
Comparison of the polyoma virus early and late promoters by transcription *in vitro*

Parmjit Jat, Jeffrey W. Roberts*, Alison Cowie and Robert Kamen

Transcription Laboratory, Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields,
London WC2A 3PX, UK

Received 9 November 1981; Revised and Accepted 5 January 1982

ABSTRACT

Polyoma virus DNA was transcribed in the HeLa whole cell extract *in vitro* system (1). Early region transcripts with the same 5'-ends as *in vivo* mRNAs, located 31+2bp from 'TATA'-boxes, were synthesised by RNA polymerase II. Sequences sufficient for efficient expression of the early promoter were present in a substitution mutant lacking viral DNA from a position 55bp before the principal cap sites. Late region transcripts were synthesised inefficiently. Only one (at nt5129+2) of the many late mRNA cap sites functioned as an *in vitro* initiation point. This was the one 5'-end located 31+2bp from a sequence resembling the 'TATA' consensus. The proportion of late to early region RNA polymerase II transcripts decreased dramatically at suboptimal template concentrations. An hypothesis to explain the regulation of late gene expression *in vivo* based on these results is proposed. Although linear templates were transcribed only by RNA polymerase II, transcripts with the same sense as late mRNAs and 5'-ends at nt5076+2 were produced from superhelical templates by an α -amanitin resistant enzyme.

INTRODUCTION

The development of *in vitro* systems (1,2) which faithfully initiate the transcription of eucaryotic genes has facilitated analysis of the DNA sequence elements comprising transcriptional promoters in higher organisms (reviewed in refs. 3-4). We have been studying the transcription of polyoma virus (Py) DNA *in vivo* for a number of years. The genetic structure of Py DNA is now well understood (5), and the viral RNAs synthesized during productive infection and in transformed cell lines have been characterized in molecular detail (5-9). The viral DNA has two transcription units, the early and the late, extending in opposite directions from near the origin of DNA replication around the circular genome (Figure 1). The early transcription unit is expressed throughout lytic infection and in transformed cells. At late times of productive infection, its activity is negatively regulated by one of its gene products, the large T-protein (7,10). The principal capped 5'-termini of early region mRNAs map (see Figure 1) at several alternative sites within a

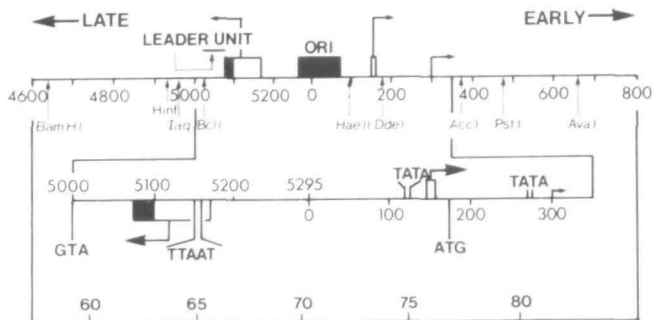


Figure 1. The 1.5kb of Py DNA spanning the replication origin (ORI). The upper coordinate is the nucleotide numbering system (modified [Tyndall et al [12]], from Soeda *et al* [11]) used in this paper. Restriction endonuclease cleavage sites relevant to this paper are shown. The bottom scale is standard map units (5,13). Locations of *in vivo* mRNA cap sites (9,14-18; A.C., P.J., & R.K., submitted for publication) are shown with arrows above the nucleotide number coordinate, and on an expanded scale below; the filled portion of the late cap site region demarcates the positions of more abundant 5'-ends. The "leader unit" is the DNA sequence amplified in the tandemly repeated leader structures of the late region mRNAs (14, 19-20). Positions of sequences related to the 'TATA' box consensus, and of the translational initiation codons (5), are indicated along the expanded nucleotide number scale.

short sequence at 73.3 mu (nt148-153 in the DNA sequence numbering system proposed by Soeda *et al*, [11]), but other less abundant 5'-termini, including one within coding sequences at 76.1 mu (nt300₊₂), have been reported (9; A.C. *et al*, submitted). The late transcription unit, by contrast, is not normally expressed in transformed cell lines, and functions during productive infection predominantly after the onset of viral DNA replication (5). It has two unusual features: the capped 5'-termini of late region mRNA are highly heterogeneous, with a minimum of 15 different purine cap sites (14-18; see Figure 1) localised in the 94 bp region from 66.4-68.1 mu (nt5075-5168); each late mRNA, moreover, has a tandemly repeated leader structure comprising multiple copies of the 57 bp region from nt5076-5020 (14). The leader repeat is thought to result from repeated splicing within the giant tandem transcripts of the entire viral genome which are the nuclear precursors of late mRNAs (17).

In this study we use the HeLa whole cell extract system (1) to compare the expression of the Py early and late transcription units *in vitro*. We find that early region *in vitro* products have the same 5'-termini as *in vivo* mRNAs. By contrast, late region *in vitro* transcripts have only one of the many 5'-ends characteristic of late mRNAs. The late transcription unit also

expresses very poorly in vitro, particularly at low DNA template concentrations. The implications of these results with regard to the regulation of viral gene expression are discussed.

METHODS

Preparation of viral DNA: Plaque purified polyoma virus (strain A2, 13) was used to infect 3T6 cells at a multiplicity of 10 pfu/cell. After 48-60 hours, cells were harvested and DNA prepared as described previously (21).

