouse cells (3, 4, 40). In the description of the results obtained with these mutants, we shall refer to the four subdivisions of the 5'-flanking sequences indicated by A through D in Fig. 1 and 2. Positions of deletions and of the boundaries of the subdivisions are measured from the midpoint of the principal in vivo mRNA (Kamen et al., in press) cap site region at nt150 (nucleotide +1). Region A (−425 to −178) includes remote upstream sequences. Region B (−177 to −46) extends across the origin of viral DNA replication and includes the sequence homologous to GGPyCAATCT. Region C (−45 to −5) spans the TATA box. Region D (−4 to +23) contains the principal cap sites and ends just before the coding sequences.

Sequences within region A are important for early gene expression in vivo. Recent experiments with simian virus 40 (SV40), which is rather closely related to Py, demonstrated that sequences far upstream of the early cap sites, on the late side of the replication origin, were required for efficient early gene expression in vivo. The critical sequence is a 72-bp direct repeat, but either copy is sufficient (6). The sequences of the Py genome immediately on the late side of its replication origin are designated region A in Fig. 1. There is no highly significant nucleotide sequence homology between Py region A and SV40 DNA; in the A2 Py strain (57), there are no long tandem repeats. To determine whether sequences within Py region A were nevertheless important for early gene expression, we constructed the recombinants shown in Fig. 1 and tested their abilities to transform rat cells in culture. The first seven constructions were substitution mutants in which the vector

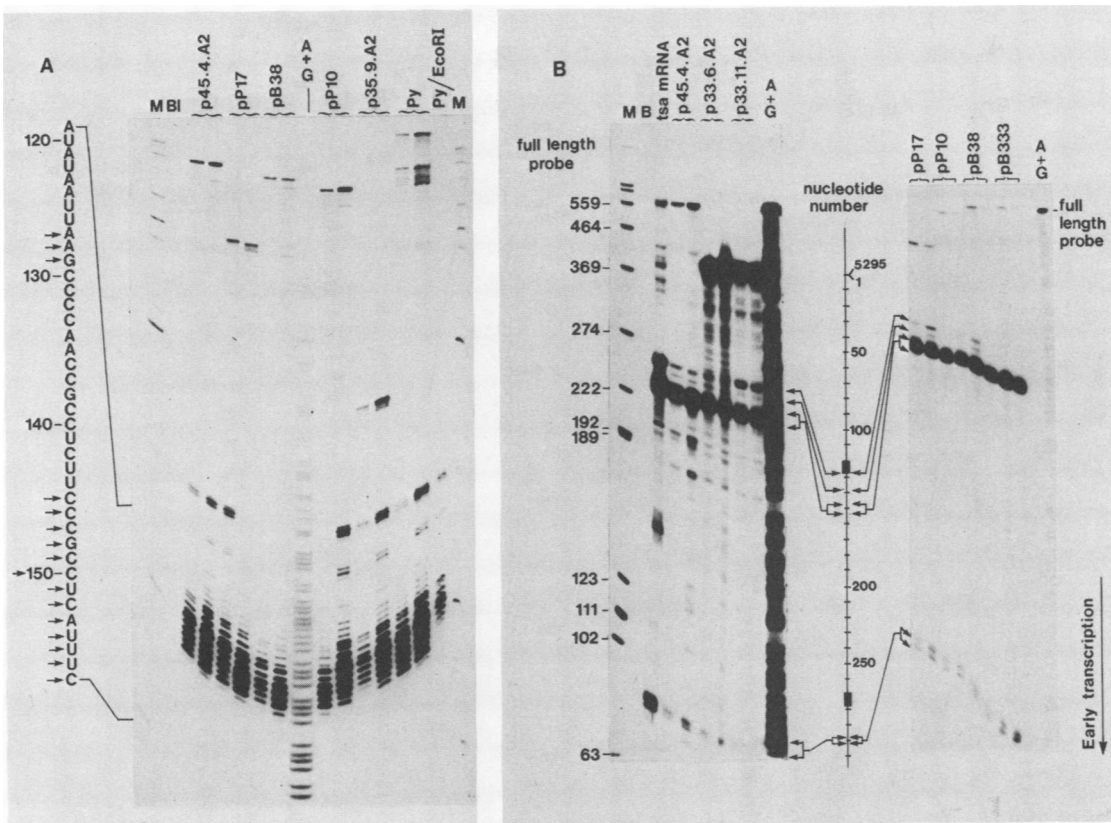

FIG. 3. Localization of 5′ ends of in vitro transcripts of region A substitution mutants by S1 nuclease gel mapping. (A) Each hybridization contained ca. 5 ng of the Py *Dde*I fragment (see the text; 2 μCi/pmol) and RNA transcribed in vitro from a 37.5-μg/ml sample of the indicated DNA template in amounts corresponding to 10-μl (lefthand track of each pair) and 20-μl (righthand tracks) transcription reactions except track Py/*Eco*RI, where the equivalent of a 10-μl transcription reaction was annealed. The DNA templates were uncut plasmids or *Eco*RI-cleaved linears as indicated. "Bl" is hybridization with carrier RNA alone; "A+G" is a purine-specific cleavage of the probe fragment; "M" are 5′-^{32}P-labeled *Bcl*I + *Dde*I fragments of Py DNA used as size markers. The S1-resistant hybrids were fractionated on an 8% polyacrylamide-urea gel. The correspondence between the bands on the autoradiogram and position on the genomic sequence is indicated. (B) Each hybridization contained ca. 3 ng of the Py *Hin*fI fragment (see the text; 1.5 μCi/pmol) and RNA synthesized in vitro from the templates indicated at the tops of the gel tracks. Plasmids p45.4.A2, p33.6.A2, and p33.11.A2 were cleaved with restriction endonuclease *Aval* and were used as templates at 30 μg/ml; RNA corresponding to 62.5-μl (left-hand track of each pair) and to 125-μl (right-hand track of each pair) transcription reactions was annealed. p33.6.A2 and p33.11.A2 contain the same viral DNA insert as plasmid pP17 shown in Fig. 1. The pP17, pP10, pB38, and pB333 templates were also cleaved with *Aval*, but the DNA concentration in the transcription reactions used was 12.5 μg/ml and amounts of RNA corresponding to reactions of 50 μl (left-hand tracks) and 25 μl (right-hand tracks) were hybridized. The "tsa mRNA" track is hybridization with cytoplasmic RNA extracted from mouse cells grown at the permissive temperature for 36 h after infection with mutant *tsA* (21) virus and shifted to the nonpermissive temperature 4 h before harvesting. The deduced positions of the principal termini on the genomic DNA sequence are indicated in the center. The lengths of the largest S1-resistant DNA products generated with RNA synthesized in vitro in (A) and (B) correspond to the viral plasmid junctions in the templates; they represent a summation of all initiations in the vector sequences. The DNA products apparently mapping 5′ termini at nt128±2 are the S1 artifacts described previously (18; Kamen et al., in press).

*Bam*HI site was fused to viral DNA at the beginning of region A (the viral *Bcl*I site) and at six further internal positions. Since the viral *Bcl*I-*Eco*RI fragment is known to transform cells (49; J. Hassell, personal communication) with an efficiency similar to full-length DNA, we used a recombinant plasmid containing this fragment (p45.4.A2) as the positive control. We found (Fig. 1) that progressive removal of sequence from region A caused an increasingly severe inhibition of transforming activity which reached a 30-fold decrease when the entire region was substituted. Complementary results were obtained with internal deletions (Fig. 1).

