Volume 16 Number 18 1988 Nucleic Acids Research

Deletions in the SV40 late polyadenylation region downstream of the AATAAA mediate similar effects on expression in various mammalian cell lines

Edward R.Gimmi[1,2], Kenneth J.Soprano[1], Martin Rosenberg[2] and Mitchell E.Reff[1,2]

[1]Department of Microbiology and Immunology, Temple University School of Medicine, Philadelphia, PA 19140 and [2]Department of Molecular Genetics, Smith, Kline & French Laboratories, 709 Swedeland Road, King of Prussia, PA 19406, USA

Received March 15, 1988; Revised and Accepted August 11, 1988

ABSTRACT

A series of deletions in the SV40 late polyadenylation region was assayed by transient expression in a hamster fibroblast cell line. Because of differences in expression data between our results and the published results of another laboratory using a similar set of deletions introduced into a monkey kidney cell line, we studied our deletions in cells of different tissue-types and species (1). Deletion of the SV40 late polyadenylation region to 49 nucleotides downstream of the hexanucleotide AATAAA showed a small effect on gene expression, while further truncation of the region to 6 nucleotides downstream of the AATAAA showed an 85% drop in marker enzyme activity, protein levels and steady-state message levels. Another deletion in the same region, from base pair 10 to 15 past the AATAAA, which removes the wild-type site of RNA cleavage, showed a 50% drop in marker gene expression. The effects of these mutants on gene expression were similar in all of the cell lines tested and agree with other studies that DNA downstream of the AATAAA plays a role in efficient gene expression.

INTRODUCTION

Maturation of most polymerase II transcribed RNA requires a series of processing events which includes cleavage at the 3'-end of the primary transcript (2) with the subsequent addition of adenylate residues (3). Associated with this event is the consensus sequence 5'-AAUAAA-3', which is located 10-30 nucleotides upstream from the site of cleavage (4). This highly conserved motif, has been shown to be necessary for the proper 3'-end formation of mRNA in cells (5-10). It has also been demonstrated to be required for cleavage and polyadenylation *in vitro* (11). However, this sequence alone is not sufficient for such processing as it may be found in message coding sequence where no such RNA processing events occur (12-14). Other sequences downstream of this hexanucleotide also appear to be essential for accurate and efficient cleavage at the site of polyadenylation (15-31). Moreover, it has been observed that specific downstream sequences are needed for the formation of polyadenylation-specific complexes on substrate RNAs in nuclear extracts (32). Several

conserved nucleotide motifs have been suggested as essential downstream elements in vivo as well as in vitro (1, 17, 18). An octonucleotide YGTGTTYY has been reported to be present in 67% of the mammalian RNA polymerase II transcribed genes analyzed (17). It has also been suggested that RGTTTTYRR, T(A/G)TTTTT, or other thymidylate-rich sequences serve a role in 3'-end processing, and a related sequence in the SV40 late poly(A) region, AGGTTTTTT, is proposed to be necessary for the accurate and efficient processing at the 3'-end (1,18,30). However, other studies do not find this sequence to be needed for efficient cleavage (24,30). In general, none of these downstream sequences is present in all functional polyadenylation regions. In addition to the downstream sequences, the sequence at the actual cleavage site, which may be part of a diffuse signal, has been implicated as necessary for efficient processing, although no specific sequence is required (24,30).

In this paper we report the effects of deletions downstream of the AATAAA in the SV40 late polyadenylation region on the expression of a marker gene in several mammalian cell lines. This is the first direct comparison between polyadenylation region mutants in mammalian cell lines of different species and tissues. Also, these analyses augment the literature which contains conflicting data on the effect of deleting sequences downstream of the AATAAA in the SV40 late polyadenylation region. It is clear from our data that these polyadenylation region mutants are utilized with similar efficiencies in hamster and monkey cell lines and in cell lines derived from different tissues.

MATERIALS AND METHODS

Endonucleases and Bal 31 exonuclease were obtained either from New England Biolabs, BRL or BMB and used according to the manufacturers specifications. The pDP vector was created by D. Pfarr as a modified version of pDSP1 (33). pDSP1BGH has been described previously (34).

DNA sequencing

The method of Maxam and Gilbert was employed to sequence 5'-end labelled single-stranded fragments (35).

Transfections.

10 µg. of purified vector DNA was used to transfect R1610 cells (galactokinase minus/XGPRT minus hamster fibroblasts) (36), HAK cells, CV-1 cells and COS-1 cells by calcium phosphate coprecipitation (37). After 48 hours the cell lysates were prepared as described (38). Galactokinase (39)

and XGPRT (40) enzyme levels were analyzed by filter assays. For each data set two sets of reactions were carried out at two time points (15 and 30 min.) in order to demonstrate linearity of enzyme activity (33).

Construction of vectors.

The parental pDP vector is outlined in figure 1. The SV40 early polyadenylation region flanking the galactokinase gene was removed by digestion with Nhe I and Xho I. Subsequent filling-in and ligation of this vector was done to produce pDPminus. Two different fragments of the SV40 late polyadenylation region were ligated into the filled-in Nhe I-Xho I digested pDP. A filled in Bcl I-BamH I fragment from pSV40 containing the SV40 late polyadenylation region was used to create pDPSV. This fragment contains 133 base-pairs upstream of the AATAAA and 108 base-pairs past the AATAAA hexanucleotide. A smaller fragment of the SV40 late polyadenylation region was obtained from pSVKGPT by Stu I-EcoR V digestion to create pDPΔ116 (33). This fragment contains 129 base-pairs upstream of the AATAAA and 49 base-pairs downstream. In addition, it contains 24 base-pairs of 5' residual sequence from the SV40 early promoter and 12 base-pairs of residual gal K sequence 3' to the polyadenylation region. These flanking sequences are identical and present in all of the deletion constructs.

Deletion mutations Δ27, Δ5B, Δ23B and Δ117 were created in the SV40 late polyadenylation region of pSVKGPT by Hpa digestion, Bal 31 digestion, and insertion of a Hpa I linker. Recreation of the upstream sequence was achieved by the insertion of a Stu I-Hpa I fragment from the parental pSVKGPT vector into the deletion vector after cutting it with Stu I and Hpa I. pDPΔ44 was made by Hpa I-Sma I digestion of pSVKGPT followed by ligation. All of the deletions created in pSVKGPT were removed from the vector by Stu I-EcoR V digestion and subcloned into pDP vector as described above. pDP118 was produced by the insertion of a filled-in Ava II-Hph I fragment, obtained from the pBr322 ß-lactamase gene, into the Sma I site of pSVKGPT. A slightly larger 108 base pair Stu I-EcoR V fragment from this vector was then inserted into pDP as described above. Both orientations of this insert were selected to obtain pDP118 and pDP118R.

