

5' Termini of Polyoma Virus Early Region Transcripts Synthesized *in vivo* by Wild-type Virus and Viable Deletion Mutants

ROBERT KAMEN, PARMJIT JAT, RICHARD TREISMAN†
JENNIFER FAVALORO

*Transcription Laboratory, Imperial Cancer Research Fund
London WC2A 3PX, England*

AND

WILLIAM R. FOLK

*Department of Biological Chemistry, The University of Michigan
Ann Arbor, Mich. 48109, U.S.A.*

(Received 22 October 1981, and in revised form 23 March 1982)

The 5' termini of polyoma virus early region transcripts synthesized during the productive infection of permissive mouse cells by wild-type or tsa virus, and those expressed in a variety of transformed rodent cell lines, have been mapped on the viral genome. Results were obtained using the S₁ nuclease and primer extension gel mapping procedures. Principal 5' ends of cytoplasmic polyadenylated mRNAs in every instance mapped between nucleotides 145 and 156 (numbering according to Soeda *et al.*, 1980) in the DNA sequence, 17 to 28 base-pairs before the translational initiation codon for early proteins and 26 to 37 base-pairs after a sequence agreeing with the TATA box consensus. Our data implied a minimum of two different termini with this region of the genome, at nucleotide 147 ± 2 and at 152 ± 2. The 5' ends at 147 ± 2 were particularly common in the mRNA overproduced after thermal inactivation of the large T protein at late times during infection. The sequences determining the principal 5' termini, unlike the analogous sequences in the closely related simian virus 40 (SV40), are distinct from those involved in the viral origin of DNA replication.

A number of minor alternative 5' termini of cytoplasmic mRNAs, located both before and after the principal 5' ends, were also detected. Of those downstream from the principal termini, one at nucleotide 300 ± 2 was prominent. Although this apparent 5' end is well within the early protein coding sequence, it occurs at a position 31 ± 2 base-pairs after a second TATA box. The several minor apparent 5' termini mapping upstream of the principal termini occurred primarily in the vicinity of the highly conserved papovavirus DNA replication origin sequence. This sequence includes a third TATA box. Two of the minor 5' termini, at 14 ± 2 and at 20 ± 2, were near the consensus distance from this TATA box, but the others mapped within or before it. Early region mRNA extracted at late times from the cytoplasm of cells infected with wild-type virus, or with tsa virus at the permissive

† Present address: Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Mass. 02138, U.S.A.

temperature, was usually (but not invariably) enriched for RNA species with apparent 5' termini mapping in the replication origin region, as well as for even longer RNAs. Such RNAs were correctly spliced and had the normal polyadenylated 3' ends. They were very minor in the cytoplasmic mRNA overproduced after thermal inactivation of the large T protein at late times during infection, but the nuclear RNA from these cells comprised giant species with highly heterogeneous apparent 5' ends, including predominantly those in the origin region and others further upstream. Nuclear viral RNA from most transformed cell lines lacked the giant species and had principal 5' termini in the nt145 to 156 region. We consider two models to account for the presence of the longer mRNAs in the cytoplasm of infected cells at late times during infection. The first postulates a shift in transcriptional initiation sites to upstream positions because of the repressor action of the large T protein. The second, which we favour, proposes that the longer species occur because of inefficient transcriptional termination, which leads to transcription around the entire circular genome and consequently to the eventual accumulation of long cleavage products in the cytoplasm.

We further studied the mRNAs expressed by three viable deletion mutants (Bendig *et al.*, 1980) that lack all or part of the sequence determining the principal 5' termini and the TATA box that precedes it. These efficiently synthesized early region mRNA with slightly or highly heterogeneous 5' ends. Two of the mutants (dl-75 and dl-17) produced mRNAs with principal alternative 5' ends located slightly before or after the deletions. The 5' ends of these mutant mRNAs corresponded to those of very minor transcripts of wild-type templates. These results suggest that sequence information other than the TATA box has an important role in specifying the approximate position of mRNA 5' ends.

1. Introduction

The early region of polyoma virus DNA is a single transcription unit determining three principal mRNAs (Fig. 1; Kamen *et al.*, 1980b). These messenger RNAs differ in the location and extent of an internal region removed by RNA splicing, but share common 5' and 3' termini. We recently described the cloning and sequencing of partial complementary copies of the mRNAs, which allowed precise assignments of the three different splice junctions (Treisman *et al.*, 1981). Consideration of available open codon reading frames permitted us to determine which messenger encoded each of the three known viral early proteins (Fig. 1). As the first step in the identification of DNA sequence elements that comprise the early region transcriptional promoter, we now present the detailed mapping of the several 5' termini of the mRNAs with respect to the known viral DNA sequence (Deiningner *et al.*, 1979; Friedmann *et al.*, 1979; Soeda *et al.*, 1980).

We mapped the 5' termini by three independent but complementary approaches. Two indirect procedures, nuclease S₁ gel mapping (Berk & Sharp, 1977; Sollner-Webb & Reeder, 1979; Weaver & Weissmann, 1979) and reverse transcriptase primer extension (Ghosh *et al.*, 1980a), were used to compare cytoplasmic and nuclear transcripts from cells infected by wild-type virus or by an early region temperature-sensitive mutant under different growth conditions. The sensitivity of the indirect procedures also permitted us to map accurately the 5' ends of viral transcripts in a variety of polyoma virus-transformed cell lines. These results are presented here. No indirect method, however, exactly determines the first nucleotide of a transcript or distinguishes capped from other 5' termini. This was

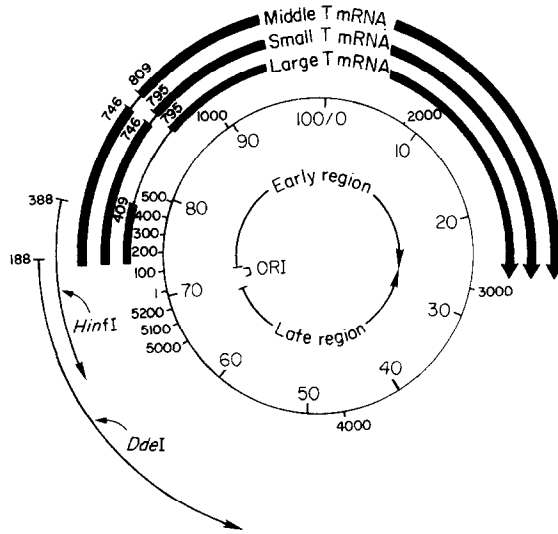


FIG. 1. Map of the principal polyoma virus early region mRNAs. The 3 major spliced mRNAs are oriented around the standard physical map of the viral genomic DNA, with map units (numbers inside the circle) and nucleotide numbers (outside the circle, Soeda *et al.* (1980) numbering system) indicated. Splice site assignments are those originally predicted from S_1 mapping data (Kamen *et al.*, 1980*a,b*) and more recently confirmed by sequencing cloned cDNA copies (Treisman *et al.*, 1981). The two ($5'$ - ^{32}P)-labelled, *E*-DNA strand, restriction fragments used as hybridization probes (see the text) are positioned with respect to the physical map.

particularly important in the case of the polyoma virus early region transcripts, because the results obtained with indirect procedures suggested terminal heterogeneity. We therefore have further analysed directly the ends of ($5'$ - ^{32}P)-labelled viral RNA using the procedure described by Cowie *et al.* (1981); these experiments will be presented separately (Cowie *et al.*, 1982, accompanying paper). The principal capped 5' termini localized by all three approaches mapped within a genomic region previously identified (Bendig *et al.*, 1980) as dispensable for productive viral infection or transformation. It was therefore important to analyse the 5' ends of the early region mRNAs synthesized by several of the viable deletion mutants that defined this non-essential region. These experiments showed that neither the Hogness-Goldberg TATA box, which usually occurs approximately 31 bases before cap sites (for a review, see Breathnach & Chambon, 1981), nor the wild-type cap sites, are required for efficient viral early gene expression *in vivo*. The significance of this result in relation to similar findings with other transcription units (Grosschedl & Birnstiel, 1980*a,b*; Benoist & Chambon, 1981; McKnight *et al.*, 1981), and with regard to the identity of DNA sequence elements specifying transcriptional initiation, will be discussed.

2. Materials and Methods

Much of the methodology used in our laboratory for the study of polyoma virus transcripts has been described in considerable detail (Favaloro *et al.*, 1980). Only further procedures are presented here.

(a) *High-resolution S₁ gel mapping using 5'-labelled, single-stranded viral restriction fragments*

The method used evolved from the Weaver & Weissmann (1979) modification of the Berk & Sharp (1977) procedure. In a typical experiment, 10 µg of purified polyoma virus DNA (strain A2) was digested to completion with either restriction endonuclease *Hinf*I (Biolabs) or *Dde*I (BRL) using the suppliers' recommended conditions. After extraction with phenol mixture (Favaloro *et al.*, 1980), the DNA was recovered by precipitation with ethanol, washed with 75% ethanol, and redissolved in 80 µl of distilled water to which were added 10 µl of 10 × alkaline phosphatase buffer (100 mM-Tris·HCl (pH 9.5), 10 mM-spermidine·HCl, 1 mM-EDTA) and 10 µl of a solution of calf intestine alkaline phosphatase (4 units/ml; Boehringer grade 1). After incubation at 37 °C for 30 min, the phosphatase was inactivated at 65 °C for 60 min. Dephosphorylated restriction digestions were stored at -20 °C without further purification. When required for 5' labelling, 10-µl portions (1 µg viral DNA) were mixed with 2 µl of 10 × kinase buffer (450 mM-Tris·HCl (pH 9.5), 100 mM-MgCl₂, 50 mM-dithiothreitol, 5 mM-spermidine·HCl, 0.5 mM-EDTA), 50 to 100 µl of [γ -³²P]ATP (Radiochemical Centre: >5000 Ci/mmol, 1 mCi/ml in 50% ethanol), lyophilized to dryness, and redissolved in 20 µl of distilled water. Phage T4 polynucleotide kinase (8 to 15 units; P.L. Biochemicals) was then added and, after incubation for 1 h at 37 °C, the reaction was terminated by addition of 2 M-ammonium acetate (100 µl), 100 mM-EDTA (4 µl) and 10 mg yeast RNA carrier/ml (2 µl). Protein was removed by extraction with phenol mixture, and the labelled DNA was recovered from the aqueous phase by precipitation with ethanol. After reprecipitation from 0.3 M-sodium acetate (pH 5.3) and washing in 75% ethanol, it was dissolved in 14 µl of denaturation buffer (Maxam & Gilbert, 1980: 30% (v/v) dimethylsulphoxide, 1 mM-EDTA (pH 7.5), 0.05% (w/v) xylene cyanol, 0.05% (w/v) bromphenol blue) to which was then added a further 6 µl of undiluted dimethylsulphoxide. The DNA was denatured at 90 °C for 2 to 3 min and then loaded into a 2 cm slot on a pre-run strand separation gel (0.8 mm × 23 cm × 43 cm) containing 5% (w/v) acrylamide/methylene bisacrylamide (50:1, w/w, prepared as described by Maxam & Gilbert, 1980). The complementary strands of both the 603-nucleotide *Hinf*I and the 1043-nucleotide *Dde*I fragments used as hybridization probes (see Fig. 1) separated well after electrophoresis at 300 to 400 V for 24 h. Hybridization of DNA eluted from the gels with asymmetric polyoma virus complementary RNA (Kamen *et al.*, 1976) demonstrated that the slower migrating strand of each of these fragments was the *E*-DNA strand†. The yield of labelled DNA fragment was normally 10⁶ to 1.5 × 10⁶ disintegrations/min per single-stranded fragment, corresponding to a spec. act. of 1500 to 2400 Ci/mmol 5' end (approx. 50% of theoretical maximum).

Hybridizations and S₁ nuclease digestions were as previously described (Favaloro *et al.*, 1980), except that the annealing temperature was 30 °C, the annealing time was 12 to 16 h, and the S₁ digestions were normally for 2 h at 12 °C. S₁-resistant DNA products were fractionated on 5% or 12% (19:1 (w/w) acrylamide/methylene bisacrylamide) thin area/polyacrylamide gels (0.3 mm × 23 cm × 43 cm; Sanger & Coulson, 1978) after denaturation in the formamide buffer used for DNA sequencing (Maxam & Gilbert, 1980). Size markers used were either (5'-³²P)-labelled *Pv* DNA *Dde*I fragments, or samples of the hybridization probe randomly cleaved before purine residues (Maxam & Gilbert, 1980).

(b) *Reverse transcriptase primer extension*

Hybrids between (5'-³²P)-labelled primer fragments and viral RNA were formed as described above for S₁ nuclease experiments. Details of primer and RNA quantities used are provided in the Figure legends. After annealing, the hybrids were recovered by precipitation with ethanol and redissolved in 18 µl of a buffer containing 83 mM-HEPES (pH 8.3), 6.6 mM-

† Abbreviations used: *E*-DNA strand, the DNA strand complementary to early region mRNAs; *L*-DNA strand, the DNA strand complementary to late region mRNAs; m.u., map unit(s); nRNA, nuclear RNA; cRNA, the product of transcription of polyoma virus DNA *in vitro* by *Escherichia coli* RNA polymerase; nt, nucleotide number; bp, base-pair.

dithiothreitol and 66 mM-KCl. This solution was diluted to a final volume of 30 μ l containing, in addition, 8 mM-MgCl₂, 0.2 mM each deoxynucleotide triphosphate and 19 units of avian myeloblastosis virus reverse transcriptase (a gift from Dr J. W. Beard). After incubation at 37°C for 1 h, the reaction was terminated by addition of EDTA to 10 mM, sodium dodecyl sulphate to 0.1%, ammonium acetate to 1 M and 10 μ g of carrier yeast RNA. Protein was extracted with phenol mixture, and the nucleic acids were precipitated from the aqueous phase with ethanol. Subsequent analysis by polyacrylamide/urea gel electrophoresis was as described above for S₁ nuclease experiments.

3. Results

(a) *The principal 5' termini of polyoma virus early region mRNAs*

Expression of the polyoma virus early region (Cogen, 1978; Kamen *et al.*, 1980*a,b*), and that of the closely related simian virus 40 (SV40: Tegtmeyer *et al.*, 1975; Reed *et al.*, 1976; Alwine *et al.*, 1977; Khoury & May, 1977), is negatively regulated by one of its gene products, the large T protein. During the early phase of productive infection (conventionally defined as that period preceding the onset of viral DNA replication), the predominant nuclear RNA species are *E*-DNA strand transcripts, which are the precursors of early region mRNAs; similarly, only these messengers are detectable in the cytoplasm (see Tooze (1980) for a review of viral transcription in the lytic cycle). During the late phase of productive infection, which follows the onset of viral DNA synthesis, transcription of the complementary *L*-DNA strand becomes increasingly dominant. At 30 hours post-infection, there is at least ten times more late than early region cytoplasmic mRNA. The strong differential expression of *L*-DNA strand transcripts is, at least in part, caused by a regulatory activity of the large T protein. Thermal inactivation of altered forms of this protein containing temperature-sensitive mutations derepresses *E*-DNA strand transcription (Cogen, 1978). Within a few hours after a shift from the permissive to the non-permissive temperature, the abundance of early region mRNAs in the cytoplasm approaches that of the late messengers (Kamen *et al.*, 1980*a,b*). The repressor activity of purified SV40 large T protein has been demonstrated directly using transcription systems *in vitro* (Rio *et al.*, 1980; Myers *et al.*, 1981; Hansen *et al.*, 1981).

We exploited this regulatory system to obtain cytoplasmic polyadenylated RNA enriched for viral early region transcripts. Mouse cells were infected with the large T protein mutant *tsa* (Fried, 1965) and, after 36 hours of growth at the permissive temperature, the infected cultures were shifted to the non-permissive temperature four hours before harvesting and RNA extraction. Growth at the permissive temperature allows viral DNA replication to occur, amplifying the transcriptional templates. The subsequent incubation at the non-permissive temperature inactivates the repressor activity of the large T protein, resulting in the rapid accumulation of early region mRNA in the cytoplasm. We refer to such preparations as *tsa* shift-up RNA.

