Volume 10 Number 10 1982 Nucleic Acids Research

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

Peter R.Young*, Richard W.Scott*, Dean H.Hamer+ and Shirley M.Tilghman*

*Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111, and +Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, MD 20205, USA

Received 1 March 1982; Revised and Accepted 30 April 1982

ABSTRACT

An α-fetoprotein (AFP)* 'minigene' was constructed by joining the first two exons along with 0.9 kilobases of 5' flanking sequence of the mouse AFP gene to its last exon and 0.4 kilobases of 3' flanking sequence. This 'minigene' was tested for activity by inserting it into an SV40 vector and infecting African green monkey kidney cells. Correct initiation, termination, polyadenylation and splicing were observed. Furthermore a 220 nucleotide transcript was detected and mapped to a mouse Alu-like or B1 repeat on the opposite strand to that encoding the AFP gene in the first intervening sequence.

INTRODUCTION

In any developmentally regulated gene system, one of the questions of considerable import is the region or regions of DNA at which control is exerted. The most promising method at this time for approaching this question is to introduce exogenous genes into eukaryotic cells (1). Then, in conjunction with in vitro mutagenesis (2,3), the various regulatory regions of genes can be identified. This has already proved to be a powerful method for studying the control of constitutive RNA transcription (4,5,6,7,8). However, to extend these methods to a developmentally regulated system, one would like to reintroduce a gene into the same cell that regulates it in vivo. In so doing, a means to distinguish the exogenous gene transcript from the endogenous one must be developed. One could introduce a heterologous gene in which a sequence that could be distinguished by hybridization has been inserted. It is also possible to retain the 5' untranscribed region of the homologous gene but replace the structural gene itself with a different coding sequence. This second approach has been adopted for studies of the glucocorticoid regulation of MMTV transcription, whereby the viral sequences have been replaced by mouse DHFR (9) or the Ha-MuLV p21 protein (10). This approach is particularly attractive when the

0305-1048/82/1010-3099$ 2.00/0

gene product is a selectable marker.

We have adopted a new approach which is particularly suited to genes that are encoded by long stretches of DNA. A 'minigene' that retains potential 5' and 3' control regions but dispenses with most of the coding blocks and their corresponding intervening sequences, can be constructed in vitro. The assumption we have made is that deletion of coding segments within a gene does not remove sequences necessary for regulation. Expression of this gene after reintroduction into a cell is therefore detected by virtue of the fact that its transcript is shorter than the host gene product.

This approach is illustrated in the present paper for the mouse α-fetoprotein (AFP)* gene. AFP, which is the major serum protein in mouse fetuses, is encoded by 15 exons spanning a total length of 22 kilobases (Kb)* of DNA which is 10 times longer than the mature RNA transcript, 2.2 Kb (11). The gene is subject to developmental regulation in fetal liver, where it is synthesized in increasing amounts until birth, at which point its rate of transcription decreases by several orders of magnitude (S.M. Tilghman, unpublished results.) The gene is also expressed in the visceral endoderm layer of early mouse embryos, a situation which can be studied using mouse teratocarcinomas (12,13).

Before proceeding with studies using these two regulated systems, it is essential to show that the AFP mini-gene is functional in vivo. Hence the present paper describes both the construction of an AFP minigene and its expression in an SV40 vector.

MATERIALS AND METHODS

1) Construction of the AFP minigene. Two genomic subclones of the mouse AFP gene inserted into pBR322 were used to construct the minigene: (1) pAFP14Z.1, which contains a 3.4Kb BamHl/EcoRl fragment which begins 950 bp upstream of the cap site of AFP mRNA (14) and extends through the first three coding blocks, terminating 400 bp into the third intervening sequence; (2) pAFP7E, which contains a 1.1Kb EcoRl fragment encompassing the last two exons (14 and 15) of the AFP gene plus 450 bp of 3' flanking sequence. Both pAFP14Z.1 and pAFP7E were digested with EcoRl, the mixture ligated and used to transform HB101 cells (15). Recombinant DNA clones likely to contain both inserts were selected by colony hybridization to the [^{32}P] labelled pAFP7E insert (16,17). Positive clones were picked and checked for the presence of both inserts and their correct 5'-3' orientation by restriction mapping. One of these, named pAFPZE and shown in Fig. 1, was used for further constructions.

