Volume 9 Number 23 1981 Nucleic Acids Research

A region of the polyoma virus genome between the replication origin and late protein coding sequences is required in cis for both early gene expression and viral DNA replication

Chiara Tyndall, Girolama La Mantia+, Colin M.Thacker, Jennifer Favaloro and Robert Kamen

Transcription Laboratory, Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX, UK

Received 2 November 1981

ABSTRACT

Deletion mutants within the Py DNA region between the replication origin and the beginning of late protein coding sequences have been constructed and analysed for viability, early gene expression and viral DNA replication. Assay of replicative competence was facilitated by the use of Py transformed mouse cells (COP lines) which express functional large T-protein but contain no free viral DNA. Viable mutants defined three new nonessential regions of the genome. Certain deletions spanning the PvuII site at nt 5130 (67.4 mu) were unable to express early genes and had a cis-acting defect in DNA replication. Other mutants had intermediate phenotypes. Relevance of these results to eucaryotic "enhancer" elements is discussed.

INTRODUCTION

The genetic organization of polyoma virus (Py) DNA is well characterised (1-4). Far less is known about sequence elements which regulate the expression of the viral genes and the replication of the viral DNA. Such regulatory signals most likely occur within the non-coding sequences shown diagramatically in Figure 1. This region is already known to include (1) the origin of viral DNA replication (5-7), probably located within a palindromic sequence highly conserved among papovavirus DNAs (8, "SV40 homology" in Fig. 1). (2) A region with high affinity for the Py large T-protein, containing the sequence AGAGGC three times (9). (3) The sequences specifying the capped 5'-ends of both early and late mRNAs, and the promoter elements involved in their transcription (10-13). (4) The sequence determining the repeated leader of the late mRNAs (11; "leader unit"). (5) A domain hypersensitive to DNase I in viral chromatin ("DHSR", 14).

Non-essential regions within this portion of the genome have been identified by the isolation of viable deletion mutants. NER-E1 is eleven base pairs within the repeated sequences which comprise the high affinity binding site for the large T-protein. Strain A2 virus has three AGAGGC sequences,

while strain A3 virus has only two (15). A large number of viable deletions have been obtained between the origin region and the beginning of the early coding sequences (16-19); these define the limits of NER-E2.

We have been particularly interested in the region designated "A" in Figure 1, between the replication origin and the beginning of late protein coding sequences. Region A includes the late region transcriptional promoter elements, the late mRNA leader sequence and the DNaseI hypersensitive domain. Within it are also the sequences tandemly duplicated in certain Py variants which, unlike wild type virus, are able to grow in undifferentiated embryonal teratocarcinoma cell lines (20-23, see legend to Fig. 1). We report here the construction and characterisation of a number of mutants lacking sequence from region A. The viability of certain mutants defined the three nonessential regions, NER-L1, NER-L2 and NER-L3 (Fig. 1). Results of experiments in which the phenotypes of nonviable mutants were determined using assays which separately measured viral DNA replication and early gene expression identified a sequence within region A absolutely required, in cis, for both functions. Insertion of Py region A into plasmids also containing rabbit β-globin genes dramatically enhanced the expression of these genes from their own promoters (J. de Villiers and W. Schaffner, submitted for publication). Its relationship to other eucaryotic "enhancer" sequences (24-28) is discussed.

METHODS

DNA replication assays: Subconfluent cultures (50mm) of 2^0 mouse embryo fibroblasts or Py transformed mouse cells (COP cells, see Results) were transfected with up to 100ng of recombinant plasmid DNA by the DEAE-Dextran method (31). When the viral DNA inserts were excised from the vector before transfection, Py anti-serum was included in the culture medium to prevent reinfection. Viral DNA was extracted 72 hrs post-infection by the selective SDS method (32). After ribonuclease treatment and deproteinization, the DNA was digested with restriction endonuclease MboI (Biolabs). The digests were fractionated by agarose gel electrophoresis, transferred to nitrocellulose (33) and probed with nick-translated recombinant DNA containing the entire viral genome inserted into the BamHI site of plasmid pAT153 (34).

Gene expression assays: HeLa, F2408 rat1 (35), or rat2 (the gift of W. Topp) cells (90mm cultures) were transfected with 20μg of recombinant DNA by the calcium phosphate method (36) as modified by Banerji et al (27). At 60-72 hours post-transfection, cells on coverslips were stained for Py large T-protein by indirect immunofluorescence (37) using tumour-bearing rat serum kindly provided by S. Dilworth (ICRF), or cytoplasmic RNA was extracted (38) for subsequent hybridisation analysis. To measure early region protein synthesis, the cultures were labelled with ^{35}S-methionine (0.5mCi per culture) for three hours. Proteins were extracted, immunoprecipitated, and separated on 15% SDS-polyacrylamide gels as described by Ito et al (37).

Construction of deletion mutants around PvuII sites by transfection of mouse cells with shortened linear viral DNA: DNA of d128, a viable Py mutant (16)