The 5' ends of polyoma virus early region mRNAs were localized on the viral genome, in the first instance, by high-resolution nuclease S₁ gel mapping. The RNAs were hybridized to two different single-stranded, (5'-³²P)-labelled, viral

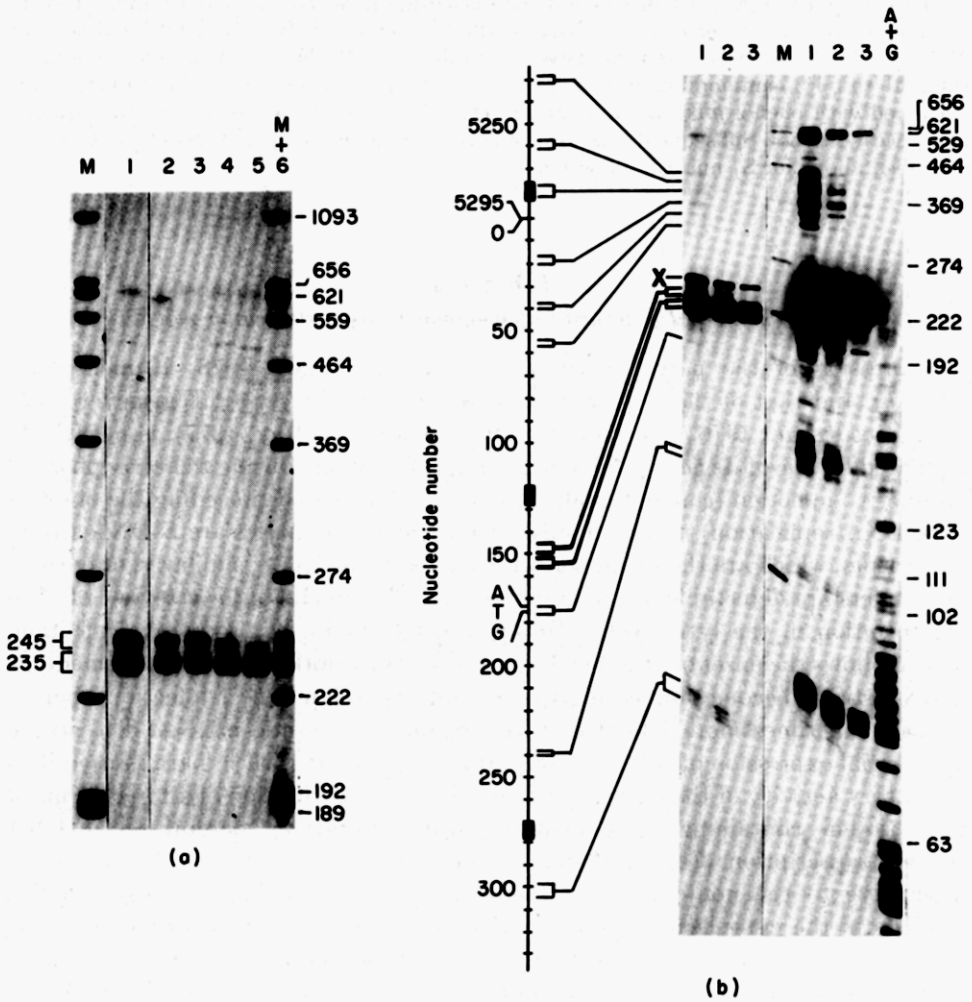


FIG. 2. Localization of the 5' termini of polyoma virus early region mRNAs by S_1 nuclease gel mapping. (a) and (b) Autoradiographs of 5% polyacrylamide/urea gels fractionating S_1 nuclease-resistant ($5'$ - ^{32}P)-labelled DNA products resulting after the hybridization of various RNA samples from polyoma virus-infected mouse 3T6 cells to the *Hinf*I *E*-DNA strand probe (1.5 to $2.0 \mu\text{Ci}/\text{pmol}$) extending clockwise from nucleotides 388 to 5077 (see Fig. 1). The RNA samples used for (a) were: tracks 1 to 3, 3 different preparations of *tsa* shift-up (see the text) cytoplasmic polyadenylated RNA (0.5 , 0.2 and $0.5 \mu\text{g}$, respectively); track 4, wild-type mRNA from infected cells grown at 37°C for 30 h ($1.5 \mu\text{g}$); track 5, mRNA from wild-type-infected cells grown at 39°C for 30 h ($3 \mu\text{g}$); track 6 (co-electrophoresed with the size markers), mRNA from *tsa*-infected cells grown continuously at 39°C for 30 h ($5 \mu\text{g}$). Tracks M are ($5'$ - ^{32}P)-labelled *Dde*I fragments of Py DNA used as size markers. Marker lengths are indicated to the right of the gel, and the calculated lengths of S_1 -resistant products are indicated to the left. (b) *tsa* shift-up polyadenylated polysomal mRNA (2.5 , 1.0 and $0.5 \mu\text{g}$, respectively, for tracks 1 to 3). Lighter and darker autoradiographic exposures of the same gel are shown. Alignment of the S_1 -resistant products with the sequencing track was used to deduce the positions of 5' termini with respect to the viral DNA sequence, as shown to the left of (b). Sequences related to the TATA box consensus (Breathnach & Chambon, 1981) are indicated by filled boxes.

DNA restriction fragment probes. The map positions of the *Hinf*I and *Dde*I probes used are indicated in Figure 1. After nuclease S_1 digestion, the resulting hybrids were denatured and the protected single-stranded DNA products were fractionated on polyacrylamide/urea gels. Lengths were determined either from mobilities relative to DNA markers of known size or, more precisely, by comparison with the products of partial degradation of a portion of the hybridization probe by the purine-specific reaction of chemical DNA sequencing (Maxam & Gilbert, 1980). Data obtained with the *Hinf*I probe are shown in Figures 2, 5 and 6; those obtained with the *Dde*I probe are in Figure 4.

Two major S_1 -resistant DNA products were obtained after hybridization of *tsa* shift-up mRNA to the *Hinf*I DNA probe (Fig. 2(a), tracks 1 to 3). These had lengths of approximately 245 and 235 nucleotides, but inspection of the original films clearly indicated that each "band" comprised a group of products with single nucleotide spacing. The mRNA extracted from *tsa*-infected cells grown continuously at the non-permissive temperature (Fig. 2(a), track 6) yielded a similar result. By contrast, the mRNA from wild-type-infected cells yielded predominantly the shorter of the two sets of DNA products whether the cells were grown at 37°C (Fig. 2(a), track 4) or at 39°C (Fig. 2(a), track 5). The quantities of mRNA used in this experiment were adjusted to obtain approximately equal protection of the DNA probe, compensating for the difference in abundance between early region mRNAs from wild-type and *tsa* virus-infected cells (see the legend to Fig. 2).

The same principal S_1 -resistant products were obtained when increasing amounts of polyadenylated RNA purified from the polysomes of *tsa* shift-up cells were annealed to the *Hinf*I probe (Fig. 2(b)). The band designated X in this and subsequent Figures is an artifact caused by incomplete nuclease digestion. It can be eliminated preferentially by use of higher enzyme levels (cf. Fig. 6, tracks 1 to 4). We attribute its occurrence to the DNA probe folding back on itself to form a short stretch of imperfect DNA-DNA duplex extending the correct length of a DNA-RNA hybrid. The significance of other minor S_1 -resistant products, particularly apparent in the overexposure of the gel pattern shown to the right of Figure 2(b), is considered separately in section (b), below.

Alignment of the principal S_1 -resistant products with the known viral DNA sequence, as shown in Figure 2(b), indicated that major 5' termini mapped at nucleotides 145 to 150 and at 150 to 155, using the numbering system proposed by Soeda *et al.* (1980). This is some 20 bases before the initiation codon thought to be used for translation of all three viral early proteins (Fig. 2(b)). These approximate assignments were confirmed by the results of the primer extension experiments shown in Figure 3. With a primer DNA fragment labelled at the same *Hinf*I site as the probe used in the S_1 nuclease experiments, a pair of extension products with lengths indistinguishable (within the accuracy of this measurement) from the S_1 -resistant products were obtained with *tsa* shift-up mRNA (Fig. 3, track 2). Correspondingly longer extension products were found using as primer a fragment labelled at a *Dde*I site 55 bases downstream from the *Hinf*I site (Fig. 3, tracks 8 and 8'). Note that no product corresponding to band X in Figure 2(b) was seen in the primer extension studies.

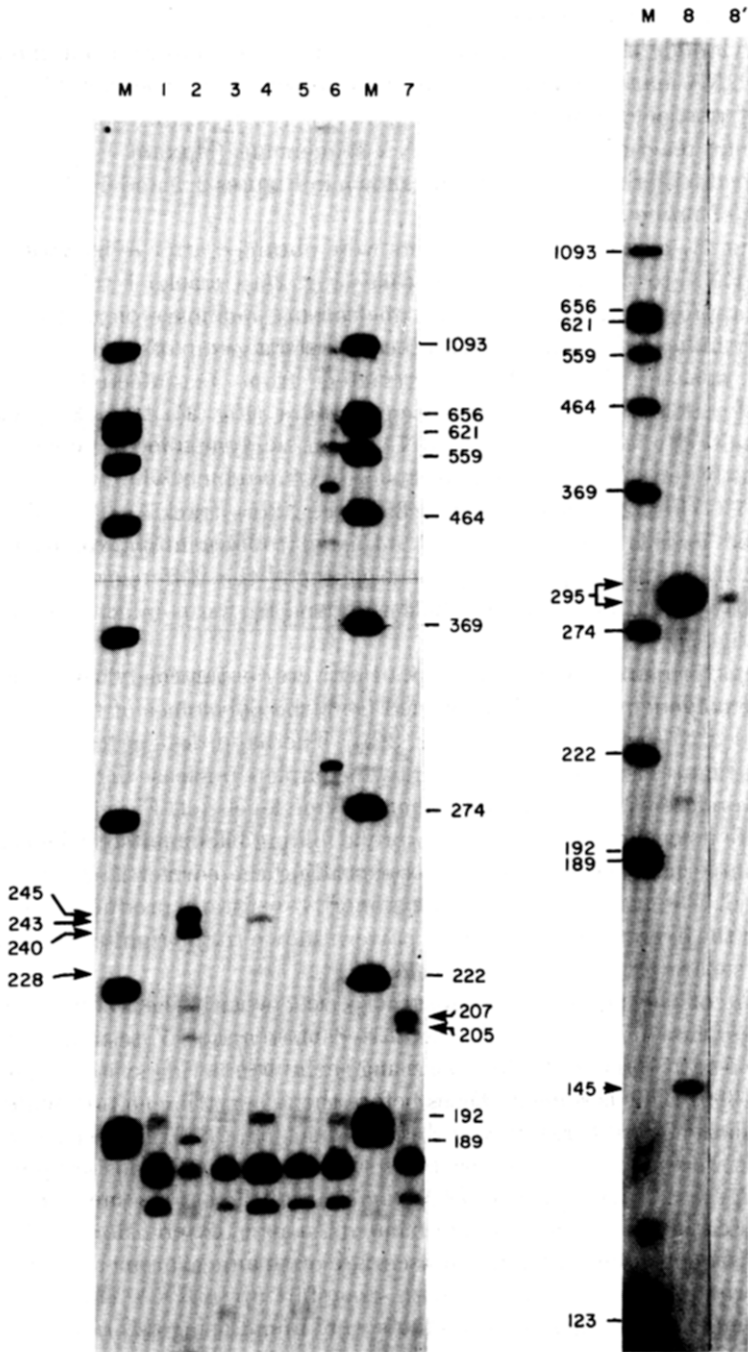


FIG. 3.

To localize the principal 5' termini more accurately, we used a DNA probe (the *DdeI* fragment indicated in Fig. 1) with a ($5'$ - ^{32}P) label much closer to the expected positions. We found a bimodal distribution of products with single nucleotide spacing (Fig. 4, track 3). While it is possible to specify precisely from this experiment the sites of S_1 nuclease cleavage, exact assignments of mRNA 5' ends cannot be made because of the ragged DNA 3' end of one to five bases normally observed after S_1 nuclease digestion (Sollner-Webb & Reeder, 1979; Weaver & Weissmann, 1979; Hentschel *et al.*, 1980). The data with *tsa* mRNA (Fig. 4, tracks 3, 4 and 8) strongly suggest, however, that there must be at least two alternative 5' ends at $\text{nt}147 \pm 2$ and at 154 ± 2 . mRNA from wild-type-infected cells grown at 37°C or 39°C (Fig. 4, tracks 1 and 2) apparently had much less of the 5' end at $\text{nt}147 \pm 2$, in agreement with the *HinfI* results discussed above (Fig. 2(a), tracks 4 and 5). In this experiment (Fig. 4, tracks 5 to 7) we also include results obtained with mRNA extracted from *tsa*-infected cells after only 12 hours of growth at 32°C , well before viral DNA replication or late region transcription can be detected. The low abundance of viral sequences in this preparation necessitated the use of rather large amounts of RNA in the annealing reactions. The pattern of S_1 products closely resembles that of *tsa* shift-up mRNA (compare tracks 7 and 3), but the lengths of the products corresponding to termini at $\text{nt}154 \pm 2$ increased by about two nucleotides with increasing RNA level (tracks 5 to 7). A similar effect was obtained with *tsa* shift-up RNA by lowering the S_1 nuclease levels (data not shown). This suggested that the nuclease, under our standard digestion conditions, removed some of four $\text{rA} \cdot \text{dT}$ or $\text{rU} \cdot \text{dA}$ base-pairs from $\text{nt}153$ to $\text{nt}156$ at the ends of the hybrid structures (see the sequence in Fig. 4), implying that the actual principal 5' end might be the A residue corresponding to nucleotide 153.

Several conclusions can be inferred from the data presented thus far. The principal 5' ends of polyoma virus early region mRNAs occur approximately 26 to 37 base-pairs, 3' to a T-A-T-A-A-T-T-A sequence ($\text{nt}120$ to $\text{nt}127$; see Figs 2 and 4) that strongly resembles the Hogness-Goldberg box consensus of T-A-T-A(A/T)A(T/A) (Breathnach & Chambon, 1981). The terminal microheterogeneity suggested by our results has also been found for other viral (Baker & Ziff, 1981; Kahana *et al.*, 1981; Haegeman & Fiers, 1980; Ghosh & Lebowitz, 1981) and cellular (Malek *et al.*, 1981; Gretz *et al.*, 1981) genes that have TATA boxes. Curiously, when early region transcription is derepressed at late times by the inactivation of the large T protein, the distribution of 5' termini among the

FIG. 3. 5' End mapping by reverse transcriptase primer extension. ($5'$ - ^{32}P)-labelled viral DNA fragments were hybridized to mRNA and extended with avian myeloblastosis virus reverse transcriptase as described in Materials and Methods. The extension products were fractionated on 5% polyacrylamide/urea gels. Tracks 1 to 7, extension of the *E*-DNA strand of the *HinfI* to *MboII* DNA fragment from $\text{nt}388$ to $\text{nt}267$ (approx. 10,000 disints/min per sample), annealed to carrier RNA alone (tracks 1, 3 and 5), or to $5 \mu\text{g}$ of *tsa* shift-up mRNA (track 2), dl 17 mRNA (track 4), dl 2-19 mRNA (track 6), and dl 75 mRNA (track 7). The 3 short extension products seen in the carrier only tracks presumably result from self-priming. Track 8 extension of the *E*-DNA strand of the *DdeI* fragment from $\text{nt}443$ to $\text{nt}380$ (approx. 30,000 disints/min) annealed to $0.5 \mu\text{g}$ of *tsa* shift-up mRNA. Track 8' is a shorter exposure of track 8 to show that the 295-nucleotide band is a doublet. Lengths of *DdeI* marker fragments are indicated to the right in (a) and to the left in (b). Calculated lengths of extended primers are indicated with arrows.

alternative positions appears to change in favour of those at nt147 \pm 2, closer to the TATA sequence.

(b) *Minor 5' termini of polyoma virus early region mRNAs*

Several S₁-resistant DNA products other than those mapping the principal 5' ends in the nt145 to nt156 region of the genome were consistently found. The more abundant of these are discussed in this section; consideration of the others is deferred to section (e), below.