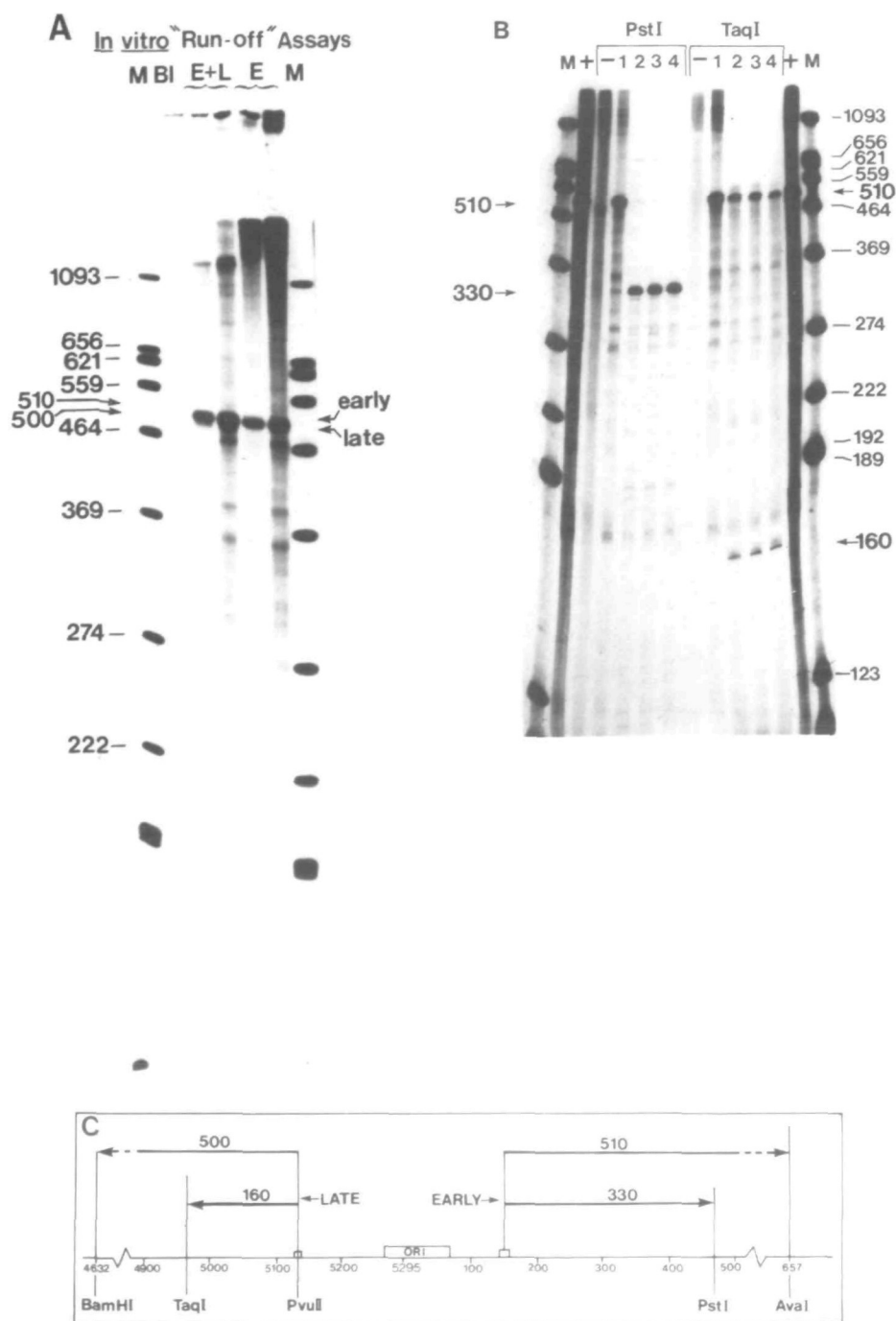
Recombinant plasmids: The vector in all cases was pAT153 (22). Plasmid p37.3.A2 contains wild-type Py strain A2 BamHI linear DNA inserted at the homologous vector site such that the standard nucleotide numbering systems for the viral and vector DNAs are opposed; p35.9.A2 contains the viral HaeII-EcoRI fragment (nt96-1565) between the vector HaeII site at nt547 and the EcoRI site; p45.8.A2 contains viral sequence from BamHI (nt4632) to EcoRI inserted at the homologous vector sites; pP15 (the gift of U. Novak) contains viral sequence from nt5131-1565 inserted between vector BamHI and EcoRI.

Preparation of HeLa cell extracts: HeLa cells were grown in suspension culture in Jocklicks medium supplemented with 10% fetal calf serum and glutamine to a density of approximately 5×10^5 cells per ml. Extracts were prepared by the Manley *et al* (1) procedure.

In vitro incubations and purification of RNA products: Standard 20 μ l reactions contained 12mM Hepes pH 7.9, 60mM KCl, 7.5mM MgCl₂, 0.06mM EDTA, 1.2mM DTT, 10% glycerol, 10mM creatine phosphate, 0.5mM unlabeled NTP's and 0.05mM labeled NTP (either α -³²P-UTP or GTP at 5 Ci/mMole), 12 μ l HeLa extract and 10-30 μ g/ml DNA. Preparative incorporations (either 100 or 200 μ l final volume) for S1 nuclease or primer extension analysis were done without labeled triphosphate. After incubation at 30°C for 60 minutes, reactions were terminated by adding an equal volume of a solution containing 100mM Tris.HCl pH 7.5, 2% (w/v) SDS, 20mM EDTA and 400 μ g/ml proteinase K (Merck). Reactions were incubated for a further 10 minutes at 37°C and deproteinised by extraction with phenol mixture (21). Unincorporated triphosphates were removed by two successive ethanol precipitations (23). For subsequent analysis by hybridisation, template DNA was hydrolysed by incubation with 20 μ g/ml pancreatic DNase I (purified as described by Favaloro *et al*, [21]) in 100 μ l of 10mM Tris.HCl pH 7.5, 10mM MgCl₂ for 30 minutes at 37°C followed by phenol mixture extraction and ethanol precipitation.