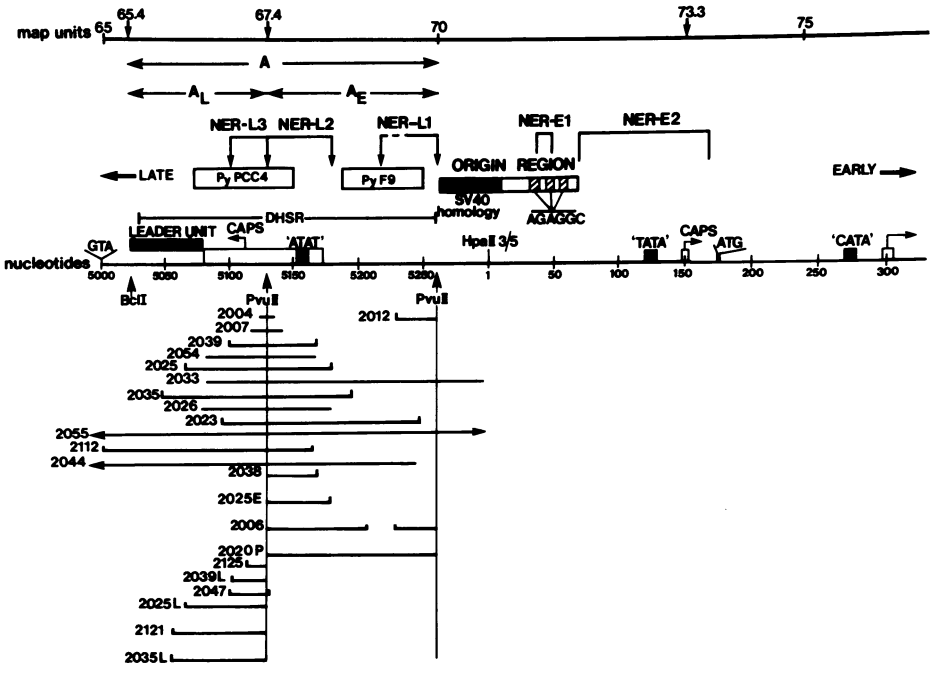


Figure 1. The positions of deletion mutants described in this paper relative to landmarks within the 600 bp of Py DNA spanning the replication origin. Coordinates are given in approximate map units (top line, 29) and in nucleotide number (middle line, 4). The directions of transcription of the early and late regions (1) are shown by arrows. Locations of mRNA cap sites (12, R.K., P. Jat, R. Treisman & W. Folk, submitted for publication), Hogness-Goldberg (30) boxes, and the translational initiation codons (2-4) are shown along the nucleotide number coordinate. "Leader unit" indicates the position of the sequence tandemly repeated near the 5'-termini of late mRNAs (11). "DHSR" indicates the DNaseI hypersensitive region of viral chromatin (14). The locations of DNA sequences tandemly repeated in Py variants able to grow in undifferentiated embryonal teratocarcinoma cells are shown by the boxes around "PyPCC4" and "PyF9" (20-23). Other PyF9 variants (21,23) have only a single base change at nt5233 (see Fig. 3). NER's (nonessential regions) define the limits of sequences deleted among different viable mutants (16-19 and this paper; the dashed line in NER-L1 indicates that the origin-distal limit is not yet defined. Two PyPCC4 variants (20) have in addition to a tandem duplication, a deletion within NER-L1. However, it is not known whether virus with the deletion but not the duplication would be viable.) "Origin Region" demarcates the limits of continuous sequence required for viral DNA replication, and includes the high affinity binding site for the large-T protein (comprised of the sequence AGAGGC repeated three times in the A2 strain, [9]) and a palindromic region homologous to the SV40 origin sequence (8). Lines below the nucleotide number coordinate show the deletions in the indicated mutants; those with vertical bars at their ends have been sequenced. Deletions 2004-2007, and 2026 are shown centred on the PvuII site because their extents into region A_L and A_E are not known. Large deletions extending beyond the region shown are represented with arrows.

lacking sequences around the BglI site (within NER-E2, see Fig. 1), kindly provided by B. Griffin (ICRF), was digested with PvuII under conditions (30µg/ml form I DNA, 20µg/ml ethidium bromide [39], 95 units/ml enzyme) empirically determined to yield maximal proportions of single-cut linear molecules. These were slightly shortened by sequential treatment with T4 DNA polymerase (in the presence of 10µM dTTP) and nuclease S1. Linear molecules of nearly full length were purified by low melting point agarose gel electrophoresis. The excised gel band was melted and diluted with at least ten volumes of phosphate buffered saline. Serial dilutions containing ca 0.2-200ng of linear DNA were used to transfect 2^0 mouse embryo cells, with and without 500ng of tsa (40) helper DNA. Plaques were picked after 8-10 days of incubation at 37^0 (no helper) or 39^0 (plus helper) and screened for those containing deletion mutants by restriction enzyme analysis of small scale ^{32}P-labelled viral DNA preparations (41).

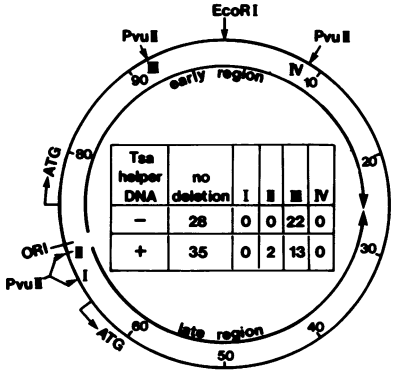
Construction of deletion mutations in cloned viral DNA: Full length Py strain A2 DNA was inserted into plasmid vector pAT153 (34) at the unique BamHI sites. After partial PvuII digestion (see above), XhoI linkers were added by blunt end ligation. A plasmid with a unique XhoI site at 67.4mu (nt 5130) was selected (p 43.2b.67). This DNA was cleaved with XhoI, treated with excess amounts of nuclease Bal31 for different times (30" to 5'), XhoI linkers were added back, the DNAs were recircularized and cloned into E.coli HB101. Single ampicillin resistant colonies were picked, grown into 5ml liquid cultures, and 1ml volumes were used for plasmid isolation using a modification (D. Ish-Horowitz, personal communication) of the Birnboim and Doly method (42). The locations of deletions were determined by restriction with XhoI plus each of the following enzymes: BamHI, PvuII, AccI and SacI. Plasmids of interest (pdl2000 series) were grown on large scale (400ml) and purified for biological testing. The deletions in molecules which proved important were chemically sequenced (43) from a 5'-label at the PvuII site (70mu, nt 5264) after secondary cleavage with KpnI (59.2mu, nt 4693) for dl2025, dl2035, dl2038, dl2039 and dl2121. Certain of the deletions which extended bidirectionally from the original XhoI site were used to construct unidirectional mutants by ligation of XhoI + ClaI fragments (ClaI cuts once in pAT153 but doesn't cut Py DNA) from the mutant plasmids to the appropriate XhoI + ClaI fragment of the parental plasmid p43.2b.67. The unidirectionals were called either L or E to indicate whether the deletion extended towards the late (L) or early (E) region.

A further deletion mutant (dl2020P) lacking the small PvuII fragment from 67.4-70.0mu was constructed, starting with a plasmid [pG20, the gift of J. Jenkins and A. Cowie] in which the Py SacI fragment from 52.1-81.3mu was inserted into the PstI site of pAT153 with GC tails. pG20 was cleaved with PvuII, circularised at low concentration, and recloned to yield pG20P, lacking the small PvuII fragment. The viral BamHI-BglI (58.0-72.2mu) fragment from pG20P was ligated to a BamHI + BglI digest of Py DNA, and after BamHI redigestion, the mixture of products was ligated to BamHI digested pAT153 and transfected into E.coli HB101. A plasmid containing Py linear DNA lacking the PvuII fragment, inserted into the BamHI site of pAT153 (p dl2020P), was isolated. Plasmid p2020 was constructed in parallel, but contained the wild-type BamHI to BglI fragment from pG20. Plasmid pdl2006, containing a viral genome with deletions around both PvuII sites at 67.4 and at 70mu was constructed by cloning dl2012 (a mutant lacking the PvuII at 70.0mu isolated in mouse embryo cells) into pAT153, partial cleavage with PvuII, exonuclease III digestion followed by S1 nuclease treatment, blunt end ligation and recloning in HB101. The deletions in pdl2006 and pdl2020P were sequenced from a 5'-label at the HpaII site (70.5mu, nt1) after secondary digestion with BamHI or KpnI (58.0 and 59.2mu). That in pdl2012 was sequenced from the PvuII site at 67.4 mu after secondary cleavage with BglI (72.2mu).

RESULTS

Deletions at 67 map units were not recovered when modified Py DNA was grown in mouse cells.