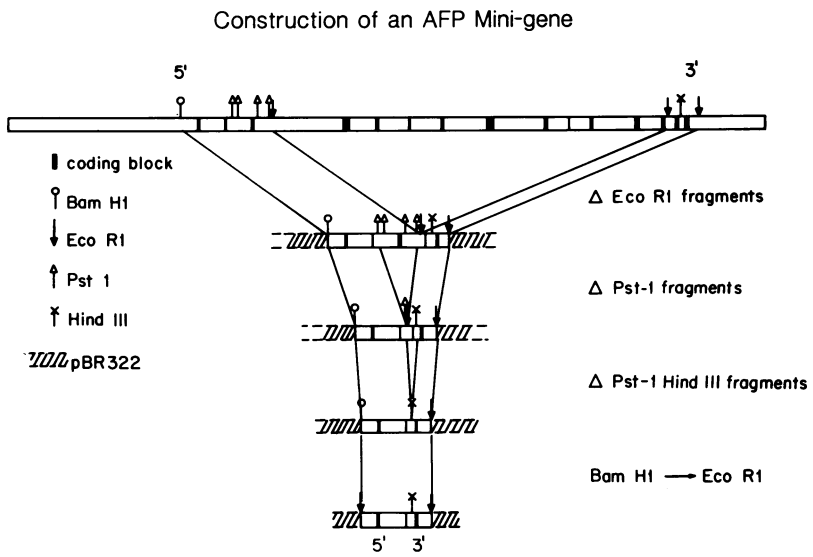

Figure 1. The construction of the AFP minigene. The top line represents 25 Kb of the AFP gene, with the 15 exons outlined by black rectangles. The chimeric clones derived from the gene are designated from top to bottom as follows: pAFPZE, pAFPZE.1, pAFPZE.2 and pAFPZE.2(R1).

The next step was to make a large internal deletion so that the final recombinant would fit into the late region of SV40. This was achieved in two stages (Fig. 1). First, the three internal PstI fragments, clustered in the second and third intervening sequences, were deleted by treating pAFPZE with PstI, religating the fragments at a concentration of 25 μg/ml, and transforming HB101 cells. The identity of the desired recombinant clone, pAFPZE.1, was confirmed via colony hybridization and restriction mapping.

The second stage involved deletion of the third exon of pAFPZE.1 (i.e., exon 14 of the AFP gene) and fusion of the two surrounding intervening sequences. This was achieved by partial digestion of pAFPZE.1 with PstI to obtain a significant proportion of linearized plasmid. Blunt ends were generated at the PstI site by treating with a mixture of S1 nuclease and DNA polymerase I (18). The products were blunt-end ligated to HindIII linkers, (Collaborative Research) digested with an excess of HindIII, and then religated at a concentration of 23 μg/ml. The ligation products were used to transform HB101 cells. Restriction enzyme mapping of the plasmids generated identified one clone, pAFPZE.2, with the correct PstI/HindIII deletion. This deletion also removes the internal EcoR1 site, thus allowing

the insert to be removed as one piece from pBR322.

2) Insertion of AFP minigene into the late region of SV40. pAFPZE.2 was digested with BamHl followed by Sl nuclease and DNA polymerase (18) to make blunt ends. The linearized plasmid was blunt-end ligated to EcoRl linkers and digested with an excess of EcoRl. The product was ligated into one of the two EcoRl sites available in a pBR322-SV40 recombinant plasmid described earlier (19). This plasmid is a fusion of the HpaII/BamHl restriction fragment of the early region of SV40 dl2005, which bears a 230 bp deletion in the intervening sequence of large T antigen (20), and the complete pBR322. The HpaII and BamHl sites have been converted to EcoRl sites and are inserted at the single EcoRl site of pBR322. The pBR322-SV40 DNA had been partially cut with EcoRl and the linear plasmid purified via agarose gel electrophoresis. The desired recombinants, which contained both SV40 DNA and the AFP minigene, were further characterized for the site of insertion and orientation of the AFP minigene via restriction enzyme mapping.

Two recombinants, pAFPZES-A and pAFPZES-C, which contained the AFP minigene inserted in opposite orientations at the same EcoRl site, were partially digested with EcoRl and the linear 5.5 Kb fragment, which contains just the SV40 early region and the AFP minigene, purified on a 1.2% agarose gel. After electroelution and ethanol precipitation, the fragment was circularized at a concentration of 2 μg/ml with 150 units T4 DNA ligase for 2 hr at 20°C. These mixtures were designated SVAFP-A and SVAFP-C, corresponding to the expected recombinant viruses (Fig. 2).

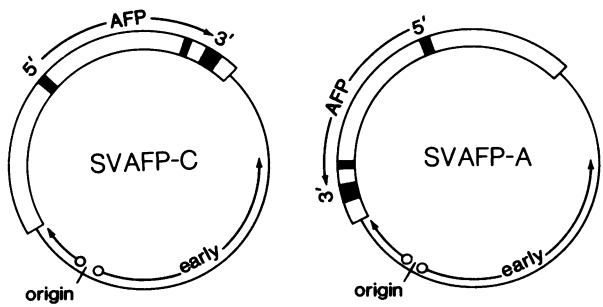

Figure 2. The Structures of the Chimeric SV40-AFP viruses. The two SV40/AFP minigene recombinants, SVAFP-C and SVAFP-A, are drawn with the 2.7 Kb AFP insert outlined by double lines. The black boxes illustrate the positions of the 3 exons and the direction of transcription is indicated by the 5' to 3' symbols. The 3.0 Kb SV40 vector is drawn with a single line, with the origin of replication at 0.67 map units and the early and late transcription start sites (circles) indicated.