Analysis of ³²P-labeled in vitro products: ³²P-labeled RNAs were analysed directly by electrophoresis on ultra-thin (21) 5% polyacrylamide (19:1 acrylamide-methylene bisacrylamide) containing 50% w/v urea, after denaturation for 1-5 minutes at 90°C in 5 μ l of a loading buffer containing 80% w/v deionised formamide (Analar), 50mM Tris-borate pH 8.3, 1mM EDTA, 0.1% w/v bromophenol blue and xylene cyanol. Alternative analysis using hybridisation to cellulose-immobilised Py DNA fragments was done as described elsewhere (15,18).

S1 nuclease gel mapping (25-27): In vitro products were annealed to 5'-³²P-labeled single-stranded viral DNA fragments and the resulting hybrids were digested with 400 u/ml S1 nuclease for 2hrs at 12-14°C, as described previously (21). S1-resistant DNA products were fractionated on urea-polyacrylamide sequencing gels (24), specifications of which are described in the legends. The 5'-ends of early region transcripts were mapped using as probes the E-DNA strand of the BclI + DdeI fragment (labeled at nt188 and extending to nt5026) of Py DNA or an AccI fragment from recombinant plasmid p35.9.A2. The probe fragment (labeled at nt371 [the AccI site] in viral DNA sequence) extends to Py nt96 (the HaeII site) and then continues with 100



bases of plasmid sequence to the AccI site at nt650 in the vector. The 5'-ends of late region transcripts were mapped using as probe the L-DNA strand of the BclII + DdeI fragment of Py DNA (labeled at nt5022 and extending to nt185). Complementary strands of the DNA probes were separated on 4% polyacrylamide strand separation gels (23). The slower migrating strands were complementary to early mRNAs.

Reverse transcriptase primer extension: The modifications of the primer extension procedure (28) recently described (14) were used. The Py DNA fragment primer for late region transcripts was labeled with ^{32}P at nt4932 (a HinfI site) and ended at nt5021 (the BclI site).

RESULTS

Specific in vitro transcription of Py DNA: A Py DNA fragment expected to include both the early and the late region promoters (the BamHI to AvaI fragment, see Figure 1) was tested in the HeLa cell extract (1) using the run-off assay (2). High resolution fractionation of the products on a urea-polyacrylamide gel (Figure 2A, tracks "E+L") revealed a major 510 nucleotide and a minor 500 nucleotide RNA, as well as a variety of shorter species particularly prominent at higher template concentration. Transcription of viral DNA cut only at a site in the early region (the AvaI site shown in Figure 2C) resulted in the selective disappearance of the 500 nucleotide RNA, suggesting that it alone resulted from transcription towards the late region (Figure 2A, track "E"). Evidence substantiating this

Figure 2. Analysis of ^{32}P -labeled RNA synthesized *in vitro*. Panel A: Gel (5% polyacrylamide-urea, 0.3mm x 23cm x 43cm) fractionation of α - ^{32}P -UTP-labeled *in vitro* run-off products synthesised with no added template (B1), 10 or 25 $\mu\text{g}/\text{ml}$ of the BamHI to AvaI Py DNA fragment (E+L), and 10 or 25 $\mu\text{g}/\text{ml}$ of Py DNA restricted only with AvaI (E). Lanes "M" are 5'- ^{32}P -labeled DdeI fragments of Py DNA (lengths indicated to the left of the autoradiogram) used as size markers. Deduced lengths of major RNA products are indicated with arrows. Panel B: Hybridisation mapping of *in vitro* "run-off" transcripts. α - ^{32}P -GTP labeled RNA transcribed from AvaI cleaved Py DNA was hybridised to either a PstI fragment of recombinant plasmid pP15 (which includes the viral sequence clockwise from nt5131-484) or the TaqI fragment of Py DNA extending from nt4965-1316. The hybrid RNA samples eluted from the DNA celluloses were fractionated on an 8% polyacrylamide-urea gel. Lanes "M" are DNA fragment chain length markers. Lanes "+" are samples of the RNA prior to hybridisation, and correspond to darker exposures of the result shown in panel A, lane E. Lanes "-" are the material which did not hybridise to the viral DNA cellulose. The RNAs were eluted from the celluloses without ribonuclease T1 digestion (lanes 1), or after treatment with 5 units/ml for 30 minutes at 37 $^{\circ}$ (lanes 2), 5 units/ml at 20 $^{\circ}$ (lanes 3), or 1 unit/ml at 20 $^{\circ}$ (lanes 4). Marker lengths are indicated on the right, and the calculated lengths of RNA products are shown by arrows. Panel C: Diagram aligning the RNA products indicated in panels A and B with the nucleotide sequence of the origin region of Py DNA. The mutually consistent data position one end of the early region transcript near nt150 and one end of the relatively minor late region transcript near nt5130.

conclusion and positioning both run-offs more accurately was obtained by hybridising the in vitro products to viral DNA fragments immobilized on cellulose. Greater than 95% of the 510 nucleotide RNA annealed to a DNA fragment extending from nt5130 clockwise to the PstI site at nt484 (Figure 2B, Pst track 1). Digestion of the hybrids with ribonuclease T1 to remove single stranded regions shortened it to 330 nucleotides (Figure 2B, Pst tracks 2-4). This suggested that it was an early region run-off product initiated in the vicinity of the principal in vivo early region cap sites at nt150 (Figure 2C). Since the minor RNAs also visible in Figure 2B (Pst track 1) were not truncated by the ribonuclease digestion, some of them must result from premature termination and/or RNA cleavage rather than from initiation at other sites. The 510 nucleotide RNA, as well as a product longer than 1 kb, hybridised to the TaqI fragment from nt4965-1316. Ribonuclease treatment of these hybrids did not alter the size of the 510 nucleotide RNA but truncated the larger transcript to 160 bases (Figure 2B, TaqI 2-4). As illustrated in Figure 2C, these run-off data indicate that a minor late region transcript initiated near nt5130, as well as the major early region transcript from near nt150, are synthesized in vitro.