In experiments using the *Hinf*I probe (Fig. 2(b), tracks 1 to 3; cf. also Fig. 8(a), mRNA track) a group of products approximately 90 bases long was found with every mRNA preparation examined, whether wild-type or mutant. These would map a 5' end at nt300 \pm 2 in the DNA sequence. The reverse transcriptase primer extension experiment shown in Figure 3 (track 8, the 145-nucleotide band) confirmed that this was a 5' end rather than a site of internal S₁ cleavage. Inspection of the preceding DNA sequence revealed C-A-T-A-T-A-A (nt270 to nt276) at the usual distance for a Hogness-Goldberg box before a cap site. RNAs starting from this alternative site would lack the initiation codon (which is at nt173 to nt175) believed to be used for translation of the early proteins. They could, however, encode truncated polypeptides not detected thus far. The recovery of minor RNAs with this 5' end from polysomes (Fig. 2(b)) indicates that it would be worthwhile to look for such proteins.

Experiments with *tsa* shift-up mRNA preparations also generated very minor S₁-resistant DNA products longer than the principal ones. These include a complex set of DNA products that would map at least nine 5' ends between nt5230 and nt5260 (Fig. 2(b), particularly the darker exposure of track 1, where the RNA was in excess relative to the probe DNA; see also faint bands near the top of the gel shown in Fig. 4, tracks 3 and 4). The locations of these putative 5' ends can be deduced more accurately from the gels shown in Figure 12, and are indicated in Figure 14. There was also RNA that protected nearly the full length of the *Hinf*I

FIG. 4. Localization of the 5' termini of polyoma virus early region mRNAs by S₁ nuclease gel mapping using the *Dde*I probe (see Fig. 1). Cytoplasmic polyadenylated mRNA from polyoma virus-infected 3T6 cells was annealed to the (5'-³²P)-labelled *Dde*I E-strand probe extending from nt188 to nt4389. S₁-resistant products were fractionated on a 12% polyacrylamide/urea gel. Each hybridization contained approx. 1.5 ng of DNA fragment (spec. act. 2.2 μ Ci/pmol) and the following RNA samples: track 1, 5 μ g of RNA from wild-type-infected cells grown for 30 h at 37 °C; track 2, 5 μ g of RNA from wild-type-infected cells grown for 30 h at 39 °C; track A+G, the products of the chemical sequencing A+G reaction (Maxam & Gilbert, 1980) done on a portion of the probe fragment to provide homologous length markers; track 3, 1 μ g of *tsa* shift-up mRNA; track 4, 5 μ g of mRNA from mRNA from *tsa*-infected cells grown continuously for 30 h at 39 °C; tracks 5 through 7 are 5, 10 and 20 μ g, respectively, of mRNA from cells infected with *tsa* virus, grown at 32 °C and harvested 12 h post-infection; track 8, 1 μ g of polysomal polyadenylated mRNA from *tsa* shift-up cells; tracks M, polyoma DNA *Dde*I size markers. The sequence shown to the left of the autoradiogram uses the A+G track marker fragments to align the S₁-resistant DNA fragment lengths with the DNA sequence of the strand with the same sense as the early region mRNAs. Arrows indicate the positions of putative mRNA 5' termini. In making these assignments, we have corrected (Sollner-Webb & Reeder, 1979) for the fact that the chemical reactions cleave 5' to the purine nucleotides leaving 3'-phosphate termini, whereas the S₁-resistant DNA products have 3'-hydroxyls (the difference in termini accounts for the fact that sequence track bands and S₁-resistant products do not comigrate precisely).

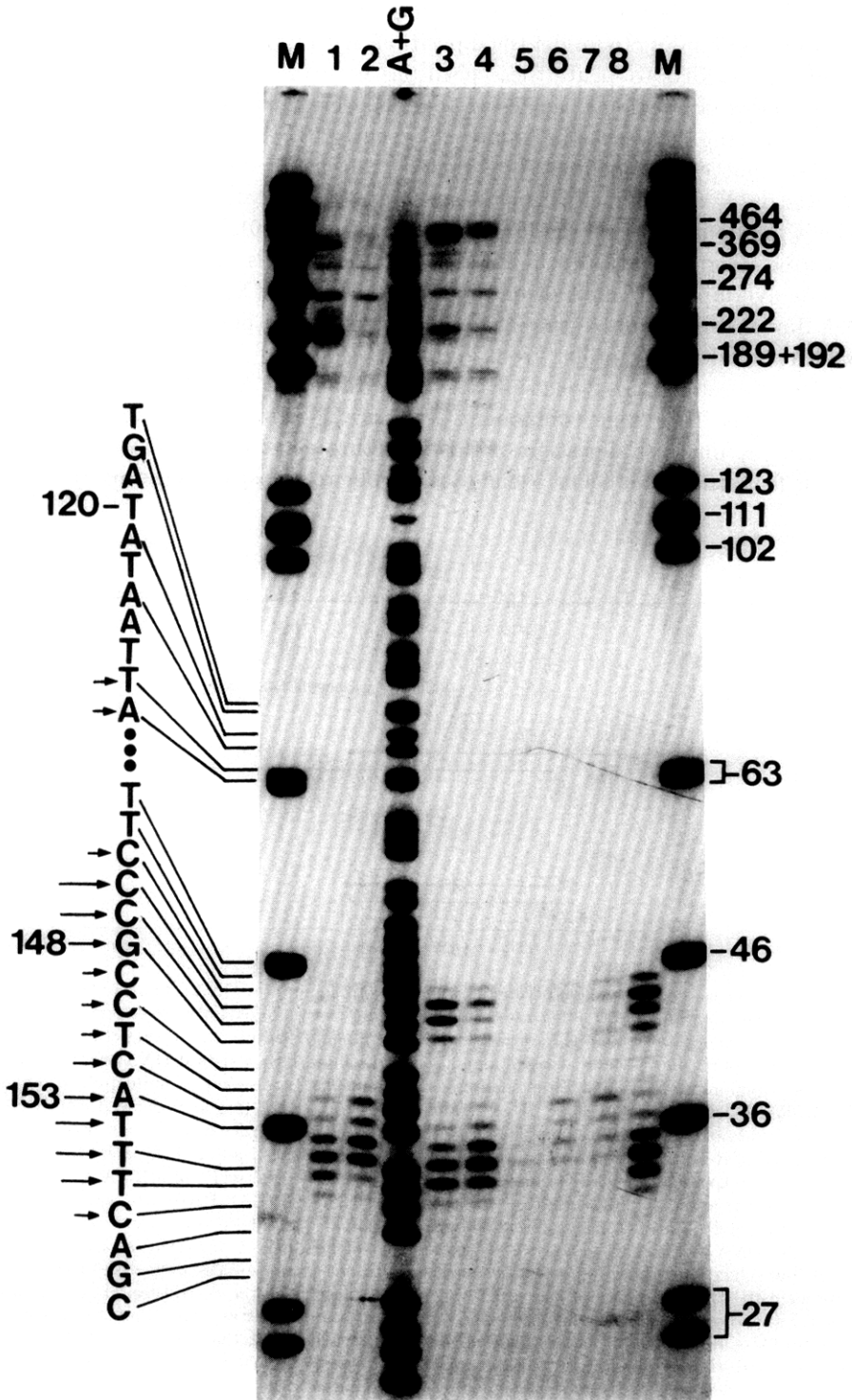


FIG. 4.

probe (Fig. 2(b), dark exposure), indicating that there are further minor species with 5' ends extending to unidentified positions within the late region of the genome (see Fig. 1). Efforts to locate these positions with the *DdeI* probe (Fig. 4), which contains more late region sequence (cf. Fig. 1), were unsuccessful because of RNA-RNA annealing between the minor long early RNAs and the 5' portions of abundant but complementary late region transcripts. Such RNA-RNA partial duplexes, in S_1 mapping experiments, generate S_1 -resistant DNA products, which we have previously termed shadows (Favaloro *et al.*, 1980). The origin of shadows is illustrated diagrammatically in Figure 5. If the 5' ends of the minor long early mRNAs overlap the 5' ends of more abundant late mRNAs, the region of RNA-RNA duplex that forms during the annealing reaction prevents the ^{32}P -labelled, single-stranded DNA probe from hybridizing up to the 5' end of its RNA complement (see Fig. 5). Because RNA-RNA duplexes are more stable than DNA-RNA hybrids, the triple component structure shown in the diagram is highly favoured. S_1 nuclease digestion of this structure generates a DNA product that extends from the ^{32}P -labelled 5' end of the probe to a position determined by the 5'

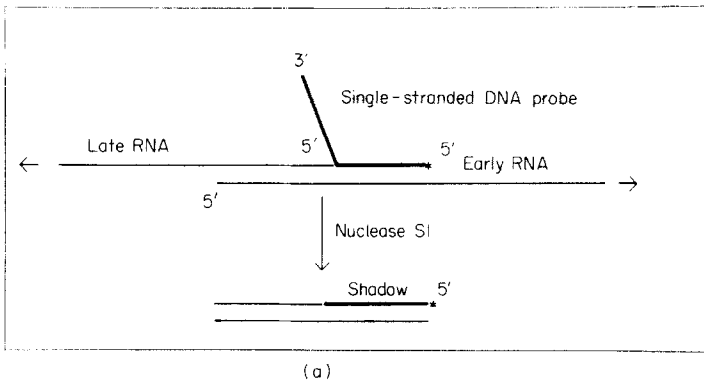


FIG. 5. S_1 nuclease gel mapping experiments with the complementary DNA strands of the *HinfI* DNA fragment. The diagram in (a) illustrates how annealing between partially overlapping symmetric transcripts can generate erroneous S_1 -resistant DNA products (shadows, as explained in the text). (b) The autoradiograph shows control experiments used to distinguish S_1 -resistant DNA products mapping actual early region mRNA 5' ends from shadows. Each annealing contained, as indicated, either the (^{32}P)-labelled *E*-DNA strand or the *L*-DNA strand of the *HinfI* DNA fragment derived from wild-type or deletion mutant DNAs (probes homologous to the RNAs were used; the deletion mutants (cf. Fig. 10) are discussed in Results, section (e) and the following mRNAs: lanes 1 and 2, 1 μg each from 2 different *tsa* shift-up preparations; lanes 3, 4, 5 and 6, 1, 5, 2.5 and 5 μg , respectively, of mRNA from wild-type-infected cells harvested 30 h post-infection; lanes 7 and 8, 0.5 μg of dl 75 mRNA harvested at 30 h; lanes 9 and 10, 1 μg of dl 17 mRNA harvested at 30 h. A 12% polyacrylamide/urea gel was used to fractionate S_1 -resistant DNA. Lanes marked S are A + G sequencing tracks for the *E* or *L*-DNA strands (SE and SL) of each hybridization probe. Lanes M are markers with the chain lengths indicated to the left. The positions in the DNA sequence of late mRNA 5' ends deduced from the S_1 -resistant DNA products in lanes 6, 7 and 10 are shown with arrows to the right. The data demonstrate that there exist insufficient amounts of late mRNA with 5' ends extending closer to the early region than nt5172 to produce shadows interfering with the mapping of early region mRNA 5' termini downstream (with respect to the early region) of this position. Note that the apparent low abundance of late region mRNA is because the major late 5' ends (Cowie *et al.*, 1981) occur too close to the end of the probe to form stable hybrids.

end of the late RNA strand of the hybrid, rather than to the position of the 5' end of the early mRNA that hybridized to the probe. We therefore identify this product as a shadow of the late mRNA 5' end. Its detection would clearly demonstrate the occurrence of an early mRNA that overlapped the 5' end of the late mRNA, but its length would not position the 5' end of this early RNA. In order to distinguish shadows from DNA products mapping 5' ends of RNAs complementary to the

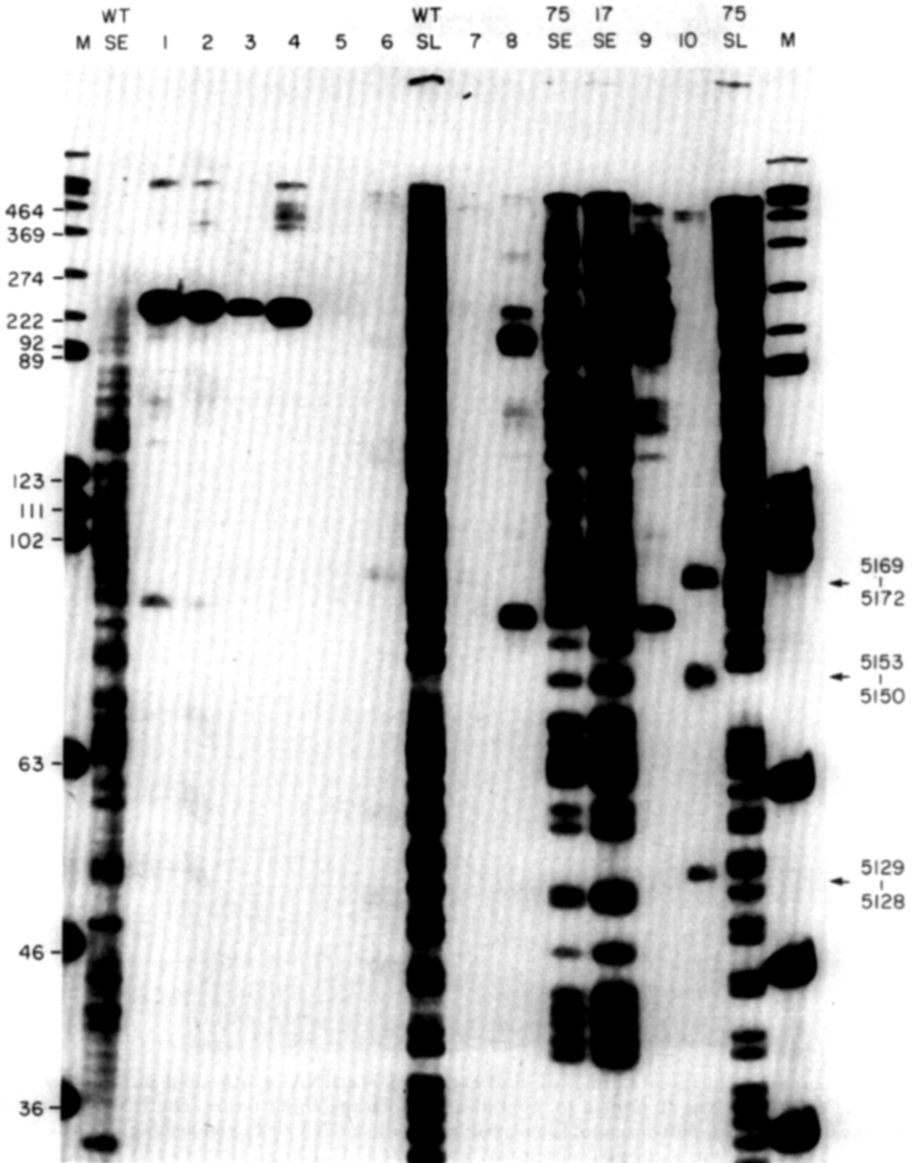


FIG. 5(b).