Vectors used to synthesize antisense RNA probes were prepared as follows. pGEM-3 (Promega Biotec) was digested with Hind III and Hinc II and a 465 base pair EcoR V-Hind III fragment from the E. coli XGPRT gene or a 659 base pair fragment of the E. coli gal K gene was inserted. Both the gal K and XGPRT probes were engineered to protect regions of coding

sequence and therefore do not map 3'-ends. The inserts were both oriented to produce antisense RNA from the bacteriophage T7 promoter.

RNA isolation.

Cytoplasmic RNA was isolated from cells 48 hours post-transfection by the method of Tushinski (41).

RNase protection assays.

RNase protection assays were performed by the method of Melton as modified by Toscani, et. al (42,43). The probes synthesized from gal K and XGPRT gene constructions are distinguishable by size on acrylamide denaturing gels. 5 μg. of cytoplasmic RNA was hybridized with both or a single probe at 55°C. After 12 hours, the hybridized RNA is digested with RNase A and T1, run on a 10% acrylamide-8M urea gel. Autoradiographs were scanned densitometrically with an LKB Ultroscan XL laser densitometer at various exposures to ensure linearity of the film response. Peak areas were determined by integration and final tabulations show the relative ratio of the galactokinase to XGPRT peak area.

Western analysis.

Cell lysates normalized to 10,000 cpm of XGPRT activity were run on 15% acrylamide-SDS gels and electro-blotted onto nitrocellulose. The filters were incubated with rabbit-anti-galK (IgG fractionated) antibody and probed with ^{125}[I]-protein A (ICN). Dilutions of purified galactokinase from E. coli was Western blotted and blots were scanned as described above. These data were used to generate standard curves to compare with the experimental data (data not shown).

High molecular weight DNA isolation and Southern blot analysis.

High molecular weight DNA was prepared as described (44).

Cell nuclei were isolated by resuspending the washed cells in ten times cell pellet volume of Buffer A (20 mM Tris, 5mM EDTA, 0.5 mM EGTA, pH 8.0) and left on ice for 10 minutes to swell. NP40 was added to 0.3%, and the cells were dounced. An equal volume of 0.66 M sucrose in Buffer A was added and the nuclei were pelleted (1000 rpm, 5 minutes, Beckman TJ6 centrifuge). The nuclear pellet was then resuspended in NTE (100mM NaCl, 10mM Tris, ImM EDTA, pH 7.5) and the remainder of the HMW DNA isolation protocol was followed.

DNA was electrophoresed on 0.6% agarose gels and transferred onto nitrocellulose (44). The vector pDSP1 was nick-translated and hybridizations were done at 65°C as described (44).

RESULTS

The vector system used allows measurement of gene expression in mammalian cells when polyadenylation regions are ligated downstream of an E. coli galactokinase gene (33, 34). The parental vector also contains a second marker gene, xanthine-guanine phosphoribosyl transferase (XGPRT) from E. coli. The XGPRT transcriptional unit provides an internal control for the efficiency of independent transfections. The XGPRT activity from a particular vector correlates linearly with the amount of DNA transfected (data not shown).

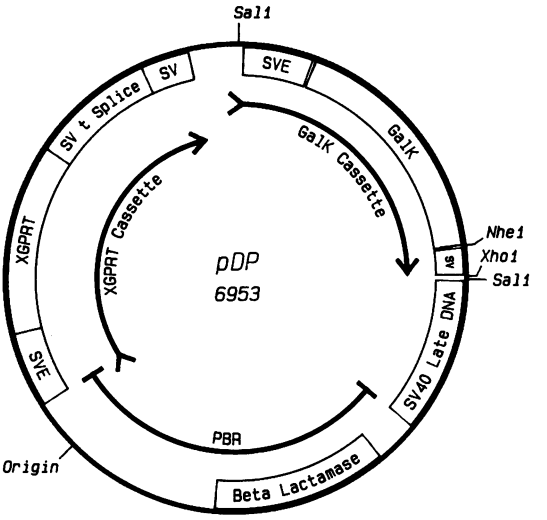
The XGPRT gene is placed several kilobases downstream of the gal K transcriptional unit in order to minimize effects of the gal K unit on XGPRT expression. Vectors can be cut into two separate pieces by Sal I, each DNA piece containing a transcriptional unit. Because transfection of Sal I cut vectors yields identical relative expression data as supercoiled vectors, it is probable that the effects of the gal K on XGPRT are minimal or non-existant in supercoiled vectors. The effects of the XGPRT unit on gal K expression in supercoiled molecules are assumed to be constant since changes are only being made at the 3'-end of the gal K gene.

All measurements are expressed as a ratio of gal K to XGPRT expression. This system of using the relative ratio of expression levels of two bacterial genes has been employed previously to quantitate transient expression (33, 34). The relative enzyme expression has been shown to correlate with the relative steady state mRNA levels present in the cell (34). The relative ratios of gene expression observed in transient experiments with these vectors are identical in pools of cells stably transfected and selected for integration of the XGPRT gene (data not shown).

pDP, the parental vector used in this study, is diagrammed in figure 1. The SV40 late polyadenylation region and various deletion mutants were inserted downstream of the gal K gene in order to measure the relative expression mediated by these regions.

Following transfection, relative expression data was obtained by enzyme analysis. Relative protein data from Western blots and relative RNA data from RNase protection assays was quantitated by densitometric scanning of exposed film.

Our gene expression data generated from mutants in the SV40 late and other polyadenylation regions implies that the efficiency with which the 3'-end processing occurs has an effect on the levels of steady-state



SV= SV40 polyA early
SVE= SV40 early promoter

Figure 1. General vector structure of pDP: The SV40 early promoter is driving the galactokinase gene from E. coli and mutants of the SV40 late polyadenylation region were substituted for the early polyadenylation region at the Nhe I-Xho I site. The SV40 early promoter is also driving the xanthine-guanine phosphoribosyl transferase gene from E. coli and is polyadenylated at SV40 early site. The Sal I sites indicate where the vector was sectioned in one of the studies.

message and protein in the cell (33, 34). There are several reasons for this belief:

1) Efficient gene expression is seen when a polyadenylation region is inserted downstream of the marker gene, but not when it is absent; 2) A sequence which contains an AATAAA from prokaryotes or other heterologous DNAs do not mediate efficient expression; 3) The relative expression levels seen between two different polyadenylation regions remains the same even when the marker genes are switched (34); and 4) Deletions made within polyadenylation region-specific sequence can lower the marker gene expression. It is possible, however, that these effects on expression are due to alterations in message stability rather than to the efficiency of 3'-end processing.

