MOLECULAR AND CELLULAR BIOLOGY, July 1983, p. 1295–1309
0270-7306/83/071295-15$02.00/0

Vol. 3, No. 7

Transient Expression of a Mouse α-Fetoprotein Minigene: Deletion Analyses of Promoter Function

RICHARD W. SCOTT AND SHIRLEY M. TILGHMAN*
Institute for Cancer Research, Philadelphia, Pennsylvania 19111

Received 12 January 1983/Accepted 21 April 1983

The constitutive transcription of a mouse α-fetoprotein (AFP) minigene was examined during the transient expression of AFP-simian virus 40-pBR322 recombinant DNAs introduced into HeLa cells by $Ca_3(PO_4)_2$ precipitation. We tested three constructs, each of which contains the AFP minigene and pBR322 DNAs inserted in the late region of simian virus 40 and found that the relative efficiency of AFP gene expression was dependent on the arrangement of the three DNA elements in the vector. The transcripts begin at the authentic AFP cap site and are properly spliced and polyadenylated. To define a sequence domain in the 5′ flanking region of the AFP gene required for constitutive expression, sequential 5′ deletion mutants of the AFP minigene were constructed and introduced into HeLa cells. All AFP deletion mutants which retained at least the TATA motif located 30 base pairs upstream from the cap site were capable of directing accurate and efficient AFP transcription. However, when the TATA sequence was deleted, no accurately initiated AFP transcripts were detected. These results are identical to those obtained from in vitro transcription of truncated AFP 5′ deletion mutant templates assayed in HeLa cell extracts. The rate of AFP transcription in vivo was unaffected by deletion of DNA upstream of the AFP TATA box but was greatly affected by the distance between the simian virus 40 control region and the 5′ end of the gene. The absence of any promoter activity upstream of the TATA box in this assay system is in contrast to what has been reported for several other eucaryotic structural genes in a variety of in vivo systems. A sequence comparison between the 5′ flanking region of the AFP gene and these genes suggested that the AFP gene lacks those structural elements found to be important for constitutive transcription in vivo. Either the AFP gene lacks upstream promoter function in the 5′ flanking DNA contained within the minigene, or the use of a viral vector in a heterologous system precludes its identification.

α-Fetoprotein (AFP) and albumin are the major serum proteins of fetal and adult mammals, respectively (1, 42). In mice, AFP and albumin are encoded by unique copy genes closely linked on chromosome 5 (9). Both genes are in the same orientation relative to the direction of transcription, with the albumin gene 13.5 kilobases (kb) 5′ to the AFP gene (26). In developing mouse livers, AFP and albumin mRNAs accumulate in parallel until birth. After birth, a rapid decline in AFP mRNA commences that continues until approximately 2 weeks of postnatal life, when a basal mRNA level is reached and thereafter maintained. During this time, the intracellular concentration of albumin mRNA remains high and relatively unchanged (47).

The accumulation and selective decline in AFP mRNA in the liver has been shown to occur via alterations in its rate of transcription (47). To identify the DNA sequences responsible for the modulation of AFP gene transcription, it is first necessary to identify the sequences involved in the constitutive expression of the gene, that is, those elements which affect both the accuracy and efficiency of transcription. Once these regions are defined, it is possible to ask whether they can be distinguished from regions involved in tissue-specific regulation. To date, two approaches have been used to analyze constitutive gene transcription. Initially, cloned templates were transcribed in heterologous soluble cell-free extracts by RNA polymerase II (31, 32, 51). In most instances, the only sequence required for proper in vitro initiation appeared to be the TATA or Goldberg-Hogness box (8, 20, 33, 50), a highly conserved sequence located approximately 30 base pairs (bp) upstream from the transcription initiation site of many eucaryotic genes (5). However, with the use of homologous cell extracts (48) or circular DNA templates (17), sequences upstream of the TATA box could be discerned which quantitatively affected in vitro transcription rates.

There is now evidence which suggests that in

vitro transcription systems overemphasize the role of the TATA sequence in the function of the eucaryotic promoter. Studies with various in vivo transcriptional systems have shown that sequences 5′ to the TATA box, as well as the TATA box itself, demonstrate promotor function. In vivo, deletion of the TATA box generally causes a change in the specificity of the transcription initiation sites (4, 15) and in some instances reduces transcriptional frequency as well (4, 19). Deletion or sequence substitution of 5′ flanking DNA located between 40 and 120 bp upstream of the structural gene (6, 18, 36, 37) and in some cases in regions further upstream (4, 16, 21) results in a sharp decrease in the transcriptional efficiency but does not affect initiation specificity.

Because these in vivo systems are capable of defining a broader range of promotor functions, we analyzed the constitutive transcription of the AFP gene in intact animal cells. Previously, Banerji and co-workers (3) described a system wherein the rabbit β-globin gene, inserted into a pBR322-simian virus 40 (SV40) recombinant vector, is expressed from its own promoter during a transient period in transfected HeLa cells. We used similar pBR322-SV40 vectors containing intact and 5′-deleted AFP templates to define the 5′ flanking regions required for accurate and efficient AFP transcription in HeLa cells.

MATERIALS AND METHODS

DNA sequencing. The nucleotide sequences of overlapping restriction enzyme fragments of the mouse AFP 5′ genomic subclone pAFP14Z.1 (53) and the albumin 5′ genomic subclone pa1b5E (28) were determined by the method of Maxam and Gilbert (34).

In vitro transcription. Whole cell extracts from exponentially growing spinner cultures of HeLa cells were prepared as described by Manley et al. (32). The 50-μl reaction mixtures contained 12 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid), (pH 7.9); 60 mM KCl, 7.5 mM $MgCl_2$; 0.6 mM EDTA; 1.2 mM dithiothreitol; 10% glycerol; 600 μM each CTP, ATP, and UTP; 50 μM GTP; 50 μCi of [α-^{32}P]GTP (ICN; 350 Ci/mmol); 2.5 μg of DNA; and 30 μl of extract. The reactions were incubated at 30°C for 60 min and terminated by adjusting to 50 mM sodium acetate (pH 5.2)–0.5% sodium dodecyl sulfate in a total volume of 200 μl. After two extractions with equal volumes of phenol-chloroform (1:1) and one each with chloroform and ether, the aqueous phase was adjusted to 0.3 M sodium acetate (pH 7.2), carrier yeast tRNA was added to 50 μg/ml, and the samples were precipitated by the addition of 2 volumes of ethanol. The pellets were washed by resuspension in 0.4 M sodium acetate and reprecipitated with ethanol. The final dried pellets were suspended in 15 μl of 70% formamide–1× TBE (90 mM Tris-borate [pH 8.3], 4 mM EDTA)–0.1% xylene cylanol and heated at 68°C for 2 min. The denatured RNA transcripts were resolved on a 7 M urea–5% polyacrylamide gel in 1× TBE buffer.

Transient expression assays. Transient expression assays were done as described by Grosveld et al. (18) with several modifications. On day 1, HeLa cells were seeded at 3×10^6 cells per 10-cm petri plate in Dulbecco modified Eagle medium–10% fetal calf serum (GIBCO). On day 2, 15 μg of vector DNA per plate in the presence of $Ca_3(PO_4)_2$ was added to the cells for 10 h. The monolayer cultures were then treated with 15% glycerol in 1× Hanks buffered saline for 3 min at room temperature, washed once with Dulbecco modified Eagle medium, and maintained in Dulbecco modified Eagle medium–10% fetal calf serum for 55 h. The cells were harvested by trypsinization, and RNA was prepared by the hot phenol method (43) and enriched for polyadenylated [poly(A)$^+$] species by oligodeoxythymidylate-cellulose chromatography (2).

S1 nuclease hybridization assays. The S1 nuclease protection assay to map the 5′ terminus of the AFP transcripts was done as described by Favoloro et al. (11). Briefly, the RNA samples were suspended in 40 μl of the hybridization mixture containing 80% formamide, 40 mM PIPES [piperazine-*N*,-N′-bis(2-ethanesulfonic acid)] (pH 5.2), 400 mM NaCl, 1 mM EDTA, and 100 μg of the heat-denatured *Sau*3A/*Pvu*II probe labeled at the *Sau*3A site with [γ-^{32}P]ATP (New England Nuclear Corp.; 1,000 Ci/mmol) and polynucleotide kinase. The samples were incubated at 52°C for 15 to 16 h, and 20-μl portions were digested with S1 nuclease (*Aspergillus oryzae*; Miles Laboratories) in 0.3-ml digestion mixtures containing 30 mM sodium acetate (pH 4.5), 300 mM NaCl, 3 mM $ZnCl_2$, and 10 μg of denatured calf thymus DNA per ml. The reactions were stopped by the addition of ammonium acetate and EDTA to final concentrations of 800 and 15 mM, respectively, and precipitated with isopropanol. After several washes, the pellets were suspended in 80% formamide–1× TBE–0.1% bromophenol blue-xylene cylanol, denatured at 90°C for 2 min, and electrophoresed on a 7 M urea–10% polyacrylamide gel in 1× TBE buffer.

The S1 nuclease hybridization assay with unlabeled AFP minigene DNA as the hybridization probe was performed as described previously (53).

Preparation of AFP 5′ deletion mutants. The AFP 5′ deletion mutant ZE2Δ1 was constructed by partially digesting the AFP minigene pZE2 with *Pvu*II to obtain a significant proportion of linearized plasmid. *Bam*HI linkers were blunt-end ligated onto the free ends, the fragments were digested with an excess of *Bam*HI, and the products were ligated at 5 μg/ml. The ligation samples were used to transform HB101 cells (49), and ampicillin-resistant colonies were screened by colony hybridization for plasmids in which the 400-bp *Bam*HI/*Pvu*II fragment of pZE2 had been deleted.