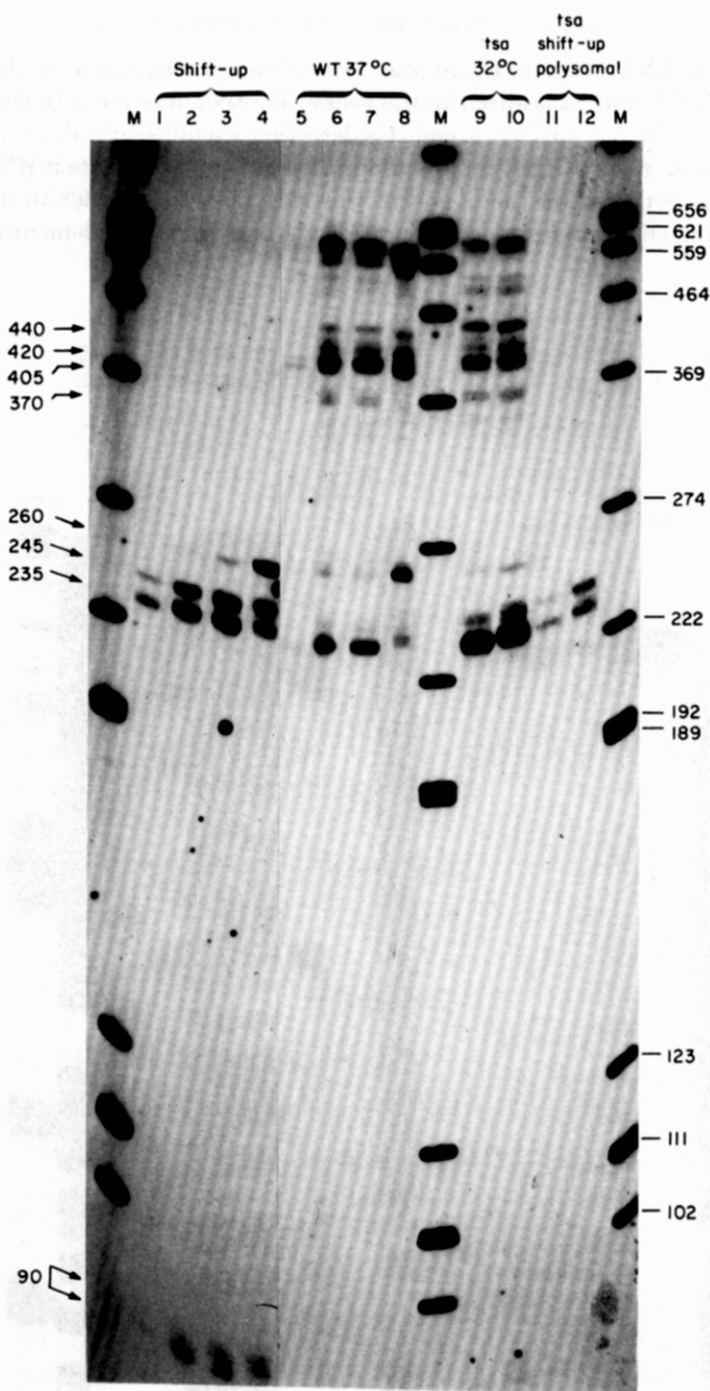


FIG. 6. Comparison of the 5' termini of viral early region mRNAs synthesized in the absence and presence of functional large T protein by hybridization with the *Hinf*I probe. The RNA preparations were: tracks 1 to 4, *tsa* shift-up cytoplasmic polyadenylated RNA (0.2 μ g/annealing); tracks 5 to 8, mRNA from wild-type virus-infected cells grown for 30 h at 37°C (1.5 μ g); tracks 9 and 10, mRNA from *tsa*-infected cells grown at 32°C extracted before the temperature shift (1 μ g); tracks 11 and 12, *tsa* shift-up polysomal polyadenylated mRNA. All S_1 nuclease digestions were at 12°C for 2 h, using the following enzyme concentrations: tracks 1 and 5, 1000 units/ml; tracks 2, 6, 9 and 11, 400 units/ml; tracks 3, 7, 10 and 12, 200 units/ml; tracks 4 and 8, 66 units/ml. Tracks M are *Py Dde*I fragment size markers. Marker lengths are listed on the right, and deduced lengths of S_1 -resistant products on the left.

probe DNA strand, it is necessary to separately hybridize both strands of the DNA probe to the RNA sample. An example of such an experiment is also shown in Figure 5. Here we annealed the *E* and the *L*-DNA strand of the *Hinf*I probe (see Fig. 1) to *tsa* shift-up RNA (tracks 1 and 2) or to RNA from wild-type virus-infected cells harvested 30 hours post-infection (tracks 3 to 6). The results obtained with the *E*-DNA strand (tracks 1 to 4) are equivalent to those shown above in Figure 2. With the *L*-DNA strand probe (tracks 5 and 6), we detected only three relatively minor DNA products (the smallest is best seen in track 10). These map the 5' ends of late mRNAs closest to the early region at nt5169 to nt5172, nt5150 to nt5153 and nt5128 to nt5129 (Treisman, 1980). The several major 5' ends of late mRNAs are further from the early region (Cowie *et al.*, 1981) and are not detected with the *Hinf*I probe, either because they occur beyond it or too close to the 5' label to form nuclease-resistant hybrid. We can conclude from the experiment shown in Figure 5 that there are no 5' ends of late mRNAs closer to the early region than nt5180 that are sufficiently abundant to cast shadows. However, we would expect to detect them beyond nt5170. We also include in Figure 5 similar control experiments for RNA extracted from cells infected with two of the viable deletion mutants (tracks 7 to 10) discussed in section (e), below.

Returning to our discussion of the localization of the 5' termini of the minor longer mRNAs detected in *tsa* shift-up RNAs, we infer that with the *Dde*I probe (Fig. 4, track 3), the longest S_1 -resistant product (apparently mapping a 5' end at approx. nt5075) is a shadow of the 5' end of the repeated leader unit (Treisman, 1980) present at the beginning of each late mRNA; similarly, the two fainter bands below the longest one on this gel are shadows corresponding to the late mRNA 5' ends detected in Figure 5, track 6. We can therefore conclude from these experiments only that some minor RNA molecules have 5' ends that extend into the late region of the genome.

Early region mRNAs with 5' termini mapping upstream of the principal ones were always very minor in *tsa* shift-up mRNA preparations. Indeed, they were detected in S_1 experiments only under conditions of relative RNA excess (Fig. 2(b), track 3) and were not apparent in other experiments (Fig. 2(b), track 1; Fig. 6, tracks 1 to 4, 11 and 12) where band intensity was proportional to RNA abundance because the probe DNA was in excess. By contrast, early region polyadenylated RNA extracted at 30 hours post-infection from the cytoplasm of cells infected with wild-type virus was commonly enriched for RNA with the upstream 5' termini (Fig. 6, tracks 5 to 8), as was the RNA from *tsa*-infected cells harvested after 36 hours at 32°C, just before the temperature shift (Fig. 6, tracks 9 and 10). We established by the two-dimensional S_1 gel mapping experiments shown in Figure 7 (see the legend for a detailed description of these and their interpretations) that at least a substantial proportion of the longer RNAs were correctly spliced and had the normal polyadenylated 3' end. The possible origin and significance of these RNA species is considered further in the Discussion. We note here, however, that among a large number of RNA preparations made in his laboratory during the past few years, the yield of the longer species has been rather variable, and indeed in some preparations of late RNA from wild-type cells harvested at 30 hours post-infection these RNAs were minor (for example, Fig. 5, tracks 3 and 4).

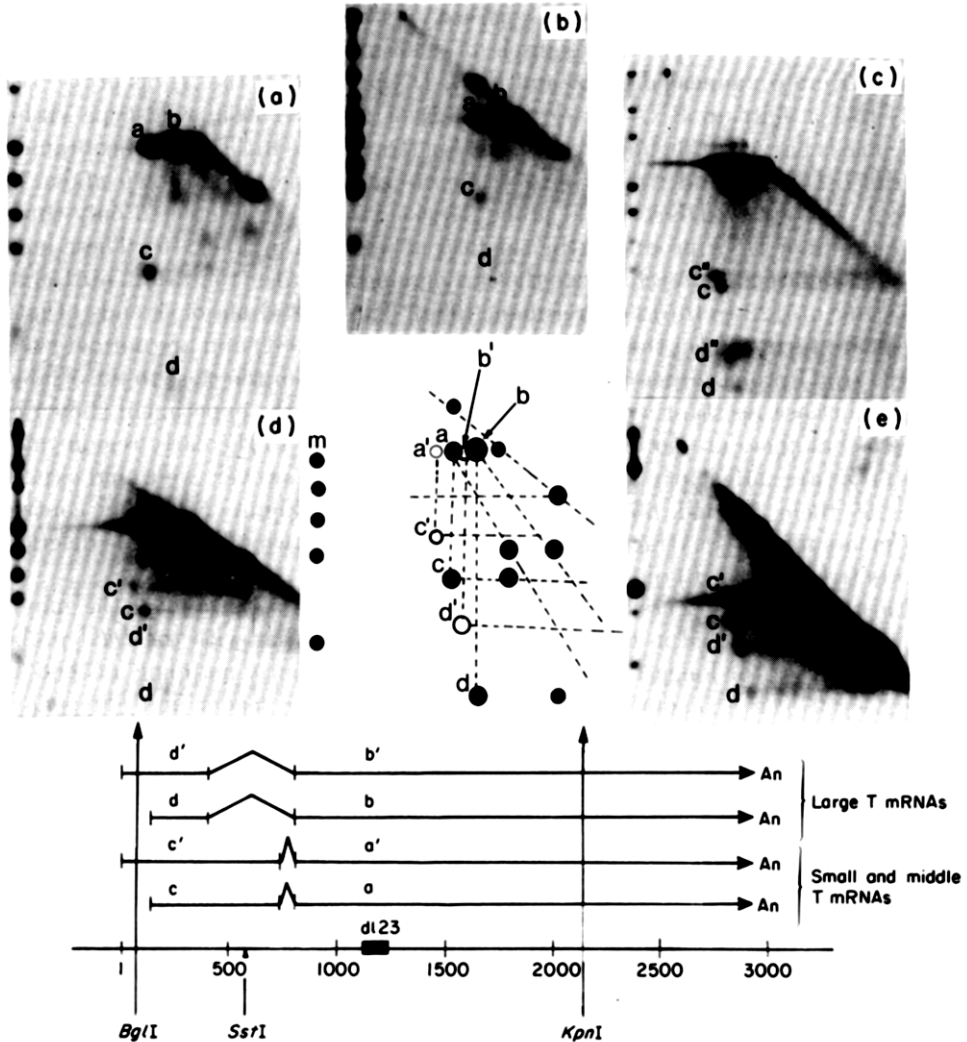


FIG. 7. Two-dimensional S_1 gel mapping demonstrating the specific occurrence of longer 5' colinear segments in the early region mRNAs produced at late times of lytic infection. The theory and practice of the 2-dimensional method were described by Favalaro *et al.* (1980). (a) to (e) Autoradiograms of nitrocellulose filter transfers from 2-dimensional agarose gels separating S_1 -resistant DNA products, visualized by annealing with various ^{32}P -labelled virus-specific DNA probes. Only those aspects of the patterns pertinent to the 5'-terminal colinear segments are discussed here, although all features of the autoradiograms can be interpreted to generate structural maps of the principal and minor early region mRNAs (see Treisman *et al.*, 1981). (a) *tsa* shift-up polysomal RNA hybridized to unlabelled restriction fragment *KpnI*-1 (nt4693 clockwise to nt2176; cf. Fig. 1); transfer annealed to nick-translated viral DNA. The spots along the left edge of this and all other panels are *HpaII* viral restriction fragments loaded in the second (alkaline) dimension as size markers. Spots designated a and c correspond to the 3' and 5' colinear regions of the small and middle T proteins mRNAs, respectively, as shown in the line diagram below. Note that the 14-nucleotide difference in the lengths of the 3' colinear regions of the small and middle T protein mRNAs (Treisman *et al.*, 1981) is not resolved on the agarose gels. Spots b and d correspond to the 3' and 5' colinear regions of the large T protein mRNA. The undesignated spot adjacent to spot b is caused by S_1 cleavage of the RNA strand of hybrids opposite the excised DNA loop. Other spots relate to minor species (Kamen *et al.*, 1980*a,b*; Treisman *et al.*, 1981) irrelevant to the present concerns. (b) mRNA from *tsa*-infected cells grown at the non-permissive temperature (39°C) for

(c) *The 5' termini of nuclear RNA molecules*

Nuclease S_1 gel mapping experiments similar to those already described were also attempted with nuclear RNA. At late times of wild-type infection, there is such a large excess of continuous transcripts of the entire *L*-DNA strand that RNA-RNA self-annealing effectively prevents detection of *E*-DNA strand transcripts (Fig. 8(b), tracks 4 to 7; cf. Kamen *et al.*, 1974). At very early times of infection, there was too little viral nuclear RNA for fine-structure S_1 mapping experiments. Useful results, however, were obtained with *tsa* shift-up nRNA, because under these conditions transcription of both DNA strands occurs at similar rates. The data shown in Figure 8 reveal significant quantitative and qualitative differences between *tsa* shift-up nRNA and mRNA. We were surprised to find that very similar amounts of S_1 -resistant DNA were obtained with 0.1 μ g of total nuclear RNA (nRNA track in Fig. 8(a)) and 0.3 μ g of poly(A)-selected cytoplasmic RNA (mRNA track in Fig. 8(a)). More extensive quantitation of the S_1 -resistant DNA produced as a function of RNA concentration (data not shown), after correction for the relative recovery of cellular RNA in the total nuclear and cytoplasmic polyadenylated fractions, demonstrated that there was at least ten times more viral nRNA than polyadenylated cytoplasmic mRNA. This result did not reflect inefficient polyadenylation, because virtually all of the viral cytoplasmic RNA was recovered in the polyadenylated fraction. The predominant S_1 products obtained

30 h, hybridized to the *Kpn*I-1 DNA fragment; transfer annealed to 32 P-labelled *L*-strand DNA (cDNA complementary to asymmetric cRNA; Favalaro *et al.*, 1980). The pattern is the same as in (a) (except for the spot along the diagonal above spot a, which corresponds to unspliced RNA). Comparison of (a) and (b) demonstrates that all transcripts detected were complementary to the *E*-DNA strand and, further, that the mRNA overproduced by inactivation of large T protein have the same structures as those synthesized in the absence of large T protein before the initiation of DNA replication. (c) mRNA prepared at a late time during infection hybridized to full-length linear DNA cleaved by restriction endonuclease *Bgl*I at a point (nt93) about 55 nucleotides upstream of the map positions of the principal 5' termini of mRNAs (see line diagram) synthesized under *tsa* shift-up conditions. The pattern was visualized with 32 P-labelled *L*-strand DNA. The features of interest, in comparison with the results in (a) and (b), are the 2 spots, *c''* and *d''*, just above and to the left of spots *c* and *d*. These additional spots represent longer alternative 5' colinear segments of normally spliced mRNAs. The lengths of products *c''* and *d''* (some 50 nucleotides longer than corresponding products *c* and *d*) suggest that they extend from the *Bgl*I site to the splice joins. The experiment shown in (c) used RNA from cells infected with viable deletion mutant dl 23 (map position shown in the line diagram; Griffin & Maddock, 1979) because the pattern obtained was particularly photogenic. However, the same results were obtained with RNA extracted at late times during infection from cells infected with wild-type virus and have nothing to do with the deletion mutation. This point is proved in (d) and (e). (d) mRNA from *tsa*-infected cells grown at the permissive temperature for 37 h, hybridized to *Kpn*I-1 DNA; the probe was nick-translated viral DNA. Spots *c'* and *d'* correspond to 5' colinear segments of small and middle T mRNAs (*c'*) and large T mRNAs (*d'*) some 150 nucleotides longer than those represented by *c* and *d*. Comparison of this result with that shown in (c) (spots *c''* and *d''*) proves that DNA products *c'* and *d'* extend across the *Bgl*I site. (e) mRNA from wild-type virus-infected cells grown at 37°C for 30 h, hybridized to *Kpn*I-1 DNA; the probe here was a nick-translated plasmid containing viral restriction fragment *Sst*I-2 (nt4341 clockwise to the site at nt569 shown in the line diagram, including only the sequences complementary to 5' portion of the early region). Spots *c'* and *c*, corresponding to alternative 5' segments of small plus middle T mRNAs, and *d'* and *d*, corresponding to those of large T mRNAs, are apparent. The detection of these spots with the *Sst*I-2 probe confirms the map positions shown in the line diagram; note that the 3' colinear portions of the mRNAs are not detected with this probe, and that other spots along the overexposed diagonal correspond to the late region mRNAs.