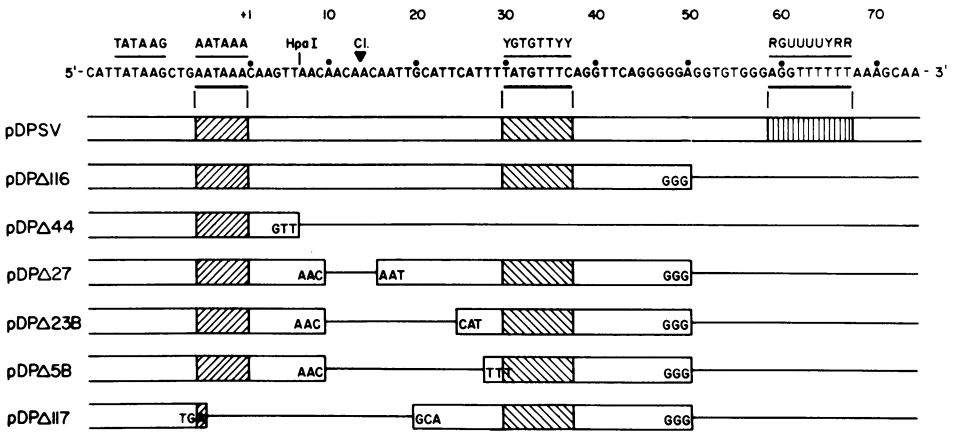


Figure 2. Map of deletions of the SV40 late polyadenylation region. Nucleotides indicated in the text are numbered as shown. Bars indicate the remaining sequence. The AATAAA consensus hexanucleotide, YGTGTTYY analog and RGUUUUYRR analog are indicated in each mutant by various shading. The wild-type site of cleavage is indicated by an arrow.

TABLE 1.
Relative Expression of Polyadenylation Region Mutants in R1610 Hamster Fibroblasts

Vector	Average Relative Expression and Std. Dev.	Transfection Number 1	2	3	4	Sal I digested* 5
pDPSV	100*	100	100	100	100	100
pDPΔ116	82/2.8	85	79	80	85	86
pDPΔ44	17/3.7	12	18	21		
pDPΔ27	59/5.3	56	65	59		
pDPΔ23B	39/4.6	33	41	44		
pDPΔ5B	38/0.9	39	37	38	39	42
pDPΔ117	13/3.2	9	17	11		17
pDP118	3/0.7	3	4	3		
pDP118R	5/0.6	6	5	5		
pDPminus	5/1.2	6	6	5	3	

*Sal I data not averaged in with supercoiled data

Deletion of some sequence downstream of the AATAAA dramatically reduces galactokinase expression in R1610 hamster lung fibroblasts although deletion of the AGGTTTTTT does not.

Figure 2 diagrams the set of deletions in the SV40 late polyadenylation region. In this paper nucleotide 1 is the first nucleotide downstream of the AATAAA. R1610 cell extracts can be assayed for gal K expression by enzyme analysis as these cells are gal K negative. Table 1 shows the results of three or more sets of enzyme assays from separate transfections into R1610 cells.

These data show that truncation of the polyadenylation site to nucleotide 49 in mutant pDPΔ116 gave a reproducible 20% drop in relative gal K activity as compared to the reference polyadenylation site in pDPSV which contains 108 nucleotides downstream of the AATAAA. Further truncation of the polyadenylation region to nucleotide 6 (pDP Δ44) showed a marked drop of 85% in relative gal K activity.

Since the pDPΔ44 mutant does not contain the natural site of cleavage, cleavage site deletion mutants were assayed to examine the effect of its removal. The mutant pDPΔ27 which deletes nucleotides 10 through 15, including the natural cleavage site, shows a 40% drop in relative activity compared to pDPSV. The mutants pDPΔ23B and pDPΔ5B which delete nucleotides 10 through 24 and 10 through 27 respectively, showed 60% decreases in activity as compared with pDPSV. Since pDPΔ116 shows a 20% drop in relative activity due to the removal of the last 60 nucleotides, the drops in expression observed from pDPΔ27, Δ23B and Δ5B are equivalent to a 20% to 40% drop in activity due to the removal of the sequence around the cleavage site or a change in the spacing of the downstream sequences from the AATAAA. Moreover, sequence between the cleavage site and nucleotide 49 probably cause some of the drop in relative gal K activity observed in pDPΔ44.

The negative control pDPminus gave relative background enzyme activities for these experiments of approximately 5%. An additional negative control, pDP118, which contains an AATAAA from the ß-lactamase gene of pBR322, also gives relative background levels of 5%. Another mutant prepared as a negative control for these experiments was pDPΔ117 which deletes 5 of the nucleotides of the AATAAA consensus sequence as well as 19 base pairs downstream of it. This mutant still shows activity which is about 10% above background.

To confirm these results and show that protein levels correlate to

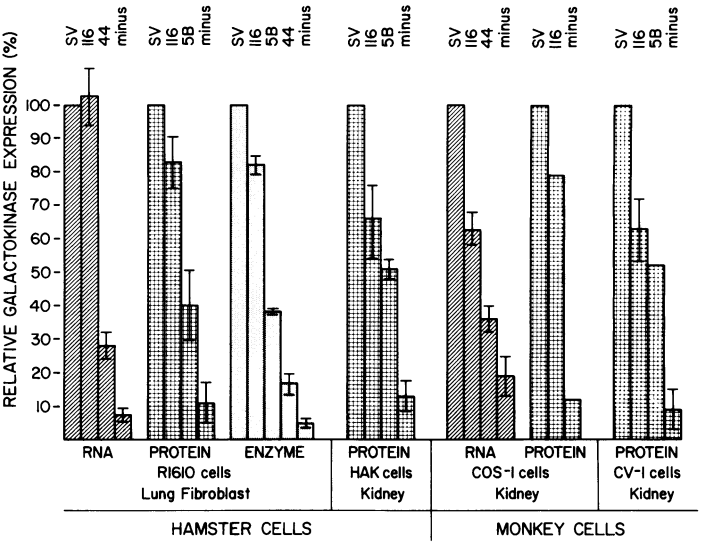


Figure 3. Bar chart relating the relative galactokinase marker gene expression between the SV40 late poly (A) region mutants in hamster and monkey cells. Bars shaded with diagonal lines represent a compilation of relative densitometric data from RNase protection assays. Cross-hatched bars represent a compilation of relative densitometric data from Western blots. Stippled bars represent a compilation of relative galactokinase enzyme levels. All assay data shown has been normalized to pDPSV vector data which is considered 100% for these experiments, therefore no error bars are shown for pDPSV. All other vector data is shown with error bars if it has been compiled from 3 or more independent transfections. No error bars are shown for data compiled from 2 or 1 transfections (except for pDPSV which has been normalized).

steady-state RNA levels in our system, cytoplasmic RNA was isolated from R1610 cells at 48 hours post-transfection and subjected to RNase protection assays in order to determine the steady-state levels of the gal K and XGPRT messages (41, 42: see methods). Figure 3 shows densitometry data tabulated from three independent RNase protection assays. Figure 4 shows a typical autoradiograph obtained from this procedure. The steady-state level of gal K message is compared to that of XGPRT in each lane. The ratios obtained may then be used to compare gal K message levels between vectors carrying various mutations in the SV40 late polyadenylation region. The relative steady-state levels of pDPSV and pDPΔ116 gal K message are approximately equal in this assay.