The AFP 5′ deletion mutants ZE2Δ3, ZE2Δ7, ZE2Δ8, ZE2Δ9, and ZE2Δ10 were constructed by *Bal*31 digestion of ZE2Δ1. Twelve micrograms of ZE2Δ1 DNA was digested with 2 U of *Bal*31 (Bethesda Research Laboratories) in a total volume of 100 μl at 30°C for 2, 4, and 5 min. The reactions were stopped by the addition of EDTA to 50 mM and sequentially extracted with equal volumes of chloroform, phenol, and ether. The fragments were blunt-end ligated to *Bam*HI linkers, digested with an excess of *Bam*HI,

and religated at 10 μg/ml. HB101 cells were transformed with the ligation products, and plasmids derived from ampicillin-resistant transformants were screened on the basis of size. Plasmids containing deletions of interest were precisely mapped by DNA sequencing from the newly constructed *Bam*HI site.

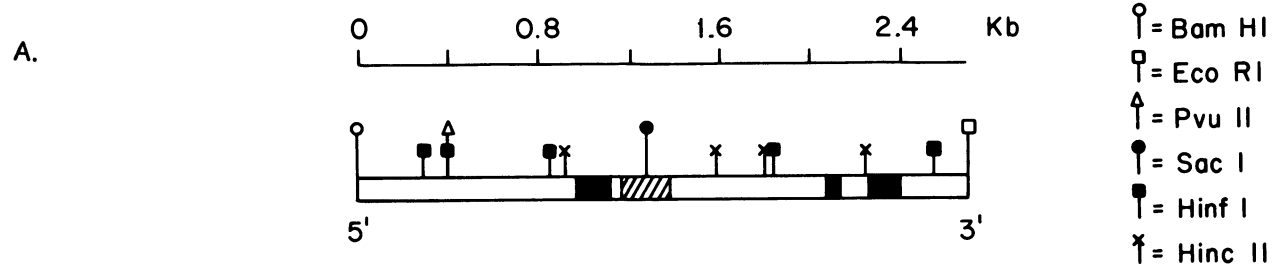
Preparation of transient expression vectors. The SV40-pBR322 recombinant plasmid used in the construction of the pSV1ZE2 transient expression vector is termed pSV1 and was prepared as follows. The plasmid pSV (kindly provided by R. Baserga, Temple University) contains the 4,490-bp *Eco*RI-*Bam*HI fragment of SV40 inserted into pBR322 (12). This plasmid was partially digested with *Pvu*II so that a large proportion of the plasmid was linearized. In a second reaction, pBR322 DNA was digested to completion with *Pvu*II. *Sal*I linkers were blunt-end ligated onto the free ends of both DNA samples, and the fragments were digested with *Bam*HI and an excess of *Sal*I. Equimolar concentrations of the two samples were ligated at a final concentration of 25 μg/ml, and the ligation products were used to transform HB101 cells. Ampicillin-resistant colonies were screened by colony hybridization for plasmids which contained the 2,672-bp pBR322 fragment (extending from the *Bam*HI site at 375 nucleotides [nt] [45] counterclockwise to the *Sal*I-converted *Pvu*II site at 2067 nt) linked to the 2,978-bp SV40 fragment (extending from the *Sal*I-converted *Pvu*II site at map position 0.711 to the unique *Bam*HI site at position 0.144 [29]). Proper orientation of these fragments was confirmed by restriction enzyme analysis. The minigene template ZE2 was inserted into the pSV1 vector by replacing the 375-bp *Eco*RI/*Bam*HI pBR322 fragment with the *Bam*HI/*Eco*RI minigene fragment.

The SV40-pBR322 recombinant plasmid used in the construction of the pSV2ZE2 transient expression vector is termed pSV2 and was prepared in two stages. The first stage involved construction of a recombinant plasmid containing unique *Bam*HI and *Eco*RI sites at the late side of SV40 so that AFP minigene templates could be conveniently inserted in the desired orientation. pSV DNA was partially cleaved with *Pvu*II, *Bam*HI linkers were blunt-end ligated onto the free ends, and the fragments were digested to completion with *Pvu*II and *Bam*HI. *Sal*I linkers were then blunt-end ligated onto the newly generated *Pvu*II-cleaved termini. After digestion with an excess of *Sal*I, the fragments were religated at a concentration of 25 μg/ml and used to transform HB101 cells. Colony hybridization and restriction enzyme mapping of ampicillin-resistant transformants identified a plasmid (pSVt) with a 2,297-bp pBR322 fragment extending from the *Eco*RI site at 0 counterclockwise to the *Sal*I-converted *Pvu*II site at 2067, linked to an internally deleted SV40 fragment which included DNA from the *Sal*I-converted *Pvu*II site at map position 0.33 clockwise to the *Bam*HI-converted *Pvu*II site at 0.711. The *Pvu*II fragment C between 0.711 and 0.987 had been deleted, leaving a 66-bp SV40 *Bam*HI/*Eco*RI fragment which could be replaced with the *Bam*HI/*Eco*RI minigene fragments. This recombinant plasmid, however, has a truncated early region, and because we have obtained maximum AFP expression in HeLa cells from vectors with an intact early region, the second stage of the vector preparation involved the insertion of the entire SV40 early region into the pSVt plasmid. This was done by cleaving pSV DNA with *Bam*HI, filling in the *Bam*HI site with reverse transcriptase, and blunt-end ligating *Sal*I linkers onto the free ends. The fragments were digested with *Sal*I and *Bgl*I and ligated at equimolar concentrations to a partial *Bgl*I digest of pSVt DNA at a final concentration of 25 μg/ml. The ligation products were used to transform HB101 cells, and ampicillin-resistant colonies were screened by colony hybridization for plasmids in which the 1,729-bp SV40 fragment of pSVt extending from the *Bgl*I site to the *Sal*I-converted *Pvu*II site at map position 0.330 had been replaced by the 2,702-bp SV40 fragment which extends from the *Bgl*I site at position 0.661 to the *Sal*I-converted *Bam*HI site, which contains the entire SV40 early region. The identity of the final SV40-pBR322 recombinant (pSV2) was confirmed by restriction enzyme analysis.

The SV40-pBR322 recombinant plasmid (pSV3) used in the construction of the pSV3ZE2 transient expression vectors has been described earlier (39) and was kindly provided by D. Hamer (National Cancer Institute). This plasmid contains the *Hpa*II/*Bam*HI fragment of the early region of SV40 *dl*2005, which has a 230-bp deletion in the intervening sequence of the large T antigen gene (44). The *Hpa*II and *Bam*HI sites have been converted to *Eco*RI sites and are inserted into the *Eco*RI site of pBR322. The AFP minigene templates were inserted in the pSV3 vector by replacing the 375-bp *Eco*RI/*Bam*HI pBR322 fragment with the *Bam*HI/*Eco*RI minigene fragments.

Construction of pSV2ZE2Δ7A and pSV2ZE2Δ7B transient expression chimeras. The AFP deletion mutant Δ7A was constructed by deleting all AFP sequences 5′ of the *Hin*fI site at position −85 in the AFP 5′ flanking region. The deletion mutant Δ7B was constructed by deleting the AFP sequences 5′ of the *Hin*cII site at position −52. pZE2 DNA was digested with either *Hin*fI or *Hin*cII, and in the former case the ends were rendered blunt ended by filling in the 5′ extended termini with reverse transcriptase. *Bam*HI linkers were blunt-end ligated onto the free ends, and, after digestion with *Sac*I (which cleaves the plasmid DNA only once in the first intervening sequence of the minigene) and *Bam*HI, each sample was ligated at equimolar concentrations to *Bam*HI/*Sac*I-cleaved pSV2ZE2 DNA at a final concentration of 25 μg/ml. The ligations were used to transform HB101 cells, and ampicillin-resistant colonies were screened by colony hybridization for plasmids in which the 880-bp *Bam*HI/*Sac*I fragment of pSV2ZE2 had been replaced by either the 370-bp *Bam*HI-converted *Hin*fI/*Sac*I fragment (Δ7A) or the 335-bp *Bam*HI-converted *Hin*cII/*Sac*I fragment (Δ7B) of pZE2. The final identities of pSV2ZE2Δ7A and p5V2ZE2Δ7B were confirmed by restriction enzyme analysis.

Preparation of UE#1 and UE#2. The upstream element (UE) fragments UE#1 and UE#2 were prepared from pZE2Δ1 by subcloning derivatives of the *Bam*HI/*Hin*cII fragment of the AFP 5′ flanking DNA which had been deleted in a 3′ to 5′ direction from the *Hin*cII site by *Bal*31 digestion. pZE2Δ1 DNA was cleaved with *Hin*cII and digested with *Bal*31, and *Bam*HI linkers were ligated onto the free ends. After digestion of the samples with an excess of *Bam*HI, the products were ligated into the *Bam*HI site of pBR322 and used to transform HB101 cells. The *Bam*HI fragments of interest were identified by screening ampicil-