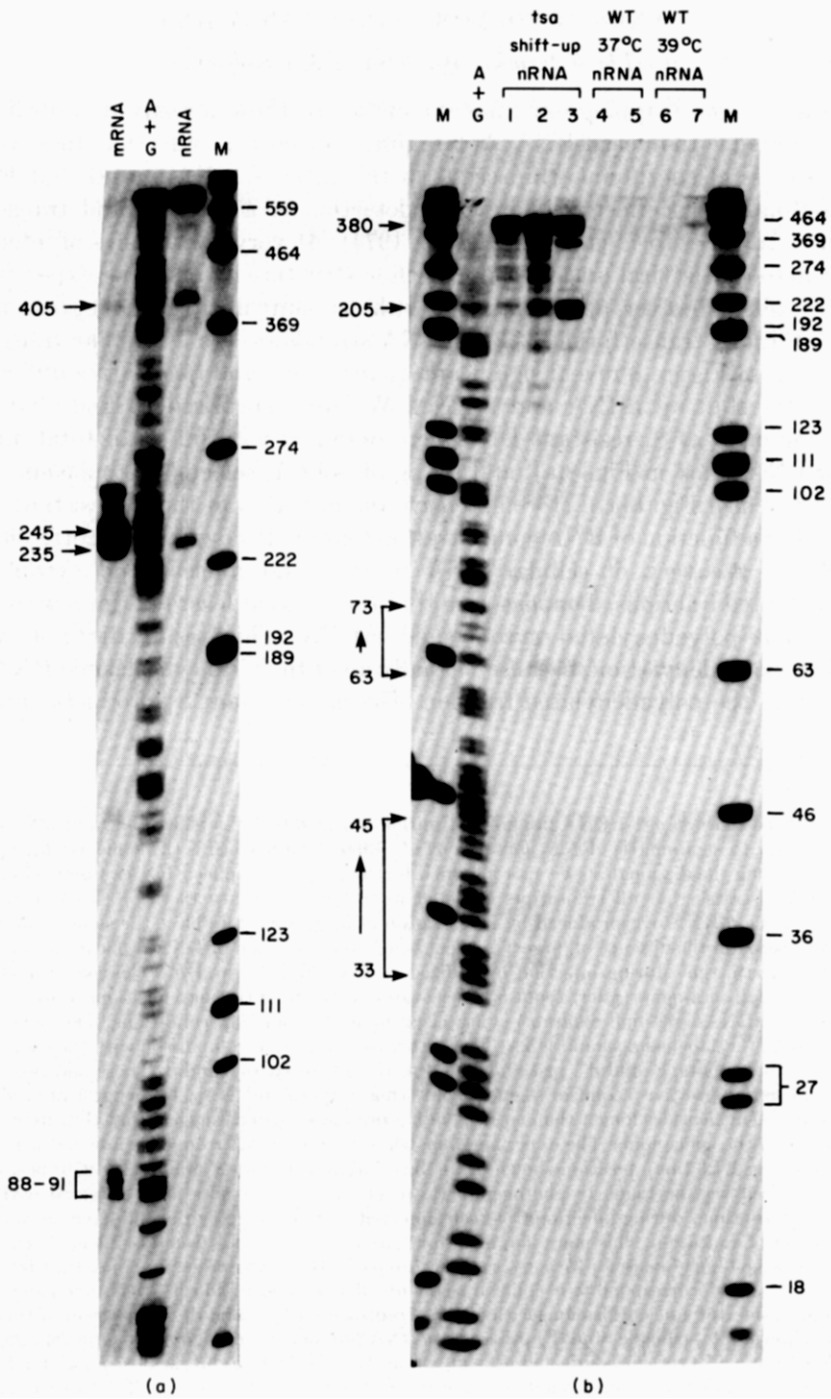


FIG. 8. Comparison of the 5' termini of cytoplasmic mRNA and total nuclear RNA by S_1 gel mapping using (a) the *Hinf*I or (b) the *Dde*I DNA probes. The quantities of tsa shift-up RNA used in (a) were 0.3 μ g of cytoplasmic polyadenylated mRNA and 0.1 μ g of total nuclear RNA. (b) Tracks 1 to 3, 0.06, 0.125 and 2.0 μ g tsa shift-up total nuclear RNA; tracks 4 and 5, 2.5 and 5 μ g of total nuclear RNA from wild-type virus-infected cells grown for 30 h at 37°C; tracks 6 and 7, 1 and 5 μ g of total nuclear RNA from wild-type virus-infected cells grown at 39°C for 30 h.

with the abundant nRNA were full-length protection of the *HinfI* DNA probe (Fig. 8(a)) and a 380-nucleotide band (Fig. 8(b)) with the *DdeI* probe. The 380-nucleotide product is the shadow of the repeat unit in the leaders of late mRNA precursors. DNA products (the 245 and 235 nt bands in Fig. 8(a)), nRNA track, and the 63 to 73 nt bands in Fig. 8(b)) corresponding to the principal 5' ends of mRNAs were detected with nRNA, but they were minor and enriched for the nt153 \pm 2 5' end. The 205 nt band seen principally under conditions of nRNA excess (Fig. 8(b), track 3) with the *DdeI* probe, and the corresponding 405 nt band found with the *HinfI* probe (Fig. 8(a)) may result from S₁ cleavage at eight rU·dA pairs back from nt5268 to nt5275 (see Fig. 14) within continuous DNA-RNA hybrids (Hansen *et al.*, 1981).

Our interpretation of these nRNA results is based on the observation that, at late times of productive infection, the entire *E*-DNA strand of the viral genome is transcribed to yield giant tandemly repeated transcripts (Kamen *et al.*, 1974; Acheson & Mieville, 1978). From the ratio of cytoplasmic to nuclear RNA, we conclude that only a small fraction of this RNA is exported. We presume, but have not proved, that the original 5' end is capped and that only the first copy of the early region, after cleavage and polyadenylation, matures to become mRNA. The rest of the RNA must be retained in the nuclei for eventual degradation, perhaps through a series of endonucleolytic cleavages. Therefore, the RNA we detect in nuclei could be predominantly these residual sequences rather than mRNA precursors. The complex pattern of minor 5' ends seen with nRNA, under this hypothesis, would be cleavage rather than start sites. To test this possibility, we recently examined the heterogeneous 5' ends of tsa shift-up nRNA by direct RNA analysis. We found that only those 5' ends corresponding to the principal mRNA cap sites were capped (Cowie *et al.*, 1982, accompanying paper).

(d) *The 5' ends of viral cytoplasmic and nuclear RNA in transformed cell lines*

We concluded in a previous report (Kamen *et al.*, 1980b) that the 5' ends of viral mRNAs expressed in polyoma virus-transformed rodent cell lines mapped within viral sequences at approximately the same positions as the early region mRNAs synthesized during productive infection. This conclusion was confirmed by the high-resolution S₁ mapping data shown in Figure 9. All of the transformed rat or mouse cell lines examined had mRNAs with 5' ends in the nucleotide 145 to 156 region. In other experiments, not shown, we obtained the same results with mRNA from a set of rat cell lines transformed by recombinant plasmids containing all or part of the polyoma virus early region. This demonstrates that the polyoma virus early promoter functions in the integrated state.

The analysis of viral nRNA in productively infected cells, described in the previous section, was complicated by the circularity of the viral genome and the apparent absence of efficient transcriptional termination. Transformed cell lines, particularly those in which the structure of the integrated viral DNA is well-known, afford a simpler system with which to ask whether the mRNA 5' termini correspond to those of nuclear precursors. We therefore used the S₁ nuclease method to map the 5' ends of nuclear viral RNA in a number of transformed rat and

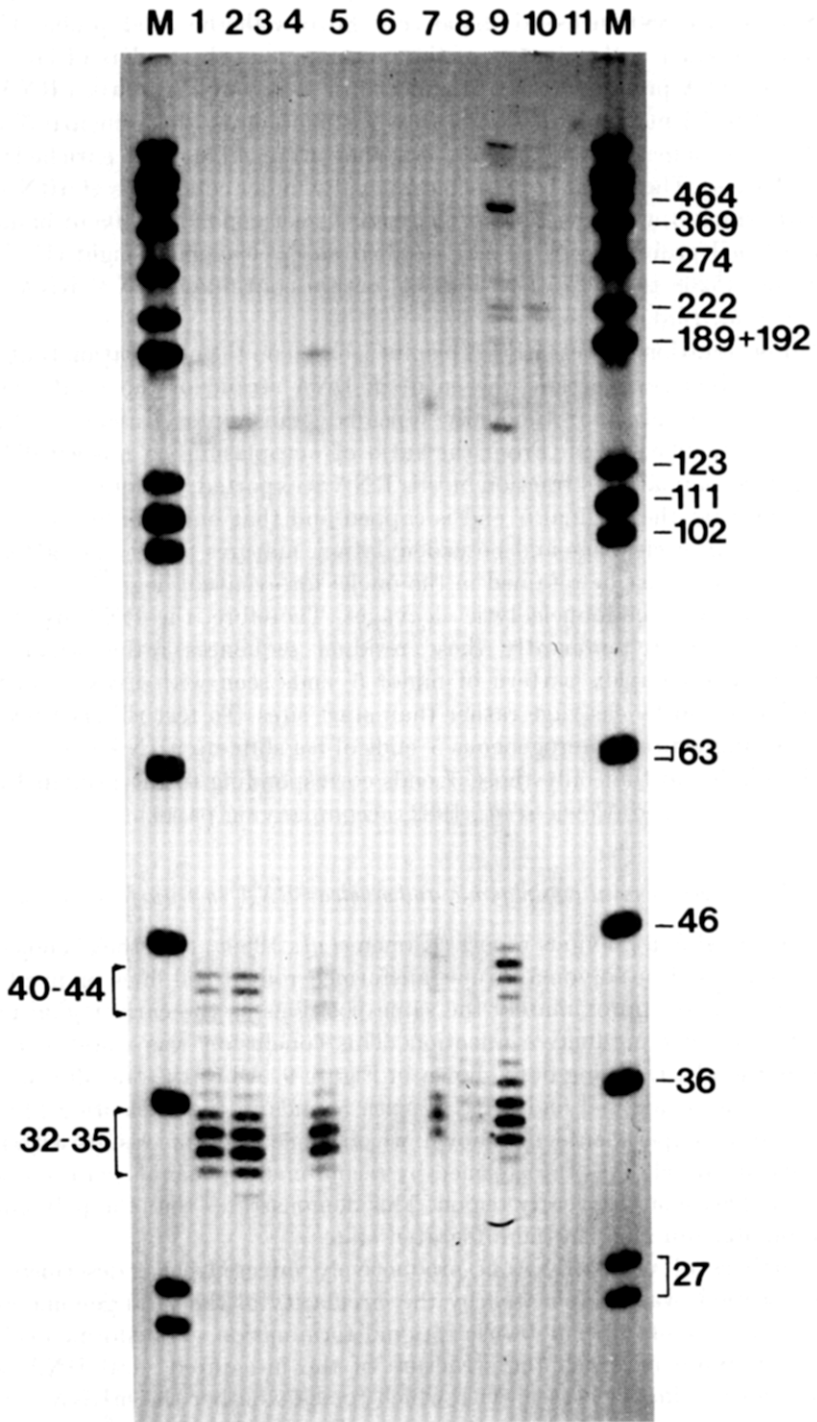


FIG. 9.

mouse cell lines. The results of the two experiments shown in Figure 10 demonstrate, with one exception, that the predominant nRNA 5' termini correspond to those of the principal mature mRNAs in the nt145 to nt155 region.

The exceptional instance was rat line BE-1, which yielded a pattern of S₁ products strikingly similar to that of nRNA from productively infected cells, as shown in the adjacent gel track (Fig. 10). The BE-1 cell line contains two copies of the viral DNA fragment from *Bcl*I to *Eco*RI (nt5022 clockwise to nt1565, see Fig. 1) integrated in tandem (Novak *et al.*, 1980). Each of these fragments includes a functional early region promoter (Treisman, 1981), but lacks the major viral early region polyadenylation site. Transcripts initiated at the first promoter would continue across the second promoter region. The second promoter region therefore is in a position topographically similar to the single promoter region present in a circular genome at late times of infection. The results obtained with line BE-1 therefore support the hypothesis that the heterogeneous pattern of 5' termini found with nuclear RNA from infected cells results from readthrough transcription rather than multiple initiation sites.

(e) *Early region mRNAs produced by viable deletion mutants*

A diagram relating the positions of the wild-type polyoma virus early mRNA 5' ends to other features of the beginning of the early coding region and its 5'-flanking sequences is shown in Figure 11. The principal 5' termini (nt145 to nt156) lie approximately 30 bases downstream from a TATA box, and about 90 bases from the limit of sequences known to be important for viral DNA replication (indicated as ORIGIN in Fig. 11). This limit was defined by a number of viable deletion mutants isolated in several laboratories (Bendig & Folk, 1979; Griffin & Maddock, 1979; Magnusson & Berg, 1979; Wells *et al.*, 1979), which together define a 'non-essential region' extending from after the high-affinity large T protein binding site (Caudray *et al.*, 1981) to just before the translational initiation codon (Fig. 11). Surprisingly, this non-essential region includes the principal mRNA 5' end sites and their associated TATA box. It was therefore of considerable interest to map the 5' ends of the mRNAs produced by several of the viable deletion mutants in order to assess the consequences of removing a sequence element strongly implicated as an important component of eukaryotic promoters (Breathnach & Chambon, 1981).

FIG. 9. Polyacrylamide (12%)/urea gel fractionation of S₁-resistant DNA after hybridization of cytoplasmic polyadenylated RNA from various polyoma virus-transformed rat or mouse cell lines to (5'-³²P)-labelled *Dde*I-1 *E*-strand DNA. Tracks 1 and 2, transformed rat cell lines obtained from Y. Ito (Ito & Spurr, 1980); track 1, line tsA-RE-B3, 7.0 µg; track 2, line T1A1, 8.4 µg. Tracks 3 to 7, transformed rat and mouse lines obtained from M. Fried (Lania *et al.*, 1980; Kamen *et al.*, 1980*a,b*); track 3, rat 53, 8.2 µg; track 4, rat 82, 10 µg; track 5, mouse tsC-3T3-3, 10 µg; track 6, mouse 3T3-WT-1, 10 µg; track 7, mouse tsC-3T3-1, 8.4 µg; track 8, mouse tsA-3T3-1, 5 µg; track 9, mouse line Py6 (T. Benjamin, unpublished results; see also Kamen *et al.*, 1974). Track 11, RNA from wild-type virus-infected 3T6 cells, 5 µg. Track 12, carrier RNA only. Tracks M are size markers as in Fig. 2. The lengths of principal S₁-resistant DNA products are indicated to the left of the autoradiogram. The bands in tracks 3, 5 and 6 may not be clearly visible in the reproduced photograph, but were clearly the same as those in the other tracks on the original film.

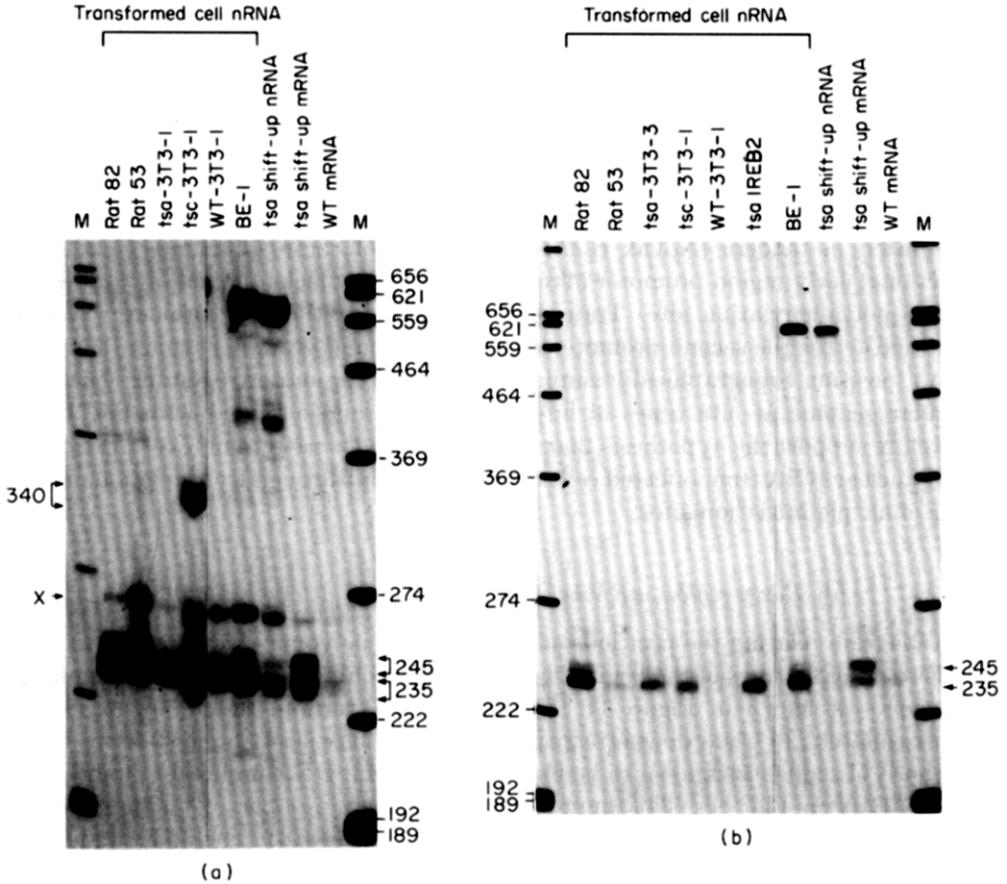


Fig. 10. Localization of the 5' ends of viral nuclear RNA in a variety of virally transformed rodent cell lines by hybridization to the *Hinf*I DNA probe and fractionation of S_1 -resistant products on a 5% polyacrylamide gel. The quantities of RNA used were: rat 82, 40 μ g; rat 53, 100 μ g; tsa3T3-1, 40 μ g; tsc3T3-1, 100 μ g; WT3T3-1, 200 μ g; tsa-IREB2, 80 μ g; BE-1 (Novak *et al.*, 1980), 30 μ g; tsa shift-up total nuclear RNA, 0.2 μ g; tsa shift-up polyadenylated cytoplasmic mRNA, 0.2 μ g; wild-type mRNA, 0.2 μ g. (a) and (b) show duplicate experiments. The large quantities of RNA used in these annealings required RNAase A digestion of S_1 -resistant material before electrophoresis to avoid overloading the thin gels. The minor bands at approx. 340 nucleotides seen with tsc3T3-1 nRNA map the position of an 11-base deletion in the RNA relative to the probe DNA present because the cell line was transformed with strain A3 polyoma virus (M. Fried, personal communication; see also Fig. 11).