Accurate mapping of 5'-ends of in vitro transcripts: The 5'-ends of RNAs synthesised in vitro were more accurately localised by S1 nuclease gel mapping (9, 25-27). Figure 3A shows representative results mapping the 5'-ends of early region transcripts with a complementary probe derived from a recombinant plasmid in which viral sequences upstream of a position 55 bp before the principal in vivo cap sites were replaced with plasmid DNA (see Methods). The DNA template used for transcription in vitro, in the first instance, extended from BamHI (nt4632, 810 bp before the in vivo cap sites) to AvaI (nt657). The S1-resistant DNA products obtained with the RNA synthesised in vitro were nearly identical to those protected by mRNA extracted from Py infected mouse cells (Figure 3A, -810 tracks and in vivo tracks). They map principal 5'-termini at nt145-155 and minor termini at nt300+2. The DNA product ca 275 nucleotides long, detected with both in vitro and in vivo RNAs, results from the viral/plasmid DNA junction in the probe sequence; its presence indicates only that some minor RNA molecules extend 5' to nt95 in the viral DNA sequence. The DNA product ca250 bases long is an S1 nuclease artifact described previously (9). The complete inhibition of in vitro transcription by low levels of α -amanitin, demonstrated in Figure 3, shows that the active enzyme is RNA polymerase II (30). We also transcribed in vitro the plasmid DNA in which viral sequences from 55 bp upstream of the cap sites were

replaced with vector DNA; this substitution had no detectable qualitative or quantitative effect (Figure 3A, - 55 tracks).

We next compared the early region 5'-ends of in vitro RNA with those of in vivo transcripts in further detail by using a shorter DNA probe (Figure 3B). The results suggested that RNAs synthesised in vitro have the same heterogeneous distribution of 5'-termini as the viral RNA extracted from nuclei of infected cells. This distribution, however, is slightly different from that of in vivo polyadenylated cytoplasmic RNA. Exact assignment of 5'-termini cannot be made because of the usual ragged DNA overhang of 1-5 bases found in S1 mapping (26,31) as well as the possibility of overdigestion through the AU or TA base pairs from nt153-156 in the DNA sequence. Direct RNA analysis (A.C. et al, submitted), however, showed that the in vivo and in vitro transcripts both had capped 5'-ends located at multiple points from nt 148 to 153.

The 5'-termini were also localised by the primer extension (28) method. These results (data not shown) confirmed that early region transcripts synthesised in vitro, like those present in virus infected cells, have principal 5'-ends in the nt148-153 region and minor 5'-ends at nt300+2.

The late region transcripts synthesised in vitro were accurately mapped with an L-DNA strand probe labeled at nt5022 (the BclI site) and extending to nt185 (a DdeI site). The S1 nuclease mapping results shown in Figure 4A demonstrate that the 5'-ends of greater than 95% of in vitro transcripts synthesised at the optimal template concentration (25µg/ml tracks in Fig. 4A; cf Fig. 6) were at nt5129+2. The S1 products mapping this terminus occurred as a doublet, but results of primer extension analysis (see below) suggested that there was only one 5'-end at this position. Other minor S1-resistant DNA products detectable, particularly at higher DNA template concentrations and when the probe used in the S1 analysis was saturated (Figure 4A, 37.5µg/ml tracks). Most of these corresponded to internal S1 nuclease cleavages, particularly within U-rich regions of hybrids, and were also seen with complementary RNA synthesised in vitro by E.coli RNA polymerase (cRNA track in Fig. 4B). Figure 4A further shows that late region transcription is inhibited by low levels of α -amanitin. When supercoiled instead of linear DNA was transcribed, an additional 5'-terminus at nt5076+2, was detected (Figure 4B). This 5'-end corresponds exactly to the 5'-end of the reiterated leader segments of cytoplasmic mRNA (Figure 4B). Whereas the generation of RNA with 5'-ends at nt5129+2 was inhibited by either 1µg/ml or 100µg/ml of α -amanitin, the synthesis of molecules with 5'-ends of colinear sequence at 5076+2 was