B.

```
-600                -580                -560                -540                -520                -500
ACCCATGCACCTGTGACATACATATGGAATTATTCTTTGGCACATAAAATTAGAATGGGAGCTGGCTCATCAGGTTTTGTGCTGTAAGTTTTCTATGTTAAACCAGATGCGATACACTAA
|Δ1
-480                -460                -440                -420                -400                -380
ATAAAATAAAATATACTTGACCGATGGTTTTGAGCGAAATAATAACTGGATAATCAAGAAATATATCCACTAATGAATAGCCTGACTACTGAAACAATTTGTTCAGTGCCTAGCATATGG
                                                                                                UE#2|
-360                -340                -320                -300                -280                -260
TGTGCATTTTATTATTTCTTTCAAAAAGAATGTATTTGGAGTTACATAGTAAGTCTGCTACCTTTTCTTTATGGCTATATCTATGTCTTATGTTGAGATGAATGAATTATTCTTCAGGGG
                                                                                                                    |Δ3
-240                -220                -200                -180                -160                -140
AAATAATCTATTTGAACAGTTTAGATGGTGAAGAACATTTGCAGCATTTGCAAGATTTTTTTCCACTCTGAAGTGGTCTTTGTCCTTGAACATAGGATACAAGTGACCCCTGCTCTGTTA
                                                                                    UE#1|
-120                -100                -80                 -60                 -40                 -20
ATTATTGGCAAATTGCCTAACTTCAACGTAAGGAAATAGAGTCATATGTTTGCTCACTGAAGGTTACTAGTTAACAGGCATCCCTTAAACAGGATATAAAAGGACTTCAGCAGGACTGCT
|Δ7                 |Δ7A                |Δ7B                |Δ8                 |Δ9                 |Δ10
1                   20                  40                  60                  80
cap site
CGAA ACATCCCACTTCCAGCACTGCCTGCGGTGAAGGAACCAGCAGCC ATG AAG TGG ATC ACA CCC GCT TCC CTC ATC CTC CTG CTA CAT TTC GCT GCG
                                                  Met Lys Trp Ile Thr Pro Ala Ser Leu Ile Leu Leu Leu His Phe Ala Ala
100                 120                 140                 160                 180                 200
                                    3' splice junction                  3' alu
TCC AAA GCA TTG CAC GAA AAT GAG TTT GGG ATA G GTAAGATGCTTTGTGACTTGGGAGCTATTTTCCTATAAC AAAATTCTTAAAAATTCCCATATTTGCATCTCCA
Ser Lys Ala Leu His Glu Asn Glu Phe Gly Ile G
220                 240                 260                 280                 300                 320
NGTATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAGACAGGGTTTCTCTGTGTAGCCCTGCTGTCCTGAGCTCACTTTGTAGACCAGCTGGCCTCGAACTCAGAAATC
340                 360                 380                 400                 420
                                    5' alu
CGCCTGCCTCTGCCTCCCGAGTGCTGGGATTAAAGGTGTGCGCCACCACACCCGGCTCTCN AGGTATCTTGACTTANGCATTCTATATACTGTTGAAGACAGGTAT
```

C.

```
                    -80                                 -40                                 +1
CTTC AACGTAAGGAAATA-GAG TCATATGTTTGCTCACTGAA GGTTACTAGTTAACAGGCATCCCTTAAACAGGATATAAAA GGACTTCAGCAGGACTGCTCGAAA   AFP
TAGG AACCAAT-GAAATGCGAG      GTAAGTAT       GGTTAATGATCTACAGTTATTGGTTAAAGAAG-TATATTA GAGCGAGTCTTTCTGCACACAGA   albumin
-80                                 -40                                 +1
```

lin-resistant colonies by colony hybridization with an AFP 5′ flanking DNA probe. UE#1 and UE#2 were chosen on the basis of their size, and the precise 3′ borders of the fragments were determined by DNA sequencing.

RESULTS

Sequences at the 5′ region of the AFP gene. The sequences surrounding the 5′ end of the mouse AFP gene were determined (Fig. 1A). The cap site for the mature AFP mRNA, labeled +1, had been previously determined (10), and, based on evidence obtained for other eucaryotic and viral genes (7, 22, 24), we assume that this represents the initiation site of transcription. The sequence of the first coding block (10, 14) and a portion of the first intron is included to show the position of an *Alu*-like repeat element located within the AFP gene (53). This repeat element is transcribed from the opposite strand of the AFP gene in vivo, as discussed below.

The 5′ flanking sequence contains the sequence TATAAAA located at position −30, but another conseerved sequence, a CCAAT pentamer, found approximately 80 bp upstream from many genes transcribed by RNA polymerase II, is not present at this position in the AFP 5′ flanking region, although similar sequences, such as GCAAAT at −117 and GGAAAT at −93, are found within the same general region. In an effort to identify any further sequence elements of potential regulatory significance, the 5′ flanking sequence of mouse AFP was compared to that of mouse albumin to determine whether there had been conservation of any sequence within this region. As shown by the boxes in Fig. 1B, significant sequence homologies between the two related genes are evident: one extending for 30 bp 5′ of the TATA boxes and a second that includes the CCAAT motif of the albumin gene at −84. Thus, although the AFP gene lacks a CCAAT motif itself, it exhibits significant homology at −100 to this region of the albumin gene.

Transcription of the AFP minigene in HeLa cells. To define a sequence domain in the 5′ flanking region of the AFP gene required for its constitutive expression, cloned AFP templates were introduced into intact HeLa cells via a pBR322-SV40 vector. The mouse AFP gene is composed of 15 coding blocks spanning 20 kb of DNA (28) and is therefore not represented in its entirety in any one phage clone. To obtain an AFP transcription template in a form that could be easily introduced into a eucaryotic cell, an internally deleted AFP minigene was constructed by joining the first two coding blocks of AFP along with 1.0 kb of 5′ flanking sequence to the last coding block and 0.4 kb of 3′ flanking DNA (53). By using SV40-AFP viral recombinants, we have shown that the minigene transcripts are correctly initiated, spliced, and polyadenylated in lytically infected African green monkey kidney cells. Because the generation of the SV40-AFP recombinant viral stocks is relatively time consuming and frequently results in significant rearrangement of the inserted AFP DNA, we chose instead to study AFP transcription during the transient expression of AFP-SV40-pBR322 recombinants after their introduction into HeLa cells.

The chimeras, which were constructed as outlined in Materials and Methods and are diagrammed in Fig. 2, are similar to those used by Banerji et al. (3) and Grosveld et al. (18). All three contain AFP minigene and pBR322 DNAs inserted into the late region of SV40 but differ from one another in the relative orientations of the various components. Each contains the entire early region of SV40, the SV40 origin of replication, both copies of the 72-bp repeats, SV40 late transcription initiation sites, and the late polyadenylation site.

The first series of experiments utilized the pSV2 vector, in which the AFP minigene is inserted into the vector adjacent to and in the same transcriptional orientation as the SV40 late promotor (Fig. 2 and 3). A deletion of the minigene, termed Δ1, which contains 600 bp of 5′ flanking DNA, was constructed and introduced into HeLa cells by $Ca_3(PO_4)_2$ precipitation (52), and the cells were harvested 55 to 60 h later. Cellular RNA was extracted and analyzed for specific AFP transcription by S1 nuclease hybridization assays with a DNA probe extending from the 5′ end of Δ1 to a ^{32}P-labeled *Sau*3A site within the first exon (Fig. 3). The DNA

FIG. 1. Nucleotide sequences of the 5′ region of the AFP and albumin genes. (A) Restriction map of the 2.7-kb AFP minigene, with the three AFP coding blocks indicated by the black rectangles and the *Alu*-like gene indicated by the striped box. The restriction sites used in the construction of various mutants are indicated. (B) The contiguous sequence of 600 bp of AFP 5′ untranscribed DNA (−1 to −604), the first exon of the transcribed gene (+1 to +129), and a portion of the first intervening sequence (+130 to +428) are depicted. The TATAAA box at −30 is underlined. The translated codons of the first exon are arranged in triplet array. The positions of the cap site at +1, the 3′ splice junction of the first exon at +129, and the 5′ and 3′ borders of the *Alu*-like gene on the anticoding strand (+381 and +168, respectively) are also shown. The extent of 5′ flanking DNA in deletions Δ1 to Δ10 are shown, as well as the 3′ ends of UE#1 and UE#2. (C) The 5′ untranscribed regions of the mouse AFP and albumin are aligned with the cap sites at +1 on the right. The boxes delineate regions of significant homology between the genes, as illustrated by the hatch marks.

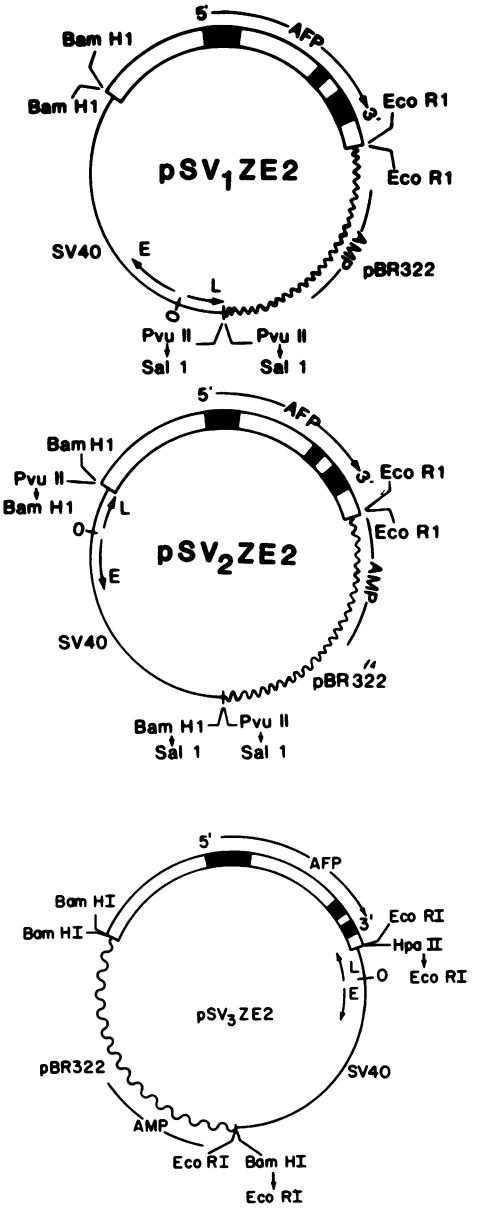


FIG. 2. Structures of three SV40-pBR322-AFP chimeras. The structures of the three chimeras used in the transient expression assays are illustrated. In each case, the solid lines represent SV40 DNA, with O delineating the origin of replication and E and L the direction of early and late transcription. The wavy lines represent pBR322 DNA, with the ampicillin (AMP) resistance gene designated. The AFP minigene is represented by the boxes, with the positions of the three exons indicated by the closed boxes. The direction of transcription is shown by the 5′ → 3′ symbols. The restriction sites in each segment used to construct the vectors are illustrated before the arrows; the linkers used follow each arrow. Details of construction are given in the text.