The three mutants selected for study (Fig. 11) all lack the TATA box. Mutants dl 17 and dl 75 also lack the mRNA 5' end site sequences. dl 17 and dl 2-19 grow nearly as well as wild-type virus, but dl 75 is partially defective in early gene expression (Bendig *et al.*, 1980). In preliminary experiments, we examined the early region mRNAs extracted from mutant-infected mouse cells 30 hours post-infection by one and two-dimensional S_1 nuclease/agarose gel mapping. We found normally spliced mRNAs with all three mutants. The 5' colinear segments of dl 17 and dl 75 early region mRNAs were approximately the same length as those of wild-type virus. Mutant dl 2-19 produced highly heterogeneous 5' colinear sequences, most of

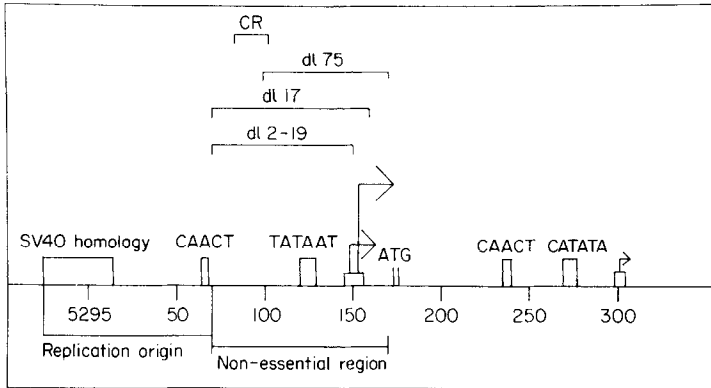


FIG. 11. Map of viable deletion mutants within the 5' flanking sequences of the polyoma virus early region. Nucleotides are numbered according to Soeda *et al.* (1980). The maximal extent of contiguous sequences essential for viral DNA replication, defined by viable deletion mutants are indicated by Replication origin. Putative sequence elements (see Discussion) comprising the early region promoter(s) are shown with boxes. Arrows show the positions of principal mRNA 5' ends mapped in this paper. Deletions dl 75, dl 17 and dl 2-19 were described by Bendig *et al.* (1980). The smaller deletion CR was isolated by M. Fried (personal communication) and sequenced by one of us (R. Kamen, unpublished results). The non-essential region defined by the extent of the viable deletions is indicated.

which were longer than normal. Mutants dl 17 and dl 2-19 synthesized normal levels of early region mRNAs, but dl 75 overproduced these molecules by about tenfold at late times during infection.

The 5' ends of the deletion mutant mRNAs were positioned more accurately by high-resolution S_1 gel mapping. The probes used were homologous single-stranded DNA fragments, labelled at the *HinfI* site (cf. Fig. 1). Results obtained with dl 17 mRNA demonstrated two principal 5' termini (Fig. 12(a), track 1; Fig. 12(b), tracks 1 and 2) at nt164 \pm 2 and nt58 \pm 2. These positions are just before and just after the deleted sequences (see the diagram in Fig. 12(b)). There were a multitude of other less prominent S_1 -resistant DNA products, indicative of a number of alternative 5' termini at the positions indicated in Figure 11(b). The locations of the principal dl 17 5' termini and the occurrence of minor alternatives were confirmed by primer extension analysis (Fig. 3, track 4). To ask whether any of these termini corresponded to those of minor species also present among transcripts of non-deleted viral genomes, we annealed wild-type and *tsa* shift-up RNA preparations to the deletion mutant *HinfI* DNA probe. Because S_1 nuclease does not efficiently cleave the DNA strand of a DNA-RNA hybrid opposite a single-stranded RNA loop, S_1 -resistant DNA products generated with non-deleted RNA would have the same lengths as those produced by transcripts of the deletion mutant DNA with 5' ends at the corresponding positions. Hybridization of wild-type mRNA to the dl 17 DNA probe (Fig. 12(a), track 2) yielded, in addition to a very prominent product mapping the 3' limit of the deletion (generated by mRNAs with the principal 5' ends at nt145 to nt156 or other 5' ends within the deleted sequences), a distribution of minor ones consistent with the results presented in Figures 2 and 4. Comparison of the results obtained with dl 17 mRNA and wild-type mRNA (Fig. 12(a), tracks 1 and 2; Fig. 12(b), tracks 1, 2 and 6) indicated that only the principal dl 17 5'

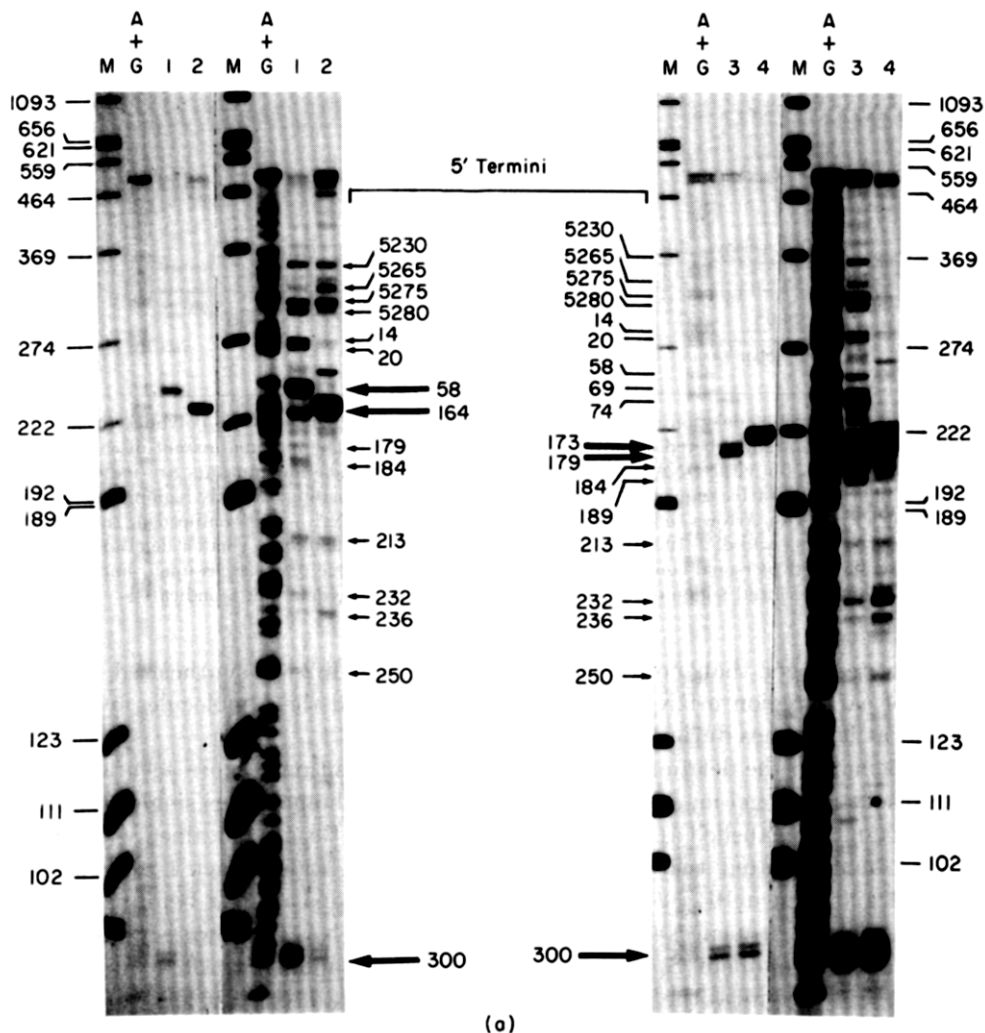


FIG. 12. S_1 nuclease gel mapping of the 5' ends of dl 17 and dl 75 mRNAs extracted 30 h post-infection by hybridization to homologous deletion mutant *HinfI* DNA probes. (a) Track 1, 1 μ g dl 17 mRNA; track 2, 1 μ g wild-type mRNA; track 3, 0.5 μ g dl 75 mRNA; track 4, 0.5 μ g tsa shift-up mRNA. (b) Tracks 1 and 2, 1 and 2 μ g dl 17 mRNA; track 3, 0.5 μ g tsa shift-up mRNA; tracks 4 and 5, 0.125 and 0.3 μ g tsa shift-up mRNA; track 6, 2 μ g wild-type mRNA.

termini at nt58 \pm 2 and nt164 \pm 2 were highly over-represented in the mutant mRNAs. All of the other dl 17 products align with wild-type equivalents. The mutant mRNA was also moderately enriched for termini at approximately nt14, nt20 and nt300 \pm 2. Careful inspection of Figure 12(a) (track 2) and Figure 12(b) (tracks 3 and 6) revealed that the principal dl 17 mRNA 5' end at nt58 \pm 2 is present at a very low level in wild-type or tsa shift-up mRNA preparations. Moreover, RNA with 5' ends at this place is relatively abundant in tsa shift-up nuclear RNA (Fig. 12(b), tracks 4 and 5). Products corresponding to the dl 17 5' end at nt164 \pm 2

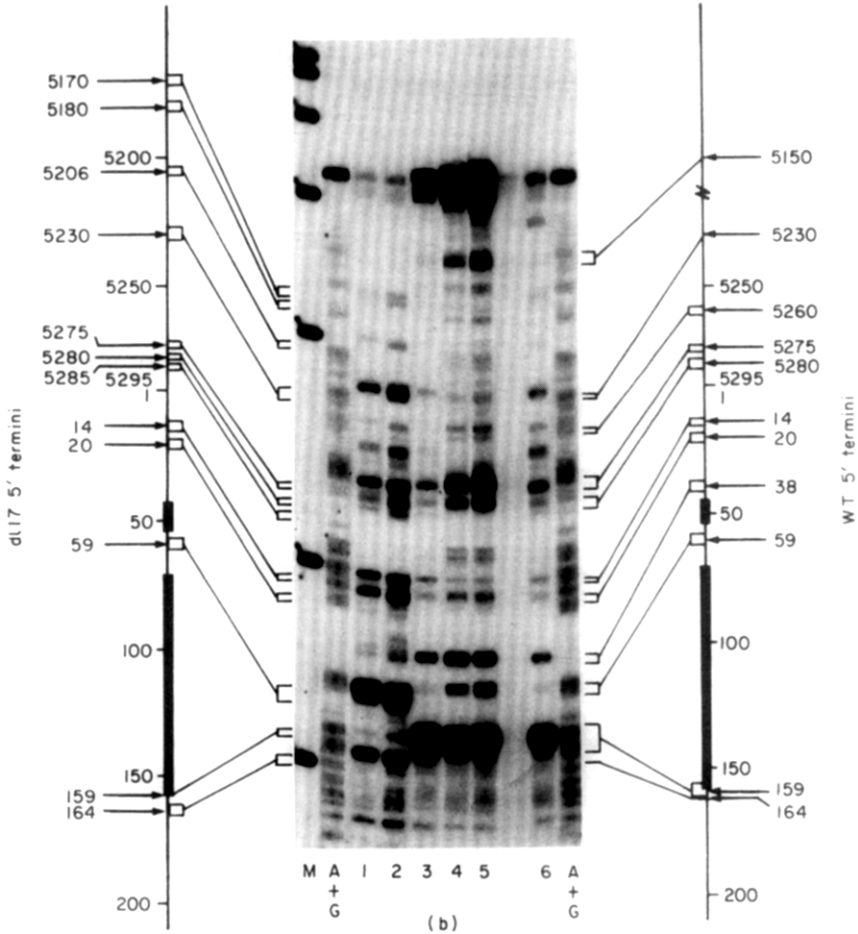


FIG. 12(b).

are also present as faint bands in control mRNA tracks, but these are difficult to see in the photographs because they lie just below the very dark bands corresponding to the 3' limit of the deletion.

Similar analysis of dl 75 mRNA (Fig. 12(a), tracks 3 and 4) demonstrated principal 5' termini mapping 3' to the deletion at positions (± 2 nucleotides) 173, 179, 184 and 189. The primer extension result shown in Figure 3 (track 7) is consistent with this conclusion. Other 5' ends found with the S_1 technique mapped 5' to the deletion at positions 74, 69 and 58, but these were relatively minor. In addition, all of the products corresponding to minor ones in tsa shift-up mRNA were somewhat over-represented (Fig. 12(a), tracks 3 and 4). The majority of the early region mRNA produced by dl 75 thus start within and after the protein synthesis initiation codon at positions 173 to 175 (cf. Fig. 11). This explains a paradox. Bendig *et al.* (1980) observed that dl 75 synthesized very reduced amounts of early proteins. However, we found that dl 75 over-produced early region mRNA by tenfold relative to the wild-type virus. We infer from the mapped

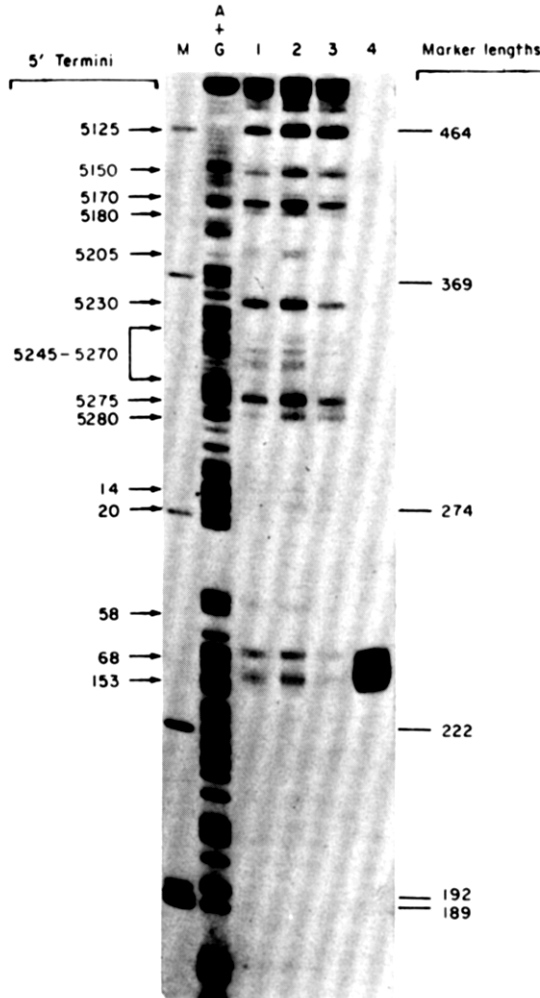


FIG. 13. S_1 nuclease gel mapping of the 5' ends of dl 2-19 mRNA extracted 30 h post-infection by hybridization to the dl 2-19 *Hinf*I DNA probe. Tracks 1 to 3, 1, 2.5 and 5.0 μ g dl 2-10 mRNA; track 4, 2.7 μ g wild-type mRNA, 30 h post-infection.

positions of the principal 5' termini of dl 75 early region cytoplasmic transcripts that the majority of these "mRNAs" lack the initiation codon and thus could not be normally translated. Therefore, normal levels of large T protein are not attained, and early strand transcription is not repressed.