Deletion of all sequence downstream of nucleotide 6, as in pDPΔ44, which yielded relative gal K activities averaging 15%, showed similar

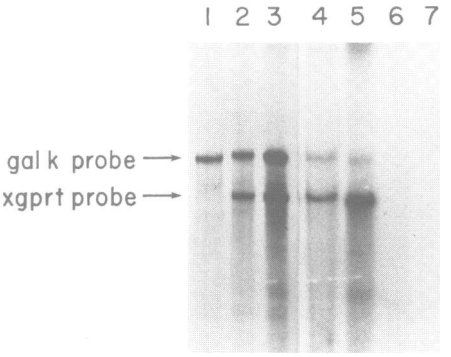


Figure 4. RNase protection assay of galactokinase and XGPRT steady-state mRNA levels from SV40 late poly (A) region mutants in R1610 hamster lung fibroblasts. 5 µg. of cytoplasmic RNA from R1610 cells isolated 48 hrs. post-transfection was either hybridized to an anti-sense galactokinase RNA probe or both galactokinase and XGPRT mixed probes. The hybridized RNAs were subjected to RNAse T1 and A and the protected fragments were run on a 10% acrylamide-8M urea denaturing gel. These fragments represent probe regions which protect message coding sequence. Lane 1: pDPΔ116, galactokinase probe only. Lane 2: pDPΔ116, galactokinase and XGPRT probes (mixed probes). Lane 3: pDPSV, mixed probes. Lane 4: pDPΔ44, mixed probes. Lane 5: pDPminus mixed probes. Lane 6: mock transfected R1610 cells, mixed probes. Lane 7: galactokinase and XGPRT probes.

relative steady-state gal K message levels of 25%. Background levels of gal K message seen in pDPminus are 10%.

Therefore, the differences seen in the relative enzyme data correlate with the differences seen in the relative RNase protection assay data.

Vector sequence downstream of the polyadenylation region mutants exerts no effect on relative galactokinase expression in hamster lung fibroblasts.

Several of the deletion mutants were examined in order to see what effect the removal of vector sequences downstream would have on their expression when compared to the supercoiled vector data. Sal I digestion cuts the vector into two pieces, separating the gal K marker gene from the SV40 late region DNA as well as the XGPRT cassette (fig. 1). The mutants still contain 28 base pairs of residual prokaryotic sequence and polylinker sequence downstream from the polyadenylation region mutants(5'-GGGAGCTTGGATTCGAGGGGGGATCCCG-3').

Following transfection, the relative gal K levels of the Sal I digested vectors were identical to those shown for supercoiled vectors (Table 1, Lane 5). The steady-state gal K mRNA levels from cells transfected with Sal I digested pDPSV and pDPΔ116 were also identical to the RNase

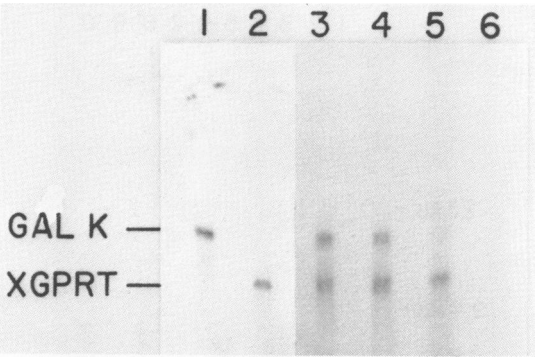


Figure 5. RNase protection assay of galactokinase and XGPRT steady-state mRNA levels from R1610 hamster fibroblasts transfected with Sal I digested vectors. 5 µg. of cytoplasmic RNA was isolated and treated as described in the legend for figure 4. Lane 1: pDPSV galactokinase probe only. Lane 2: pDPSV XGPRT probe only. Lane 3: pDPSV, mixed probes. Lane 4: pDP 116, mixed probes. Lane 5: pDPminus, mixed probes. Lane 6: mock transfected R1610 cells, mixed probes.

protection assay data obtained using supercoiled vectors (fig. 5).

Therefore, the DNA downstream of the mutant polyadenylation regions, including the SV40 late DNA, exerts no effect on the relative gal K enzyme or message levels in this system.

In order to prove that vectors transfected into R1610 hamster fibroblasts remained linear during the course of this study, high molecular weight DNA was isolated from cells transfected with Bam HI digested vectors (Bam H1 cuts the vector once) and analyzed by Southern blotting. pDSP1 and pDSP1BGH were used in this analysis. pDSP1 is the parental vector for all of the pDP mutants and pDSP1BGH is a construct which contains the bovine growth hormone polyadenylation region downstream from the gal K marker gene. HMW DNA was digested with restriction enzymes that either would not digest in the vector DNA (Bgl II) or digest once in each vector at a site distant from the Bam HI site (Eco RV site in the middle of the XGPRT gene). Southern blot analysis of the DNA showed that the Bam H1 digested vector DNA introduced into R1610 cells remained linear at 2 days post transfection (fig. 6, lanes 3 through 6). To directly compare stable versus transient vector DNA structure, cells were selected in mycophenolic acid for the stable expression of the XGPRT gene. Resistant colonies were harvested and used to prepare HMW DNA for Southern blot analysis. In contrast to the transient data, in stably selected R1610 cells at 12 days