Our initial goal was to identify sequences involved in Py late gene expression by the isolation of deletions which were either viable or complemented by early gene ts mutants. To this end, viral DNA was modified around any of the four cleavage sites for restriction endonuclease PvuII (see Fig. 2) and transfected into mouse embryo cells, with and without helper tsa (40; a large T-protein ts mutant) DNA. Two of the PvuII sites (at 67.4 and 70.0mu, see Figs. 1 and 2) occur between the replication origin and the beginning of late protein coding sequences. The site at 67.4mu was of particular interest because it lies between the sequences which determine the most abundant of the heterogeneous capped 5'-ends of late region mRNAs and the only potential Hogness-Goldberg 'TATA'-box (11,12,30) in the late 5'-flanking sequences (Fig. 1). As this site is some 300 bp upstream of the major early mRNA cap sites (Fig. 1), we did not originally expect deletions around it to affect early gene expression. The data shown in Figure 2 summarises the rather surprising distribution of deletion mutants obtained. A number of mutants lacking the PvuII site at 92.1mu were isolated. These are similar to known viable mutants (16,17), and were not further characterised in this study. Some of them subsequently proved useful in mapping regions of the Py middle-T protein important for oncogenic transformation (Nilsson, S., Tyndall, C. and Magnusson, G., manuscript in preparation). No mutants lacking the PvuII site at 8.9mu were anticipated or found, because this occurs within sequences essential for large T-protein function (3,4). Viable mutants deleting the site at 70.0 mu (nt5267) were isolated in very low yield. This



Tsa helper DNA	no deletion	I	II	III	IV
−	28	0	0	22	0
+	35	0	2	13	0

Figure 2. Analysis of PvuII deletion mutants isolated in mouse cells. Py DNA was partially cut with PvuII, modified and used to transfect WME cells, alone or with tsa DNA helper, as described in Methods. Virus plaques (50 with helper, 50 without) were picked and screened for DNA deletions around the PvuII sites. The map shows the locations of the four PvuII sites (I) on the circular Py genome with respect to landmarks in the origin region shown in Figure 1. The table summarizes the yield of mutations lacking each of the four sites.

was expected because it is very near the origin of viral DNA replication (5-7). The larger mutant obtained (dl2012) proved to extend entirely away from the origin, deleting nt5235-5267 and preserving the T8 tract from nt5271-5278 (cf Fig. 3). This agrees with the recent demonstration that deletions

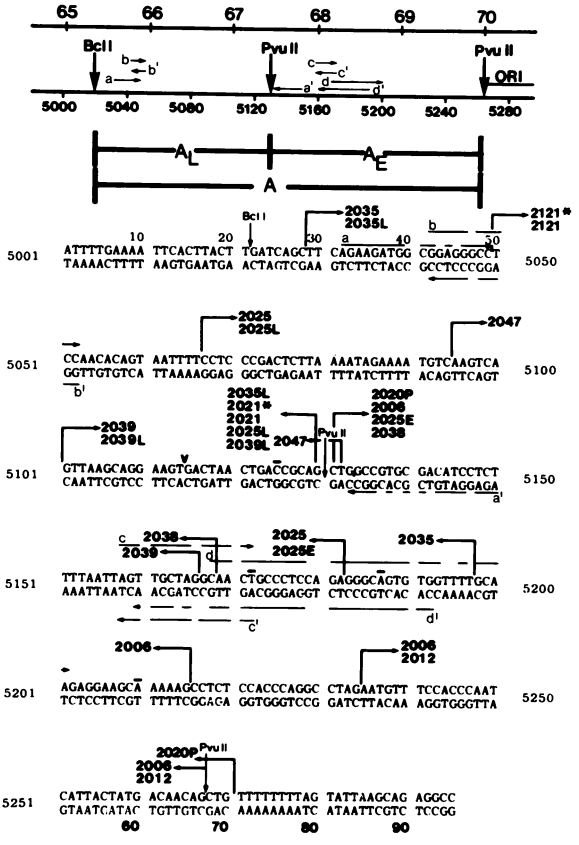


Figure 3. Positions of deletions in the DNA sequence. Bent arrows indicate the limits of deleted sequences. XhoI linker sequences (CCTCGAGG), inserted at the site of deletion, are not shown. The sequence is the one we have determined for this region of the viral insert in our original cloned DNA (p43.2b.67). It differs from that published for the A3 virus strain (2) at one point (no C between A5155 and T5156), and from that for the A2 strain (4) at five points (these differences are indicated by (V) for a bp deleted, and by overlining for a bp inserted with respect to the previously published [4] A2 sequence). The A2 numbering system is modified to accomodate the sequence changes (total Py sequence length now 5295). At the top is shown the lefthand portion of Figure 1, with regions A, A_L and A_E (see text) demarcated. Arrows represent dyad symmetries, with imperfections marked with spaces. Pairs of letters (a, a'; b, b' etc) correlate the components of each dyad.

impinging on the T tract are cis defective for viral DNA replication (G. Magnusson, personal communication). The major surprise was that we were unable to isolate any deletions spanning the site at 67.4mu. Control experiments demonstrated that this site was represented in the partial PvuII digests. The absence of the expected mutants could merely reflect bad luck, but we decided instead to test the more optimistic hypothesis that this region contains a cis acting sequence important for early gene expression and/or for viral DNA replication.

Construction and characterisation of deletion mutants using cloned viral DNA

Further deletion mutations were constructed (see Methods) within full length viral DNA cloned in the BamHI site of plasmid pAT153 (34) and propagated in bacteria. This simplified the isolation of mutants, particularly those potentially defective in viral DNA replication. The mutants are shown diagrammatically in Figure 1 and in more detail in Figure 3. Except where noted in the figure legend, the mutants all have an XhoI linker inserted at the site of deletion. Most are within the region designated A in Figures 1 and 3 (from the BclI site at 65.4 mu to the PvuII site at 70.0). The PvuII site at 67.4 mu divides region A into subregions A_L and A_E (Figure 1 and 3). Some mutants extend across this site and lack sequence from both subregions, whereas others extend only from the PvuII site into either A_L or A_E. One mutant (pdl 2012) lies within A_E but extends instead from the PvuII site at 70.0 mu; pdl 2006 is a double deletion derived from pdl 2012 which lacks further sequence from A_E proximal to the PvuII site at 67.4 mu.