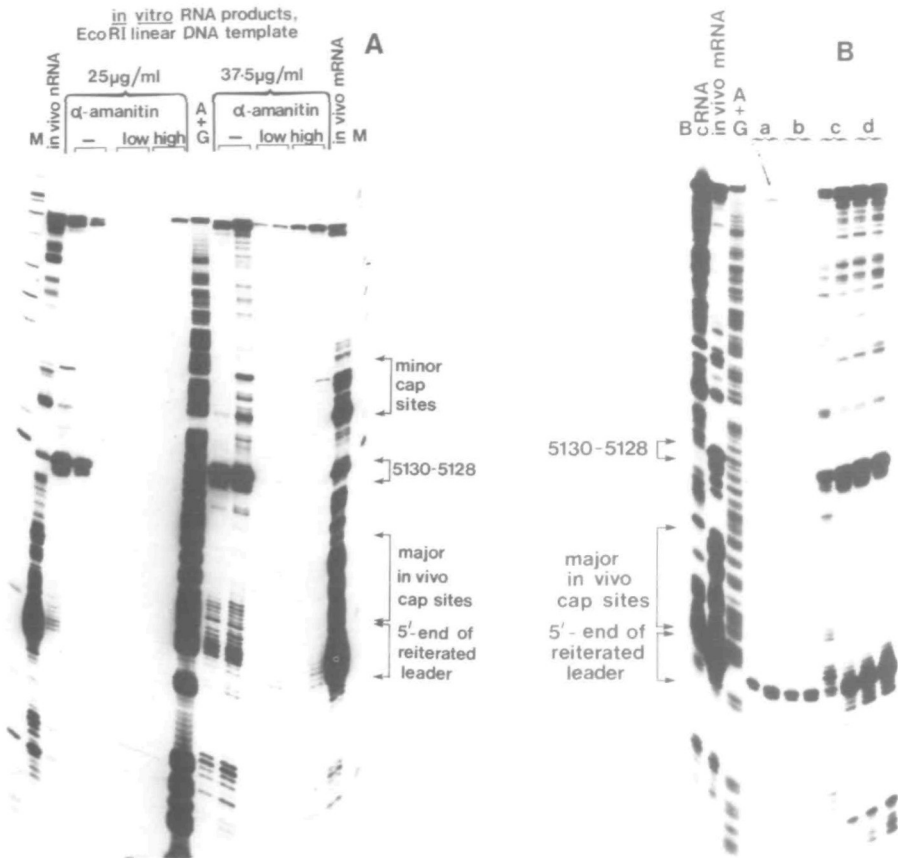


Figure 4. Localisation of 5'-ends of late region *in vitro* transcripts by S1 nuclease gel mapping. Panel A: 0.8ng of *Bcl*I + *Dde*I fragment (2.0µCi/pmole) was hybridised to RNA transcribed *in vitro* from the indicated concentration of EcoRI linear Py DNA templates, without ("–") or with 1 µg/ml ("low") or 100 µg/ml ("high") α-amanitin. The pairs of tracks are different amounts of RNA hybridised, which corresponded to 10 µl or 20 µl *in vitro* transcription reactions (the lower quantity is the lefthand track of each pair, except for the "–" tracks at 25µg/ml where it is the righthand track). The deduced position of the 5'-termini on the genomic sequence are indicated. Panel B: 1.5ng of *Bcl*I + *Dde*I fragment (2.0µCi/pmole) was hybridised to the following RNA samples: tracks a and b, RNA transcribed from 75µg/ml of uncut plasmid p37.3.A2 (see Methods) with 1 and 100µg/ml α-amanitin in amounts corresponding to 12.5 and 25µl transcription reactions; tracks c and d are RNA transcribed from uncut plasmid p45.8.A2 (see Methods); cRNA track, is complementary RNA synthesised by transcribing Py DNA with *E.coli* RNA polymerase under conditions where the L-DNA strand is preferentially transcribed (29). This RNA proved to be highly heterogeneous, but the result is included to demonstrate the S1 nuclease cleavages within the oligo U sequences just 5'-to the leader unit. "In vivo mRNA" track shows the S1 result obtained after hybridisation to cytoplasmic polyadenylated RNA from infected mouse cells.

insensitive to the drug (Figure 4B, tracks a and b).

The position and template dependence of the late region 5'-termini was further assayed by the primer extension (28) method. These results confirmed the conclusions of S1-mapping experiments, namely that the principal 5'-termini of RNA synthesised from supercoiled templates are at nt5129+2 and 5076+2 (Figure 5, tracks a,b,d & e) whereas those at 5129+2 alone are synthesised from linear templates (Figure 5, track c). As the 5'-termini at 5076+2 were detected by primer extension, they probably represent an RNA polymerase I or III initiation site rather than a 5'-end of colinear sequence caused by RNA splicing. The data shown do not exclude an abortive splicing event involving cleavage but not ligation. However, efforts to detect the predicted 3'-end were unsuccessful. The biological significance of an *in vitro* initiation site for an enzyme other than polymerase II is unknown. Condit et al (32) purified transcription complexes from the nuclei of infected cells and found that >98% of the *in vitro* elongation activity could be inhibited by 0.2µg/ml α -amanitin. However, as RNA polymerase III (unlike RNA polymerase II) can terminate transcription *in vitro* (33,34), a transcription unit producing a small RNA might not have been detected. We think that it is more likely that the *in vitro* result is an artifact resulting from initiation in a readily denaturable AT-rich region of superhelical templates (the 23bp from nt5074-5096 are 78% AT, and the 16bp from 5076-5091 are 87.5%).

Relative efficiency of transcription from the early and late promoter regions:

The initiation of transcription at various template levels was determined by quantitating the yield of S1-resistant hybrids resulting from hybridisation of RNA to separated strands of the BclI+DdeI fragment. As shown in Figure 6, RNA synthesis from both the early and late promoter regions was saturated at template concentrations higher than 7×10^{-9} M (25µg/ml) and, under these conditions, transcripts initiated at the principal early region cap sites were 3 times more abundant than those of the late region. A four fold decrease in the template concentration below the saturation level increased this proportion to 10.5; further decrease in the template concentration markedly exaggerated this differential effect. At template concentrations below 1.4×10^{-9} M (5µg/ml), gel bands corresponding to initiation in both the early and late regions were clearly visible on the autoradiogram after exposure for 7 days, but there was insufficient radioactivity in the S1-resistant products from the late region for quantitation in the scintillation counter. We conclude that the promoter in the late region is inherently less active than the early region principal promoter, particularly at low template DNA