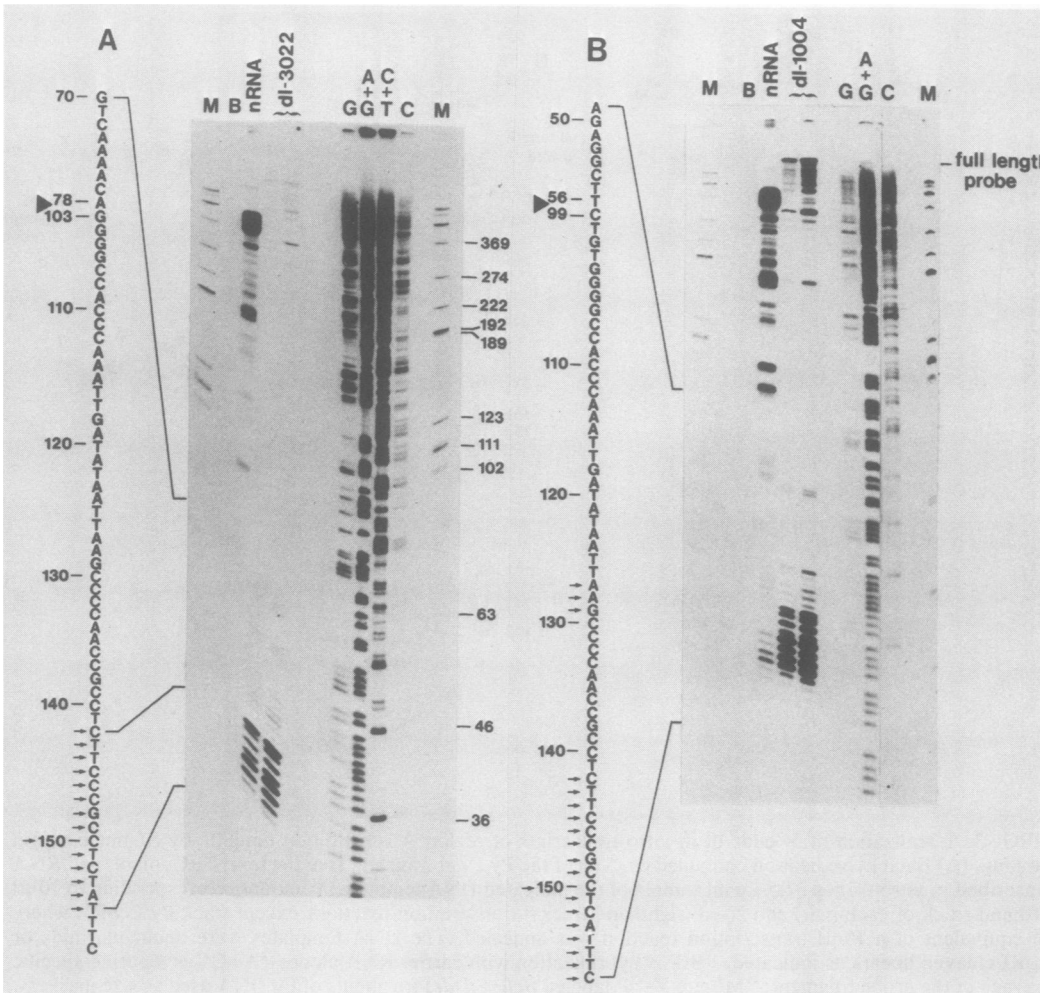

FIG 4. Localization of 5′ ends of in vitro transcripts of region B deletion mutants by S1 nuclease gel mapping. Each hybridization contained ca. 3 to 4 ng of the homologous *Dde*I fragment (see the text; 1–2.5 μCi/pmol) and the following RNA samples: B, carrier RNA; nRNA, nuclear RNA (0.3 μg) extracted from *tsA* shift-up cells as described in legend of Fig. 3; in vitro RNA corresponding in amount to 10- and 30-μl transcription reactions transcribed from 12.5 μg/ml of the indicated plasmids restricted with either *Ava*I (pdl3022 and pdl1004) or *Eco*RI (pdl3027). Lanes G, A+G, C+T, and C are base-specific cleavages of the probe fragment. The deduced sequence of the E DNA strand of each mutant is shown on the left of each autoradiogram; the numbering system corresponds to the position on the wild-type (strain A2 [25]) sequence. The location of the deletion (▶) and the 5′ ends of in vitro transcripts (→), on the mutant E strand sequence, are indicated; the insertion of the *Xho*I linker at the site of the deletion in mutant *dl*3027 is also shown. In (C), gel bands corresponding to initiation at the principal start sites are not visible in the nRNA track but were clearly visible on a darker autoradiographic exposure (not shown).

Deletion mutant pdl2020P, lacking the part of region A which was retained in substitution mutant pP10, had the same four- to fivefold reduction in transformation efficiency. A shorter deletion (pdl2005) spanning the common endpoint of these two mutants had the same very low efficiency as a mutant lacking all of region A (pP17). This suggested that sequences located on both sides of the *Pvu*II site (nt5130; see Fig. 1) within region A are important for viral transformation and, by implication, possibly for viral gene expression. We therefore constructed and characterized a more extensive set of deletion mutants (68). The phenotypes of these mutants determined in transient assays by T-antigen immunofluorescence, immunoprecipitation of ^{35}S-labeled early proteins, and nuclease S1 gel mapping of early region mRNAs (68) identified sequences within region A essential for early gene expression. The progressive decrease not-

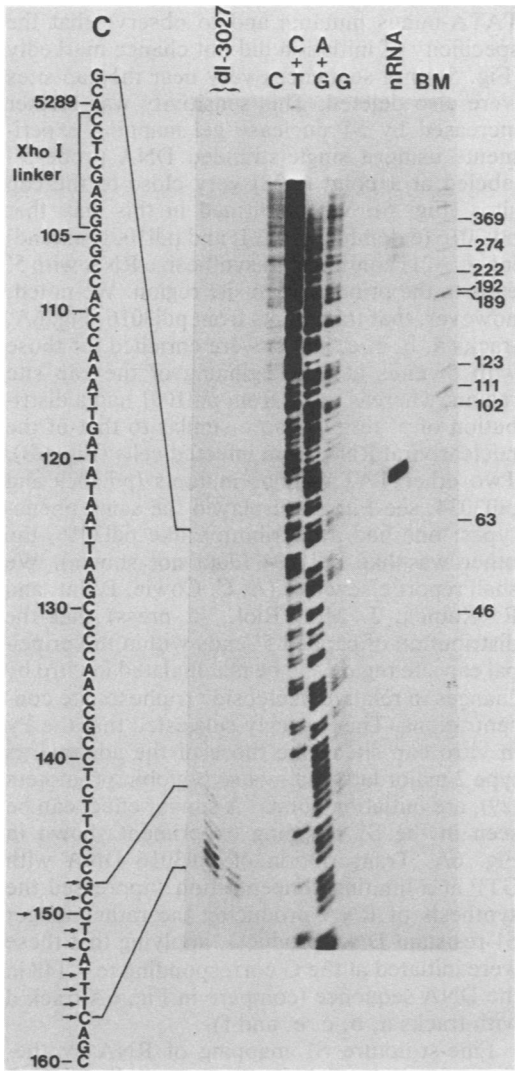

FIG. 4C.

ed in transforming activity was reflected by intermediate effects in transient expression assays. The data obtained (68) suggested the existence of at least two important elements located on either side of the *Pvu*II site. Precise localization of the critical sequences will require analysis of further mutants. However, it is clear that deletions such as pdl2005 (see Fig. 1) express early region RNA very poorly; they are also *cis* defective in viral DNA replication (68).