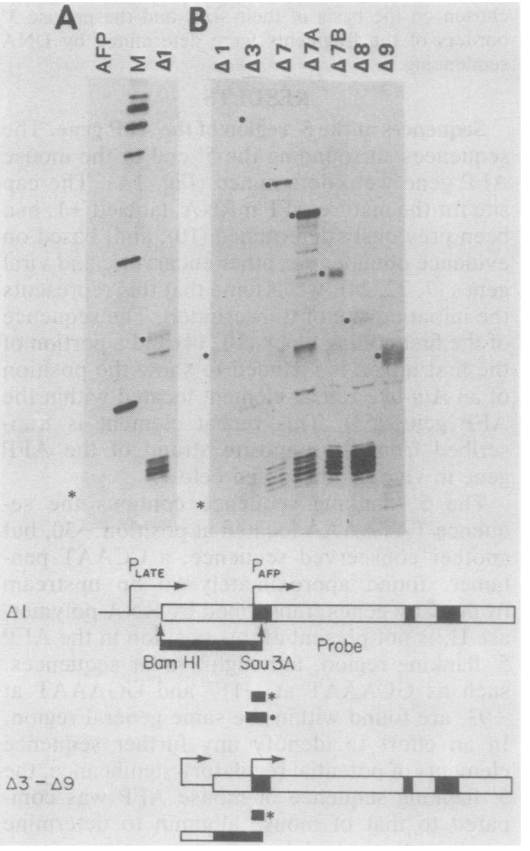


FIG. 3. Transient expression of AFP minigene derivatives in HeLa cells. (A) pSV2ZE2Δ1 DNA was transfected into HeLa cells, and poly(A)$^+$ RNA was isolated 60 h later. Twenty micrograms of RNA was hybridized to an excess of a *Pvu*II-*Sau*3A fragment labeled at the *Sau*3A site with [γ-^{32}P]ATP and polynucleotide kinase. As shown in the diagram, this fragment contains 600 bp of 5′ untranscribed DNA of Δ1 and a portion of the first exon. After S1 nuclease digestion, the labeled protected fragments were resolved in a denaturing polyacrylamide gel. The asterisks denote the fragments protected by authentic AFP mRNA or Δ1 RNA. The closed circle denotes a 34- to 36-bp longer fragment protected by Δ1 RNA, which is discussed in the text. The marker (M) lane contains labeled *Hae*III-digested single-stranded ϕX174 DNA, and the fragment sizes from top to bottom are 310, 271/269, 234, 194, 118, and 72 nt. (B) Three micrograms of poly(A)$^+$ RNA from transfection of Δ1 through Δ9 was treated as described in (A). In the case of Δ3-Δ9, the closed circle denotes a band corresponding to an unspliced read-through late SV40 transcript whose size exactly equals the distance from the 5′ terminus of the deletion to the labeled *Sau*3A site of the probe, as shown by the closed block in the diagram.

fragments protected from S1 nuclease were resolved on a denaturing polyacrylamide gel, and the radiolabeled fragments were identified by autoradiography. The results in Fig. 3A show that pSV2ZE2Δ1 transcripts contain the same 5′ terminus as AFP mRNA. The heterogeneous set of S1-protected bands indicated by the asterisk is always observed with AFP mRNA (10, 53) and either represents true microheterogeneity at the 5′ terminus and/or imprecise S1 nuclease digestion.

Two less abundant S1-protected fragments which migrate more slowly than the AFP protected fragments were also observed in experiments with Δ1 RNA (Fig. 3). The upper band, indicated by a closed circle, maps to positions −33 to −36. It was previously observed in SV40-AFP chimeric viruses (53) and is the result of either initiation within the AFP 5′ flanking region or possibly a spliced transcript initiating at the SV40 late promoter. The relative abundance of this transcript varies in transfections with different AFP minigene derivatives and in duplicate transfections with the same DNA. On the other hand, variations in the relative abundance of AFP and SV40 early transcripts are not found in duplicate transfections (see below). The middle band is an S1 nuclease digestion artifact, since its levels are not dependent on RNA concentration and it disappears with higher S1 nuclease concentrations.

Splicing of the individual coding blocks and transcription from an *Alu*-like template encoded within the first intervening sequence (53) were examined by an alternate S1 protection assay. Poly(A)$^+$ RNA was hybridized to AFP minigene DNA. After S1 nuclease digestion and alkali treatment, the S1-protected, single stranded DNA fragments were resolved on a nondenaturing polyacrylamide gel, transferred to nitrocellulose, and hybridized to probes that span the regions of interest (Fig. 4). When the S1-protected DNA fragments from transient expression assays with pSV2ZE2Δ1, as well as a series of deletion mutants discussed below, were hybridized with a probe which spans the first AFP coding block, proper splicing of the first coding block was indicated by the presence of a 129-nt band which comigrates with the fragment protected by AFP mRNA.

The *Alu*-like gene in the first intervening sequence of the AFP gene, which is illustrated in Fig. 4 by the hatched box, is also transcribed from pSV2ZE2Δ1, as shown by the protection of a 220-nt band detected by probe 1 (asterisk in Fig. 4). We have previously shown that this gene is transcribed in SV40-AFP recombinant viruses from the opposite strand to that encoding the AFP gene. It initiates 383 bp downstream from the AFP cap site and terminates in a T-rich region 39 bp away from the 3′ border of the first AFP exon (Fig. 1A).

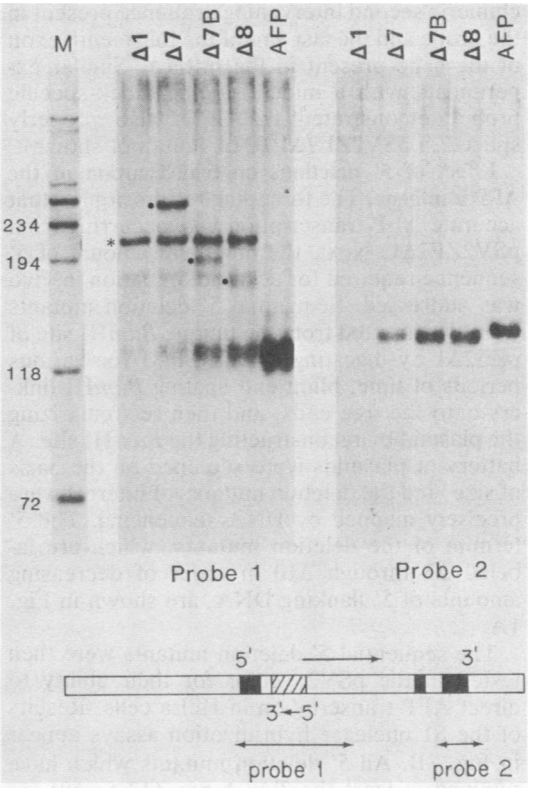


FIG. 4. Structure of the AFP minigene transcripts. Twenty micrograms of poly(A)$^+$ RNA from transfection of HeLa cells with the indicated AFP minigene derivatives was hybridized to an excess of minigene DNA and treated with S1 nuclease, and the protected fragments were resolved on a denaturing acrylamide gel. The DNA was transferred to a nitrocellulose filter and hybridized to either a 660-bp *Hin*cII fragment that spanned the first coding block (probe 1) or a 350-bp *Hin*dIII/*Kpn*I fragment that contained the third coding block (probe 2). The AFP lane contains the result obtained with authentic AFP mRNA. The asterisk denotes a 220-nt band that corresponds to the transcript of an *Alu*-like gene encoded on the anticoding strand of the minigene in the first intervening sequence (hatched box in diagram). The closed circles denote fragments protected by unspliced read-through transcripts of the late SV40 promoter (see diagram in Fig. 3). The marker lane (M) is as described in the legend to Fig. 3.

Splicing of the third coding block of the AFP minigene was analyzed by hybridization of a duplicate filter with probe 2 (Fig. 4). Protection of a 141-nt band, the actual exon size, showed that the third coding block is spliced in AFP transcripts. The larger size (144 bp) of this exon in authentic AFP mRNA is the result of a 3-nt overlap in sequence between the last 3 nt of the

chimeric second intervening sequence present in the probe and the last 3 nt of the fourteenth exon of the gene present in the mRNA. Similar experiments with a middle coding block-specific probe demonstrated that it is also properly spliced in pSV2ZE2Δ1 RNA (data not shown).

Effect of 5′ deletions on transcription of the AFP minigene. The foregoing results argued that accurate AFP transcription was occurring from pSV2ZE2Δ1. Next, the minimum amount of 5′ sequence required for accurate initiation in vivo was addressed. Sequential 5′ deletion mutants were constructed from the unique *Bam*HI site of pZE2Δ1 by digesting it with *Bal*31 for various periods of time, blunt-end ligating *Bam*HI linkers onto the free ends, and then recircularizing the plasmid by reconstructing the *Bam*HI site. A battery of plasmids were screened on the basis of size, and the deletion mutants of interest were precisely mapped by DNA sequencing. The 5′ termini of the deletion mutants, which are labeled Δ3 through Δ10 in order of decreasing amounts of 5′ flanking DNA, are shown in Fig. 1A.

The sequential 5′ deletion mutants were then tested in the pSV2 vector for their ability to direct AFP transcription in HeLa cells. Results of the S1 nuclease hybridization assays appear in Fig. 3B. All 5′ deletion mutants which have retained at least the TATA box (Δ3 to Δ8) are capable of directing accurate AFP transcription in HeLa cells. However, once the TATA sequence is removed (Δ9), no accurately initiated AFP transcripts can be detected. Rather, S1-protected bands corresponding to downstream initiation events are now observed.