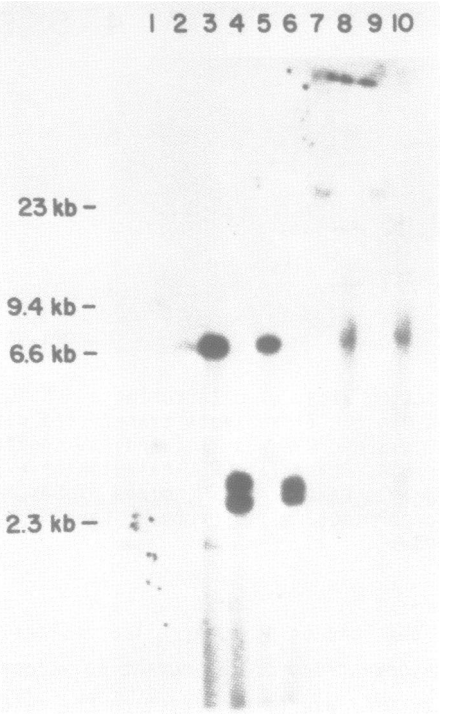


Figure 6. Southern blot analysis of total DNA isolated from R1610 hamster fibroblasts following transfection of linearized vectors. The blot was probed with nick-translated pDSP1. Lanes 1 and 2 are marker lanes - Lane 1: 25 pg. pDSP1 supercoil and nicked circular form (4 and 9 Kb respectively) mixed with 1 µg. Bgl II digested R1610 DNA. Lane 2: 25 pg. pDSP1 (7 Kb) Bam HI linear mixed with 1 µg. Bgl II digested R1610 DNA. Lanes 3 through 6 are from R1610 cells transiently transfected with vectors. DNA was isolated 2 days post-transfection - Lane 3: 1 µg. Bgl II digested DNA from R1610 cells transfected with Bam HI digested pDSP1, Lane 4: DNA as in lane 3 digested with Eco RV. Lane 5: 1 µg. of Bgl II digested DNA from R1610 cells transfected with Bam HI digested pDSP1BGH, Lane 6: DNA as in lane 5 digested with Eco RV. Lanes 7 through 10 are R1610 cells transfected with vector and selected for stable colonies in mycophenolic acid. DNA was isolated 12 days post-transfection - Lane 7: 10 µg. of Bgl II digested DNA from R1610 cells transfected with Bam HI digested pDSP1. Lane 8: DNA as in lane 7 digested with Eco RV. Lane 9: 10 µg. of Bgl II digested DNA from R1610 cells transfected with Bam HI digested pDSP1BGH. Lane 10: DNA as in lane 5 digested with Eco RV.

post transfection, all of the vector DNA had integrated into very large (>30 kb.) pieces of DNA that could not be resolved on the gel (fig. 6, lanes 7 and 9). Eco RV digestion of the same DNA produced primarily linear-sized vector molecules (fig. 6, lanes 8 and 10), indicating that the integrated vector is probably tandemly repeated.

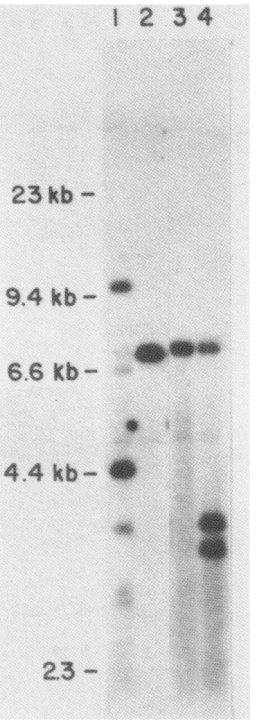


Figure 7. Southern blot analysis of DNA isolated from the purified nuclei of R1610 cells 2 days post-transfection with linearized vector. Lanes 1 and 2 are marker lanes - Lane 1: 25pg. pDSP1 supercoil and nicked circular form mixed with 1 µg. Bgl II digested R1610 DNA. Lane 2: 25pg. pDSP1 Bam HI linear mixed with 1 µg. Bgl II digested R1610 DNA. Lane 3: 1 µg. of Bgl II digested DNA isolated from the nuclei of R1610 cells transfected with Bam HI digested pDSP1. Lane 4: DNA as in lane 3 digested with Eco RV; digestion did not go to completion so higher linear band is still seen.

Since previous reports indicated that introduction of linear vectors into mammalian cells should result in the formation of recircularized or concatomerized DNA (45-50), we found it necessary to confirm our results. Nuclei were prepared from cells transiently transfected with Bam H1 restricted pDSP1 or pDSP1BGH two days post transfection. Data from Southern blot analysis proved to be identical to the results obtained from DNA prepared from whole cells (fig. 7, lanes 3 and 4). Apparently, the Bam H1 restricted vector present in the nucleus of R1610 cells following transfection remains linear for at least 2 days.

Therefore, the above comparison between vectors cut into two pieces (with Sal I) and supercoiled vectors, is probably valid because there is no

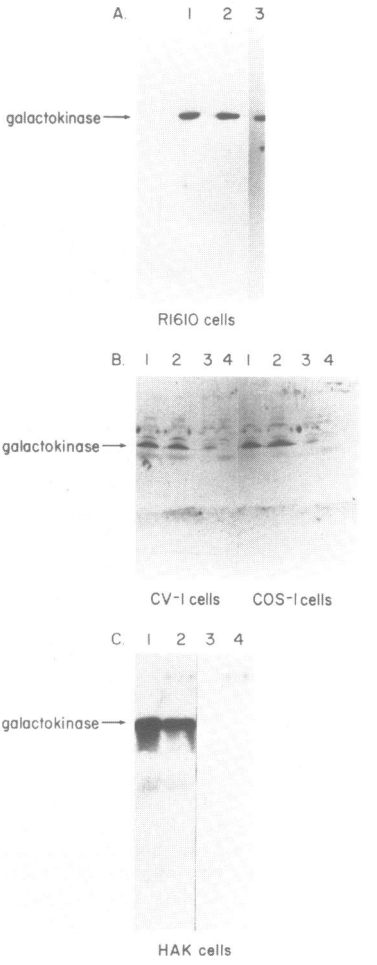


Figure 8. Western blot analyses of galactokinase levels from SV40 late polyadenylation mutants in hamster and monkey cells. Protein lysates were prepared 48 hr. post-transfection. Samples were normalized to the XGPRT enzyme levels and an equivalent of 10,000 cpm were run on 15% acrylamide-SDS gels and electro-blotted onto nitrocellulose. Blots were probed with rabbit anti-E. coli gal K IgG fraction and bound with a 125[I]-protein A reporter. Plate (A): galactokinase protein levels in R1610 hamster lung fibroblasts, Lane 1: pDPSV, Lane 2: pDPΔ116, Lane 3: E. coli galactokinase standard (10 ng.). Plate (B): galactokinase protein levels in CV-1 monkey kidney cells, Lane 1: pDPSV, Lane 2: pDPΔ116, Lane 3: pDPminus, Lane 4: mock transfected CV-1 cells. Second half of plate (B) shows galactokinase protein levels in COS-1 monkey kidney cells, Lane 1: pDPSV, Lane 2: pDPΔ116, Lane 3: pDPminus, Lane 4: mock transfected COS-1 cells. Plate (C): galactokinase protein levels in HAK hamster kidney cells, Lane 1: pDPSV, Lane 2: pDPΔ116, Lane 3: pDPminus, Lane 4: mock transfected HAK cells.