Sequences before position −46 (regions A and B) are not required for accurate and efficient expression in vitro. We have previously shown that the Py early region is faithfully transcribed in the HeLa cell crude extract in vitro system (34), producing RNA polymerase II transcripts with principal 5′ ends at nt148–153 and minor 5′ termini at nt300±2 (Fig. 2). A substitution mutant, in which plasmid DNA replaced viral sequences 5′ to position −55, appeared to function in vitro like wild-type DNA (34). We extended this observation by transcribing plasmids with substitutions or deletions within regions A and B shown in Fig. 1 and 2. Products were analyzed by S1 nuclease gel mapping. Representative results shown in Fig. 3B indicate that mutants lacking all or part of region A (pB333, pB38, p33.6.A2, p33.11.A2, pP17, and pP10) were efficiently transcribed in vitro: initiation of RNA synthesis was at the normal start sites, nt148–153 and nt300±2. Moreover, the results in Fig. 3A show that deletion of region A sequences did not change the 5′-end microheterogeneity in the principal cap site region (compare tracks 1 through 4 with track 6). Results mapping the 5′ termini of RNA synthesized from *dl*3022 (Fig. 4A), *dl*1004 (Fig. 4B), and *dl*3027 (Fig. 4C) indicated that sequences upstream of −46 (the limit of the deletion in *dl*3027; Fig. 2 and 4C) could be deleted without altering the specificity of transcription. Thus, region A (essential for expression in vivo) and region B (which includes the GGPyCAATCT homology) are not required for specific transcription in vitro.

The effect of region B sequences on the transcriptional efficiency was determined more quantitatively by examining the transcripts of selected mutants by the primer extension procedure (Fig. 5). We were able to use the minor 5′ end at nt300±2 as an internal standard to minimize experimental variation because deletions within regions A through D had no measurable effect on initiation at nt300±2. The results summarized in Table 1 show that elimination of sequences before −46 reduces transcriptional efficiency from the principal cap sites by no more than about twofold. We hesitate to interpret the small differences (from 45 to 82% of control values) observed among mutants *dl*3033, *dl*1004, *dl*3022, and *dl*3027 (Table 1) as significant. Further work using a more extensive set of smaller deletions would be necessary to establish the importance of such small effects.

The TATA box in region C is important for efficient in vitro expression but is not essential for accurate initiation. We next describe the expression in vitro of deletion mutants removing the TATA box (−30 to −22) and further sequences from region C. The results for three such mutants (pdl3016, pdl1001, and pdl3012), shown in Fig. 5 and summarized in Table 1, demonstrate that removal of most (pdl3016) or all (pdl1001) of the TATA element reduced transcription from the principal cap sites by 8- and 20-fold, respectively. Further extension of the deletion to within 5 bp of the cap site region midpoint (pdl3012) caused a more severe inhibition, and in this case

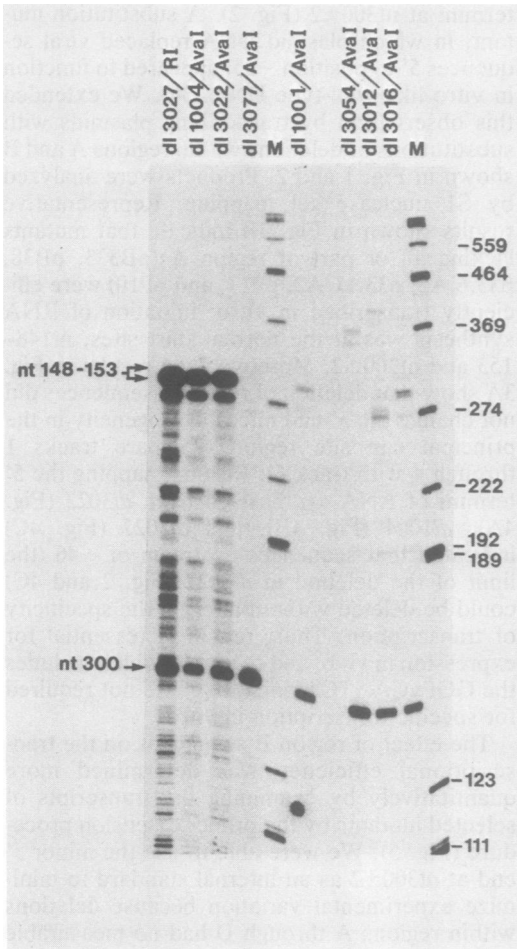

FIG. 5. Localization of 5′ ends of in vitro transcripts by primer extension. Each hybridization contained ca. 20,000 dpm of the *DdeI* fragment (labeled at nt443 and extending to nt381) and in vitro-synthesized RNA samples corresponding to 20-μl incubations using 12.5 μg of the indicated template DNA per ml. The correspondence between bands on the autoradiogram and the deduced position of 5′ ends on the DNA sequence is indicated on the left. The DNA product ca. 274 nucleotides in length was not detected by S1 gel mapping and therefore represents a reverse transcriptase stop site, rather than a bona fide 5′ end. The many faint bands were not reproducibly found with the primer extension method and were not represented by corresponding bands in S1 experiments; they are also therefore reverse transcriptase stops.

other cryptic apparent initiation sites were detected.

Results obtained with a number of other genes (13, 28, 33, 37, 39, 55, 60, 65, 66, 70) had led us to expect substantial decreases in efficiency upon removal of the TATA box. The analytical method we have used, however, afforded the sensitivity to detect the transcripts produced by TATA-minus mutants and to observe that the specificity of initiation did not change markedly (Fig. 5) until sequences very near the cap sites were also deleted. This sensitivity was further increased by S1 nuclease gel mapping experiments using a single-stranded DNA probe 5′-labeled at a point (+38) very close to the cap sites (Fig. 6). We confirmed in this way that pdl3016 (extending to −24) and pdl1001 (extending to −21) continued to synthesize RNA with 5′ ends in the principal cap site region. We noted, however, that transcripts from pdl3016 (Fig. 6A, tracks a, b, c, e, and f) were enriched for those with 5′ ends at the beginning of the cap site region, whereas those from pdl1001 had a distribution of 5′ termini more similar to that of the nuclear viral RNA from infected cells (Fig. 6B). Two other TATA-minus mutants (pdl3029 and pdl3034; see Fig. 2) displayed the same phenotypes: one had a distribution like pdl3016, the other was like pdl1004 (data not shown). We shall report elsewhere (A. C. Cowie, P. Jat, and R. Kamen, J. Mol. Biol., in press) that the distribution of capped 5′ ends within the principal cap site region can be manipulated in vitro by changes in relative nucleoside triphosphate concentrations. This strongly suggested that the Py in vitro cap sites, like those of the adenovirus type 2 major late and mouse β-globin promoters (29), are initiation points. A similar effect can be seen in the S1 mapping experiment shown in Fig. 6A. Transcription of pdl3016 DNA with GTP at a limiting concentration suppressed the synthesis of RNA producing the rather longer S1-resistant DNA products, implying that these were initiated at the G corresponding to nt148 in the DNA sequence (compare in Fig. 6A track d with tracks a, b, c, e, and f).

Fine-structure S1 mapping of RNA synthesized from the deletion mutant ending at −5 (pdl3012) confirmed that correct initiation is further inhibited, but still occurs at a very low level (Fig. 6C). It is apparent in the experiment shown here that far more RNA chains were initiated at upstream positions, mostly within flanking plasmid sequence, than at the principal cap sites. The prominent S1-resistant product corresponding to an apparent 5′ end at nt5129 (Fig. 6C) is caused by RNA-RNA annealing. It in fact maps the 5′ termini of late region in vitro transcripts complementary to the opposite DNA strand which are known to occur at this position (34). We have previously termed such S1-resistant DNA products "shadows" and described the methodological artifact which produces them (18; Kamen et al., in press).