3) Propagation of recombinant virus in African green monkey kidney cells. Approximately 1 μg of recombinant virus DNA (SVAFP-A or SVAFP-C) plus 0.2 μg of SV40 tsA239 helper DNA were used to transfect 2 X 10^7 confluent African green monkey kidney (AGMK) cells according to methods previously described (21). The resulting virus stocks were used in subsequent infections of AGMK cells for analysis of cellular DNA and RNA.

4) Analysis of cellular DNA and RNA. Viral DNA was prepared by Hirt extraction (22) followed by CsCl density gradient centrifugation. RNA was prepared by the hot phenol method (23) and enriched for poly A^+ sequences via oligo dT cellulose chromatography (24). The RNA was analyzed by suspending it in 75% formamide containing xylene cyanol, heating at 65°C for 2 min, quenching on ice and loading on a prerun 5% polyacrylamide/7M urea gel. Following electrophoresis at 40mA the gel was washed for 30 min in 50mM NaOH, 100mM NaCl to remove the urea, followed by two 15 min washes in 10 X SSC*(1 X SSC is 0.15M NaCl, 0.015M Na citrate). The RNA was transferred to a 0.45μ Millipore filter in 10 X SSC overnight and baked for 2 hr at 80°C. Prehybridization and hybridization conditions are standard (25), and the filter could be reused by washing off the probe according to the method of Thomas (26). The heterogeneous poly A tails were removed by digestion of poly A^+ RNA with RNAseH in the presence of oligo dT under conditions suggested by the supplier, Bethesda Research Labs.

A modification of the S1 protection experiment was used in which RNA was hybridized to unlabelled pAFPZE.2 DNA, previously cleaved with BamH1 and EcoR1, under conditions favouring RNA-DNA hybridization (27). The hybrids were treated with S1 nuclease, isopropanol precipitated, alkali treated, and the single strand DNA fragments were denatured in 80% formamide at 90°C and separated on a 7.5% polyacrylamide gel. The gel was then treated according to Southern (28), but with washes of 30 min only, and the DNA was transferred overnight in 10 X SSC to 0.45μ Millipore filters. The filters were hybridized successively to more than one [^{32}P] labelled probe by washing in 0.1N NaOH, 2 X SSC for 20 min at 25°C, followed by 2 washes in 2 X SSC.

5) Materials: Restriction endonucleases, T4 polynucleotide kinase, and T4 DNA ligase were purchased from New England Biolabs, DNA polymerase I from Boehringer Mannheim, S1 nuclease from Miles, and RNAse H from Bethesda Research Lab. Suppliers' reaction conditions were adhered to in most cases. Primary African green monkey kidney cells were obtained from Flow Laboratories and were maintained and propagated using medium and serum from Grand Island Biochemical Company. [α-^{32}P] dCTP and [γ^{32}P] ATP were obtained

from New England Nuclear.

RESULTS

Construction of the AFP minigene. In constructing a truncated gene, we took into consideration three features that seemed essential for constitutive synthesis of stable RNA polymerase II transcripts. Evidence to date clearly indicates the requirement for some 5' flanking sequence, although estimates vary as to how much, depending on the gene being tested. Recent experiments with both viral (29,30) and cellular (6,31,32) genes suggest that levels of transcription can be affected by sequences up to 500 base pairs (bp)* 5' to the initiation site. At the 3' end, deletion mapping studies indicate a requirement for the sequence AATAAA some 11-19 bp 5' to the polyadenylation site (33). No sequences 3' to the gene have yet to be shown to be required, at least for constitutive gene expression.

The third apparent requirement for stable transcripts is the presence of at least one intervening sequence. This was suggested by early experiments using SV40 vectors (34,35). However the universality of these observations is suspect, since the use of an alternative initiation site in SV40 apparently circumvents the need for an intron (36,37). Furthermore, genes such as those coding for histones (38) and interferons (39), are now known not to contain introns and yet are functional. Hence the question is clearly an unsettled one, but one that can be circumvented by including at least one intact homologous intron.

Fig. 1 summarizes the construction and final structure of the AFP minigene (see Materials and Methods for details). It contains 950 bp of 5' untranscribed sequence, followed by the first two exons and first intron, thus providing the gene with a homologous intron. The second intron is then fused to the last intron of the gene, generating a chimeric intron of approximately 120 bp. The last exon, containing only 3' untranslated DNA is followed by 400 bp of flanking DNA. The total length of the gene is ~2.7 Kb. This plasmid was designated pAFPZE.2.