There is a striking correlation between transcriptional efficiency and the size of the 5′ flanking deletion, with Δ8, which contains only 33 bp of 5′ flanking DNA, being approximately 120 times more efficient than Δ1. To compare the amounts of steady-state transcripts of the various mutants, the transfections were standardized with respect to the amount of transcriptionally competent DNA in the HeLa cell nuclei by assaying the levels of two other transcripts encoded by the vectors. The *Alu*-like gene and the SV40 early T antigen gene are both encoded on the opposite strand from the AFP gene, and their transcription would presumably not be affected by deletions in the AFP 5′ flanking region. As shown in Fig. 4, the levels of the *Alu*-like transcript are very similar in transfections of Δ1, Δ,7, Δ7B, and Δ8, indicating that the DNA uptake is very reproducible. This was further substantiated by analyzing the levels of SV40 early RNAs on a denaturing agarose gel, transferring to nitrocellulose, and hybridizing with a labeled SV40 early region probe (Fig. 5). The levels of the SV40 early transcripts did not vary more than twofold among transfections, as judged by densitometric scanning. This was seen for transient expression assays with all three vectors, and all RNA samples used in the S1 nuclease hybridization assays displayed in this report have been shown to contain nearly identical *Alu* and/or SV40 early transcript levels.

The 5′ deletion mutant RNAs derived from Δ3 to Δ8 DNA lacked the S1-protected fragment corresponding to position −33 observed for Δ1, but instead displayed strong bands equivalent to the distance from the ^{32}P-labeled *Sau*3A site of the probe to the 5′ termini of the deletions (closed circles in Fig. 3B). Appropriate S1 nuclease hybridization assays have positioned the 5′ termini of these transcripts to a series of sites located within the late region of the SV40 com-

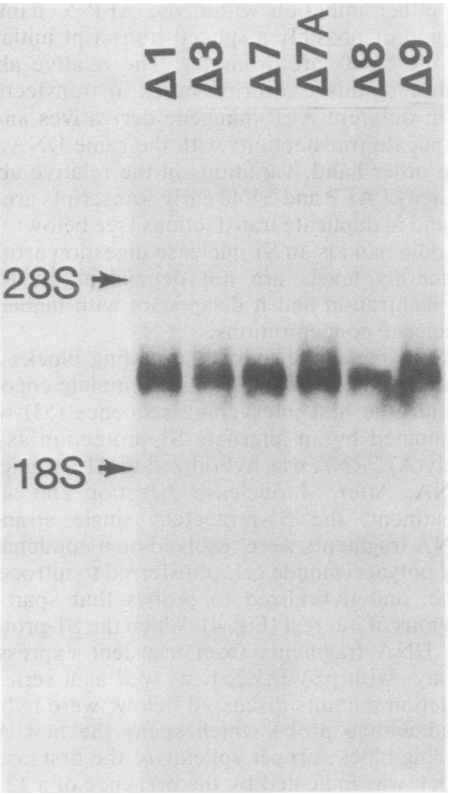


FIG. 5. Quantitation of SV40 early transcripts in transient expression RNA. Three micrograms of poly(A)$^+$ RNA from transfections of various AFP minigene derivatives were electrophoresed in a 1.5% agarose gel containing 1× MOPS buffer (20 mM MOPS [morpholine propanesulfonic acid] [pH 7.0], 5 mM NaOAc, 1 mM EDTA) and 6.3% formaldehyde. The RNAs were denatured in 1× MOPS–62.5% formamide–6.3% formaldehyde at 68°C for 5 min before electrophoresis. The RNAs were transferred to nitrocellulose in 20× SSC (1× SSC, 0.15 M NaCl plus 0.015 M sodium citrate) (46) and hybridized to a labeled SV40 *Pvu*IIA fragment from 0.33 to 0.711 map units, which contains a major portion of the early region.

ponent between map coordinates 0.010 and 0.035. Late SV40 read-through transcripts are apparently not detected with pSV2ZE2Δ1, a pattern consistently observed in transfections with a series of AFP minigene derivatives (see below). That is, deletion of AFP sequences between Δ1 and Δ3 correlated with the disappearance of the S1-protected fragment corresponding to position −33 and the appearance of the SV40 late read-through transcripts. The simplest explanation for these results is that the transcripts defined by the −33 S1-protected band are also late SV40 read-through transcripts which are spliced from a 5′ splice junction located between Δ1 and Δ3 to a 3′ splice junction located at position −33. Examination of the sequence at the cryptic 3′ junction TCCCTTAAACAG-G shows an overall but not exact homology with the consensus 3′ sequence TPyTPyPyPyTCAG-G (30).

For transfections with Δ8, these two types of transcripts cannot be distinguished from one another, since the point of sequence divergence of Δ8 DNA with the probe maps to within 1 bp of the putative cryptic splice site at −32. These read-through transcripts can also be seen in Fig 4 (closed circles), where the lengths of the additional bands obtained with probe 1 in Δ7, Δ7B, and Δ8 RNAs (250, 185, and 170 bases) correspond very closely to the known distance from the 5′ deletion border to the 3′ border of the first coding block (248, 182, and 165 bases). This result argues that these read-through transcripts are spliced at the first AFP coding block 5′ splice junction.

The data in Fig. 4 also emphasize the striking increase in the amount of stable transcripts that include the correctly initiated first coding block as 5′ flanking DNA is deleted from Δ1. In contrast, levels of transcripts which contain a spliced third exon are more similar in Δ7, Δ7B, and Δ8 transfections. The likely explanation is that the transcripts from the SV40 late promoter are reading through the entire minigene and that the two AFP introns are being correctly spliced.

Effect of distance from the SV40 control region on AFP transcription. One major concern about attributing any biological significance to the actual marked enhancement of transcriptional efficiency observed upon deletion of 5′ flanking DNA is whether this reflects a transcriptional control mechanism inherent in the AFP 5′ flanking region or rather is a consequence of the association of the minigene with the SV40 vector component. Previous studies had shown that the SV40 control region, in particular the 72-bp repeat, could have dramatic effects on transcription rates of distal promoters in *cis* (3, 25, 38). This possibility seemed likely, given that insertion of AFP minigene templates with progressively larger 5′ deletions into the pSV2 vector causes a corresponding decrease in the distance separating the AFP cap site from the SV40 control element.

One test of this possibility is to insert the mutants in the pSV3 vector, where transcription from the minigene is occurring in the opposite direction to the SV40 late promotor and the 5′ AFP deletions have no effect on the position of the AFP cap site relative to the SV40 control regions, except in the counterclockwise direction at a distance of ~4,500 bp. The AFP minigene ZE2 and its derivatives Δ3 and Δ8 were inserted into the pSV3 vector and introduced into HeLa cells by calcium phosphate precipitation. The absolute AFP transcript levels in the pSV3 chimera are 5- to 10-fold less than the same deletion mutant in pSV2. However, as shown in Fig. 6, the efficiency of AFP transcription is once again inversely correlated with the amount of 5′ flanking DNA, analogous to the results obtained with the pSV2 vector. That is, there is a twofold enhancement of transcription of Δ3 compared to ZE2 and a further fivefold enhancement of Δ8 relative to Δ3.

Read-through transcripts were also detected

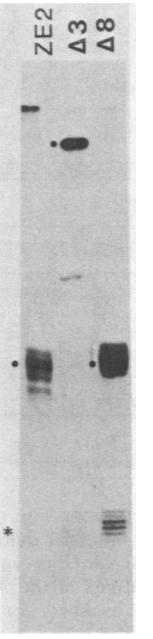


FIG. 6. Transient expression of AFP minigene derivatives in the pSV3 vector. Twenty micrograms of poly(A)$^+$ RNA from transfections of ZE2 and Δ3 and Δ8 minigenes inserted into the pSV3 vector were hybridized to the labeled *Pvu*II-*Sau*3A fragment and treated as described in the legend to Fig. 3. The asterisk denotes the fragments protected by authentic AFP transcripts, and the closed circles denote either spliced (ZE2) or unspliced (Δ3 and Δ8) read-through transcripts.

with the pSV3ZE2 chimera which share some of the same properties as those described for pSV2ZE2. The S1-protected fragments corresponding to position −33 were once again observed with ZE2 but disappeared in the mutants with less than 600 bp of flanking DNA (ZE2 in Fig. 6). In Δ3 and Δ8 RNA, S1-protected fragments corresponding to the point of sequence divergence between probe and template were detected instead. The initiation site of the read-through transcripts has not been determined, but it is likely that it occurs within the pBR322 component. Many discrete pBR322 transcripts can be identified by Northern blot analysis, and there is no evidence of SV40 early transcripts which have failed to properly terminate or have been aberrantly spliced (unpublished observation).

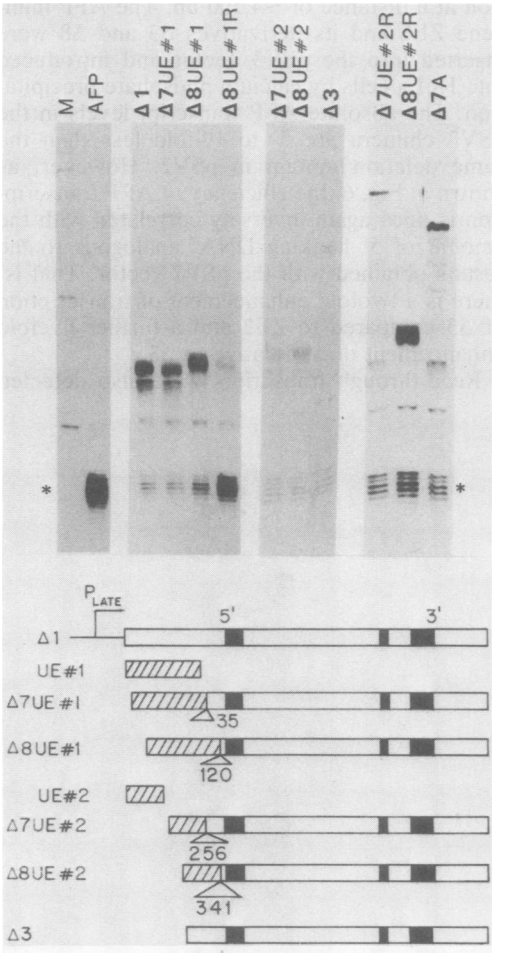


FIG. 7. Transient expression of AFP minigene derivatives containing internal deletions. Three micrograms of poly(A)$^+$ RNA from transient expression of the indicated derivatives of the AFP minigene in the pSV2 vector were assayed for AFP-specific transcripts (asterisk) as described in the legend to Fig. 3. The structure of pSV2Δ1 is drawn to scale below the gel, with the coding blocks of the minigene indicated by the dark boxes. The position of the SV40 late promoter is also shown. The 5′ flanking DNA contained within UE#1 (−604 to −154) and UE#2 (−604 to −375) is shown by the hatched boxes, either alone or adjacent to Δ7 and Δ8. The size of the subsequent internal deletion is indicated by the numbers below the line diagrams. The length of Δ3 is drawn to scale to illustrate its similarity to Δ7UE#2 and Δ8UE#2.