Mutants lacking both the TATA box and principal cap sites are transcribed very inefficiently in vitro but still express well in vivo. Py viable deletion mutants (3, 4) without the principal cap

site region and the TATA box which precedes it (region C+D mutants in Fig. 2) express early region mRNA efficiently during the productive infection of mouse cells but produce RNAs with heterogeneous 5′ ends (Kamen et al., in press; see also Fig. 7). These 5′ ends correspond to those of very minor species also present among transcripts of wild-type templates. When DNA of the three mutants studied in vivo (*dl*17, *dl*75, and *dl*12–19; cf. Fig. 2) was transcribed in vitro, we found by contrast that they, like other deletions extending into region D (pdl3077 and pdl3054 in Fig. 5 and Table 1), functioned very inefficiently. Figure 7 shows representative S1 nuclease results obtained with in vitro transcripts of recombinant plasmids containing different fragments of *dl*75 DNA (the deletion is from −57 to +21). The only major S1-resistant DNA products obtained corresponded to the 5′ termini at nt300±2 and to the vector/viral DNA junction in the template DNA. The former indicate that initiation at this minor site is unaffected by the *dl*75 deletion; this cannot account for the mutant's viability because nt300 is well within early region coding sequences. The latter only represent the summation of all background initiations in flanking vector sequence. However, the very minor S1-resistant DNA products protected by *dl*75 in vitro transcripts and the heterogeneous pattern of products obtained with RNA from *dl*75-infected cells showed remarkable similarities (compare the faint bands larger than 190 nucleotides in Fig. 7, track D, with the prominent bands in the *dl*75 in vivo track).

To quantitate the in vitro transcriptional efficiencies of mutants which produced highly heterogeneous 5′ termini, we annealed in vitro products to a small DNA probe spanning the minor initiation site at nt300±2 and determined the ratio between S1 products corresponding to full-length protection and those representing starts at nt300±2. This method integrates initiations at any point within the beginning of the early region or upstream into a single gel band. The results summarized in Table 2 demonstrate that even after such integration, *dl*17 and *dl*75 were found to express very poorly. The efficiency of mutant *dl*2–19 was somewhat better, but this deletion ends at position −1 with respect to the cap sites. Other S1 mapping experiments (not shown) indicated that the residual cap site sequences in *dl*2–19, like the cap sites in region C mutant pdl3012 (Fig. 6C), function at low level.

Restriction endonuclease cleavage of the template DNA within the 100 bp before the cap sites inhibits transcription in vitro. Several of the deletion mutants used in the experiments described above have *Xho*I linkers inserted at the site of deletion. These allowed us to determine the length of upstream DNA sequence required for efficient initiation in vitro by comparing the activity of the templates with and without *Xho*I digestion. The S1 nuclease method was used for this experiment, and once again data were normalized to the minor start at nt300±2 (Table 3). We found that cleavage at positions −52 to −61 reduced efficiency by about fivefold. The effect moderated when 89 and 103 bp were present in the template upstream of the cap sites. Restriction cleavage at −119 and −181 afforded a small stimulation relative to uncut template, but this is unlikely to be beyond experimental error. Thus, although DNA sequence upstream of −46 can be replaced with other sequences, removal of the DNA entirely has a major inhibitory effect. We noted with interest that the Py early region sequence homologous to GGPyCAATCT occurs between positions −80 and −87. Our result suggested that the polymerase or some other transcriptional factor may make contact with the DNA up to and including the position of this homology sequence. We therefore examined three further constructions (pdl2018, p35.9.A2, pdl3027) with different sequences other than wild-type DNA upstream of the cap sites and found that in each case restriction endonuclease cleavage at −50 to −55 inhibited transcription, but cleavage further than 120 bp upstream did

TABLE 1. Efficiencies of in vitro initiation at the principal cap sites measured by the primer extension assay[a]

DNA template[b]	Endpoints relative to cap site[c]	Efficiencies relative to the wild-type
p37.3.A2	Wild type	100
pdl3033*	−61, −55	55
pdl1004	−93, −52	82
pdl3022	−71, −48	45
pdl3027*	−155, −46	50
pdl3016	−77, −24	13
pdl1001	−75, −21	4
pdl3012	−224, −5	<0.7[d]
pdl3077	−155, +9	0[e]
pdl3054	−198, +55	0[e]

[a] Data are averages from repeated experiments like that shown in Fig. 4. Gel bands corresponding to 5′ termini at nt148–153 and at nt300 ± 2 were excised and counted. The ratio between the two was determined and used to calculate normalized efficiencies of initiation at nt148–153 relative to wild-type DNA.

[b] All templates were restricted with *Ava*I except those indicated by *, which were restricted with *Eco*RI.

[c] Distances are in base pairs from the midpoint of the principal cap site region (nt150).

[d] The 5′ ends at nt148–153 were hardly detectable, but other heterogeneous 5′ termini occurred (Fig. 4).

[e] The sequence at nt148–153 was deleted. Inefficient initiation at other cryptic sites was found (see Fig. 4).