Expression of the putative minigene in an SV40 vector. To test the constitutive expression of the AFP minigene, it was inserted into an SV40 vector which had previously been used successfully for other genes (21,34) and could accommodate the 2.7 Kb minigene (19). The minigene was inserted in both orientations into the late region of SV40 between the HpaII (0.725 map units) and BamHl (0.144 map units) sites. The virus carrying the minigene in the clockwise orientation, i.e. in the same orientation as late

transcription, is designated SVAFP-C, whereas SVAFP-A refers to the virus in which AFP transcription is in the same direction as SV40 early transcription (Fig. 2). The entire SV40 early region, the late promotor and polyadenylation site, the origin of replication and the 72 bp repeat, found to be essential for SV40 early gene expression (40), are included in the vector. By examining the expression of the minigene in both orientations relative to these SV40 control elements, possible effects of the SV40 control elements on AFP gene expression could be identified. The recombinant viruses were propagated by cotransfecting cells with tsA239 helper virus, which lacks early region functions at the nonpermissive temperature, but can supply the late proteins required for encapsidation and subsequent cell lysis.

Intracellular viral DNA prepared by Hirt extraction (22) was examined by restriction endonuclease cleavage and agarose gel electrophoresis. The results are shown in Fig. 3. Lane 2 contains an EcoRl digest of pAFPZES-A, where the 3 EcoRl fragments correspond to the 4.36 Kb pBR322 vector, the 3.0 Kb SV40 vector and the 2.7 Kb AFP minigene insert. EcoRl digestion of viral DNA should generate a helper tsA239 5.2 Kb band and two bands, 3.0 and 2.7 Kb for the chimeric viruses. In the ethidium bromide stain of lanes 3 and 4, the helper and SV40 vector bands can both be clearly seen, and given the relative molecular weights there would appear to be more vector DNA copies than helper DNA copies. However the 2.7 Kb band of the AFP minigene is only faintly visible for SVAFP-C (lane 4) and not at all for SVAFP-A (lane 3), although some smaller bands can be seen in the latter case.

The autoradiogram of the same gel, after Southern transfer and hybridization of the filter with labelled AFP minigene sequences, shows a wide dispersion of fragments containing AFP sequences, especially in SVAFP-A (lane 6), in addition to the expected 2.7 Kb band. Thus there has been fairly extensive alteration of the integrated AFP sequences, probably as the consequence of deletions, within the insert to generate smaller EcoRl fragments as well as deletions across one of the EcoRl sites that would result in linear viral genomes that migrate close to the helper viral genome (see lanes 4 and 7 for SVAFP-C). Nevertheless, at least some of the original recombinant template has been retained in both cases.

RNA was extracted from infected cells by the hot phenol method (23) and the poly A^+ fraction isolated by oligo dT cellulose chromatography (24). This RNA was electrophoresed through a denaturing polyacrylamide gel, transferred to a Millipore filter and analyzed for AFP sequences. The results are shown for SVAFP-C in Fig. 4. The expected length of the minigene

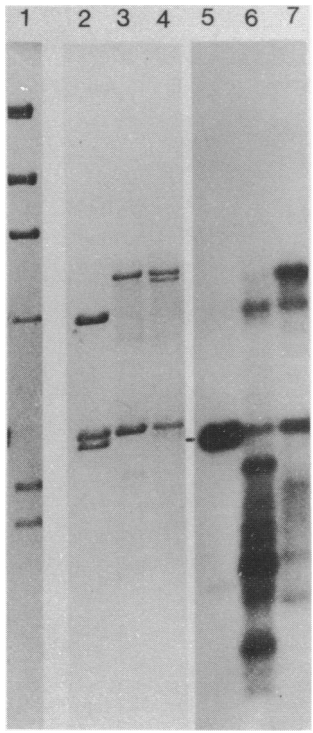

Figure 3. Analysis of viral DNA: Supercoiled DNA obtained from AGMK cells infected with recombinant virus. Lane 1, 1 μg λDNA [Miles] digested with HindIII: the sizes of the fragments, from top to bottom are 22.0, 9.4, 6.6, 4.4, 2.3 and 1.9 Kb: Lanes 2-7 EcoRl digestions of: Lanes 2 and 5, 1 μg SV40-pBR322-mini AFP (pAFPZESA), Lanes 3 and 6, 1 μg SVAFP-A + tsA239; Lanes 4 and 7, 1 μg SVAFP-C + tsA239. Lanes 1-4 are an ethidium bromide stained 1% agarose gel. Lanes 5-7 are an autoradiogram of lanes 2-4 hybridized to probe 3 (see Fig. 6).