A second protocol to control for the influence of the SV40 control elements as well as a means of testing more limited sequence deletions is shown in the diagram in Fig. 7. Internal deletion mutants were constructed by inserting DNA fragments isolated from the 5′ flanking region of AFP (UE) at the 5′ borders of Δ7 and Δ8.UE#1 contains 450 bp of the 5′-most flanking DNA of Δ1. It was inserted in its original orientation at the 5′ border of Δ7 and Δ8 through linkage at the in vitro-constructed *Bam*HI site. This generated mutants with small internal deletions extending from positions −153 to −118 (Δ7.UE#1) and −153 to −33 (Δ8.UE#1). UE#2 contains the 5′-most 229 bp of flanking DNA of Δ1, and linkage of it in its original orientation to the 5′ border of Δ7 and Δ8 creates derivatives with much larger 5′ flanking internal deletions extending from positions −374 to −118 (Δ7.UE#2) and −374 to −33 (Δ8.UE#2). Thus, these UE now separate the AFP cap site from the SV40 control elements, so that the differences in the amount of 5′ flanking DNA between Δ7 and Δ8 are minimized.

Figure 7 displays the results of the S1 nuclease hybridization assays after the transient expression of Δ7.UE#1 and Δ8.UE#1 in HeLa cells. Both vectors were capable of directing accurate AFP transcription and did so with similar efficiencies, unlike Δ7 and Δ8, which exhibited a five- to sixfold difference in transcriptional efficiency. The efficiencies of Δ7.UE#1 and Δ8.UE#1 were very similar to that of pSV2ZEΔ1, a vector containing a comparable amount of 5′ flanking DNA. Identical results were also obtained for Δ7.UE#2 and Δ8.UE#2 (Fig. 7); the levels of accurate AFP transcripts were equivalent to those of Δ3, a chimera with a comparable amount of 5′ AFP flanking DNA.

Thus, these results argue that the rate of AFP transcription in pSV2-derived chimeras is at least in part a function of the distance between the SV40 control elements and the AFP transcriptional start site. However, distance alone cannot be the sole determining factor, as insertion of either UE#1 or UE#2 in the opposite orientation upstream of Δ7 or Δ8, labeled Δ8.UE#1R, Δ7.UE#2R, and Δ8.UE#2R in Fig. 7, results in an increase in AFP transcription, as compared to the same elements in their correct orientation. For example, the transcriptional ef-

ficiencies of Δ7.UE#2R and Δ8.UE#2R are approximately fivefold greater than those for Δ7.UE#2 and Δ8.UE#2, and the same increase holds true for Δ8.UE#1 and Δ8.UE#1R.

One question which arises from the foregoing is whether AFP transcription absolutely requires the presence of the SV40 control region, in particular, the 72-bp repeat. A direct test of this possibility by deletion of only the 72-bp repeat cannot be performed in this system, as studies have shown that AFP transcription cannot be detected in the absence of SV40 early gene expression, which in turn is dependent on this region. That is, a vector construction similar to pSV2 that contained a partial deletion of the 3′ end of the early region exhibited no AFP transcription. This activity could be restored in *trans* upon cotransfection with an SV40-pBR322 recombinant containing an intact early region (data not shown). However, no AFP transcripts could be detected with transfections of pSV1ZE2 (Fig. 2), in which the distance between the 72-bp repeat and the AFP promoter was at a minimum of ~3,300 bp in the clockwise direction, as compared to 1,000 and 1,700 bp in pSV2ZE2 and pSV3ZE2, respectively. This result suggests that a sufficient separation of the 72-bp repeat from the AFP promoter results in the extinction of AFP transcription. Another possibility unrelated to distance that could explain the absence of AFP transcription in pSV1ZE2 is interference generated by the presence of pBR322 DNA between the minigene and the SV40 control elements. Although an example of plasmid interference in transient expression experiments has been reported (3), a very similar vector construct to pSV1ZE2 was used successfully to assay β-globin gene expression (18).

In vitro transcription of the AFP gene. The foregoing set of experiments suggests that deletion of the 5′ flanking DNA between Δ7 at −118 up to the TATAAA box at −30 has little effect on AFP transcription in vivo. In this regard, the in vivo results are identical to those obtained with an in vitro transcription system using a whole cell extract derived from HeLa cells. A plasmid containing the minigene clone pZE2 was restricted with *Ava*I, truncating the AFP transcription template 339 bp downstream of the putative cap site. After incubation in a soluble whole cell extract prepared from HeLa cells as described by Manley et al. (32), the radiolabeled transcription products were resolved on a denaturing polyacrylamide gel. The arrow in Fig. 8, lane 1, identifies a 330-base transcript which is sensitive to 2 μg of α-amanitin per ml, indicating that it is an RNA polymerase II product. That this 330-base transcript actually results from accurate initiation at the AFP cap site is supported by several additional experiments. The use of templates truncated at other downstream sites resulted in the disappearance of the 330-base transcript and the appearance of a transcript of the predicted size (data not shown). Also, an S1 nuclease hybridization assay with an end-labeled DNA probe which spans the AFP cap site demonstrated that the sizes of the DNA fragments protected by either the in vitro transcripts or authentic AFP mRNA are identical (data not shown). The abundant higher-molecular-weight species in all the lanes of Fig. 8 most likely represent nonspecific end-to-end transcription

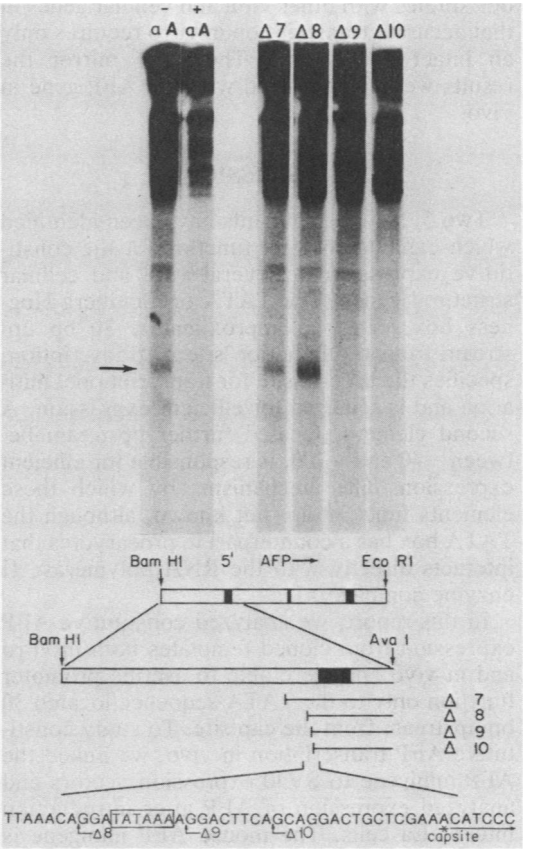


FIG. 8. In vitro transcription of the AFP gene. The structure of the *Bam*HI-*Eco*RI subclone pAFP14Z.1 is shown in the diagram, with the first three coding blocks in black. The sequences in the deletion mutants Δ7 to Δ10 are indicated below, along with the *Ava*I site used to generate the plasmid templates. The nucleotide sequence of the 5′ end of the gene and the precise locations of the deletions are shown in the bottom line. The TATAAA box is outlined, and the putative cap site is indicated. In the autoradiogram of the denaturing polyacrylamide gel above, the first two lanes show the in vitro transcription products of pAFP14Z.1, cut by *Ava*I, in the absence or presence of 2 μg of α-amanitin (αA) per ml. The transcripts of the four deletion mutants are shown in the last four lanes. The arrow refers to the AFP-specific transcript.

products and/or endogenous end-labeled RNA/DNA species.

The in vitro transcripts synthesized from the deletion mutants truncated at the *Ava*I site are also shown in Fig. 8. Both Δ7 and Δ8 are capable of directing accurate AFP transcription, based on the presence of the 330-base run-off transcript, but Δ9 and Δ10 have lost this ability. Although the deletion mutants have different pBR322 sequences flanking the 5′ AFP terminus, identical results were obtained with plasmids having the same pBR322 sequence joined to the 5′ terminus of the AFP deletion mutant (data not shown). These results agree with previous studies with other viral and cellular genes in that accurate transcription in vitro requires only an intact TATA box. They also mirror the results we have obtained with the AFP gene in vivo.

DISCUSSION

Two 5′ flanking elements have been identified which exhibit promoter function for the constitutive expression of several viral and cellular structural genes. The TATA or Goldberg-Hogness box, centered approximately 30 bp upstream from the initiation site of transcription, specifies the precise site for transcriptional initiation and is required for efficient expression. A second element, located further upstream between −40 and −100, is responsible for efficient expression. The mechanisms by which these elements function are not known, although the TATA box has a counterpart in procaryotes that interacts directly with the RNA polymerase II enzyme complex (41).