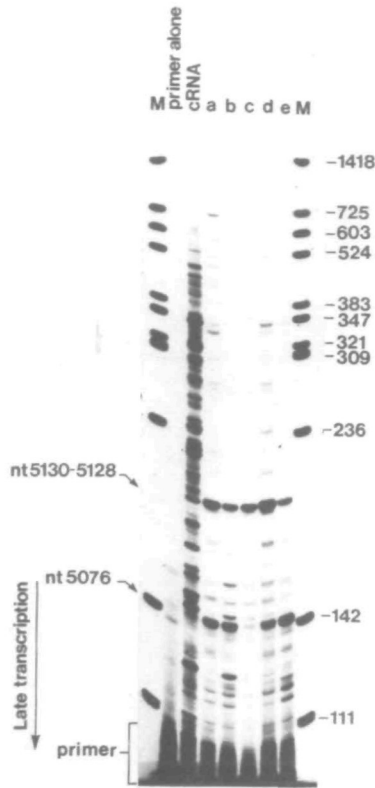


Figure 5. Localisation of 5'-ends of late region transcripts by primer extension. Each hybridisation contained 8,000 dpm of *Hinf*I-*Bcl*I fragment and the following RNA samples: tracks a, b and d, RNA transcribed from uncut plasmid p37.3.A2. Track c, RNA transcribed from an *Eco*RI linear plasmid p37.3.A2 template. Track e, RNA transcribed from uncut plasmid p45.8.A2. The deduced position of the major bands on the genomic sequence are indicated by arrows. The results demonstrate that transcripts with 5'-ends at nt5076+2 are synthesised only with circular templates added to the reactions as supercoils.

concentrations.

DISCUSSION

The experiments described in this paper used the crude HeLa cell *in vitro* system (1) to characterise and compare the expression of the polyoma virus early and late region promoters. In interpreting our results, we shall assume that 5'-ends of *in vitro* transcripts correspond to transcriptional initiation points. This has been shown with other transcription units (35);

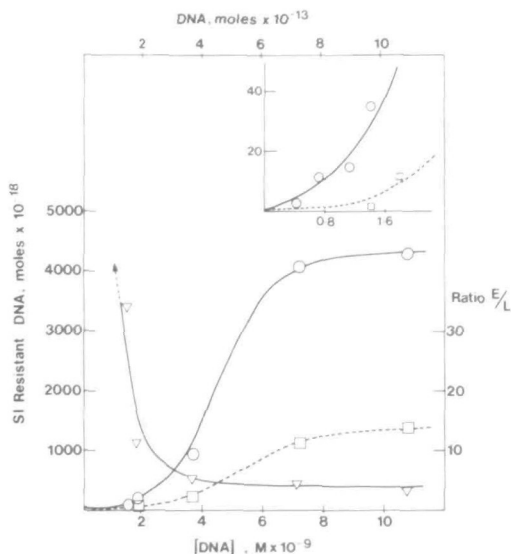


Figure 6. The effect of template concentration on initiation of transcription from the early and late promoter regions. RNA synthesis from the early and late regions was determined by quantitating the yield of S1 resistant DNA products after hybridisation of RNA synthesised at varying template concentrations to the separated strands of the *Bcl*I + *Dde*I fragment as described above. The S1 resistant hybrids were size fractionated and after a period of autoradiography, appropriate parts of the gel were excised and the yield of S1 resistant DNA determined by Cerenkov counting. The yield of S1 resistant DNA obtained with RNA corresponding to 100 μ l reactions at different template concentrations from the early region (o-o) and late region (\square - \square) are shown; the ratio of transcription (E/L) from the early and late regions at the different template concentrations was determined (∇ - ∇). Results at the lower template levels are shown in the insert on expanded scales.

we present elsewhere evidence directly demonstrating the principal for Py early region transcripts (A.C. *et al.*, submitted).

The Py early region promoter closely resembles that of cellular genes. It functions throughout productive infection and when the viral DNA is integrated into the genome of transformed cell lines (8). The DNA preceding the principal early mRNA cap sites includes sequences homologous to the Hogness-Goldberg 'TATA' consensus and to the 'CAAT' box (3,4). The principal cap sites are somewhat heterogeneous, but they are all at positions 31 \pm 3bp from the beginning of the 'TATA' box (9). The minor 5'-end of *in vivo* mRNAs at nt 300 \pm 2, although occurring within coding sequences, is also at the usual distance from 'TATA' and 'CAAT' elements (9). The results reported in this paper demonstrate that *in vitro* transcripts of the early region have 5'-ends

corresponding accurately, within the precision of the methods used, to the principal and minor cap sites of in vivo mRNAs. Microheterogeneity at the principal cap sites very similar to that noted in vivo was also found. A substitution mutant template, in which vector DNA replaced viral sequence from a point 55bp upstream of the principal cap sites, functioned like wild-type viral DNA in vitro. This implied that sequences sufficient for in vitro RNA chain initiation are located proximal to the cap sites; these will be analysed in more detail separately (P.J., U. Novak, A.C., C. Tyndall and R.K., submitted for publication). A rather different result was obtained in vivo. Tyndall et al (12) demonstrated that a remote upstream region [between approximately nt 5030-5200] contains sequences essential for early gene expression. A 246 bp fragment of Py DNA including this region, inserted into a recombinant plasmid containing rabbit β -globin genes, dramatically stimulated the in vivo expression of the globin promoters in transient assays (36). This effect was analogous to the activity of the SV40 72bp repeat sequence, which has been shown to enhance the expression of several eucaryotic promoters in an orientation independent manner and to act over considerable distances (37,38). The present results would suggest that "enhancers" are not important for transcription in vitro.