product in which all three exons are spliced together is 129 + 52 + 141 = 322 nucleotides, but what is observed in lane B is a heterogeneous distribution of RNA lengths with a lower limit of 320 nucleotides plus a single band at 220 nucleotides. Since this heterogeneity could result from different extents of polyadenylation, which would have a pronounced effect on short transcripts, the RNA was pretreated with RNAseH in the presence of oligo dT, which removes the polyadenylated tails (41). Narrow bands are then obtained, as illustrated for two non-overlapping [^{32}P] labelled AFP minigene probes (probes 1 and 3 in Fig. 6) in lanes C and D of Fig. 4. Since lane D detects transcripts in the region of the 1st coding block of the minigene, and lane C those around the third coding block, it can be seen that the one transcript which strongly hybridizes to both is 320 nucleotides long, exactly corresponding to the expected size of a fully spliced transcript. Other minor higher molecular weight bands are visible but none hybridizes significantly to both probes. The other major transcript of 220 nucleotides, which is processed to a 140 base transcript after RNaseH digestion (lane D) will be discussed in detail below.

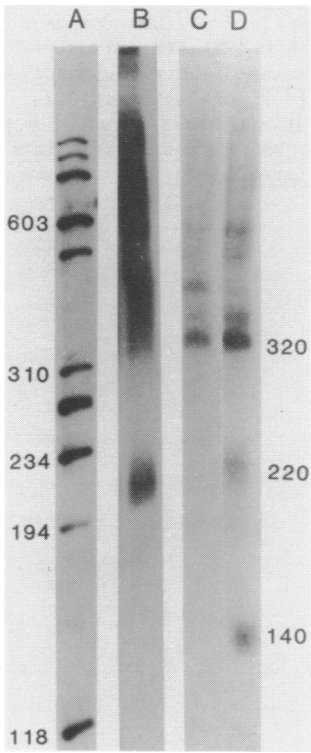

Figure 4. Acrylamide gel electrophoresis of viral RNAs. Autoradiogram of RNAs extracted from AGMK cells infected with recombinant virus, run on a 5% polyacrylamide /7M urea gel and transferred to nitrocellulose. Lane A, Single stranded ϕX174 DNA [Miles] digested with HaeIII and labelled with [γ^{32}P] ATP and polynucleotide kinase. Lane B: 2 µg RNA from SVAFP-C/tsA239 infected cells hybridized with the complete AFP minigene insert, pAFPZE.2. Lanes C and D: 3 µg RNA from SVAFP-C/tsA239 infected cells digested with 2.4 U RNAseH in the presence of oligo dT and hybridized to [^{32}P] labelled probe 3 and probe 1 respectively (see Fig. 6).

To substantiate that correct initiation, termination and splicing of the AFP minigene was indeed occurring, two kinds of S1 protection experiments were performed (27). The first experiment tested whether the AFP minigene was being correctly initiated by hybridizing poly A^+ RNA to an isolated DNA fragment which extended from a PvuII site 500 bp 5' to the first exon to a Sau3A site within the first exon (14), labelled with [γ^{32}P] ATP and polynucleotide kinase. The labelled DNA fragments protected by RNA from S1 nuclease were then characterized on a denaturing polyacrylamide gel. The results shown in Fig. 5 indicate initiation in both SVAFP-A and SVAFP-C at a site identical to that used by AFP mRNA which always exhibits a heterogeneous series of bands which may be due to microheterogeneity at the 5' terminus or to S1 nuclease imprecision (lower arrow). The amount of correctly initiated transcripts is much less for SVAFP-A than for SVAFP-C, which from previous experience could reflect either the lower copy number of intact SVAFP-A than SVAFP-C genomes (Fig. 3) or less efficient initiation (21). The present data

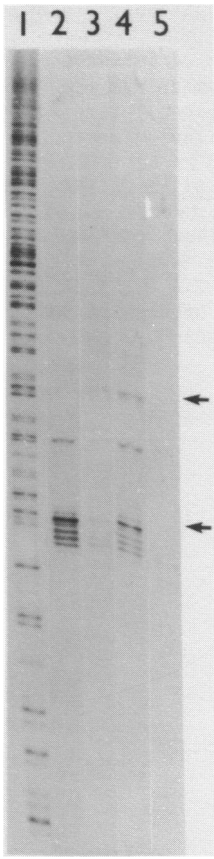

Figure 5. Mapping the 5'termini of viral RNAs. S1 protection experiment with a Sau3A/PvuII DNA fragment labelled at the Sau3A site with [γ^{32}P] ATP and polynucleotide kinase. Lane 1: Maxam-Gilbert A>C reaction of this fragment. Lanes 2-5 [^{32}P] DNA hybridized to poly A$^+$RNA and digested with S1 nuclease: Lane 2, 0.5 µg yolk sac RNA, Lane 3, 1 µg SVAFP-A/tsA239 RNA, Lane 4, 1 µg SVAFP-C/tsA239 RNA; Lane 5 1 µg RNA from mock infected AGMK cells.