In this report, we analyzed constitutive AFP expression from cloned templates both in vitro and in vivo and were able to ascribe promoter function only to the TATA sequence located 30 bp upstream from the cap site. To study constitutive AFP transcription in vivo, we linked the AFP minigene to SV40 expression vectors and analyzed expression of AFP after introduction into HeLa cells. The mouse AFP minigene is efficiently expressed during the transient expression period after DNA transfection, although the relative efficiency of AFP expression is dependent on the structure of the vector. The transcripts begin at the authentic AFP cap site and are properly spliced and polyadenylated.

The transient expression assays with a series of 5′ deletion mutants showed that AFP sequences in the −100 to −40 region have no effect on accurate or efficient AFP transcription. This conclusion is most evident when one compares the efficiencies of expression of Δ7UE#1 to Δ8UE#1 or Δ7UE#2 to Δ8UE#2 (Fig. 7). The only difference between these pairs of deletions is the removal of 5′ flanking sequence between −118 and −33, yet they are equally efficient templates for AFP transcription.

When the AFP deletion mutants were analyzed in the pSV2 vector, the AFP transcriptional efficiency actually increased as more flanking DNA was removed. This increased efficiency was shown to be a function of the distance separating the AFP cap site from the SV40 control elements rather than to the loss of 5′ AFP flanking DNA, an effect first observed by Moreau et al. (38) for several viral and cellular genes. This is demonstrated in Fig. 7, in which the AFP transcriptional efficiencies of the internal deletion mutants are equal to those of the 5′ deletion mutants that contain a comparable amount of 5′ flanking DNA. These results suggest that some component of the vector is supplying promoter function for the AFP gene. This is supported by the data with deletion mutant Δ9, which contains only 23 bp of 5′ flanking DNA. Despite the fact that the TATA box has been deleted, it is capable of directing AFP transcription, albeit at a lower efficiency and with an altered initiation specificity.

It is possible that the presence of the SV40 control elements, presumably the 72-bp repeats, overwhelms any effect of AFP upstream promoter function by performing the same function far more efficiently. The enhancer-like activity of the SV40 72-bp repeat was first reported by Banerji et al. (3), who observed that after transfection of HeLa cells, transcription of the rabbit β-globin gene was enhanced over 200-fold when it was linked to the SV40 72-bp repeat element. The enhancing effect is independent of the orientation of the 72-bp repeat (3, 25), but, as shown by Moreau et al. (38), the enhancement is reduced when the distance separating the eucaryotic promoters from the repeat element is increased. Also, a compensatory role for the 72-bp repeat has been suggested by Pelham (40) in his studies with the transient expression of a *Drosophilia* heat shock gene from an SV40 vector in COS cells. Linkage of the repeat at the 5′ terminus of the gene obscured the heat shock response normally observed when the 72-bp repeat was 3′ proximal. However, at least in one instance, the general enhancer-like activity of the 72-bp repeat can be distinguished from specific upstream promoter activity. In experiments similar to our own, a decrease in the transcription of the β-globin gene as the result of removal of 5′ flanking DNA occurred in the presence of the 72-bp repeat (18). Therefore, this system is capable of defining gene-specific regions that facilitate transcription.

In assessing the likelihood that the SV40 control elements compensate for the sequential deletion of an AFP-specific control region, one must take into account the fact that AFP transcription is not reduced upon removal of 5′

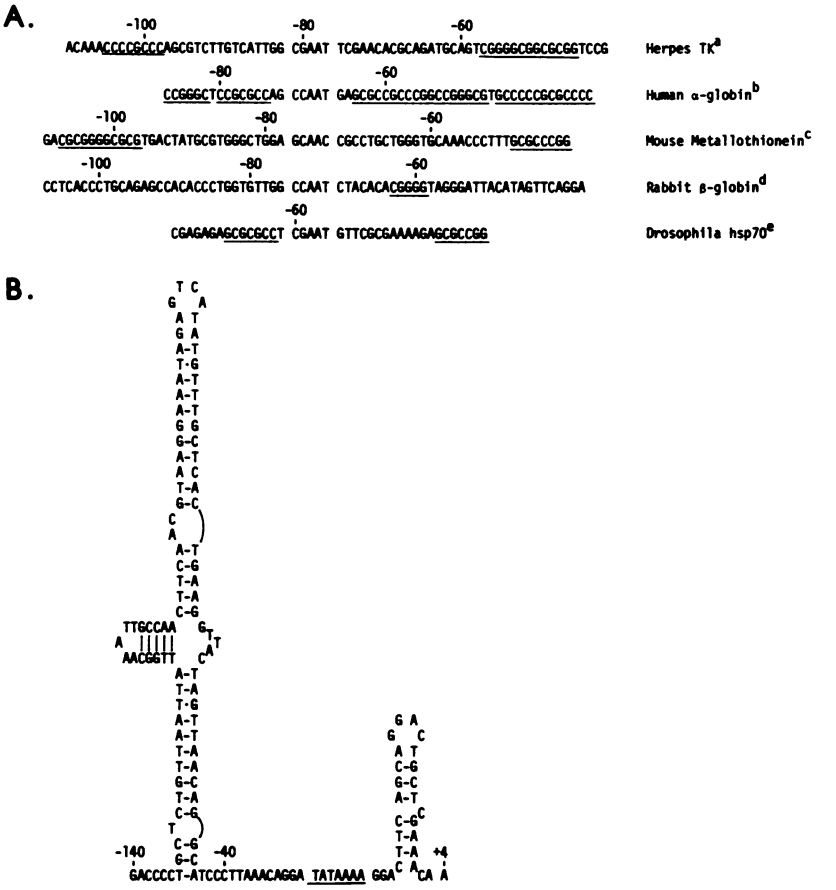


FIG. 9. Sequence comparisons of the 5′ flanking DNA of genes that exhibit upstream promoter activity. (A) The 5′ untranscribed sequences of several eucaryotic genes that exhibit upstream promoter activity in vivo are aligned via their CCAAT pentamers around −75. The guanine plus cytosine-rich regions on either side of the pentamer are underlined. References: a, 36; b, 37; c, 13; d, 23; e, 27. (B) The 5′ untranscribed region of the AFP gene in an extended cruciform structure.

flanking DNA, even in the pSV3 chimera (Fig. 6). In fact, there is a fivefold enhancement in transcription upon deletion of sequences between −247 and −33. In this chimera, differential distance of the AFP gene from the SV40 control region is difficult to invoke as a reasonable explanation for the enhanced transcription, given that the effect must act at a distance of over 4,500 bp. Thus, these results reinforce the conclusion that the upstream AFP sequences exhibit no promoter function in HeLa cells. It is possible that the pBR322 sequences that border the 5′ end of the AFP gene in the pSV3 chimera are contributing enhancer-liker activity, especially as there is active transcription within the pBR322 DNA. Grosschedl and Birnsteil (17) recently observed an enhancing effect of pBR322 sequences on sea urchin histone H2A transcription in vitro.

Our inability to detect any promoter function upstream of the TATA box sets the AFP gene apart from other genes whose constitutive expression has been examined to date. In an effort to understand this distinction, the 5′ flanking sequences of several genes that demonstrate upstream promoter activity were examined for features they might hold in common. They are aligned in Fig. 9A with the CCAAT pentamer. Two common elements are apparent: the CCAAT box at approximately −75 and guanine plus cytosine-rich regions which flank it, which are underlined. In a careful study of this region utilizing the herpesvirus thymidine kinase gene, McKnight and Kingsbury (36) were able to implicate these guanine plus cytosine-rich regions in the specific control of thymidine kinase expression. They went on to propose that these regions form a cruciform structure that facilitates the interaction of RNA polymerase II and DNA.

As shown in Fig. 1B, the 5′ flanking region of the AFP gene does not contain a CCAAT pen-

tamer, although it shares significant homology at −100 to the region of the albumin gene at −80 that does contain this conserved sequence. In addition, there are no guanine plus cytosine-rich regions in the 5′ flanking DNA until position −138, where the sequence CCCCTGC is found. Thus, the flanking sequence of the AFP gene exhibits neither of these characteristics found to be important for upstream promoter function, which may explain why it behaves differently in the transient expression assay.

A second possible explanation for our failure to detect upstream promoter function could be that it resides further upstream than the 1.0 kb that is included in the minigene. We cannot exclude this possibility, but there is no precedent for such action at this distance in other genes whose constitutive expression has been analyzed by these techniques. Also, one cannot argue that the region between −100 and −40 is dispensable for the regulation of AFP expression in homologous cells until the gene is assayed in such a system. Of potential importance is the fact that the sequences from −103 to −45 can be drawn in an extended cruciform structure (Fig. 9B). Thus, although we cannot attribute any biological activity to this region of the gene, it exhibits those elements (conserved sequence relative to the albumin gene and potential secondary structure) that one associates with regulatory function.

One final word of caution must be noted with regard to the use of the pBR322-SV40 vector system to study heterologous gene expression. First, the transcription pattern from the chimeras is not simple. Significant levels of upstream read-through transcripts have been identified with both the pSV2 and pSV3 chimeras. In addition to the read-through transcripts, the chimeras also direct the production of SV40 early transcripts, an RNA polymerase III directed *Alu*-like transcript, and discrete pBR322 transcripts. The effect these various transcriptional processes have on the absolute rate of AFP transcription is not known. Second, seemingly neutral sequence perturbations around the 5′ AFP flanking DNA have been shown to dramatically affect the AFP transcription rate. As shown in Fig. 8, AFP transcription from the chimeras where UE#1 and UE#2 had been inserted in the reverse orientation at the 5′ border of Δ7 and Δ8 in pSV2 is significantly enhanced over what is found for the corresponding chimeras where the element has been inserted in its original orientation. It is not likely that this reflects some transcriptional control mechanism inherent in the 5′ AFP flanking DNA, since deletion of the DNA within the UE has little or no apparent affect on AFP transcription. Thus, one must be conservative in attributing enhancing activity to DNA juxtaposed to potentially regulatory DNA.