The Py late "promoter" is unusual. We have identified in vivo (14-18) a minimum of 15 different cap sites within the 94bp (nt5168-5075) proximal to the DNA sequence ("leader unit" in Figure 1) determining the repeated leader of the late mRNAs (13). An AT-rich tract (TAATTA AAA, nt5158-5150), possibly related to the 'TATA' consensus, is included within this region, but only one of the many alternative mRNA 5'-ends (at nt5129+2; 13) is at the usual distance from it. A proportion of Py minichromosomes (10-20%) in productively infected cells have an apparently nucleosome-free domain hypersensitive to DNase I (39) which spans the heterogeneous late mRNA cap sites. The results presented here showed that the initiation site at nt5129+2 is almost uniquely recognised by the in vitro system. The other, apparently 'TATA'-box independent, cap sites of in vivo mRNAs did not function as initiation sites in vitro. This is consistent with the inactivity in vitro of several other genes lacking 'TATA'-boxes (40-42). Utilisation of the one functional late region site was less efficient (ca 30% at saturating template levels) than initiation at the principal early region promoter. When relative promoter efficiency was measured as a function of DNA concentration, we found that the proportion of late to early region initiations dramatically decreased at suboptimal template levels. A similar effect has been reported for SV40 early

and late promoters in vitro (43), but here the initiation sites involved were not accurately positioned.

These results suggest a model for the regulation of viral transcription in vivo. Our hypothesis is based on two assumptions. The first is that "enhancer" sequences are, as suggested by Moreau et al (38), chromatin RNA polymerase II entry sites. Such sites must be bi-directional because the SV40 and Py (36-38) "enhancers" function independent of their orientations with respect to the transcription unit enhanced. Having bound to the DNA in chromatin at the "enhancer" element, the polymerase would slide along the DNA in either direction, scanning the sequence, like E.coli RNA polymerase (44,45), for other promoter elements, including the 'TATA' box, which function to specify RNA chain initiation at particular nucleotides. However, initiation can occur at a variety of other positions, with low relative efficiency, because deletion mutants lacking the sequence elements proximal to normal start sites still function in vivo (46-49). In the Py early region, the alternative start sites functional in such deletion mutants in fact correspond to sites used very occasionally in wild-type DNA (9). Moreau et al (38) also noted that the SV40 72bp repeat can enhance transcription in vivo in the absence of any other known promoter elements. Efforts to reproduce the in vivo requirements for "enhancer" sequences in vitro have in general been unsuccessful (50, see above), suggesting that RNA polymerase II entry is not a rate limiting step during the transcription of purified DNA templates added to cell free systems. The second assumption is that the DNaseI hypersensitive domain (39) indeed represents a particularly accessible region of the viral DNA, the occurrence of which is not a consequence of transcriptional activity. Although this domain includes the Py "enhancer", as does the corresponding region in SV40 DNA (37,38,51), it may be an independent element because the SV40 enhancer function (the 72bp repeat) can be transposed to other positions in the DNA molecule without the generation of a second DNase I hypersensitive domain (P. Chambon, personal communication).

To explain the known in vivo and in vitro properties of the Py late "promoter", we propose that RNA polymerase II molecules bind to viral chromatin at the "enhancer", which overlaps (12,18) the late mRNA cap site region, and then slide in either direction. Those reaching the principal early promoter region 'TATA' box stop and initiate transcription. The homologous element in the late region, which inefficiently specifies starts at nt5129+2 in vitro, is only recognized by a minority of polymerases traversing it. We predict that the initiation site at nt5129+2 should function, albeit,

at a very low rate, both at early times of infection and in transformed cells. Other polymerases scanning the beginning of the late region initiate *in vivo* at a multitude of points. These lie principally at the late region limit of the DNase I hypersensitive domain because further sliding is inhibited by the first nucleosome encountered. Initiation at these heterogeneous sites, like initiation at the 'TATA'-box independent start sites of deletion mutants, is so inefficient that it is hardly detectable *in vitro*. At late times of infection, at least two changes may occur. Large T-protein binding to the origin region (52), and possibly to the principal early cap site region (9), may block polymerase migration towards the early region. The DNA template concentration increases dramatically, which we have shown, *in vitro*, to stimulate expression of the late region promoter. Thus there is an apparent activation of late region transcription because an abundance of polymerase molecules enter and scan a limited region including definite, but inherently inefficient, transcriptional initiation sites. This interpretation is obviously speculative, but it explains the known phenomena and is testable using methodology currently available.

ACKNOWLEDGEMENTS

We are grateful to Richard Treisman for helpful discussion, Kit Osborne for cell culture, Jennifer Favalaro for technical assistance, and to Penny Morgan, Gina Yiangou and Audrey Symons for their help in the preparation of this manuscript.

*Permanent address: Cornell University, Section of Biochemistry, Molecular and Cell Biology, Wing Hall, Ithaca, NY 14853, USA

REFERENCES

1. Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gelfand, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 3855-3859.
2. Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G. (1979) Cell 18, 469-484.
3. Breathnach, R. and Chambon, P. (1981) Ann. Rev. of Biochem. (In press).
4. Shenk, T. (1981) Curr. Topics Microbiol. Immunol. 93, 25-46.
5. Toozé, J. (ed.) (1980) DNA Tumor Viruses, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, part 2.
6. Treisman, R.H., Cowie, A., Favalaro, J.M., Jat, P. and Kamen, R. (1981) J. Mol. Appl. Gen. 1, 83-92.
7. Kamen, R., Favalaro, J. and Parker, J. (1980) J. Virol. 33, 637-651.
8. Kamen, R.I., Favalaro, J.M., Parker, J.T., Treisman, R.H., Lania, L., Fried, M. and Mellor, A. (1980) Cold Spring Harbor Symp. Quant. Biol. 44, 63-75.
9. Kamen, R., Jat, P., Treisman, R., Favalaro, J. and Wolk, W. (1981) J.