Mutant viral DNAs were tested for their ability to form plaques on mouse embryo cells after excision from their vectors (Table 1). DNA with multiple XhoI linkers inserted at 67.4 mu (p43.2b.67), an intermediate in the production of many deletions (see Methods), was wild-type for plaque formation, as was a linker insertion mutant at 70.0 mu (nt 5267). The viral DNA from mutants lacking the origin distal 38 or 51 bp of subregion A_E (pdl 2038 and pdl 2025E, see Figure 3), and the DNA from a mutant lacking the origin proximal 33 bp (pdl 2012), were viable (Table 1). Removal of all of region A_E (pdl 2020P), or introduction into pdl 2012 of a second deletion removing the origin distal 85 bp (dl2006), eliminated viability (Table 1). These data suggested that a sequence near the middle of A_E (68.4-69.0 mu, nt 5185-5215) was essential. A mutant lacking 10-15 bp from the origin proximal side of subregion A_L (pdl 2125) was viable, extension to 30 bp resulting in small plaques (pdl 2039L), but mutants lacking 36 bp or more were nonviable. This suggested the existence of a further important sequence within A_L near

Table 1

Plasmid DNA	nucleotides deleted A_L	A_E	A	Viability plaque formation	Relative amounts[a] of replicated DNA	Complementation[f] by p2020 (wt)
WT						
p2020					++++	
p43.2b.67				normal	++++	
A_E						
pdl2012		33[b]		normal	++	N.D.
pdl2038		38		normal	++	none
pdl2025E		51		normal	+++	none
pdl2006		85+33[b]		negative	+	negative
pdl2020P		139		negative	+	negative
A_L						
pdl2125[d]	ca15			normal	++	N.D.
pdl2039	30			small plaque	+++	N.D.
pdl2047[c,g]	36	2		negative	+++	none
pdl2025L	64			negative	+++	N.D.
pdl2121	82			negative	+	slightly negative
pdl2121*	82			negative	+/-	positive
pdl2035L	102			negative	++	N.D.
A						
pdl2004[c,d]			ca10[e]	N.D.	+++	none
pdl2007[c,d]			ca25[e]	N.D.	+++	none
pdl2039	30	36		negative	+/-	negative
pdl2054[d]	ca50	ca 40		N.D.	-	N.D.
pdl2025	64	51		negative	-	none
pdl2031[d]	ca65	ca 55		N.D.	-	N.D.
pdl2035	102	67		negative	-	none
pdl2089[d]	ca95	ca 80		N.D.	-	N.D.
pdl2023[d]	ca30	ca130		N.D.	-	none
pdl2112[d]	ca220	ca 30		N.D.	-	none
pPytsa31[o]				ts	++++	-
pPytsa39[o]					+/-	positive

a. An arbitrary scale of - to ++++ is used to reflect visual comparison of gel band intensities. We do not consider difference between ++ and +++ to be beyond experimental error. b. The 33bp deletion extended from the PvuII cleavage site at 5267 into A_E. pdl2006 additionally has an 85bp deletion extending into A_E from the PvuII site at 5130. c. These deletion mutants do not contain an XhoI linker. d. Sequence analysis was not done. The deletion was approximately positioned by restriction enzyme analysis (AccI, XhoI, PvuII, BamH-I). e. The deletion removes the PvuII site at 67.4 mu but its extent into A_L or A_E is not known. f. "none" indicates no effect of helper DNA on replication. "N.D." indicates that tests were not done. g. This mutant contains three XhoI linkers at the site of deletion.

66.8 mu (nt 5100). No other deletion mutant assayed formed plaques (Table 1).

Deletion mutants defective in early gene expression

We suspected that certain mutants were nonviable because of defects in early gene expression. Other experiments done in our laboratory had shown that removal of sequence from 65.4-70.0 mu decreased the oncogenic transforming activity of the viral DNA in a progressive manner as the residual viral sequence between vector DNA and the PvuII site at 70.0 was shortened (P. Jat, A. Cowie, U. Novak, C. Tyndall, and R.K., submitted for publication).

Transformation depends on early gene expression. We therefore measured this function more directly in transient assays. Deletion mutants were initially screened for their ability to express Py large-T protein by determining the percentage of cells with positive T-antigen immunofluorescence 60-72 hours after calcium phosphate mediated transfection. Although this assay is both sensitive and convenient, the quantitative relationship between expression level and the percentage of Py T-positive cells, is unknown and unlikely to be direct. We therefore subsequently checked certain deletion mutants by two less ambiguous methods, immunoprecipitation of ^{35}S-methionine labeled protein (37) and direct RNA determination using a variation of the nuclease S1 gel mapping procedures (44).

Immunofluorescence results are summarized in Table 2. We used three different recipient cell lines, which were all non permissive for Py DNA replication: HeLa, F2408 (rat1) fibroblasts (35), and their TK$^-$ derivative, rat2 fibroblasts (established by W. Topp, personal communication). Approximately 20-30% of HeLa cells, 8-15% of rat 2 cells and 1% of rat1 cells showed positive Py T-antigen nuclear immunofluorescence after transfection with a plasmid (p37.3.A2) containing wild-type DNA or a plasmid (p43.2b.67) with XhoI linkers inserted in viral DNA at 67.4 mu (Table 2). This immunofluorescence indicated the expression of the Py large T-protein because a plasmid encoding only this early protein (pPyLT1, A. Cowie, R. Treisman and R.K., unpublished construction) afforded similar results whereas one expressing only the Py middle-T protein (pPyMT1) yielded no positive cells (Table 2). A viable deletion mutant (d175 [19,46], cloned into pAT153 to yield p37.29.75) lacking much of NER-E2 (see Figure 1), produced only slightly reduced levels of positive cells (Table 2), although it is known to synthesise very reduced amounts of early proteins during the productive infection of mouse cells (46). All of the region A deletion mutants tested had reduced or negative expression levels as measured by this assay (Table 2). In general, qualitatively similar results were found among the three cell lines used, but some quantitative differences were noted among mutants with intermediate phenotypes. Removal of subregion A_E (pdl2020P) reduced the percentage of T-positive HeLa cells to about 14% of control levels; the negative effect was less marked with the rat cells, but this difference may not be significant. Removal within A_E of the origin-distal 85bp plus the origin-proximal 33bp (pdl2006) reduced expression in HeLa cells to less than 1%; this mutant in rat cells, however, expressed at the same moderate level as the one (pdl2020P) lacking all of region A_E. In this case the difference between the results

Plasmid DNA	Relative Percentage of T-positive Cells: HeLa		F2408 rat1		F2408 rat2		Extent of deletion (bp): A_L	A_E	A
p37.3.A2	100		-		-		-	-	-
p43.2b.67	100		100		100		-	-	-
p37.29.75	56	(1)	68	(3)	-		-	-	-
pPyLT1	109	(3)	-		49	(2)	-	-	-
pPyMT1	0	(3)	0	(4)	0	(4)	-	-	-
A_L									
pdl2039L	19	(1)	-		-		30	-	
pdl2025L	5	(1)	-		-		64	-	
pdl2121*	0	(1)	-		24	(1)	82		
pdl2121	<1	(3)	-		<1	(4)	82[a]	-	
pdl2035L	<1	(2)	-		-		102	-	
A_E									
dl2038	25	(4)	56	(2)	21	(2)	-	38	
dl2006	<1	(5)	36	(4)	12	(2)	-	33+85	
dl2020P	14	(3)	36	(2)	40	(1)	-	139	
A									
pdl2039	20	(3)	50	(3)	10	(2)	30	36	66
pdl2026	-		10	(1)	-				ca 90
pdl2025	0	(2)	0	(2)	4	(3)	64	51	115
pdl2005	0	(2)	0	(2)	6	(3)	64	51[a]	115
pdl2023	0	(4)	0	(4)	0	(2)	ca 30	ca130	ca160
pdl2035	0	(1)	0	(1)	2	(3)	102	67	169
pdl2089	0	(1)	0	(1)	0	(1)	ca 95	ca 80	ca175
pdl2033	0	(1)	0	(1)	<1	(1)	> 50	> 150	> 200
pdl2032	0	(1)	0	(1)	12	(1)	>200	< 10	> 210
pdl2044	-		-		0	(1)	>200	ca115	> 315
pdl2055	-		-		0	(1)			> 400