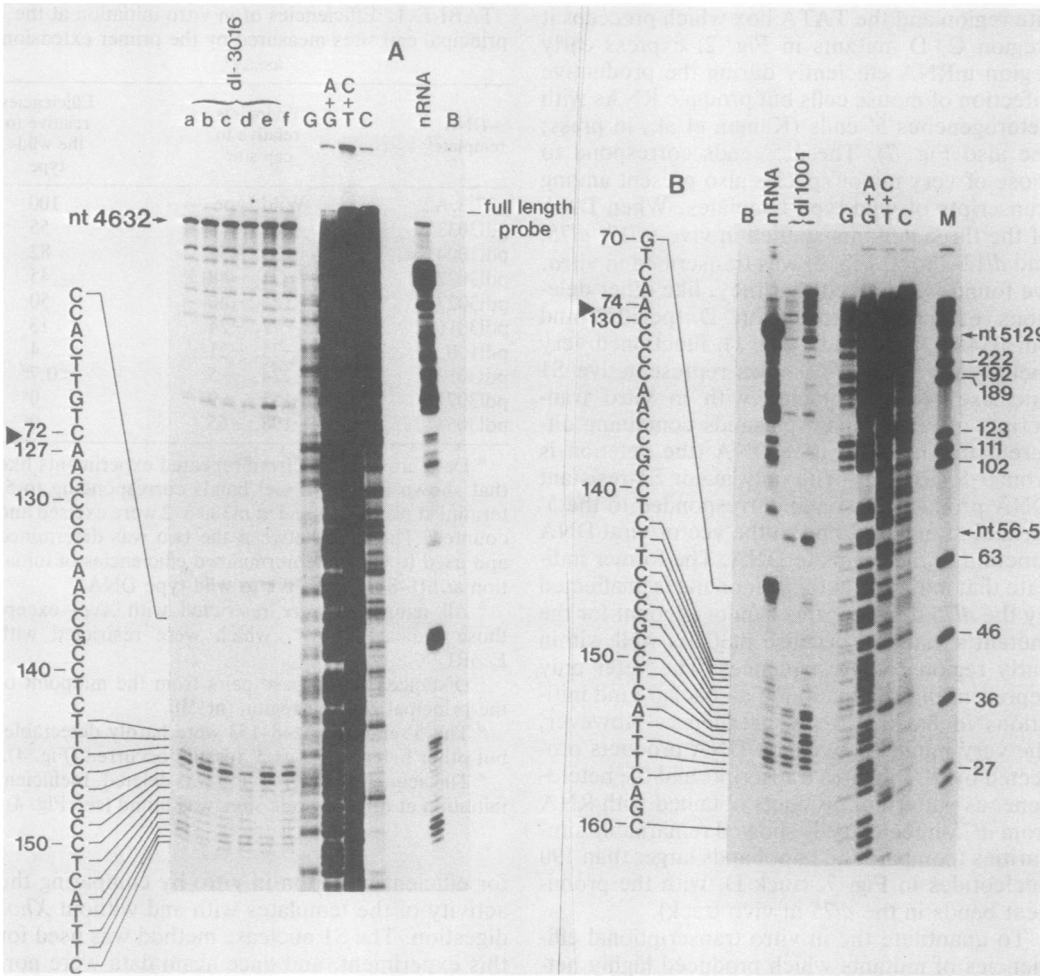

FIG. 6. Localization of 5′ ends of in vitro transcripts of region C deletion mutants by S1 gel mapping. The homologous *Dde*I fragment was used for sequence determination and mapping the 5′ termini of in vitro transcripts. (A) RNA was transcribed in vitro from 12.5 μg of *Ava*I restricted plasmid pdl3016 per ml of reaction at the following ribonucleoside triphosphate concentrations: lane a, 0.7 mM ATP, 0.4 mM UTP, 0.3 mM CTP, and 0.2 mM GTP; lane b, 0.5 mM each ATP, GTP, and CTP and 0.05 mM UTP; lane c, 0.5 mM each ATP, GTP, and UTP and 0.05 mM CTP; lane d, 0.5 mM each ATP, CTP, and UTP and 0.05 mM GTP; lane e, 0.5 mM each GTP, CTP, and UTP and 0.05 mM ATP; lane f, 0.05 mM each ATP, GTP, CTP, and UTP. RNA corresponding to a 30-μl incubation was hybridized. (B) RNA transcribed from pdl1001 in amounts corresponding to 10- and 30-μl standard reactions was hybridized. (C) RNA transcribed from pdl3012, corresponding to a 15-μl standard reaction, was hybridized. The mutant DNA sequence of the E strand is shown on the left of each panel. The positions of the deletion (▶) and the 5′ termini on the mutant DNA sequence are indicated. In (A) and (C), the gel band corresponding to a 5′ end at nt4632 results from the viral/plasmid DNA junction at the *Bam*HI site. The S1-resistant product corresponding to an apparent 5′ end at nt5129 (B and C) is caused by RNA-RNA reannealing (see text) between the late region in vitro transcript with its 5′ end at this position (34) and continuous transcripts extending in the early direction.

not. This suggests that it is the length of DNA, rather than its sequence, which is crucial.

DISCUSSION

We have identified regions of the Py genome necessary for transcription from the early promoter, both in vivo and in vitro. Substitution and internal deletion mutants were constructed in recombinant plasmids and assayed for expression in vivo by transformation of rat cells (49, 50, 64) and in vitro by transcription in the crude HeLa whole-cell extract (41). From these experiments, we conclude the following. (i) Sequences located within a remote upstream region (between 425 [−425] and 178 bp [−178] before the principal cap sites, designated region A in Fig. 1 and 2) are necessary for expression in

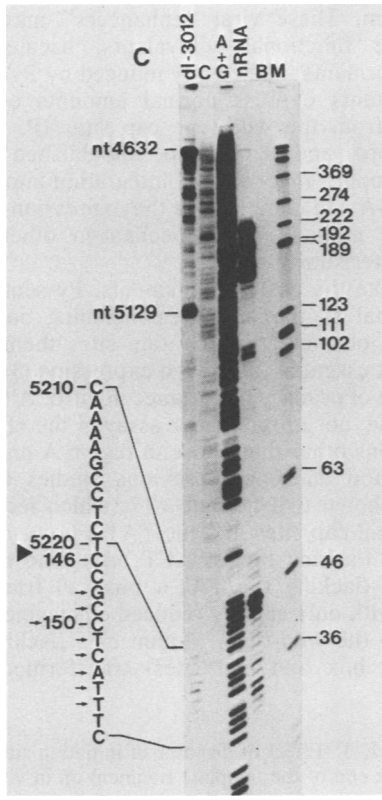

FIG. 6C.

vivo but not for transcription in vitro. (ii) Replacement of sequences 5′ to −55 bp with plasmid DNA, or internal deletion of sequences before −46 bp, has only a small (twofold or less) effect on the efficiency of accurate in vitro transcription. (iii) Deletion of all or part of the TATA box −30 to −22) reduces in vitro transcription efficiency by 10 to 20-fold, but the normal start sites are still preferentially used. (iv) Deletion of sequences up to position −5 further reduces the efficiency of correct initiation, but this is accompanied by the appearance of other heterogeneous start sites. (v) Deletions removing both the TATA box and cap sites reduce overall transcription efficiency in vitro by at least 20-fold and generate transcripts with highly heterogeneous 5′ termini. Such deletion mutants, however, are viable. (vi) About 90 to 100 bp of DNA sequence 5′ to the cap sites is required in the DNA template for optimal transcription efficiency in vitro.

The observation that sequences within region A, located some 200 to 400 bp before the cap sites, were essential for expression in vivo but are not required for transcription in vitro was a surprise. It suggested that the rate-limiting step in vivo differs from that involved when purified DNA is added to a cell-free system. We have unsuccessfully attempted to mimic the in vivo effects by transcribing in vitro recircularized mutant templates which contained only viral DNA sequences. Therefore, the difference is not caused by the presence of procaryotic vector DNA. It may, alternatively, reflect the influence of chromatin structure, but we have not as yet investigated the structure of the molecules actually transcribed in the in vitro system to ask whether they are free molecules or reassembled chromatin.

The transformation results presented here suggested that the important sequences within Py region A were located on either side of the endonuclease *Pvu*II cleavage site at nt5130 (314 bp upstream of the cap sites; see Fig. 2). These observations have been extended by Tyndall et

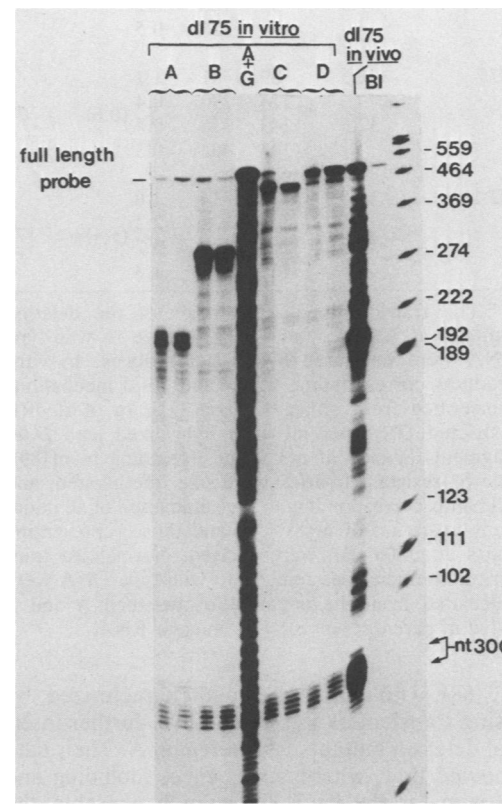

FIG. 7. Localization of 5′ ends of *dl75* in vitro transcripts by S1 nuclease gel mapping. Each hybridization contained ca. 4 ng of the *Hin*fI fragment (2 μCi/pmol) of *dl75* DNA and various RNA samples, synthesized in vitro from 30 μg of *Ava*I-restricted plasmids per ml. These plasmids contained *dl75* sequences from the *Hae*II site (nt95, lanes A), *Pvu*II sites (nt5268 [lanes B] and nt5131 [lanes C]), and *Bcl*I site (nt5022, lanes D) to the *Eco*RI site (nt1565) inserted in the vector pAT153 (67); *dl75* in vivo corresponds to hybridization with cytoplasmic RNA extracted from mouse cells infected with *dl75* virus.