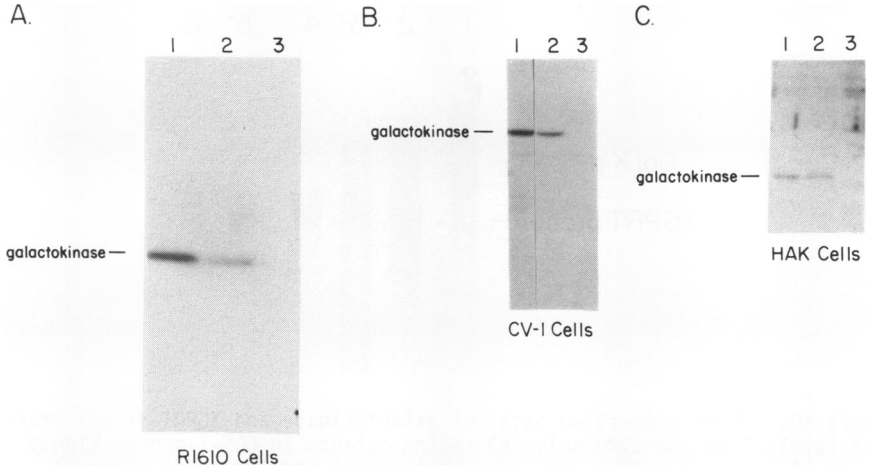


Figure 9. Western blot analyses of galactokinase levels from SV40 late poly(A) region mutants in hamster and monkey cells. Assay protocol as described previously. Plate (A): galactokinase protein levels in R1610 hamster lung fibroblasts, Lane 1: pDPSV, Lane 2: pDPΔ5B, Lane 3: pDPminus. Plate (B): galactokinase protein levels in CV-1 monkey kidney cells, Lane 1: pDPSV, Lane 2: pDPΔ5B, Lane 3: pDPminus. Plate (C): galactokinase protein levels in HAK hamster kidney cells, Lane 1: pDPSV, Lane 2: pDPΔ5B, Lane 3: pDPminus.

evidence of recombination between vectors in this cell during transient transfection.

Similar efficiencies of utilization of the SV40 polyadenylation region mutants in monkey and hamster kidney cells.

Protein lysates from COS-1 and CV-1 cells transfected with vector DNA were normalized using XGPRT assays, run on protein gels,and Western blotted onto nitrocellulose. Western blots were used to assay gal K levels in the transfected cells other than R1610 since their endogenous gal K prevented the use of enzyme assays. Figure 3 shows the tabulated results of densitometric scans from Western analyses of R1610, COS-1 and CV-1 cells. Autoradiographs of typical Western blots are shown in figure 8. In COS-1 cells deletion of downstream nucleotides 50 through 108 results in a relative drop in gal K protein to 80% of that seen in the vector pDPSV (plate B). A similar drop is seen in gal K protein data obtained from CV-1 cells with 65% of that seen in pDPSV (plate B) and HAK cells show a drop in gal K protein to 70% (plate C). The gal K levels from R1610 cells transfected with these vectors is about 80% of that seen for the control vector pDPSV (plate A). These data correlate to the enzyme activities

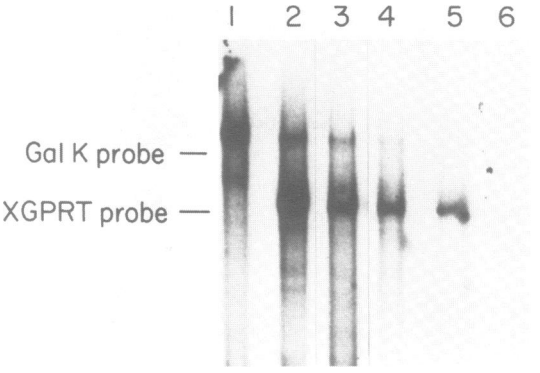


Figure 10. RNase protection assay of galactokinase and XGPRT steady-state mRNA levels from the SV40 poly (A) region mutants in COS-1 monkey kidney cells. Assay protocol as described in the text and legend of figure 4. Lane 1: pDPSV, galactokinase probe. Lane 2: pDPSV, mixed probes (gal K and XGPRT). Lane 3: pDPΔ116, mixed probes. Lane 4: pDPΔ44, mixed probes. Lane 5: pDPminus, mixed probes. Lane 6: mock transfected COS-1 cells, mixed probes. The smaller galactokinase-specific protected fragment was included in the quantitation and represents a minor stop recognized by T7 RNA polymerase (refer to lane 1). Densitometric analysis showed that this band represents 1/12 of the intensity of the upper band.

observed in R1610 cells. Therefore a comparison of pDPΔ116 to pDPSV showed a similar effect in COS-1, CV-1 and HAK cells as had previously been seen in R1610 cells.

pDPΔ5B has a 15 base-pair internal deletion which includes the natural cleavage site as well as a terminal truncation to nucleotide 49. Densitometric scanning of several Western blots was used to compare the steady-state protein levels from the parental pDPSV vector and the deletion mutant pDPΔ5B in transient transfection assays in CV-1, HAK and R1610 cells. In all of the cell lines tested the protein levels ranged between 40% and 50%. This correlates well with the gal K enzyme assays observed for R1610 cells which gave a relative expression level of approximately 40%. Figure 9, plates A through C, show typical Western blot autoradiograms of this comparison.

pDPminus gave background levels of approximately 10% for the Western blot experiments. This is similar to the 5% background level observed for the gal K enzyme assays. Untransfected cells show no band in the equivalent position to that of gal K (see Figure 8, B and C).

To affirm that the results at the protein level, comparing Western data to enzyme analyses, were also valid for steady-state cytoplasmic RNA

levels, cytoplasmic RNA was isolated from the transfected COS-1 cells 48 hours post-transfection and subjected to RNase protection analysis. In addition to comparing pDPSV, pDPΔ116, and pDPminus, another SV40 late polyadenylation region mutant, pDPΔ44, was also analyzed in order to see if the steady-state message level from this mutant was similar to the much reduced level seen in R1610 cells. Densitometric data is compiled in figure 3 and a typical autoradiograph of the analysis is pictured in figure 10. These data from COS-1 cells show a similar small drop in steady-state gal K RNA between pDPSV and pDPΔ116 as was observed in the enzyme assays and Western analyses from transfected R1610 cells. These RNA data correlate to the protein data obtained from the COS-1 cells (figure 3). We also observed an equivalent drop in relative gal K RNA for pDPΔ44 in COS-1 cells as seen in R1610 cells.

Therefore, these data confirm that the relative protein levels measured by either enzyme analysis or Western blots correlate with the relative steady-state message levels in COS-1 cells as well as in R1610 cells.