Table 2. Screening early gene expression by determining the percentage of cells with positive Py large T-protein immunofluorescence 60-72 hr post-transfection. Results are normalised to values obtained in the same experiment with the control (p43.2b.67) plasmid DNA, and are averages of independent determination where appropriate. Numbers of independent determinations are shown in parentheses. a. deletion with multiple XhoI linkers (3 in dl2121 and 8 in dl2005)

with the two species of recipient cells was statistically significant; its explanation is under investigation using the more quantitative direct analysis of viral RNA described below. Deletion of only the origin-distal 38bp from A_E (viable mutant pdl2038) resulted in a 2-4 fold decrease. Deletions extending into region A_L progressively reduced T-antigen expression, reaching a very low but still detectable level in HeLa cells when 82bp were removed. A short deletion extending 30bp into A_L and 36bp into A_E (pdl2039) moderately reduced the level of T-positive cells, whereas removal of 64 and 51bp, respectively, virtually eliminated it (pdl2025 and pdl2005). Large deletions extending across the A_L/A_E junction were similarly negative or severely depressed.

These data suggested that sequence elements located on either side of the PvuII site at 67.4 mu were important for early gene expression.

To correlate the immunofluorescence data with actual expression levels, we used the immunoprecipitation of ^{35}S-labeled T-antigens to test critical mutants. This proved feasible only in HeLa cells. The results presented in Figure 4A show that large, middle and small T-proteins were readily detectable in HeLa cells transfected with the plasmid containing viral DNA with multiple XhoI linkers at 67.4 mu (p43.2b.67). By contrast, deletion mutants which showed intermediate levels of T-positive HeLa cells (dl 2038 and dl 2039) had far less large T-protein (middle and small T antigens were undetectable). No large T-protein was found with the other mutants tested. These data qualitatively confirm the conclusions derived from immunofluorescence experiments, but imply that the latter assay rather overestimated actual expression levels.

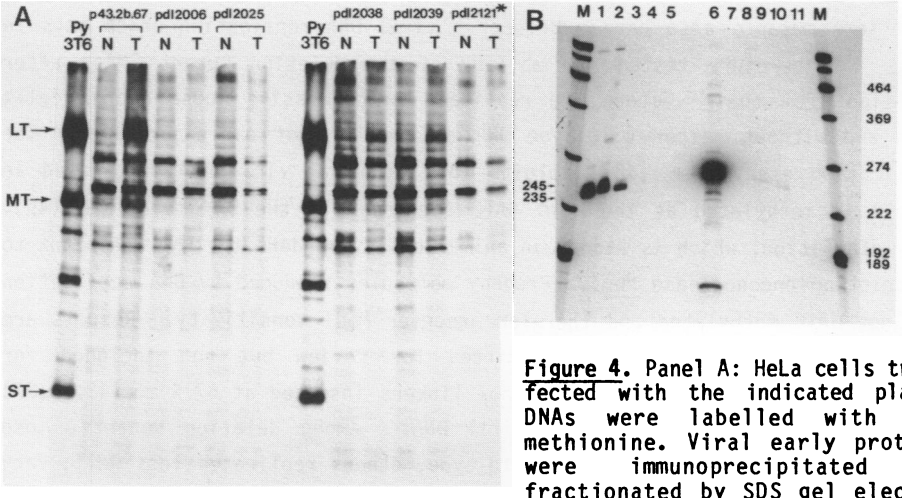


Figure 4. Panel A: HeLa cells transfected with the indicated plasmid DNAs were labelled with ^{35}S-methionine. Viral early proteins were immunoprecipitated and fractionated by SDS gel electrophoresis (see Methods). Py 3T6 lanes include markers for large T (LT), middle T (MT) and small T (ST) from mouse cells infected with virus N-nonimmune and T-immune sera. Panel B: S1-gel mapping (11,44) the 5' ends of deletion mutant early mRNAs. The 5'-^{32}P-labeled, single-stranded, HinfI fragment of Py DNA from nucleotide 388 to 5077 (4) was annealed to 10µg of cytoplasmic RNA (38) from HeLa cells transfected with each of the following DNAs: 1. plasmid p43.2b.67; 2. a plasmid encoding only large-T protein (pPyLT1); 3. a plasmid encoding only middle-T protein (pPyMT1); 4. pdl2006; 5. pdl2023; 6. pdl2025; 7. pdl2038; 8. pdl2039; 9. pdl2121*; 10. pdl2035L. Lane 6 was reference mRNA from virus infected 3T6 cells. The S1 resistant products resolved on the 5% polyacrylamide urea gel (47) of ca 235 and 245 bases map the major early region cap sites at nt 145-155 (see Fig. 1).

We further assayed the quantity and quality of the early region mRNAs synthesized after transfection using nuclease S1-gel mapping of the 5'-ends. Control plasmids (Figure 4B) produced, in HeLa cells, early region mRNAs with the same 5'-termini as those from productively infected mouse cells. Viral mRNAs with correct 5'-ends were found at drastically reduced levels with the two deletion mutants (dl 2038 and dl 2039) which produced moderate percentages of T-positive cells and detectable amounts of immunoprecipitable protein. No other deletion mutant examined synthesised quantities of viral mRNA measurable with this assay. The S1 mapping results thus agree quantitatively with those from immunoprecipitation. Further investigation of the 5'-termini of deletion mutants severely impaired in early gene expression required substantial increase in the sensitivity of the hybridisation assay and will be presented elsewhere.

Deletion mutants cis-defective in viral DNA replication

The demonstration that removal of DNA sequence from within region A impairs or eliminates early gene expression does not exclude the possibility that some mutants also have a defect in viral DNA replication which acts in *cis*. We therefore tested the ability of the deletion mutant DNAs, after excision from their vectors, to replicate in permissive mouse embryo cells with and without helper wild-type DNA as a source of early proteins. The replication assay used (48) exploits the fact that viral DNA propagated in *E.coli* is methylated at the A of GATC sequences by the DAM methylase (49). The methylation, which is stable in animal cells, renders the DNA resistant to restriction endonuclease *Mbo*I. Progeny molecules produced by DNA replication are not GATC methylated and therefore recover MboI sensitivity. Results are summarised in Table 2 (data submitted for review but not included for publication). The viral DNA with *Xho*I linkers inserted at 67.4 mu (43.2b.67) replicated as well as wild-type viral DNA. Among deletion mutants, some replicated nearly as well as the wild-type, others replicated less well, very poorly, or not at all. This range of phenotype is indicated using an arbitrary scale (++++ to -) in the table. As might be expected, there was a general correlation between replication without helper and the ability of the mutant to express its early genes. For example, viable mutant dl2038 replicated and expressed relatively well, whereas nonviable mutants dl2025 and dl2035 expressed very poorly and didn't replicate. There was, however, a curious exception. Nonviable mutant dl2039 (lacking 30bp from A_L and 36 from A_E) expressed early protein and early region mRNA as well the viable mutant dl2038 (lacking 38bp from A_E), but it replicated far less well (compare Figure