does not allow us to distinguish between these possibilities. In addition to transcripts with the expected 5' terminus, there is evidence in the recombinants for an RNA species which begins at a site 35 bp upstream from the presumptive cap site (upper arrow), that is not observed in yolk sac AFP mRNA. The alternative explanation, that an upstream splicing site is being used is unlikely, since the same S1 protected fragment is observed for both orientations of the minigene relative to the early and late SV40 promotors. Examination of the sequence of this initiation site shows it to be identical to the normally used cap site, AAACA. However the only TATA sequence within this region is located 28 bp upstream of the AFP cap site in authentic AFP mRNA (14). A third band visible in lanes 2 and 4 which migrates between the 2 fragments is unlikely to represent a true RNA transcript, as numerous experiments with AFP mRNA have shown that it is not proportional to RNA

concentration and disappears with higher S1 nuclease concentrations.

The second S1 protection experiment, to examine the splicing of the AFP transcript, involved hybridizing poly A^+ RNA to the entire minigene under conditions favouring RNA-DNA hybridization, digesting with S1 nuclease, removing the RNA by alkali treatment, and running the protected single strand DNA fragments on a non-denaturing polyacrylamide gel. A Southern blot of this gel allows the individual bands to be examined with three labelled hybridization probes, each spanning a different coding block of the minigene (Fig. 6). For all three probes, protected fragments of the known exon sizes, 129, 52 and 141 bp respectively (14), are observed with SVAFP-C RNA and comigrate with the corresponding exons in AFP mRNA. It is clear from this that SVAFP-C splices all three exons correctly. The lower quantity of

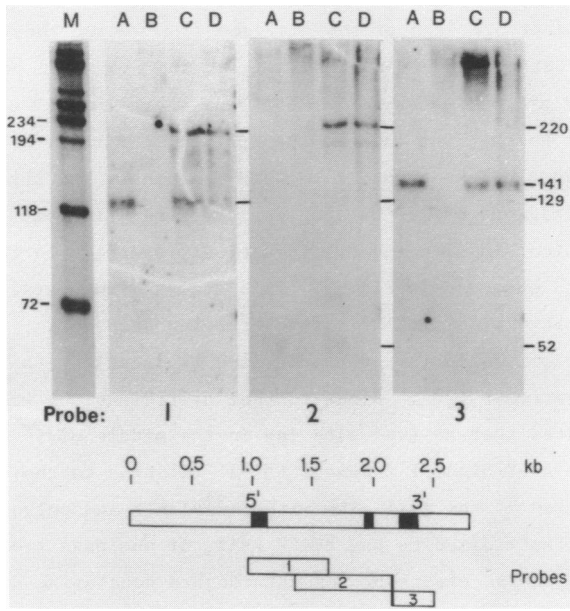

Figure 6. Determination of the internal structures of viral RNAs. S1 protection experiment of poly A+ RNAs from infected AGMK cells hybridized to 25 µg AFP minigene DNA (pAFPZE.2) and digested with S1 nuclease. After alkali treatment, a quarter of each sample was run on a 8% polyacrylamide gel: Lane A, 2.2 µg yolk sac RNA; Lane B 1 µg of RNA from mock infected cells; Lane C, 4 µg of SVAFP-C/tsA239 RNA. Lane D: 4 µg of SVAFP-A/tsA239 RNA. Lane M is a HaeIII digest of single stranded ϕX174 DNA end labelled with [^{32}P]. A blot of the gel was hybridized to the three [^{32}P] labelled probes derived from the following restriction fragments: probe 1 = HincII/HincII, probe 2 = AvaI/HindIII, probe 3 = HindIII/KpnI.

correctly initiated transcripts in SVAFP-A, as further evidenced by the relatively weak 1st coding block hybridization (probe 1), renders detection of the 2nd coding block fragment (probe 2) not possible. Note that the third exon of SVAFP-A (probe 3) is comparable in intensity to SVAFP-C, suggesting that it may also be appearing in transcripts initiated elsewhere in the vector.

This experiment and the RNA gel of Fig. 4 indicate that there are no correctly initiated transcripts detectable which fail to make either or both of the expected splices. Such transcripts would have been expected to have total lengths of 440, 1180 and 1320 nucleotides, leading to extra S1 protected fragments of 320, 1041 and 1320 nucleotides respectively. The two larger S1 protected fragments of SVAFP-C which hybridize to probe 3 have been sized at 630 bp and 760 bp on a lower percentage polyacrylamide gel. These may correspond to transcripts which are initiated at the SV40 late promotor, spliced to the 2nd or 3rd exon of the minigene from whence they run through the AFP polyadenylation signal and terminate at the SV40 late polyadenylation site. Hence they are not observed for SVAFP-A, where the AFP signals are inverted relative to the SV40 late transcriptional controls.