TABLE 2. In vitro transcription efficiencies of viable deletion mutants[a]

RNA	Template DNA concn (μg/ml)	cpm of A	cpm of B	Ratio A/B		Efficiency relative to wild type (%)
tsA mRNA		640	94	6.7		
tsA nRNA		1,077	17	63.4		
Wild type	12.5	287	40	7.1		
		717	65	11.0	(8.3)	100
	37.5	1,341	194	6.9		
*dl*17	12.5	37	85	0.4		
		57	109	0.5		
	12.5	47	78	0.6	(0.4)	5
		37	124	0.3		
	37.5	37	163	0.2		
		147	326	0.5		
*dl*75	12.5	89	81	1.1		
		123	290	0.4	(0.6)	7
	37.5	143	376	0.4		
		143	462	0.3		
*dl*2-19	12.5	108	114	1.0		
		281	131	2.2	(1.4)	17
	37.5	164	152	1.2		
		463	375	1.3		

[a] The transcriptional efficiencies of the deletion mutants *dl*17, *dl*75, and *dl*2-19 relative to wild-type DNA were measured by S1 gel analysis. In vitro products corresponding to 10- and 20-μl incubations transcribed from either 12.5 or 37.5 μg of *Eco*RI-restricted DNA per ml were hybridized to a *Dde*I fragment (labeled at nt380 and extending to nt189). The S1-resistant hybrids were size fractionated, and gel bands corresponding to the summation of all initiations upstream of nt189 (A), and those representing starts at nt300 (B), were excised. Normalized transcription efficiencies relative to wild-type DNA were calculated from the mean ratios between A and B listed in parentheses. nRNA, Nuclear RNA.

al. (68), who constructed and characterized, by using transient expression assays, further internal deletion mutants within region A. Their data showed that, whereas sequences including and proximal to the *Pvu*II site were dispensable, the simultaneous removal of two domains located on either side of this site eliminated gene expression and also caused a *cis*-acting defect in viral DNA replication. An upstream element essential for in vivo early region expression but not required for transcription in vitro has also been located in SV40 DNA (6). This element is a 72-bp direct repeat, either copy of which is sufficient. Both the SV40 72-bp sequence (2, 47) and Py region A (15) dramatically enhance the in vivo expression of other eucaryotic genes linked to them. These viral "enhancers" may have cellular functional equivalents, because the transformants inefficiently induced by Py region A mutants express normal amounts of viral RNA from the wild-type cap sites (P. Jat, J. Favaloro, and R. Kamen, unpublished data). This implies that selected integration into cellular DNA can complement the expression defect of the mutants by a mechanism other than promoter substitution.

Unlike the upstream elements, Py sequences proximal to the cap sites including both the TATA box and the initiation sites themselves are not essential for in vivo expression (3, 4, 40) but are of primary importance in vitro. Although we have not in this paper assayed the effect of deletions other than those in region A on transformation efficiency, previous studies (3, 40) have shown that mutant *dl*17 (which lacks the principal cap sites and the TATA box), mutant *dl*1004 (lacking the CAACT box), and mutant *dl*1001 (lacking the TATA box) all transform cells with only slightly reduced efficiencies relative to the wild-type. Mutant *dl*75 (lacking the TATA box and cap sites) transformed with

TABLE 3. Effect of distance of initiation site from the end of the template fragment on in vitro transcription efficiency[a]

Mutant	*Xho* cleavage point	*Xho* cut	Ratio A/B	Approx relative promoter efficiency
*dl*3027	−52	+	2.2	0.19
		−	11.7	
*dl*2018	−58	+	2.4	0.23
		−	10.4	
*dl*3033	−61	+	1.6	0.16
		−	10	
*dl*2015	−89	+	3.9	0.35
		−	11.3	
*dl*2086	−104	+	10.1	0.87
		−	11.55	
*dl*2019	−119	+	14.65	1.4
		−	10.25	
p43.34.70	−184	+	10.7	1.2
		−	9.2	

[a] Measured by S1 nuclease gel mapping. Data were obtained by hybridizing RNA transcribed in vitro from either *Xho*I-restricted (ratios designated +) or *Eco*RI-restricted (ratios designated −) plasmid DNAs to the Py DNA *Hin*fI fragment (see the text). The S1-resistant products were size fractionated, and the gel bands corresponding to initiations at nt148–153 (A) and nt300 (B) were excised. The yield of S1-resistant DNA was determined by Cerenkov counting. Initiation of transcription at the principal cap sites was normalized relative to initiation at the minor start site at nt300 for each DNA template (ratio A/B), and the pair of normalized ratios (+ and −) were used to calculate the relative promoter efficiency.

reduced efficiency (3), but this reflects the quality rather than the efficiency of gene expression because the majority of the mRNA produced by this deletion mutant starts 3′ to the translational initiation codon (Kamen et al., in press). Studies with a variety of other genes (16, 19, 26, 45, 46, 59), including the SV40 early region (5, 6, 22, 24) have demonstrated that TATA boxes and the normal cap sites are dispensable in vivo. By contrast, in vitro experiments in many systems have elegantly demonstrated that RNA polymerase II initiation sites are measured from the TATA box. In the in vitro experiments reported here, we confirmed that elimination of this element drastically reduces the efficiency of transcription in vitro. However, using sensitive mapping techniques we observed that the normal initiation sites are still preferentially utilized in TATA-minus mutants. Transcripts with heterogeneous 5′ ends were produced in vitro only when sequences 5 bp or closer to the normal cap sites were removed. This suggested that RNA polymerase II (or associated factors or both) recognizes sequence from the TATA box up to at least the initiation sites. The heterogeneity observed in other systems (6, 26, 43) can be explained by the removal of the entire region rather than just the TATA element alone. In fact, several laboratories (13, 31, 55; R. Roeder, personal communication) have observed that deletion or alteration of sequences at cap sites or further in the 3′ direction can decrease in vitro transcriptional efficiency.