4 with table 1). Results of complementation assays, in which wild-type viral DNA retained in its vector was co-transfected with the mutants, were also surprising. In these assays, the mutant and helper DNAs could in general be distinguished by their MboI restriction patterns, but in certain instances MboI and XhoI digestion was necessary. Positive complementation was obtained with only one mutant (dl2121*, see below). Deletion mutants (such as dl2025 and dl2035) which did not replicate alone also did not replicate with helper. Mutants (such as dl 2020P and dl 2039) which replicated poorly by themselves were inhibited by addition of "helper" DNA (negative complementation). Control experiments using the large T temperature sensitive mutant, tsa (40), demonstrated that cotransfection could yield positive complementation (Table 1). In addition, the one exceptional deletion mutant, pdl2121*, was positively complemented. This mutant has three XhoI linker sequences inserted at the site of a deletion removing 82bp from region A_L. Removal of two linker sequences (generating pdl 2121) restored much of the replicative ability of the mutant, but in this case added helper was slightly inhibitory such that, with helper, pdl 2121* and pdl 2121 replicated to similar extents.

We inferred from these data the tentative conclusion that manipulation of the DNA sequence within region A can cause a cis-acting impairment in viral DNA replication, in addition to the previously demonstrated defectiveness in early gene expression. The critical sequences are not continuous with the known replication origin region because at least 33bp can be removed from the origin proximal end of region A_E (viable deletion pdl 2012) without affecting viability or substantially decreasing replicative ability. The negative complementation observed with several mutants suggested that they unsuccessfully competed with helper wild-type DNA for a limiting factor. It was therefore impossible to assess the actual replicative abilities of such mutants by cotransfection with helper DNA because the source of the large T-protein was also a competitive inhibitor.

Replication of mutant DNAs in Py transformed mouse cells (COP lines)

An alternative approach to measuring the replicative abilities of deletion mutants would be to use permissive mouse cells containing a replication defective integrated viral genome as the source of the large T protein. Such cell lines exist for SV40 (the monkey cell COS, lines established by Gluzman [50]), but are not permissive for Py DNA replication (our unpublished observations). We therefore established the equivalent Py transformed mouse cell lines by transfecting C127 mouse cells (51) with a Py deletion mutant (lacking 70.33 - 7239 mu, nt 5287 to 104) which was expression

positive but replication negative. The detailed characterisation of these cell lines, which we named COP cells (by analogy to the SV40 "COS" cell lines) will be presented separately. The transformed lines selected for use here, COP-3 and COP-5, express full length Py large T-protein but produce no detectable unintegrated viral DNA molecules. The COP cells were transfected with recombinant plasmids and replication was assessed using the MboI sensitivity assay (Figure 5). A recombinant containing a Py genome with a deletion within large T-protein coding sequences (pTK Py.2.1) replicated in the COP cells, as did recombinants containing viral fragments spanning region A plus the origin region (p35.15.A2 and pG20, see Fig. 5). These data demonstrate the functionality of the large T-protein expressed by COP cells. Truncation of the viral DNA insert in one of the plasmids (p35.15.A2) by removal of region A (to form pP17) abolished replicative ability. Removal of only region A_E (pG20P and pdl2020P) markedly inhibited replication. One of the deletion mutants discussed above (pdl 2039, which has a deletion spanning the A_E/A_L junction) also replicated far less well than controls, whereas two other slightly longer deletion mutants (pdl 2025 and pdl 2035) were replication incompetent in COP cells, as was mutant pdl2121*. Experiments to directly compare the replicative abilities of pdl2038 and pdl2039 (see preceeding section) are in progress.

These results clearly demonstrate that a sequence located within region A_E is required, in cis, for efficient viral DNA replication. They further establish that removal of 115 bp around the PvuII site at 67.4 mu (64 bp from A_L plus 51 bp from A_E) eliminates the ability to replicate. As such deletion mutants also fail to express early genes, they appear to be biologically inert.

DISCUSSION

We have begun a functional analysis of the noncoding region of Py DNA to the late region side of the replication origin by the construction and characterisation of internal deletion mutations. Some mutants were viable. Others had moderate or severe defects in both viral DNA replication and early gene expression. Possible effects on late gene expression have not as yet been investigated. The use of deletion mutants to accurately localise regulatory elements is always complicated by the fact that removal of a specific sequence brings others into abnormal juxtaposition. This is particularly relevant to the results reported here because the functional elements appear to be either repeated or dispersed over a considerable