The Presence of an Alu Transcript in the AFP Minigene. The strongly hybridizing DNA fragment of 220 nucleotides observed with both probes 1 and 2 (Fig. 6) is identical in size and hybridization properties to the transcript of the same length observed in Fig. 3. This strongly suggests that it is an unspliced RNA located within the first intervening sequence.

This transcript was further characterized by locating its 5' end via an S1 protection experiment (27). Since further hybridizations with other probes had indicated that an AvaI site lay in the middle of the transcript, probe 2, a 730 bp AvaI/HindIII fragment (5'-3' relative to the minigene, see Fig. 6) was labelled at the AvaI site with [γ^{32}P] ATP and polynucleotide kinase. This was hybridized to the SVAFP RNAs, S1 nuclease treated and the protected DNA fragments electrophoresed through a denaturing sequencing gel. The labelled DNA was also sequenced by the Maxam-Gilbert method (42) and run on the same gel. Fig. 7A shows the S1 results (lane 1) alongside the A>C and G reactions (lanes 2 and 3). The small cluster of protected fragments observed in lane 1 clearly demonstrate initiation of the 220 base transcript on the noncoding strand (relative to the AFP gene) of the minigene.

The sequence of the region containing the transcript obtained by further Maxam-Gilbert sequencing is illustrated in Fig. 7B. The putative initiation site and the termination site of the transcript, based on its length, are

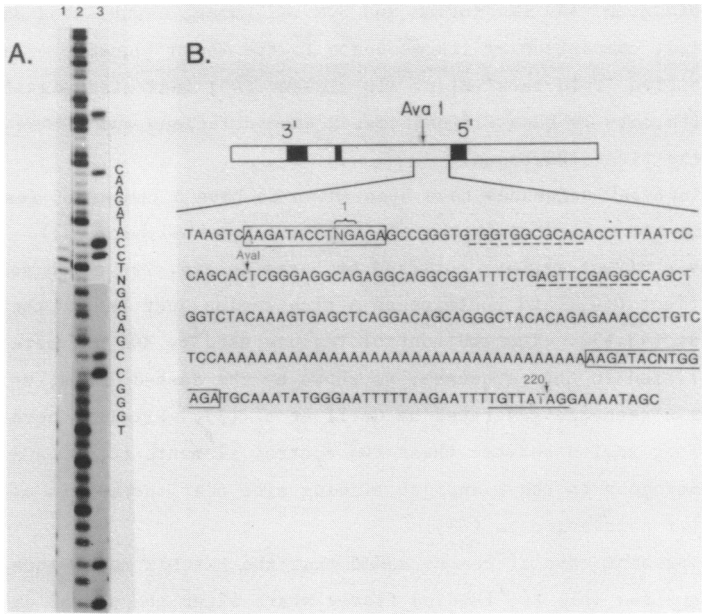

Figure 7. Mapping the 5' terminus of the Alu transcript. (A) S1 protection experiment using an AvaI/HindIII fragment (probe 2, see Fig. 6) labelled at the AvaI site with [γ^{32}P] ATP and polynucleotide kinase, and hybridized to 1 μg poly A+ RNA from SVAFP-C/tsA239 infected AGMK cells, digested with S1 nuclease and run on a denaturing polyacrylamide gel. Lane 1: S1 protected fragments. Lanes 2 and 3: Maxam and Gilbert A>C and G reactions on DNA alone. The sequence on the right is for the opposite strand.

(B) The top diagram shows the AFP mini-gene drawn in the opposite 5' to 3' orientation to that in Fig. 1. The 5' and 3' symbols refer to the AFP transcript. Sequence of the region coding for the 220 nucleotide transcript (N denotes an undetermined base). Putative initiation and termination sites are marked by the bracket (1) and underlined regions (220) respectively. The boxed region indicates the direct repeat and the dashed underlined regions are putative RNA polymerase III control regions.

marked by a bracket and underlining, respectively. Two experimental observations above now become explicable. First the long 38 residue poly A tract explains why this RNA was retained by an oligo dT cellulose column. Secondly, this tract begins 140 nucleotides from the 5' end of the transcript, which is exactly the size of RNA obtained upon RNAseH digestion (Fig. 4).