DISCUSSION

Since some of our data from the SV40 late polyadenylation region mutants examined in hamster fibroblasts differed from that generated in another lab, we wished to perform a comparative analysis of our mutants in different cell lines (1). Our data, summarized in figure 3, indicates that there are no major differences in expression between cells of different species or tissues, including cells which are permissive to the growth of the SV40 virus. We first set out to test the effect of our deletion mutants on gal K marker gene expression in R1610 cells. When we deleted all of the sequence downstream of nucleotide 49 (pDPΔ116), protein levels as assayed by Western blots and gal K activities dropped to 80% of that seen in the longer SV40 late polyadenylation region in pDPSV. This 1.2-fold difference could not be resolved by RNase protection assays of R1610 cytoplasmic RNA which gave similar results for both pDPSV and pDPΔ116. Similar effects were obtained in CV-1, COS-1 and HAK cells which, when assayed for gal K levels by Western blots, gave protein levels for pDPΔ116 in the range of 65% to 80%. RNase protection analyses of COS-1 cell RNA showed a drop to 65% for pDPΔ116. These data correlate well with data obtained in the R1610 study. So, sequences between nucleotides 49 and 108 mediate a small effect on gene expression.

It has been reported that in transient expression experiments in COS-1

cells deletion of a sequence in the SV40 late polyadenylation region, AGGTTTTTT, causes a significant lowering in correct 3'-end processing and a drop in steady-state cytoplasmic RNA levels to 25% (1). However, other experiments, conducted in Xenopus oocytes and HeLa cell nuclear extracts show that deletion of this sequence does not alter the cleavage efficiency (24,30). Furthermore, one of these studies suggests that the SV40 late polyadenylation region needs no U- or UG-rich sequences to mediate efficient cleavage of RNAs in vitro (30). Our data indicates that deletion of a large region of the SV40 late polyadenylation region, including the AGGTTTTTT, causes no observable drop in steady-state message in R1610 cells, but a small drop in gal K enzyme activity to 80% can be resolved. A similar drop can also be demonstrated at the protein level by Western blots in R1610, CV-1, COS-1 and HAK cells and at the RNA level in COS-1 cells. The reason for the discrepancy between the studies is unclear. It is possible that the residual sequence present in pDPΔ116, but not in pDPSV, contributes to lowering the effect of deleting the AGGTTTTTT. However, this sequence does little to augment the activity of the SV40 late polyadenylation region truncated to nucleotide 6 (pDPΔ44) or the AATAAA alone in pDP118. It is unlikely that our data represent accumulated message from processing within the polyadenylation region plus readthrough to a cryptic polyadenylation site, because mutants lacking a polyadenylation region or with an AATAAA sequence alone gave background levels of gal K activity and steady-state message. In addition, removal of downstream vector sequence by Sal I digestion showed no changes in relative gal K levels in hamster fibroblasts when compared to supercoiled vectors. It has also been shown that SV40 virus deletion mutants which lack the AGGTTTTTT have similar growth kinetics to the wild-type virus in CV1-P and COS-1 cells. One of these mutants showed levels of late viral mRNA and protein production indistinguishable from those of the wild-type virus (10).

A more marked decrease in gene expression was noted by deletion of all of the polyadenylation region specific sequence to nucleotide 6 which includes the wild-type cleavage site. When this mutant, pDPΔ44, was assayed in R1610 cells for enzyme activities and steady-state RNA levels, both showed drops to between 15% and 25% as compared to the control vector pDPSV. Similarly, RNase protection assays of COS-1 cell message showed drops to around 35% of pDPSV. To ascertain what role removal of the sequence embedding the wild-type cleavage site plays in deletion of the downstream sequence in pDPΔ44, we prepared and analyzed a series of

cleavage site mutants based on pDPΔ116. Removal of nucleotides 10 through 15, 24 or 27 from pDPΔ116 showed drops of 40% to 60% as compared to pDPSV. This effect was similar at the relative protein level in three cell lines from different species or tissues. Since pDPΔ116, the parental vector of this mutant series, showed a gal K activity 80% of that seen in pDPSV, it is implied that deletion of the wild-type cleavage site and sequences surrounding it contribute 20% to 40% to the gene expression in this system. However, the effect could be due to altering the position of the downstream sequences. This resembles the drop in relative cleavage efficiency observed from a deletion of 9 nucleotides including the cleavage site in the SV40 late polyadenylation region assayed in Xenopus oocyte microinjections (24). A similar effect was also seen from a linker mutant in this region assayed in HeLa cell nuclear extracts (30).

Loss of the cleavage site does not account entirely for the major drop in gene expression observed in pDPΔ44. Therefore, sequence contained between nucleotides 27 and 49 may contribute to the level of expression (see fig. 2). This region contains an AGTGTTYY sequence similar to the YGTGTTYY sequence which has been indicated to be necessary in high level marker gene expression (17). Similar downstream sequences have been described in various systems as essential 'GT-rich' (1, 21, 24, 29, 31, 32) or 'T-rich' regions (26). However, many polyadenylation regions do not contain any close variants of this sequence and are apparently functional. Also, linker mutagenesis studies have shown disparate results with regard to the function of these sequences (19, 25, 30).

pDP118 was constructed to assay if an AATAAA obtained from a prokaryotic source could mediate efficient polyadenylation in our system using R1610 cells. Another AATAAA and surrounding sequences from pBr322 has been reported to function as a cryptic polyadenylation region in vectors transfected into mammalian cells (51). Gal K activity from this mutant is equivalent to background levels. Clearly, these sequences are not sufficient for efficient gal K expression probably because of little or no processing and polyadenylation of the gal K message.

There has been speculation about whether pre-mRNA secondary structure functions in polyadenylation of cellular (7, 52) as well as viral (5, 26, 27, 29, 51, 53, 54) transcripts. We observed that the computer predicted free energy of formation for pre-mRNA secondary structures of the various SV40 late polyadenylation regions correlated to the relative gal K expression data (55, 56). This observation also held for other wild type

polyadenylation regions. The hypothesis that the extent of secondary structure at the 3'-end of the primary transcript correlates to the efficiency of 3'-end processing is currently being tested.

ACKNOWLEDGEMENTS

We are thankful for the excellent technical assistance of David Pfarr, John Trill and Antonio Toscani, and to Robert Lyons for his helpful discussion and criticism. Kenneth J. Soprano was supported by NIH grant CA 40332.