There are certain discrepancies between the results reported here concerning the consequences of TATA box plus cap site deletion and those previously described for the SV40 early promoter (43, 48). Mathis and Chambon (43) used both "run-off" transcription and S1 nuclease gel mapping of in vitro products to study SV40 mutants deleting "TATA plus cap" from the 3′ side. In the run-off assay, wild-type SV40 yields two early region transcripts differing in length by about 80 nucleotides (30, 31, 43). The longer product corresponds to initiation in the SV40 principal early cap site region, but no in vivo equivalent of the shorter one was detectable (31, 43). The TATA plus cap deletions (43) did not express detectable run-off in vitro transcripts corresponding to the longer product, but still produced the shorter one. However, Mathis and Chambon (43) found by S1 nuclease gel mapping that these mutants produced in vitro transcripts with heterogeneous 5′ termini very similar to those synthesized by the same mutants in vivo, and concluded that the sum of these different alternative initiation sites represented an overall efficiency similar to that of wild-type DNA. By contrast, our S1 nuclease analysis of Py viable mutants deleting TATA plus cap from the 5′ side indicated a drastically reduced overall efficiency (Table 2), although the 5′ ends of the in vitro RNAs did resemble those of mRNAs synthesized in vivo (Fig. 7). We have checked one of the SV40 mutants, HS105, as well as an SV40 upstream substitution mutant, PS366 (lacking a complete copy of the 72-bp sequence), in our in vitro system, using both S1 nuclease and primer extension to characterize the products. We found that the transcription efficiency of HS105 was drastically reduced relative to either wild-type SV40 DNA or PS366 when closed circular templates were transcribed (data not shown). Therefore, when assayed in the same HeLa whole-cell extract (41), Py and SV40 TATA plus cap early promoter deletions have similar phenotypes. Myers et al. (48) constructed and analyzed other SV40 3′ deletions of the same type. While they reported that several mutants were transcribed in vitro with nearly the wild-type efficiency, they did not distinguish between run-off transcripts from near the cap site region and those located 80 bases downstream, which have no in vivo equivalent. It is therefore difficult to compare their results with our own, and as we have not assayed the mutants they used ourselves, the issue remains open.

The final point we wish to discuss is our observation that substitution of sequences within Py region B (−177 to −46), which includes an element homologous to the GGPyCAATCT box, had only a small effect on transcriptional efficiency in vitro; similar results have been obtained with several other genes (31, 55, 65, 66, 72). In some of these systems (6, 16, 26, 45, 46; G. Grosveld and R. Flavell, personal communication) sequences within the analogous regions are known to be important for in vivo expression. We are therefore currently determining the phenotypes of Py region B mutants by using transient expression assays. We have observed, however, that optimal in vitro transcription involves a nonspecific requirement for 90 to 100 bp of DNA sequence upstream of the cap sites. This may imply that the polymerase, or another essential protein, interacts with a considerable length of upstream DNA sequence. During the preparation of this manuscript, Tsujimoto et al. (66) and Ostrowski et al. (51) reported similar observations with the *Bombyx mori* fibroin gene and the U3 region of Harvey murine sarcoma virus.

ACKNOWLEDGMENTS

We are grateful to Richard Treisman for helpful discussion, to Kit Osborne for cell culture, and to Penny Morgan, Gina Yiangou, and Audrey Symons for help in the preparation of this manuscript. Pierre Chambon generously provided SV40 deletion mutants HS105 and PS366 for comparative transcrip-

tional efficiency measurements. We thank William Folk and Goran Magnusson for Py deletion mutants.

U.N. thanks the Deutschen Akademischen Anstauschdienst for a fellowship.

LITERATURE CITED

1. **Baker, C. C., and E. B. Ziff.** 1981. Promoters and heterogeneous 5'-termini of the mRNA's of adenovirus serotype 2. J. Mol. Biol. **149**:189–221.
2. **Banerji, J., S. Ruseoni, and W. Schaffner.** 1981. Expression of a genomic rabbit β-globin gene is enhanced by a small segment of simian virus 40 DNA. Cell **27**:299–308.
3. **Bendig, M. M., and W. R. Folk.** 1979. Deletion mutants of polyoma virus defining a non-essential region between the origin of replication and the initiation codon for early proteins. J. Virol. **32**:530–535.
4. **Bendig, M. M., T. Thomas, and W. R. Folk.** 1980. Regulatory mutants of polyoma virus defective in DNA replication and the synthesis of early proteins. Cell **20**:401–409.
5. **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.
6. **Benoist, C., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
7. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. **8**:127–141.
8. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
9. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
10. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli* (HB101). J. Mol. Biol. **41**:459–472.
11. **Breathnach, R., and P. Chambon.** 1981. Organisation and expression of eucaryotic genes coding for proteins. Annu. Rev. Biochem. **50**:349–384.
12. **Cogen, B.** 1978. Virus-specific early RNA in 3T6 cells infected by a tsa mutant of polyoma virus. Virology **85**:222–230.
13. **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.
14. **Cowie, A., C. Tyndall, and R. Kamen.** 1981. Sequences at the capped 5'-ends of polyoma virus late region mRNAs; an example of extreme terminal heterogeneity. Nucleic Acids Res. **9**:6305–6322.
15. **de Villiers, J., and W. Schaffner.** 1981. A small segment of polyoma virus DNA enhances the expression of a cloned β-globin gene over a distance of 1400 base pairs. Nucleic Acids Res. **9**:6251–6264.
16. **Dierks, P., A. Ooyen, N. Van Mantei, and C. Weissman.** 1981. DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L-cells of β-globin RNA with the correct 5' terminus. Proc. Natl. Acad. Sci. U.S.A. **78**:1411–1415.
17. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. 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 **22**:653–668.
18. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: Analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
19. **Faye, G., D. W. Leung, K. Tatchell, B. D. Hall, and M. Smith.** 1981. Deletion mapping of sequences essential for *in vivo* transcription of the iso-1-cytochrome c gene. Proc. Natl. Acad. Sci. U.S.A. **78**:2258–2262.
20. **Freeman, A. E., R. V. Gilden, M. L. Vernon, R. G. Wolford, P. E. Hugunin, and R. J. Huebner.** 1973. 5'-Bromo-2'-deoxyuridine potentiation of transformation of rat-embryo cells induced *in vitro* by 3-methycholanthrene: induction of rat leukemia virus gs antigen in transformed cells. Proc. Natl. Acad. Sci. U.S.A. **70**:2415–2419.
21. **Fried, M.** 1965. Isolation of temperature-sensitive mutants of polyoma virus. Virology **25**:669–671.
22. **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. U.S.A. **78**:100–104.
23. **Ghosh, P. K., V. B. Reddy, M. Piatak, P. Lebowitz, and S. M. Weissman.** 1980. Determination of RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. Methods Enzymol. **65**:580–594.
24. **Gluzman, Y., J. F. Sambrook, and R. J. Frisque.** 1980. Expression of early genes of origin-defective mutants of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **77**: 3898–3902.
25. **Griffin, B. E., M. Fried, and A. Cowie.** 1974. Polyoma DNA: a physical map. Proc. Natl. Acad. Sci. U.S.A. **71**:2077–2081.
26. **Grosschedl, R., and M. L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **77**:1432–1436.
27. **Grosschedl, R., and M. L. Birnstiel.** 1980. Spacer DNA sequences upstream of the TATAAATA sequence are essential for promotion of H2A histone gene transcription *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **77**:7102–7106.
28. **Grosveld, G. C., C. K. Shoemaker, P. Jat, and R. A. Flavell.** 1981. The localisation of DNA sequences necessary and sufficient for the transcription of the rabbit β-globin gene *in vitro*. Cell **25**:215–226.
29. **Hagenbuchle, O., and U. Schibler.** 1981. Mouse β-globin and adenovirus-2 major late transcripts are initiated at the cap site *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **78**:2283–2286.
30. **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.
31. **Hansen, U., D. G. Tenen, D. M. Livingston, and P. A. Sharp.** 1981. T antigen repression of SV40 early transcription from two promoters. Cell **27**:603–612.
32. **Hentschel, C. C., and M. L. Birnstiel.** 1981. The organisation and expression of histone gene families. Cell **25**:301–313.
33. **Hu, S.-L., and J. L. Manley.** 1981. DNA sequence required for initiation of transcription *in vitro* from the major late promoter of adenovirus 2. Proc. Natl. Acad. Sci. U.S.A. **78**:820–824.
34. **Jat, P., J. Roberts, A. Cowie, and R. Kamen.** 1982. Comparison of the polyoma virus early and late promoters by transcription *in vitro*. Nucleic Acids Res. **10**:871–887.
35. **Kamen, R., J. Favaloro, and J. Parker.** 1980. Topography of the late mRNAs of polyoma virus which encode the virion proteins. J. Virol. **33**:637–651.
36. **Kamen, R., J. Favaloro, J. Parker, R. Treisman, L. Lania, M. Fried, and A. Mellor.** 1980. Comparison of polyoma virus transcription in productively infected mouse cells and transformed rodent cell lines. Cold Spring Harbor Symp. Quant. Biol. **44**:63–72.
37. **Lee, D. C., and R. G. Roeder.** 1981. Transcription of adenovirus type 2 genes in a cell-free system: apparent heterogeneity of initiation at some promoters. Mol. Cell. Biol. **1**:635–651.
38. **Legerski, R. J., J. L. Hodnett, and H. B. Gray.** 1978. Extracellular nucleases of Pseudomonas BAL31.111. Use of the double-strand deoxyribonuclease activity as the basis of a convenient method for the mapping of fragments