Mutant dl 2-19 has a deletion beginning at the same point as that in dl 17, but it retains nine more nucleotides of wild-type sequence (Fig. 11) including some of the sequences determining the principal 5' ends of wild-type mRNAs. The data shown in Figure 13 imply that dl 2-19 mRNAs have very heterogeneous 5' ends, including major species extending well into the late region of the genome. A significant fraction of molecules had 5' ends at the nt154 \pm 2 principal 5' end site. The 5' end at nt58 \pm 2, which is the major one in dl 17 mRNA, was minor, while the next most

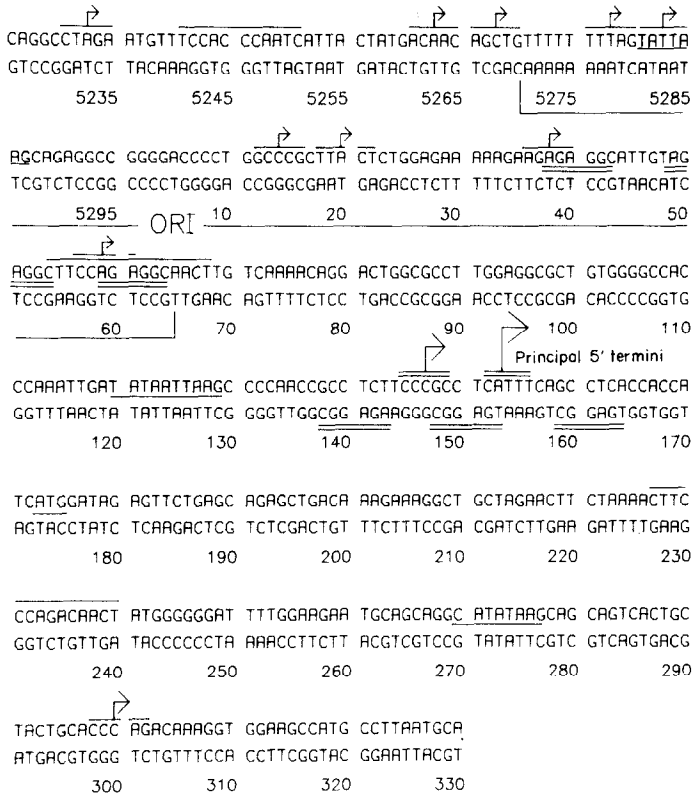


Fig. 14. The early promoter region of polyoma virus DNA. The portion of the strain A2 genomic DNA sequence (Soeda *et al.*, 1980) extending from just before the replication origin into the beginning of the early portion coding region is shown. The upper strand has the 5' to 3' polarity of early region mRNAs. Positions of 5' termini mapped in this paper are shown with arrows above the DNA sequence, with error margins of ± 2 bp. The principal termini, localized by S_1 mapping (Fig. 4) to nt147 ± 2 and nt154 ± 2 are shown with larger arrows. Minor 5' termini are indicated by smaller arrows. The data for the minor 5' termini upstream of the principal cap sites (at positions 5230, 5265, 5260, 5275, 5280, 14, 20, 38 and 59) are in Figs 2(b) (track 1), 4 (tracks 1 to 4), 6 (tracks 6 to 10), 11(a) (tracks 2 and 4), 11(b) (tracks 3 to 5). Only the most prominent of the minor 5' ends located downstream of the principal sites, at nt300 ± 2 , is included in the Figure. It was clearly detected by both S_1 gel mapping (see, for example, Fig. 2(b)) and by primer extension (Fig. 3, track 8) and, moreover, was relatively enriched among mRNAs of the viable deletion mutants (Fig. 11(a), tracks 1 and 3). Other faint S_1 -resistant DNA products smaller than the principal ones (detected, for example, in Fig. 2(b)) could represent RNA nicking or internal S_1 cleavage and have therefore not been considered significant at this time. Several DNA sequences of interest are delineated. There are 3 TATA box-like sequences (underlined), at nt5278 to nt5284, at nt120 to nt129, and at nt270 to nt278. The 3 sequences ending with C-A-A^{CT}_T boxes discussed in the text and

compared in Fig. 15, are delineated with overlining. Repeats of the sequence 5'^A_TG-A-G-G-C 3' are indicated by double underlining below the appropriate DNA strand. Two of the three A-G-A-G-G-C repeats between nt38 and nt64 are required for large T protein binding to the viral DNA and thus may constitute a high-affinity binding site for this molecule (Gaudray *et al.*, 1981). The possible significance of the related but inverted 3-fold repeat from nt164 to nt138 is discussed in the text. The approximate limits of contiguous sequence required for viral DNA replication, defined by viable deletion mutants (Bendig & Folk, 1979; Bendig *et al.*, 1980; Magnusson & Berg, 1979; Griffin & Maddock, 1979; Tyndall *et al.*, 1981; G. Magnusson, personal communication), are indicated by ORI. The translational initiation codon (positions 173 to 175) is underlined. Hyphens have been omitted for clarity.

abundant dl 17 mRNA 5' end (that at nt164 \pm 2) was not found. Other S₁-resistant DNA products generated with dl 2-19 mRNA map the 5' ends at positions 68, 20, 14, 5280, 5275 and 5230 also present in wild-type early region mRNAs at low levels (see Fig. 14). From analyses like those shown in Figure 5, we deduced that the specific products mapping apparent early mRNA 5' ends at approximately positions 5180, 5170, 5150 and 5125 (Fig. 13) are RNA-RNA annealing shadows representing the heterogeneous 5' ends of late region *L*-DNA strand transcripts: their existence proved only that 5' termini of many early region mRNAs in dl 2-19-infected cells overlap those of the late region mRNAs.

The three deletion mutants discussed above were derived from the A3 strain of polyoma virus. This strain lacks nucleotides 44 to 54 of the A2 strain (Soeda *et al.*, 1980), which is our standard wild-type virus and the parent of conditional mutant *tsa*. To demonstrate that the observed differences between wild-type or *tsa* early region mRNAs and those of the deletion mutants are not influenced by this additional deletion, we have also characterized the mRNAs produced by strain A3 viable mutant CR (Fig. 11). This is an A3 virus with a small additional deletion that removes nucleotides 86 to 97. The 5' ends of CR messengers were indistinguishable from those of wild-type A2 viruses (data not shown).

4. Discussion

This is the first of a series of papers in which we characterize the polyoma virus early region transcriptional promoter by studying its function, both *in vivo* and *in vitro*, in the wild-type state and after genetic manipulation. The primary goal of the experiments reported here was to map accurately the 5' termini of transcripts synthesized under a variety of conditions during the productive infection of permissive cells and in virally transformed cell lines. This information was essential for the interpretation of subsequent experiments to identify DNA sequence elements responsible for the specification of transcription and its regulation by the large T protein (Tyndall *et al.*, 1981; de Villiers & Schaffner, 1981; Gaudray *et al.*, 1981; Jat *et al.*, 1982; P. Jat, U. Novak, A. Cowie, C. Tyndall & R. Kamen, unpublished results). The results are of further practical utility in the construction of vectors using polyoma virus sequences to promote the expression of other eukaryotic genes.

Previous characterization of the 5' termini of polyoma virus late region transcripts and of those of SV40 late and early region mRNAs (Ghosh *et al.*, 1978*a,b*, 1980*a,b*; Lai *et al.*, 1978; Reddy *et al.*, 1978, 1979; Canaani *et al.*, 1979; Flavell *et al.*, 1979, 1980; Haegeman & Fiers, 1978, 1980; Kahana *et al.*, 1981; Treisman, 1980) established the precedents for both macro- and micro-heterogeneity. We were therefore not surprised by the rather complex results obtained, which are summarized in Figure 14. The main conclusions, discussed separately below, are (1) the principal 5' termini of early region mRNAs occur at several alternative positions between nt145 and nt156 in the DNA sequence. We show elsewhere (Cowie *et al.*, 1982, accompanying paper) that these 5' ends are capped. The principal cap sites are preceded by a TATA-like sequence. (2) Relatively minor 5' termini at nt300 \pm 2 are present in all RNA preparations. A

TATA-related sequence also precedes this position, but the putative termini would occur within the early protein coding sequence. (3) Other minor 5' termini at several points between nt5230 and nt5260 were localized by the S₁ mapping procedure. These are particularly abundant in many preparations of RNA extracted from productively infected cells at late times during infection, when functional large T protein is present. The nt5230–5260 region of the genome includes a third TATA-like sequence as well as the origin of viral DNA replication. Analysis of capped 5'-terminal T₁ oligonucleotides (Cowie *et al.*, 1982, accompanying paper) detected a minor cap site at nt16, which agrees with the position of the minor 5' end at nt14 ± 2 deduced from S₁ mapping. This 5' end occurs 31 bp downstream from the TATA-box (Fig. 14) and is located at a position analogous to that of the SV40 major early region cap sites (Ghosh & Lebowitz, 1981). (4) Homologous sequences, including a CAA^{CT}_{TC} box, occur 40 to 66 bp before each of the three TATA boxes. (5) Viable deletion mutants lacking the sequences determining the principal 5' termini and the preceding TATA sequence synthesize mRNAs with slightly or highly heterogeneous termini. Each mutant produces a different characteristic set of termini, which in all cases correspond to very minor 5' ends detectable among transcripts of wild-type viral DNA.

(a) *The principal 5' termini of polyoma virus early region transcripts*

Results of S₁ nuclease and primer extension measurements suggested that the principal 5' termini of mRNAs synthesized during productive infection occur at multiple positions between nt145 and nt156 in the genomic DNA sequence, within a region that is 23 to 36 bp after a TATA box-like element (Fig. 13). S₁ mapping data, such as shown in Figure 4 (track 3), suggested that there were a minimum of two 5' ends, at nt147 ± 2 and at nt154 ± 2 (Fig. 14). This conclusion also pertained to mRNAs and nuclear RNAs present in a variety of virally transformed cell lines (Figs 9 and 10). The present data, however, do not allow the exact assignment of terminal nucleotides or demonstrate that these termini are capped. We have directly analysed the terminal T₁ oligonucleotides of chemically decapped and (5'-³²P)-labelled tsa shift-up mRNAs. These experiments are described in the accompanying paper, but the conclusions inferred are germane to the present discussion and will therefore be summarized briefly here. We found that the cap sites correspond to the A residue at nt153, the U residue at nt151, the C residue at nt150, and to the G residue at nt148 (see Fig. 13). The two pyrimidine cap sites were relatively minor. We also observed, however, that a significant proportion of the mRNAs had capped 5' ends apparently formed by the "back-slippage" of oligonucleotide primers corresponding to different lengths of the sequence from the A residue at nt153 to the U residue corresponding to nt156. We further found that early region transcripts initiated in the HeLa cell system *in vitro* (Manley *et al.*, 1980), which faithfully transcribes the polyoma virus early promoter (Jat *et al.*, 1982), had a very similar distribution of capped 5' termini, including those produced by back-slippage during RNA chain initiation. In the system *in vitro*, the relative proportions of caps at different nucleotides from 148 to 153 could be

manipulated by varying nucleoside triphosphate concentrations. This established both that the cap sites correspond to initiation sites, and that the bias in favour of the A cap at nt153 observed *in vivo* is more probably a reflection of nuclear triphosphate levels than an inherent specificity of RNA polymerase II.

Comparison of the principal 5' ends of early region mRNAs synthesized at late times during infection in the presence or absence of functional large T protein, which is known to regulate early gene expression (Cogen, 1978; Kamen *et al.*, 1980*a,b*), demonstrated significant differences in the distribution of termini within the nt148 to nt153 region (Figs 2 and 4). In particular, only the mRNA overproduced after inactivation of the large T protein had large relative amounts of the S₁ product corresponding to the G terminus at nt148 (see Fig. 4, tracks 3 and 8). Further experiments are required to determine the mechanistic basis of this phenomenon.

(b) *Alternative minor 5' termini within the coding region*

In all S₁ nuclease or primer extension analyses of RNA from infected cells, we detected a minor 5' end located at nt300 ± 2 (Fig. 14). Since this was found with both techniques, it is probably not a methodological artifact. This conclusion is strengthened by the observations that (1) the minor 5' end occurs at the same relative level among transcripts produced *in vitro* (Jat *et al.*, 1982). (2) The terminus at nt300 ± 2 is much less abundant in nuclear RNA (Fig. 7, nRNA tracks), showing that it is not a result of S₁ nuclease "cut-through" or reverse transcriptase stops. (3) Certain deletion mutations that virtually abolish initiation at the principal cap sites *in vitro* still yield normal levels of this minor 5' end (P. Jat *et al.*, unpublished results). (4) A TATA box sequence (C-A-T-A-T-A-A; cf. Fig. 13) occurs about 30 bp 5' to position 300 (Fig. 14). We infer that this 5' end indicates a second early region initiation site. This, however, would lie some 125 bp within the early region coding sequences. Its physiological significance is unknown.

(c) *The significance of the "upstream" 5' termini*

Our initial comparisons of early region mRNAs synthesized under different conditions at late times during infection (see, for example, Figs 6 and 7) suggested that the occurrence of longer messengers with 5' ends between nt5230 and nt5260 (the replication origin region, as illustrated in Figs 11 and 14) was negatively correlated with the presence of functional large T protein. This implied the intriguing possibility that large T protein repression shifted transcription from the principal to a minor upstream promoter region. Similar conclusions have been reported with the related virus SV40 (Ghosh & Lebowitz, 1981). As we now have major reservations concerning this interpretation of the data, it is important to take this opportunity to discuss the situation in detail.

We first compare the polyoma virus early promoter region with that of SV40. The polyoma virus principal cap sites are some 150 bp remote from the replication origin (Figs 11 and 14) within a region not homologous to any sequences found in SV40 DNA. The SV40 principal cap sites (Ghosh & Lebowitz, 1981), by contrast, are very near the highly conserved papovavirus origin region, which contains a

TATA box-like element (nt5278 to nt5283 in polyoma virus DNA, see Fig. 14). The SV40 principal cap sites are about 30 bp after the counterpart of this element in SV40 DNA. Two of the minor polyoma virus upstream 5' termini indicated in Figure 14, at nt14 \pm 2 and at nt20 \pm 2 (see Fig. 2(b), track 3; Fig. 12(a), tracks 2 and 4; Fig. 12(b), track 3; for data supporting these assignments), are at positions in the polyoma sequence analogous to the SV40 principal cap sites. The 5' end at nt14 \pm 2 has been correlated with a very minor capped 5' end at the G residue corresponding to nt16 (see Fig. 14) found by analysis of terminal T₁ oligonucleotides (Cowie *et al.*, 1982). We infer that the origin region of polyoma virus DNA retains some of the promoter function of the homologous sequences in SV40 DNA, but that this polyoma promoter is very inefficient. Of the other minor 5' termini indicated by S₁ nuclease analysis, those at about nt5275 and nt5280 (Fig. 14 for relevant data; see, for example, Fig. 12(b)) probably represent the cleavage of a highly U-rich region within continuous hybrid. A similar phenomenon has been reported with SV40 (Hansen *et al.*, 1981). This demonstrates only that some RNA molecules extend to positions beyond nt5275. The minor 5' termini at nt5260 and nt5230 in the polyoma sequence (Fig. 14, cf. Fig. 12(b), track 6) may correspond to the upstream termini identified in SV40 by Ghosh & Lebowitz (1981), but we have no evidence that they are indeed cap sites. We have, however, excluded the possibility that these polyoma virus 5' termini are artifactually produced by RNA-RNA self-annealing (see Fig. 5). We know further that polyoma virus mRNA preparations that contain relatively large amounts of longer species include molecules with 5' ends extending beyond nt5230. These are not indicated in Figure 14, because RNA-RNA annealing prevented the mapping of their 5' termini by the S₁ or primer extension techniques.