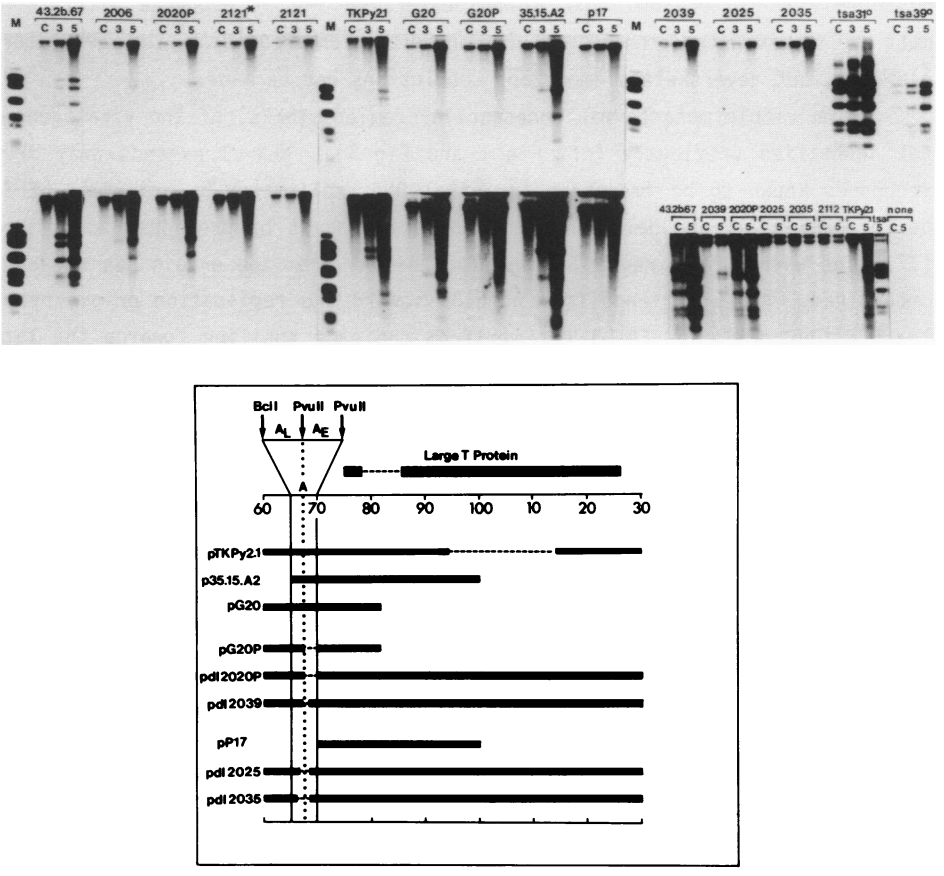


Figure 5. Replication assays with Py transformed C127 (COP) cells. The cells, either COP-3, COP-5, or a control T-antigen negative (and apparently untransformed) cell line derived in the same experiment which produced the COP lines, were transfected (31) with the indicated recombinant plasmids (100 ng). Three days post-transfection, low molecular weight DNA was extracted and analysed for progeny viral DNA sensitive to endonuclease MboI as described in Methods. Tracks M show the six fragments produced by MboI digestion of Py DNA. DNA with a temperature sensitive mutation in the large T gene (tsa) provided one of the controls for positive complementation (this DNA was not grown as a plasmid and therefore the input was also MboI sensitive). The upper panels show shorter and the lower panels show longer autoradiographic exposures, except for the gel on the far right (no further bands were visible on longer exposure), where the lower panel is a separate experiment. In this experiment, the viral DNA inserts were excised from the vectors prior to transfection. Excision enhanced the ability of the DNA to replicate in the untransformed control cell line. The diagram below the autoradiograms shows maps of the viral DNA inserts in the plasmids relative to the large T-protein coding sequence and region A. Plasmids pP17 and pTKPy2.1 (which also contains an HSV TK gene) were the generous gifts of U. Novak and M. Fried, respectively.

distance. It is obvious that a more extensive set of sequence alterations must be constructed to accurately enumerate and position the regulatory elements, but nevertheless important conclusions can be made.

The viable mutants map nonessential regions (NERs) of the viral genome not identified previously (cf. Fig 1 and Fig 3). NER-L1 extends away from sequences known to be important for viral DNA replication by a minimum of 33 bp (nt 5267-5234; one deletion mutant of less than 5 bp previously described (17) lies within this region); the limit distal from the origin has not been determined. NER-L2 extends from nt 5130 toward the replication origin by at least 51 bp, whereas NER-L3 extends from the same position towards the late region by about 30 bp. We do not as yet know whether the short deletions spanning nt 5130 are viable, but we suspect they will be because such DNAs (dl 2004 and dl 2007) replicate well in mouse cells without helper (Table 1). Larger deletions, such as dl 2039, which remove NER-L3 and much of NER-L2 (36 bp) were nonviable. For this reason we provisionally consider NER-L2 and NER-L3 as separate domains, either but not both of which can be removed without loss of viability.

We were initially surprised to find that deletion mutants within region A (Fig 1) were defective in early gene expression. The A_E/A_L junction is more than 300 bp 5'- to the principal early mRNA cap site (Kamen et al, submitted for publication), and we know that these cap sites are transcriptional initiation points (A. Cowie, P. Jat, and R. Kamen, in preparation). The cap sites and the 'TATA' box which preceeds them are within NER-E2 (Fig. 1). A viable deletion mutant (dl 75, Table 2) spanning much of this putative promoter region was tested in transient expression assays and found to be only slightly defective in early gene expression as measured by T-antigen immunofluorescence. Using the same assay, a 115 bp nonviable deletion (pdl 2025) approximately centred on the A_E/A_L junction was very severely impaired. Therefore an upstream region remote from the cap sites is far more important for efficient gene expression in vivo than proximal sequences. We have shown elsewhere, however, that deletion of the proximal sequences causes increased heterogeneity of mRNA 5'-ends (Kamen et al, submitted for publication). These results are similar to those described for other genes (24,52,53), but the Py upstream element is some 200 bp more remote.

The detailed interpretation of the effects of different mutations within region A on early gene expression is made difficult by the deletions/juxtaposition conundrum. Deletions extending either 30 bp into region A_L or 38 bp into A_E (dl 2039, dl 2039L and dl 2038) were found to express correct

early region mRNA, but at no more than 10% of the wild-type level in direct assays. Longer deletions extending into A_L were further impaired. Mutants lacking the origin-distal 85 bp from A_E, or all of A_E, were moderately impaired when assayed by T-antigen immunofluorescence in rat cells; no early region protein or mRNA was detected in HeLa cells using more quantitative assays. Such data strongly suggest that there is more than one critical sequence, but positional relationships may be important. We are now constructing further mutants to test the hypothesis that the important sequences are the striking regions of dyad symmetry indicated in Figure 3, particularly those at nt 5031-5052 and at 5158-5201.

The closely related virus, SV40, also has an upstream region critical for early gene expression on the late side of its origin (24,54). This, in strain 776, is a 72 bp direct repeat (1), one copy of which is dispensible (24,54). The 72 bp sequence can enhance expression from promoters of other eucaryotic genes linked to it (27,28). This effect is orientation independent and can act over long distances. Region A of Py DNA similarly enhances the expression of the rabbit ß-globin gene (J. de Villiers and W. Schaffner, submitted for publication) in transient assays and stimulates the transforming activity of the HSV thymidine kinase gene (U. Novak, personal communication). We searched for homology between Py region A and the SV40 72 bp sequence using the Stanford SEQ program. Three interesting regions of imperfect homology were found, from Py nt 5025-5039 (overlapping one of the dyads indicated in Fig. 1), from nt 5173-5188 (within the longest dyad region), and from nt 5212-5230, including a perfect CCCAGGC match (this region is within that duplicated in PyF9 mutants (21-23)). The statistical significance of these homologies was not high, and indeed better matches can be found elsewhere in the Py genome. However, the occurrence of sequence homologies between regions of related function is worth noting. We have also found homologies among the three sequences in Py region A, the SV40 72 bp sequence and the 73 (55) or 69 (56) bp direct repeat in the Moloney murine sarcoma virus LTR (similar observations have been made by J. Banerji, personal communication). The MSV sequence can functionally substitute for the 72 bp repeat in SV40 DNA (P. Gruss and G. Khoury, personal communication), and contains striking dyad symmetries. Although the Py A2 strain we use has no perfect tandem repeats in region A, other Py viruses, and several of the PyPCC4 (20) and PyF9 (21-23) mutants, have them (E. Ruley, unpublished sequence results and alternative interpretation of the PyPCC4 DNA structures).

The major functional differences between Py region A and the SV40 72 bp

repeat is that the former includes sequence absolutely required, in cis, for viral DNA replication. Plasmids lacking a complete copy of the SV40 72bp sequence have been shown to replicate in COS cells (57,58). While we once again cannot state exactly where within region A the critical element lies, we know that it is not continuous with the previously identified origin region. For example, mutant dl 2039, which lacks only 66 bp approximately centred on the A_E/A_L junction has a cis-acting defect, whereas dl 2012 lacks 33 bp proximal to the origin and is viable. Curiously, the replication of partially defective mutants is inhibited by "helper" wild-type DNA, implying the competition for a limiting factor. We do not know whether region A comprises common elements which regulate both gene expression and DNA replication, or distinct but intermingled regulatory signals. As only some of the possible functions for "enhancer" sequences which have been discussed (27,28) can account for a dual role in replication and transcription, it is important to distinguish between the alternatives. One approach to this problem is to study SV40/Py recombinants. Preliminary results (C. Tyndall and J. de Villiers, unpublished experiments) suggest that the origin region of SV40 (from KpnI to HindIII) can replace both the expression and replication functions present in Py region A.