of DNA produced by cleavage with restriction enzymes. Nucleic Acids Res. 5:1445–1463.
39. **Luse, D. S., J. R. Haynes, D. Van Leeuwen, E. A. Schon, M. L. Cleary, S. G. Shapiro, J. B. Lingrel, and R. G. Roeder.** 1981. Transcription of the β-like globin genes and pseudogenes of the goat in a cell-free system. Nucleic Acids Res. 9:4339–4354.
40. **Magnusson, G., and P. Berg.** 1979. Construction and analysis of viable deletion mutants of polyoma virus. J. Virol. 32:523–529.
41. **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.
42. **Marinus, M. G.** 1973. Location of DNA methylation genes on the *E. coli* K-12 genetic map (GM113). Mol. Gen. Genet. 127:47–55.
43. **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.
44. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with chemical base-specific cleavages. Methods Enzymol. 65:499–559.
45. **McKnight, S. L., E. R. Gavis, R. Kingsbury, and R. Axel.** 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell 25:385–398.
46. **Mellon, P., V. Parker, Y. Gluzman, and T. Maniatis.** Identification of DNA sequences required for transcription of the human α-globin gene using a new SV40 host-vector system. Cell 27:279–288.
47. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, and P. Chambon.** 1981. The SV40 72 base pair repeats have a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. 9:6047–6068.
48. **Myers, R. M., D. C. Rio, A. K. Robbins, and R. Tjian.** 1981. SV40 gene expression is modulated by the cooperative binding of T-antigen to DNA. Cell 25:373–384.
49. **Novak, U., S. M. Dilworth, and B. E. Griffin.** 1980. Coding capacity of a 35% fragment of the polyoma virus genome is sufficient to initiate and maintain cellular transformation. Proc. Natl. Acad. Sci. U.S.A. 77:3278–3282.
50. **Novak, U., and B. E. Griffin.** 1981. Requirement for the C-terminal region of middle-T antigen in cellular transformation by polyoma virus. Nucleic Acids Res. 9:2055–2073.
51. **Ostrowski, M. C., D. Berard, and G. L. Hager.** 1981. Specific transcriptional initiation *in vitro* on murine type C retrovirus promoters. Proc. Natl. Acad. Sci. U.S.A. 78:4485–4489.
52. **Parker, G. C., R. M. Watson, and J. Vinograd.** 1976. Mapping of closed circular DNAs by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis. Proc. Natl. Acad. Sci. U.S.A. 74:851–855.
53. **Pribnow, D.** 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proc. Natl. Acad. Sci. U.S.A. 72:784–789.
54. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. 87:107–110.
55. **Sassone-Corsi, P., J. Corden, C. Kedinger, and P. Chambon.** 1981. Promotion of specific *in vitro* transcription by excised "TATA" box sequences in a foreign environment. Nucleic Acids Res. 9:3941–3958.
56. **Shenk, T.** 1981. Transcriptional control regions: nucleotide sequence requirements for initiation by RNA polymerase II and III. Curr. Top. Microbiol. Immunol. 93:25–46.
57. **Soeda, E., J. R. Arrand, N. Smolar, J. E. Walsh, and B. E. Griffin.** 1980. Coding potential and regulatory signals of the polyoma virus genome. Nature (London) 283:445–453.
58. **Sollner-Webb, B., and R. H. Reeder.** 1979. The nucleotide sequence of the initiation and ribosomal RNA transcription in X. laevis. Cell 18:485–499.
59. **Struhl, K.** 1981. Deletion mapping of a eucaryotic promoter. Proc. Natl. Acad. Sci. U.S.A. 78:4461–4465.
60. **Talkington, C. A., Y. Nishioka, and P. Leder.** 1980. *In vitro* transcription of normal, mutant and truncated mouse α-globin genes. Proc. Natl. Acad. Sci. U.S.A. 77:7132–7136.
61. **Tooze, J. (ed.).** 1980. DNA tumor viruses, 2nd ed., part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
62. **Treisman, R. H.** 1980. Characterization of Py late mRNA leader sequences of molecular cloning and DNA sequence analysis. Nucleic Acids Res. 8:4867–4888.
63. **Treisman, R. H., A. Cowie, J. Favaloro, P. Jat, and R. Kamen.** 1981. The structures of the spliced mRNAs encoding polyoma virus early region proteins. J. Mol. Appl. Gen. 1:83–92.
64. **Treisman, R., U. Novak, J. Favaloro, and R. Kamen.** 1981. Transformation of rat cells by an altered polyoma virus genome expressing only the middle-T protein. Nature (London) 292:595–600.
65. **Tsai, S. Y., M. J. Tsai, and B. W. O'Malley.** 1981. Specific 5′ flanking sequences are required for faithful initiation of *in vitro* transcription of the ovalbumin gene. Proc. Natl. Acad. Sci. U.S.A. 78:879–883.
66. **Tsujimoto, Y., S. Hirose, M. Tsuda, and Y. Suzuki.** 1981. Promoter sequence of fibroin gene assigned by *in vitro* transcription system. Proc. Natl. Acad. Sci. U.S.A. 78:4838–4842.
67. **Twigg, A. J., and D. Sherrat.** 1980. Trans-complementable copy-number mutants of plasmid Col E1. Nature (London) 283:216–218.
68. **Tyndall, C., G. La Mantia, C. M. Thacker, J. Favaloro, and R. Kamen.** 1981. A region of the polyoma virus genome between the replication origin and late protein coding sequences is required in *cis* for both early gene expression and viral DNA replication. Nucleic Acids Res. 9:6231–6250.
69. **Van der Eb, A. J., and F. L. Graham.** 1980. Assay of transforming activity of tumour virus DNA. Methods Enzymol. 65:826–838.
70. **Wasylyk, B., and P. Chambon.** 1981. AT to A base substitution and small deletion in the conalbumin TATA box drastically decrease specific *in vitro* transcription. Nucleic Acids. Res. 9:1813–1824.
71. **Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Teoule, and P. Chambon.** 1980. Specific *in vitro* transcription of conalbumin gene is drastically decreased by single-point mutation in T-A-T-A box homology sequence. Proc. Natl. Acad. Sci. U.S.A. 77:7024–7028.
72. **Wasylyk, B., C. Kedinger, J. Corden, O. Brison, and P. Chambon.** 1980. Specific *in vitro* initiation of transcription canalbumin and ovalbumin genes and comparison with adenovirus-2 early and late genes. Nature (London) 285:367–373.
73. **Weaver, R. F., and C. Weissman.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′-termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. 7:1175–1193.