The following observations caution against the interpretation that large T protein repression causes a shift towards an upstream promoter: (1) the recovery of polyadenylated cytoplasmic mRNAs with the longer 5' regions was variable among many RNA preparations; (2) nuclear viral RNA, at late times during infection, comprised giant species produced by continuous transcription around the circular genome. These molecules had very heterogeneous 5' termini, many of which correlated with the 5' ends of the longer "cytoplasmic" mRNAs. There was far more nuclear than cytoplasmic RNA at such late times, making contamination a major problem. Only those 5' ends of nuclear RNA corresponding to the principal cap sites were indeed capped (Cowie *et al.*, 1982, accompanying paper). We therefore suspect that the other termini are processing rather than initiation sites. (3) Early region transcripts with 5' ends upstream of the principal cap sites are uncommon among the various transformed cell lines studied (cf. Figs 9 and 10). One cell line, however, produced nRNA molecules with heterogeneous 5' termini essentially identical to those present at late times during productive infection (Fig. 10, track BE-1). This cell line (BE-1; Novak *et al.*, 1980) contains two fragments of viral DNA (extending from about nt5022 clockwise to nt1560; see Fig. 1) integrated in tandem; transcripts initiated at the first promoter region would readthrough the second promoter. Moreover, although BE-1 cells produce transcripts with the longer 5' termini, they do not synthesize functional large T protein.

These observations, taken together, suggest that the production of "mRNAs" with longer 5' regions at late times during infection results from inefficient termination, transcription around the circular genome, and the subsequent cleavage of the giant transcripts at a multitude of points, some of which are in the origin region. Whether such cleavage products can serve as mRNAs if exported to the cytoplasm is not known. Fenton & Basilico (personal communication) have found that longer early region species can be detected on gels fractionating cytoplasmic RNA extracted from cells at late times during infection. They also noted that there is an open codon reading frame (from nt5255 to nt122) available in such molecules, which could encode an additional viral gene product. Several viable deletion mutants exist that interrupt this potential coding region, however, and therefore the putative protein cannot be essential.

(d) *Consensus sequences*

We have identified three regions of polyoma virus DNA that are likely to include early mRNA cap sites (Fig. 14): the "SV40-like" region from nt16 to nt20, the principal cap site region from nt148 to nt153, and the further minor 5' end at $\text{nt}300 \pm 2$. Each of these follows a sequence related to the TATA consensus by the usual distance. A second conserved sequence element, known as the CCAAT box (Efstratiadis *et al.*, 1980; Benoist *et al.*, 1980; Corden *et al.*, 1980; Breathnach & Chambon, 1981), has been found in most but not all eukaryotic RNA polII genes, usually about 65 to 80 bp upstream of the cap site. Deletion of this sequence reduces globin gene expression *in vivo* (Dierks *et al.*, 1981; Mellon *et al.*, 1981), and it is included within a region required for *Herpes simplex* virus thymidine kinase gene expression (McKnight *et al.*, 1981). The three polyoma TATA boxes are preceded by homologous sequences ending with $\text{CAA}_{\text{TC}}^{\text{CT}}$, which are located 40 to 66 bp upstream of the TATA boxes or 70 to 96 bp upstream of the cap sites (Fig. 14). As illustrated in Figure 15, there is also conservation of sequence just 5' to

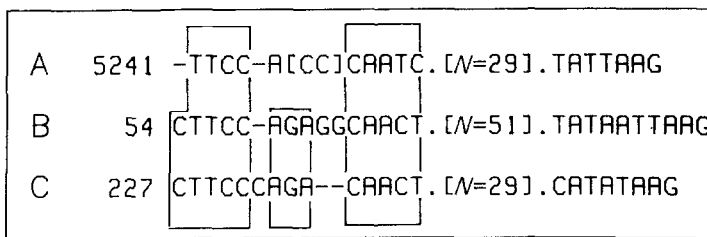


FIG. 15. Comparison of the sequences around the 3 polyoma virus $\text{CAA}_{\text{TC}}^{\text{CT}}$ boxes indicated in Fig. 14. The first nucleotide of each line is given in standard nucleotide numbers to permit relation of this diagram to the sequence shown in Fig. 14. *N* is the distance from the end of the $\text{CAA}_{\text{TC}}^{\text{CT}}$ box to the beginning of the TATA box. Hyphens have been omitted for clarity.

the three CAA^{CT}_{TC} boxes. The functional significance of these homology regions is under investigation.

(e) *The expression of viable deletion mutants*

The principal early region cap sites and their TATA box are within a region defined as non-essential for early gene expression by a number of viable deletion mutants (Bendig & Folk, 1979; Bendig *et al.*, 1980; Magnusson & Berg, 1979). This is perhaps not surprising, since it is now known that the removal of such elements from several different cellular and viral genes does not abolish transcription *in vivo* (Grosschedl & Birnstiel, 1980b; Benoist & Chambon, 1981; McKnight *et al.*, 1981). Mapping of the 5' ends of early region mRNAs produced by three of the polyoma viable deletion mutants (dl 75, dl 17, dl2-19; Bendig *et al.*, 1980) showed that each mutant synthesized messengers with a particular set of slightly (dl 75 and dl 17) or highly heterogeneous (dl 2-19) 5' termini. These termini, in all instances, corresponded to the overproduction of very minor alternative 5' ends also present among transcripts of wild-type DNA. Mutant dl 2-19 retains the DNA sequence corresponding to the A cap site at nt153. This site continued to function without the TATA box, but became one of a great number of alternative possibilities. As the polyoma virus early promoter is negatively regulated by one of its gene products, it is difficult to assess from the present data whether the deletions reduce inherent promoter efficiency. Results of transient expression assays, in which plasmids containing one of the mutants studied here (dl 75) as well as a number of further mutants (Tyndall *et al.*, 1981) were transfected into non-permissive HeLa cells, strongly suggest, however, that the efficiency of the early promoter is not substantially reduced by removal of the TATA element. Our results are thus in agreement with the view (Grosschedl & Birnstiel, 1980b; Benoist & Chambon, 1981; McKnight *et al.*, 1981) that sequence elements other than the TATA box are important in the quantitative transcription control of a certain region, whereas the TATA element serves to specify particular initiation sites from among several alternatives. The localization of the other control elements of the polyoma early promoter is the subject of a separate publication (Tyndall *et al.*, 1981).

(f) *Speculations on large T protein autoregulation*

The regulation of SV40 early gene expression by its large T protein is relatively well understood. There are three binding sites for this protein across the SV40 origin region (Tjian, 1978,1979). These span the major SV40 early region cap sites and the TATA box that precedes them. Addition of purified SV40 large T protein (or the similar D2 hybrid protein) to transcription systems *in vitro* specifically represses initiation of SV40 early region transcription (Rio *et al.*, 1980; Myers *et al.*, 1981; Hansen *et al.*, 1981). Equivalent experiments are not yet possible with polyoma virus because its large T protein has not been purified. Gaudray *et al.* (1981), however, recently localized a high-affinity binding site for polyoma virus large T on the viral DNA. This site comprises two copies of the A-G-A-G-G-C

sequence that is repeated three times between nt38 and nt64 in the strain A2 DNA sequence (see Fig. 14). As the principal polyoma virus cap sites mapped in this paper are some 90 bp distant from the high-affinity large T protein binding site, it seems likely that the mechanism of autoregulation will differ from that of SV40. We note in Figure 14 that the sequence T-G-A-G-G-C(N₅)T-G-A-G-G-C(N₄)A-G-A-G-G-C, which is strikingly homologous to the high-affinity binding site except for T to A transversions at the beginning of the first two A-G-A-G-G-C repeats, occurs across the principal cap site region (nt164 to nt138). However, the sequence is in the inverted orientation (Fig. 14). It will obviously be of considerable interest to determine whether polyoma large T protein represses transcription by binding to this inverted imperfect duplication of the high-affinity binding site.

We thank Mrs A Symons, Miss P. Morgan and Miss G. Yiangou for their assistance in the preparation of this manuscript.

Note added in proof: Recent experiments done in our laboratory (A. Cowie, F. Chaudry, S. Dilworth, S. Lupton & R. Kamen, unpublished observations) and by others (A. Scheller & C. Prives, personal communication) have confirmed the existence of the predicted second independent binding site for the large T protein in polyoma virus DNA.

REFERENCES

- Acheson, N. H. & Mieville, F. (1978). *J. Virol.* **28**, 885-894.
- Alwine, J. C., Reed, S. I. & Stark, G. R. (1977). *J. Virol.* **24**, 22-27.
- Baker, C. C. & Ziff, E. B. (1981). *J. Mol. Biol.* **149**, 189-221.
- Bendig, M. M. & Folk, W. R. (1979). *J. Virol.* **32**, 530-535.
- Bendig, M. M., Thomas, T. & Folk, W. R. (1980). *Cell*, **20**, 401-409.
- Benoist, C. & Chambon, P. (1981). *Nature (London)*, **290**, 304-309.
- Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980). *Nucl. Acids Res.* **8**, 127-143.
- Berk, A. J. & Sharp, P. A. (1977). *Cell*, **12**, 721-732.
- Breathnach, R. & Chambon, P. (1981). *Annu. Rev. Biochem.* **50**, 349-383.
- Canaani, D., Kahana, C., Mukamel, A. & Groner, Y. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 3078-3082.
- Cogen, W. (1978). *Virology*, **85**, 222-230.
- Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedingner, C. & Chambon, P. (1980). *Science*, **209**, 1406-1414.
- Cowie, A., Tyndall, C. & Kamen, R. (1981). *Nucl. Acids Res.* **9**, 6305-6322.
- Cowie, A., Jat, P. & Kamen, R. (1982). *J. Mol. Biol.* **159**, 225-255.
- Deiningner, P., Esty, A., LaPorte, P. & Friedmann, T. (1979). *Cell*, **18**, 771-779.
- de Villiers, J. & Schaffner, W. (1981). *Nucl. Acids Res.* **9**, 6251-6264.
- Dierks, P., van Ooyen, A., Mantel, N. & Weissmann, C. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 1411-1415.
- Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissmann, S. M., Slightom, C. C. & Proudfoot, N. J. (1980). *Cell*, **21**, 653-668.
- Favaloro, J., Treisman, R. & Kamen, R. (1980). *Methods Enzymol.* **65**, 718-749.
- Flavell, A. J., Cowie, A., Legon, S. & Kamen, R. (1979). *Cell*, **16**, 357-371.
- Flavell, A. J., Cowie, A., Arrand, J. R. & Kamen, R. (1980). *J. Virol.* **33**, 902-908.
- Fried, M. (1965). *Virology*, **25**, 669-671.
- Friedmann, T., Esty, A., LaPorte, P. & Deiningner, P. (1979). *Cell*, **17**, 715-724.
- Gaudray, P., Tyndall, C., Kamen, R. & Cuzin, F. (1981). *Nucl. Acids Res.* **9**, 5697-5710.
- Ghosh, P. K. & Lebowitz, P. (1981). *J. Virol.* **40**, 224-240.

- Ghosh, P. K., Reddy, V., Swinscoe, J., Choudary, P., Lebowitz, P. & Weissmann, S. (1978a). *J. Biol. Chem.* **253**, 3643-3647.
- Ghosh, P. K., Reddy, V., Swinscoe, J., Lebowitz, P. & Weissmann, S. (1978b). *J. Mol. Biol.* **126**, 813-846.
- Ghosh, P. K., Reddy, V., Piatak, M., Lebowitz, P. & Weissmann, S. M. (1980a). *Methods Enzymol.* **65**, 580-594.
- Ghosh, P. K., Piatak, M., Reddy, V., Swinscoe, J., Lebowitz, P. & Weissmann, S. M. (1980b). *Cold Spring Harbor Symp. Quant. Biol.* **44**, 31-39.
- Gretz, M., Land, H., Giesecke, K., Schutz, G., Jung, A. & Sippel, A. (1981). *Cell*, **25**, 743-752.
- Griffin, B. E. & Maddock, C. (1979). *J. Virol.* **31**, 645-656.
- Grosschedl, R. & Birnstiel, M. L. (1980a). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 1432-1436.
- Grosschedl, R. & Birnstiel, M. L. (1980b). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 7102-7106.
- Haegeman, G. & Fiers, W. (1978). *J. Virol.* **25**, 824-830.
- Haegeman, G. & Fiers, W. (1980). *J. Virol.* **35**, 955-961.
- Hansen, U., Tenen, D. G., Livingston, D. M. & Sharp, P. A. (1981). *Cell*, **27**, 603-612.
- Hentschel, C., Irmlinger, J.-C., Bucher, P. & Birnstiel, M. L. (1980). *Nature (London)*, **285**, 147-151.
- Ito, Y. & Spurr, N. (1980). *Cold Spring Harbor Symp. Quant. Biol.* **44**, 149-157.
- Jat, P., Roberts, J. W., Cowie, A. & Kamen, R. (1982). *Nucl. Acids Res.* **10**, 871-887.
- Kahana, C., Gidoni, D., Canaani, D. & Groner, Y. (1981). *J. Virol.* **37**, 7-16.
- Kamen, R. I., Lindstrom, D. M., Shure, H. & Old, R. (1974). *Cold Spring Harbor Symp. Quant. Biol.* **39**, 187-198.
- Kamen, R. I., Sedat, J. & Ziff, E. (1976). *J. Virol.* **17**, 212-218.
- Kamen, R. I., Favalaro, J. & Parker, J. (1980a). *J. Virol.* **33**, 637-651.
- Kamen, R. I., Favalaro, J. M., Parker, J. T., Treisman, R. H., Lania, L., Fried, M. & Mellor, A. (1980b). *Cold Spring Harbor Symp. Quant. Biol.* **44**, 63-75.
- Khoury, G. & May, E. (1977). *J. Virol.* **23**, 167-176.
- Lai, C. J., Dhar, R. & Khoury, G. (1978). *Cell*, **14**, 971-982.
- Lania, L., Hayday, A., Bjursell, G., Gandini-Attardi, D. & Fried, M. (1980). *Cold Spring Harbor Symp. Quant. Biol.* **44**, 597-603.
- Magnusson, G. & Berg, P. (1979). *J. Virol.* **32**, 523-529.
- Malek, L. T., Eschenfeldt, W. H., Munns, T. W. & Rhoads, R. F. (1981). *Nucl. Acids Res.* **9**, 1657-1673.
- Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Geter, M. L. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 3855-3859.
- Maxam, A. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499-560.
- McKnight, S. L., Gavis, E. R., Kingsbury, R. & Axel, R. (1981). *Cell*, **25**, 385-398.
- Mellon, P., Parker, V., Gluzman, Y. & Maniatis, T. (1981). *Cell*, **27**, 279-288.
- Myers, R. M., Rio, D. C., Robbins, A. K. & Tjian, R. (1981). *Cell*, **25**, 273-284.
- Novak, U., Dilworth, S. M. & Griffin, B. E. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 3278-3282.
- Reddy, V. B., Ghosh, P. K., Lebowitz, P. & Weissmann, S. M. (1978). *Nucl. Acids Res.* **5**, 4195-4214.
- Reddy, V. B., Ghosh, P. K., Lebowitz, P., Piatak, M. & Weissmann, S. M. (1979). *J. Virol.* **30**, 279-296.
- Reed, S. I., Stark, G. R. & Alwine, J. C. (1976). *Proc. Nat. Acad. Sci., U.S.A.* **73**, 3083-3087.
- Rio, D., Robbins, A., Myers, R. & Tjian, R. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 5706-5710.
- Sanger, F. & Coulson, A. R. (1978). *FEBS Letters*, **87**, 107-110.
- Soeda, E., Arrand, J. R., Smolar, N., Walsh, J. E. & Griffin, B. E. (1980). *Nature (London)*, **283**, 445-453.
- Sollner-Webb, B. & Reeder, R. H. (1979). *Cell*, **18**, 485-499.
- Tegtmeier, P., Schwartz, M., Collins, J. K. & Rundell, K. (1975). *J. Virol.* **16**, 168-178.
- Tyndall, C., La Mantia, G., Thacker, C. H., Favalaro, J. & Kamen, R. (1981). *Nucl. Acids Res.* **9**, 6231-6250.

- Tjian, R. (1978). *Cell*, **13**, 165–179.
- Tjian, R. (1979). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 655–662.
- Tooze, J. (1980). Editor of *The Molecular Biology of Tumor Viruses—DNA Tumor Viruses. Part 2*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Treisman, R. H. (1980). *Nucl. Acids Res.* **8**, 4867–4888.
- Treisman, R. H. (1981). Ph.D. thesis. University College, London.
- Treisman, R. H., Cowie, A., Favaloro, J. M., Jat, P. & Kamen, R. (1981). *J. Mol. Appl. Gen.* **1**, 83–92.
- Weaver, R. F. & Weissmann, C. (1979). *Nucl. Acids Res.* **7**, 1175–1193.
- Wells, R. D., Hutchinson, M. A. & Eckhart, W. (1979). *J. Virol.* **32**, 517–522.

Edited by P. Chambon