ACKNOWLEDGEMENTS

We thank Mrs. K. Osborne for cell culture assistance, M. Fried, A. Cowie and U. Novak for the gift of plasmids, and B. Griffin for the gift of a Py mutant DNA. Ms. P. Morgan, Ms. G. Yiangou and Mrs. A. Symons helped with the preparation of this manuscript.

[+]On leave from Istituto di Biologia Generale e Genetica, University of Naples, Italy

REFERENCES

1. Tooze, J., ed. (1980) Molecular Biology of Tumor Viruses, Part II: DNA Tumor Viruses (New York: Cold Spring Harbor Laboratory)
2. Deininger, P., Esty, A., LaPorte, P., and Friedmann, T. (1979) Cell 18, 771-779
3. Friedmann, T., Esty, A., LaPorte, P., and Deininger, P. (1979) Cell 17, 715-724
4. Soeda, E., Arrand, J.R., Smolar, N., Walsh, J.E., and Griffin, B.E. (1980) Nature 283, 445-453
5. Folk, W.R., and Wang, H.E. (1974) Virology 61, 140-155
6. Crawford, L.V., Robbins, A.K., Nicklin, P.M., and Osborn, K. (1974) Cold Spring Harbor Symp. Quant. Biol. 39, 219-225
7. Griffin, B.E., and Fried, M. (1975) Nature 256, 175-179
8. Soeda, E., Kimura, G., and Miura, K. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 162-166

9. Gaudray, P., Tyndall, C., Kamen, R., and Cuzin, F. (1981) Nucleic Acid Res. 9, in press
10. Kamen, R.I., Favaloro, J.M., Parker, J.T., Treisman, R.H., Lania, L., Fried, N., and Mellor, A. (1980) Cold Spring Harbor Symp. Quant. Biol. 44, 63-75
11. Treisman, R. (1980) Nucleic Acid Res. 8, 4867-4888
12. Cowie, A., Tyndall, C., and Kamen, R. (1981) Nucleic Acid Res. 9, in press
13. Treisman, R.H., Cowie, A., Favaloro, J.M., Jat, P., and Kamen, R. (1981) J. Mol. Appl. Gen. 1, in press
14. Herbomel, P., Saragosti, S., Blangy, D., and Yaniv, M. (1981) Cell 25, 651-657
15. Soeda, E., Arrand, J.R., Smolar, N., and Griffin, B.E. (1979) Cell 17, 357-370
16. Griffin, B.E., and Maddock, C. (1979) J. Virol. 31, 645-656
17. Magnusson, G., and Berg, P. (1979) J. Virol. 32, 523-529
18. Wells, R.D., Hutchinson, M.A., and Eckhart, W. (1979) J. Virol. 32, 517-522
19. Bendig, M., and Folk, W.R. (1979) J. Virol. 32, 530-535
20. Katinka, M., Yaniv, M., Vasseur, M., and Blangy, D. (1980) Cell 20, 393-399
21. Fujimura, F.R., Deininger, P.L., Friedmann, T., and Linney, E. (1981) Cell 23, 809-814
22. Sekikawa, K., and Levine, A.J. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 1100-1104
23. Katinka, M., Vasseur, M., Montreau, N., Yaniv, M., and Blangy, D. (1981) Nature 290, 720-722
24. Benoist, C., and Chambon, P. (1981) Nature 290, 304-309
25. Capecchi, M.R. (1980) Cell 22, 479-488
26. Gruss, P., Dhar, R., and Khoury, G. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 943-947
27. Banerji, J., Rusconi, S., and Schaffner, W. (1981) Cell, in press
28. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. (1981) Nucl. Acids Res. 9, in press
29. Griffin, B.E., Fried, M., and Cowie, A. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 2077-2081
30. Breathnach, R., and Chambon, P. (1981) Ann. Rev. of Biochem. 50, 349-383
31. McCuthchan, J.H., and Pagano, J.S. (1968) J. Natl. Cancer Inst. 41, 351-356
32. Hirt, B. (1967) J. Mol. Biol. 26, 365-369
33. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517
34. Twigg, A.J., and Sherratt, D. (1980) Nature 283, 216-218
35. Freeman, A.E., Gilden, R.B., Vernon, M.L., Wolford, R.G., Hugunin, P.E., and Huebner, R.J. (1973) Proc. Natl. Acad. Sci. U.S.A. 70, 2415-2419
36. Graham, F.L., and van der Eb, A.J. (1973) Virology 52, 456-467
37. Ito, Y., Spurr, N., and Dulbecco, R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 1259-1263
38. Favaloro, J., Treisman, R., and Kamen, R. (1980) Meth. Enzymol. 65, 718-749
39. Parker, G.C., Watson, R.M., and Vinograd, J. (1976) Proc. Natl. Acad. Sci. U.S.A. 74, 851-855
40. Fried, M. (1965) Virology 25, 669-671
41. Brockman, W.W., and Nathans, D. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 942-946
42. Birnboim, H.C., and Doly, J. (1979) Nucleic Acid. Res. 7, 1513-1523
43. Maxam, A., and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560
44. Berk, A.J., and Sharp, P.A. (1977) Cell 12, 721-732

45. Treisman, R., Novak, U., Favaloro, J., and Kamen, R. (1981) Nature 292, 595-600
46. Bendig, M.M., Thomas, T., and Folk, W.R. (1980) Cell 20, 401-409
47. Sanger, F., and Coulson, A.R. (1978) FEBS Lett. 87, 107-110
48. Peden, K.W.L., Pipas, J.M., Pearson-White, S., and Nathans, D. (1980) Science 209, 1392-1396
49. Geier, G.E. and Modrich, P. (1979) J. Biol. Chem. 254, 1408-1413
50. Gluzman, Y. (1981) Cell 23, 175-182
51. Langbeheim, H., Shih, T.Y., and Scolnick, E.M. (1980) Virology 106, 292-300
52. Grosschedl, R., and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 7102-7106
53. McKnight, S.L., Gavis, E.R., Kingsbury, R., and Axel, R. (1981) Cell 25, 385-398
54. Gruss, P., Dhar, R., and Khoury, G. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 943-947
55. Dhar, R., McClements, W.L., Enquist, L.W., and Van de Woude, G.F. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3937-3941
56. Reddy, E.P., Smith, M.J., Canaani, E., Robbins, K.C., Tronick, S.R., Zain, S., and Aaronson, S.A. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5234-5238
57. Myers, R.M., and Tjian, R. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 6491-6495
58. Mellon, P., Parker, U., Gluzman, Y., and Maniatis, T. (1981) Cell, in press