ACKNOWLEDGMENTS

We thank Peter Young for his assistance in the analysis of the secondary structure of the 5′ flanking regions and Jerome Freed and Mary Ellen Croke for helping us to establish the HeLa cell transfection system. R.W.S. is the recipient of a Public Health Service postdoctoral fellowship. This work was supported by grants from the National Institutes of Health and the National Foundation March of Dimes and an appropriation from the Commonwealth of Pennsylvania.

LITERATURE CITED

1. **Abelev, G. I.** 1971. Alpha-fetoprotein in oncogenesis and its association with malignant tumors. Adv. Cancer Res. **14**:295–358.
2. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin mRNA by chromatography on oligo thymidylic acid-cellulose. Proc. Natl. Acad. Sci. U.S.A. **69**:1408–1412.
3. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
4. **Benoist, C., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
5. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
6. **Brinster, R. L., H. Y. Chen, R. Warren, A. Sarthy, and R. D. Palmiter.** 1982. Regulation of metallothionein-thymidine kinase fusion plasmids injected into mouse eggs. Nature (London) **296**:39–42.
7. **Bunick, D., R. Zandomeni, S. Ackerman, and R. Weinman.** 1982. Mechanism of RNA polymerase II-specific initiation of transcription *in vitro*: ATP requirement and uncapped runoff transcripts. Cell **29**:877–886.
8. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406–1414.
9. **D'Eustachio, P., R. S. Ingram, S. M. Tilghman, and F. H. Ruddle.** 1981. Murine α-fetoprotein and albumin: two evolutionary linked proteins encoded on the same chromosome. Somatic Cell Genet. **7**:289–294.
10. **Eiferman, F. A., P. R. Young, R. W. Scott, and S. M. Tilghman.** 1981. Intragenic amplification and divergence in the mouse α-fetoprotein gene. Nature (London) **294**:713–718.
11. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
12. **Galanti, N., G. J. Jonak, K. J. Soprano, J. Floros, L. Kaczmarek, S. Weissman, V. B. Reddy, S. M. Tilghman, and R. Baserga.** 1981. Characterization and biological activity of cloned simian virus 40 DNA fragments. J. Biol. Chem. **256**:6469–6474.
13. **Glanville, N., D. M. Durnam, and R. D. Palmiter.** 1981. Structure of mouse metallothionein-I gene and its mRNA. Nature (London) **292**:267–269.
14. **Gorin, M. B., D. L. Cooper, F. Eiferman, P. van de Rijn, and S. M. Tilghman.** 1981. The evolution of α-fetoprotein and albumin. I. A comparison of the primary amino acid sequences of mammalian α-fetoprotein and albumin. J. Biol. Chem. **256**:1954–1959.
15. **Grosschedl, R., and M. L. Birnsteil.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **77**:1432–1436.
16. **Grosschedl, R., and M. L. Birnsteil.** 1980. Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence

are essential for promotion of H2A histone gene transcription *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **77**:7102–7106.
17. **Grosschedl, R., and M. L. Birnsteil.** 1982. Delimitation of far upstream sequences required for maximal *in vitro* transcription of an H2A histone gene. Proc. Natl. Acad. Sci. U.S.A. **79**:297–301.
18. **Grosveld, G. C., E. deBoer, C. K. Shewmaker, and R. A. Flavell.** 1982. DNA sequences necessary for transcription of the rabbit β-globin *in vivo*. Nature (London) **295**:120–126.
19. **Grosveld, G. C., A. Rosenthal, and R. A. Flavell.** 1982. Sequence requirements for the transcription of the rabbit β-globin gene *in vivo*: the −80 region. Nucleic Acids Res. **10**:4951–4971.
20. **Grosveld, G. C., C. K. Shewmaker, P. Jat, and R. A. Flavell.** 1981. Localization of DNA sequences necessary for transcription of the rabbit β-globin gene *in vitro*. Cell **25**:215–226.
21. **Gruss, P., R. Dhar, and G. Khoury.** 1981. Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. U.S.A. **78**:943–947.
22. **Hagenbuchle, O., and U. Schibler.** 1981. Mouse β-globin and adenovirus-2 major late transcripts are initiated at the cap site *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **78**:2283–2286.
23. **Hardison, R. C., E. T. Butler, E. Lacy, T. Maniatis, N. Rosenthal, and A. Efstratiadis.** 1979. The structure and transcription of four linked rabbit β-like globin genes. Cell **18**:1285–1297.
24. **Hofer, E., and J. E. Darnell.** 1981. The primary transcription unit of the mouse β-major globin gene. Cell **23**:585–593.
25. **Humphries, R. K., T. Ley, P. Turner, A. D. Moulton, and A. W. Nienhuis.** 1982. Differences in human α, β-, and γ-globin gene expression in monkey kidney cells. Cell **30**:173–183.
26. **Ingram, R. S., R. W. Scott, and S. M. Tilghman.** 1981. α-Fetoprotein and albumin genes are in tandem in the mouse genome. Proc. Natl. Acad. Sci. U.S.A. **78**:4694–4698.
27. **Karch, F., I. Torok, and A. Tissierès.** 1981. Extensive regions of homology in front of the two hsp 70 heat shock variant genes in Drosophilia melanogaster. J. Mol. Biol. **148**:219–230.
28. **Kioussis, D., F. Eiferman, P. van de Rijn, M. B. Gorin, R. S. Ingram, and S. M. Tilghman.** 1981. The evolution of α-fetoprotein and albumin. II. The structures of the α-fetoprotein and albumin genes in the mouse. J. Biol. Chem. **256**:1960–1967.
29. **Lebowitz, P., and S. M. Weissmann.** 1979. Organization and transcription of the simian virus 40 genome. Curr. Topics Microbiol. Immunol. **87**:43–172.
30. **Lerner, M. R., J. A. Boyle, S. M. Mount, S. L. Wolin, and J. A. Steitz.** 1980. Are snRNPs involved in splicing? Nature (London) **283**:220–224.
31. **Luse, D. S., and R. G. Roeder.** 1980. Accurate transcription initiation on a purified mouse β-globin DNA fragment in a cell-free system. Cell **20**:691–699.
32. **Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter.** 1980. DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. U.S.A. **77**:3855–3859.
33. **Mathis, D., and P. Chambon.** 1981. The SV40 early region T-A-T-A box is required for accurate *in vitro* initiation of transcription. Nature (London) **290**:310–315.
34. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
35. **McKnight, S. L., E. R. Gavis, R. Kingsbury, and R. Axel.** 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell **25**:385–398.
36. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science **217**:316–324.
37. **Mellon, P., V. Parker, Y. Gluzman, and T. Maniatis.** 1981. Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system. Cell **27**:279–288.
38. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, and P. Chambon.** 1981. The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. **9**:6047–6068.
39. **Pavlakis, G. N., N. Hizuka, P. Gorden, P. Seeburg, and D. H. Hamer.** 1981. Expression of two human growth hormone genes in moneky cells infected by simian virus 40 recombinants. Proc. Natl. Acad. Sci. U.S.A. **78**:7398–7402.
40. **Pelham, H. R. B.** 1982. A regulatory upstream promoter element in the Drosophila Hsp 70 heat-shock gene. Cell **30**:517–528.
41. **Pribnow, D.** 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proc. Natl. Acad. Sci. U.S.A. **72**:784–788.
42. **Ruoslahti, E., and M. Seppala.** 1979. Alpha-fetoprotein in cancer and fetal development. Adv. Cancer Res. **29**:275–346.
43. **Scherrer, K.** 1969. Isolation and sucrose gradient analysis of RNA, p. 413–432. *In* K. Habel and N. P. Saltzman (ed.), Fundamental techniques in virology. Academic Press, Inc., New York.
44. **Sleigh, M. J., W. C. Topp, R. Hanich, and J. F. Sambrook.** 1978. Mutants of SV40 with an altered small t protein are reduced in their ability to transform cells. Cell **14**:79–88.
45. **Sutcliffe, J. G.** 1978. pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. Nucleic Acids Res. **5**:2721–2728.
46. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. **77**:5201–5205.
47. **Tilghman, S. M., and A. Belayew.** 1982. Transcriptional control of the murine albumin/α-fetoprotein locus during development. Proc. Natl. Acad. Sci. U.S.A. **79**:5254–5257.
48. **Tsuda, M., and Y. Suzuki.** 1981. Faithful transcription initiation of fibroin gene in a homologous cell-free system reveals an enhancing effect of 5′ flanking sequence far upstream. Cell **27**:175–182.
49. **Villa-Komaroff, L., A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S. P. Naber, W. L. Chick, and W. Gilbert.** 1978. A bacterial clone synthesizing proinsulin. Proc. Natl. Acad. Sci. U.S.A. **75**:3727–3731.
50. **Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Teoule, and P. Chambon.** 1980. Specific *in vitro* transcription of conalbumin gene is drastically decreased by single-point mutation in T-A-T-A box homology sequence. Proc. Natl. Acad. Sci. U.S.A. **77**:7024–7028.
51. **Weil, P. A., D. S. Luse, J. Segall, and R. G. Roeder.** 1979. Selective and accurate initiation of transcription of the Ad2 major later promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell **18**:469–484.
52. **Wigler, M., A. Pellicer, A. Silverstein, and R. Axel.** 1978. Biochemical transfer of single-copy eucaryotic genes using total cellular DNA as donor. Cell **14**:725–731.
53. **Young, P. R., R. W. Scott, D. H. Hamer, and S. M. Tilghman.** 1982. Construction and expression *in vivo* of an internally deleted mouse α-fetoprotein gene: presence of a transcribed alu-like repeat within the first intervening sequence. Nucleic Acids Res. **10**:3099–3116.