REFERENCES

1. Sadofsky, M., Connelly, S., Manley, J. and Alwine, J.C. (1985). Mol. Cell. Biol. 5(10), 2713-2719.
2. Nevins, J.R. and Darnell, J.E. (1978). Cell 15, 1477-1493.
3. Brawerman, G. (1981). CRC Crit. Rev. Biochem. 10, 1-38.
4. Proudfoot, N. and Brownlee, G.G. (1976). Nature 263, 211-214.
5. Fitzgerald, M. and Shenk, T. (1981). Cell 24, 251-260.
6. Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1983) Nature 305, 600-605.
7. Higgs, D.R., Goodburn, S.E.Y., Lamb, J., Clegg, J.B., Weatherall, D.J. and Proudfoot, N.J. (1983). Nature 306, 398-400.
8. Simonson, C.C. and Levinson, A.D. (1983). Mol. Cell. Biol. 3(12), 2250-2258.
9. Wickens, M. and Stephenson, P. (1984). Science 226, 1045-1051.
10. Swimmer, C. and Shenk, T. (1985). Nucl. Acids Res. 13(22), 8053-8063.
11. Zarkower, D., Stephenson, P., Sheets, M. and Wickens, M. (1986). Mol. Cell. Biol. 6(7), 2317-2323.
12. Cole, C.N. and Santangelo, G.M. (1983). Mol. Cell. Biol. 3(2), 267-279.
13. McReynolds, L., O'Malley, B.W., Nisbit, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M., and Brownlee, G.G. (1978). Nature 273, 723-728.
14. Perricaudet, M., LeMoullec, J.M., Tiollais, P. and Peterson , U. (1980). Nature 288, 174-176.
15. Lai, E.C., Stein, J.P., Catterall, J.F., Woo, S.l.C., Mace, M.L., Means, A.R. and O'Malley, B.W. (1979). Cell 18, 829-842.
16. Gil, A. and Proudfoot, N.J. (1984). Nature 312, 473-475.
17. McLaughlan, J., Gaffney, D., Whitton, J.L. and Clemments, J.B. (1985). Nucl. Acids Res. 13(4), 1347-1368.
18. Hart, R.P., McDevitt, M.A., Ali, H. and Nevins, J.R. (1985). Mol. Cell. Biol. 5(11), 2975-2983.
19. Mason, P.J., Elkington, J.A., Lloyd, M.M., Jones, M.B. and Williams, J.G. (1986). Cell 46, 263-270.
20. Hart, R.P., McDevitt, M.A. and Nevins, J.R. (1985). Cell 43, 677-683.
21. Bhat, B.M. and Wold, W.M. (1985). Mol. Cell. Biol. 5(11), 3183-3193.
22. Kessler, M.M., Beckendorf, R.C., Westhafer, M.A. and Nordstrom, J.L. (1986). Nucl. Acids Res. 14(12), 4939-4952.
23. Sperry, A.O. and Berget, S. (1986). Mol. Cell. Biol. 6(12), 4734-4741.
24. Conway, L. and Wickens, M. Proc. Natl. Acad. Sci. USA (1985). 82, 3949-3953.
25. Zhang, F., Denome, R. and Cole, C.M. (1986). Mol. Cell. Biol. 6, 4611-4623.

26. McDevitt, M.A., Imperiale, M.J., Ali, H. and Nevins,J.R. (1984). Cell 37, 993-999.
27. Woychick, R.P., Lyons, R.H., Post, L. and Rottman, F. (1984). Proc. Natl. Acad. Sci. USA 81, 3944-3948.
28. Cole, C.N. and Stacy, T.P. (1985). Mol. Cell. Biol. 5(8), 2104-2113.
29. Gil, A. and Proudfoot, N.J. (1987). Cell 49, 399-406.
30. Zarkower, D. and Wickens, M. (1988). J. Biol. Chem. 263(12), 5780-5788.
31. Hart, R.P. and Green, T.L. (1988). Mol. Cell. Biol. 8(4), 1839-1841.
32. Skolnick-David, H., Moore, C. and Sharp, P. (1987). Genes & Development 1(7), 672-682.
33. Pfarr, D.S., Sathe, G. and Reff, M.E. (1985). DNA 4(6), 461-467.
34. Pfarr, D.S., Rieser, L.A., Woychick, R.P., Rottman, F.M., Rosenberg, M. and Reff, M.E. (1986). DNA 5(2), 115-122.
35. Maxam, A.E. and Gilbert, W.S. (1980). Methods Enzymol. 65, 499-560.
36. Thirion, J.P., Bannville, D. and Noel, H. (1976). Genetics 63, 137-147.
37. Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer, A., Lacey, E., Maniatis, T., Silverstein, S. and Axel, R. (1979). Cell 16, 777-785.
38. Peterson, J.L. and McBride, O.W. (1980). Proc. Natl. Acad. Sci. USA 77, 1583-1587.
39. McKenney, K., Shimatake, H., Court, D., Schmeisner, U., Brady, L. and Rosenberg, J. (1981). Gene Amplification and Analysis, Chirikjian, J.S. and Papas, T.S., Eds. (Elsevier/North-Holland, New York).
40. Johansen, H., Schumperli, D. and Rosenberg, M. (1984). Proc. Natl. Acad. Sci. USA 81, 7698-7702.
41. Tushinski, R.J., Sussman, P.M., Yu, L.Y. and Bancroft, F.L. (1977). Proc. Natl. Acad. Sci. USA 74, 2357-2361.
42. Melton, D.A., Kreig, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green,. M.R. (1984). Nucl. Acids Res. 12, 7035-7056.
43. Toscani, A., Soprano, D.R., Cosenza, S.C., Owen, T.A. and Soprano, K.J. (1987). Anal. Bio. 165, 309-319.
44. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
45. Folger, K.R., Wong, E.A., Wahl, G. and Capecchi, M.R. (1982). Mol. Cell. Biol. 2, 1372-1387.
46. Miller, C.K. and Temin, H. (1983). Science. 220, 606-609.
47. Perucho, M., Hanahan, D. and Wigler, M. (1980). Cell. 22, 309-317.
48. Robins, D.M., Ripley, S., Henderson, A.S., and Axel, R.,(1981). Cell 23, 29-39.
49. Scangos, G.A., Huttner, K.M., Juricek, D.K. and Ruddle, F.H. (1981). Mol. Cell. Biol. 1, 111-120.
50. Wilson, J.H., Berget, P.B. and Pipas, J.M. (1982). Mol. Cell. Biol. 2, 1258-1269.
51. Kessler, M.M., Westhafer, M.A., Carson, D.D. and Nordstrom, J.L. (1987). Nucl. Acids Res. 15(2), 631-642.
52. McGeady, M.L., Wood, T.G. and Vande Woude (1986). DNA 5(4), 289-298.
53. Goins, W.F. and Stinski, M.F. (1986). Mol. Cell. Biol. 6(12), 4202-4213.
54. LeMoullec, J.M., Akusjarvi, G., Stalkandske, P., Peterson , U., Chambraud, B., Gilardi, P., Narsi, M. and Perricaudet, M. (1983). J. Virol. 48, 127-134.
55. Zucker, M. and Steigler, P. (1981). Nucl. Acids Res. 9, 133-148.
56. Salser, W. (1977). Cold Spring Harbor Symp. Quant. Biol. 42, 985-1002.