- Mol. Biol., in press.
10. Cogen, W. (1978) *Virology* 85, 222-230.
 11. Soeda, E., Arrand, J.R., Smolar, N., Walsh, J.E. and Griffin, B.E. (1980) *Nature* 283, 445-453.
 12. Tyndall, C., La Mantia, G., Thacker, C. and Kamen, R. (1981) *Nuc. Acids Res.* 9, 6231-6250.
 13. Griffin, B.E., Fried, M. and Cowie, A. (1974) *Proc. Natl. Acad. Sci. USA* 71, 2077-2081.
 14. Treisman, R.H. (1980) *Nucl. Acids Res.* 8, 4867-4888.
 15. Flavell, A.J., Cowie, A., Legon, S. and Kamen, R. (1979) *Cell* 16, 357-371.
 16. Flavell, A.J., Cowie, A., Arrand, J.R. and Kamen, R. (1980) *J. Virol.* 33, 902-908.
 17. Treisman, R. and Kamen, R. (1981) *J. Mol. Biol.* 148, 273-301.
 18. Cowie, A., Tyndall, C. and Kamen, R. (1981) *Nuc. Acids Res.* 9, 6305-6322.
 19. Legon, S., Flavell, A.J., Cowie, A. and Kamen, R.I. (1979) *Cell* 16, 373-388.
 20. Zuckerman, M., Manor, H., Parker, J. and Kamen, R. (1980) *Nucl. Acids Res.* 8, 1505-1519.
 21. Favaloro, J., Treisman, R. and Kamen, R. (1980) *Meth. Enzymol.* 65, 718-748.
 22. Twigg, A.J. and Sherratt, D. (1980) *Nature* 283, 216-218.
 23. Maxam, A. and Gilbert, W. (1980) *Meth. Enzymol.* 65, 499-560.
 24. Sanger, F. and Coulson, A.R. (1978) *FEBS Lett.* 87, 107-110.
 25. Berk, A.J. and Sharp, P.A. (1977) *Cell* 12, 721-732.
 26. Sollner-Webb, B. and Reeder, R.H. (1979). *Cell*, 18, 485-499.
 27. Weaver, R.F. and Weissman, C. (1979) *Nucl. Acids Res.* 6, 1175-1193.
 28. Ghosh, P.K., Reddy, V.B., Piatak, M., Lebowitz, P. and Weissman, S.M. (1980) *Meth. Enzymol.* 65, 580-594.
 29. Kamen, R., Sedat, J. and Ziff, E. (1976) *J. Virol.* 17, 212-218.
 30. Roeder, R.G. (1976) in *RNA Polymerases* (Losick, R. and Chamberlin, M. eds.) pp.285-330 Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
 31. Hentschel, C., Irminger, J.-C., Bucher, P. and Birnstiel, M.L. (1980) *Nature* 285, 147-151.
 32. Condit, R.C., Cowie, A., Kamen, R. and Birg, F. (1977) *J. Mol. Biol.* 115, 215-235.
 33. Bogenhagen, D.F., Sokonju, S. and Brown, D.D. (1980) *Cell* 19, 27-35.
 34. Weil, P.A., Segall, J., Harris, B., NG S.Y., and Roeder, R.G. (1979) *J. Biol. Chem.* 254, 6163-6173.
 35. Hagenbuchle, O. and Schibler, U. (1981) *Proc. Natl. Acad. Sci. USA*, 78, 2283-2286.
 36. De Villiers, J. and Schaffner, W. (1981) *Nuc. Acids Res.* 9, 6251-6264.
 37. Banerji, J., Ruseoni, S. and Schaffner, W. (1981) *Cell*, in press.
 38. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. (1981) *Nucl. Acids Res.* 9, 6047-6068.
 39. Herbomel, P., Saragosti, S., Blangy, D. and Yanfu, M. (1981) *Cell* 25, 651-658.
 40. Luse, D.S., Haynes, J.R., Vanleeuwen, D., Schon, E.A., Cleary, M.L., Shapiro, S.G., Lingrel, J.B. and Roeder, R.G. (1981) *Nucl. Acids Res.* 9, 4338-4354.
 41. Talkington, C.A., Nishioka, Y. and Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* 77, 7132-7136.
 42. Lee, D.C. and Roeder, R.G. (1981) *Mol. Cell. Biol.* 1, 635-651.
 43. Rio, D., Robbins, A., Myers, R. and Tjian, R. (1980) *Proc. Natl. Acad.*

-
- Sci. USA 77, 5706-5710.
44. Bujard, H. (1980) Trends in Biochem. Sci. 5, 274-278.
 45. Gabain, A.V. and Bujard, H. (1979) Proc. Natl. Acad. Sci. USA 76, 189-193.
 46. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 1432-1436.
 47. Benoist, C. and Chambon, P. (1981) Nature 290, 304-309.
 48. McKnight, S.L., Gavis, E.R., Kingsbury, R. and Axel, R. (1981) Cell 25, 385-398.
 49. Dierks, P., van Ooyen, A., Mantei, N. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA 78, 1411-1415.
 50. Mathis, D.J. and Chambon, P. (1981) Nature 290, 310-315.
 51. Saragosti, S., Hoyne, G. and Yaniv, M. (1980) Cell, 20, 65-73.
 52. Gaudray, P., Tyndall, C., Kamen, R. and Cuzin, F. (1981) Nucl. Acids Res., in press.