This region of DNA had previously been found to be highly repetitive when hybridized to chromosomal DNA and is found to hybridize to an Alu repeat element of Chinese Hamster cells (43) and a mouse repeat contained within the

β-major globin gene (44) (R. Ingram and S.M. Tilghman, unpublished experiments). Indeed, comparison of its sequence to the known sequence of a mouse middle repetitive 'fold-back' B1 or Alu element (45) indicated significant homology, with only 10 base differences, 2 base deletions and 2 base additions over the first 129 bases.

The "Alu type" sequences have been shown to have a number of features in common, which are reflected in the Alu sequence reported here (43). It is bracketed by a direct repeat, outlined by boxes in Fig. 7B, which sometimes can be imperfect (46), and contains an A rich region just before the second direct repeat (43,45). The two control regions used by RNA polymerase III can be identified in this sequence, as shown by the dashed lines, using a tRNA consensus sequence generated by Galli et al (7). Finally there is a 14 nucleotide long region between these two control elements (GGAGGCAGAGGCAG) which is homologous to the T-antigen binding site near the origin of replication of papovaviruses (47).

In the present case it can be added that the RNA has no obvious protein product, since the only two reading frames start after the second direct repeat and have no terminator. Furthermore, the transcript presumably terminates towards the end of an A/T rich region (cf 7) only 50 nucleotides from the end of the first exon of the AFP gene.

DISCUSSION

The results presented illustrate that it is possible to shorten a gene by removal of internal introns and exons and yet still retain its ability to direct the synthesis of a stable transcript *in vivo*. Not only did the transcript in SVAFP-C initiate and terminate correctly, but it also made the expected two splices. It is likely that this is also true for SVAFP-A as well, although we could not definitively detect the presence of the second coding block due to the sensitivity of the method. Nevertheless, AFP transcription occurs in both viruses, albeit at markedly different rates.

Two features of this result bear further comment. Although there is minor use of an initiation site 35 bases upstream from the site used in yolk sac AFP mRNA, there is no evidence for its use in any tissue that expresses AFP, i.e. fetal liver, yolk sac, and gut (R.W. Scott, unpublished results). As it occurs in both SVAFP-C and SVAFP-A, it is unlikely to be due to the influence of the SV40 72 bp repeat (48,49). Alternatively it may be mediated by an effect on the chromatin structure that is vector-specific. Indeed, in DNA transfection studies with some SV40/pBR322/mini AFP recombinants, the

upstream initiation site becomes predominant (R. Scott, unpublished observations).

Secondly, the monkey kidney cells are able to splice together the second and third exons of the minigene in what is a chimeric splice of the second and fifteenth exons of AFP. Thus, there is nothing in principle to prevent non-adjacent exons from splicing together. Having started with one splice junction, the splicing complex appears to search only for the nearest specific donor or acceptor consensus sequence (50,51). If correct splicing involves any kind of secondary structure between the two exons to be spliced (52), then the complementarity would have to be a general one in order to explain the ability to make chimeric splices by design. Furthermore the removal of the first intron does not depend on the prior splicing of a particular downstream intron, supporting the view that unless the first intron is always spliced first, the order of splicing is not critical for the complete splicing of the primary transcript.

The other main finding of the present work is the presence of an "Alu-type" repeat element in the first intervening sequence of the AFP gene which is transcribed in vivo. A number of repetitive sequences have now been located not only in the flanking sequences of known genes, but also within their intervening sequences (11,53,54). In vitro studies have previously shown the ability of specific Alu sequences to be transcribed (55,56,57) and there is good evidence that transcripts of at least some of the Alu repeats are represented in the small nuclear RNA population of cells (57). To our knowledge, however, the present example constitutes the first time that such a transcript has been unequivocally identified with a specific gene in vivo. It should be added that the transcription of this gene within a gene does not seem to depend on the structure of the vector nor on the heterologous cell infected. Not only is it observed with both orientations of the minigene, but it is also seen upon transfection of HeLa cells with various SV40/pBR322/mini AFP recombinants (R. Scott, unpublished results).

What could be the role of these transcripts in constitutive or regulated AFP gene expression? Two general roles for these Alu repeats have been proposed previously (47). First, it is possible that they are involved in RNA processing, given their hydrogen bonded association with poly A^+ cytoplasmic and heteronuclear RNA (58). This association can be readily visualized in the present case, since the transcript is made from the opposite strand to that from which AFP mRNA is made and hence it is capable of hybridizing to the AFP primary transcript. However, an inhibitory role for the 220 nucleo-

tide transcript during RNA processing is not consistent with our finding the mature transcript in the presence of an excess of this RNA in heterologous cells.

Secondly, it has been proposed that the transcripts could act as DNA replication primers (47). This possibility was based on the observation that they contain a region in common with some viral origins of replication. The 220 nucleotide transcript seems well designed for the intermolecular hybridization required for either of the above two functions, since it apparently contains very little secondary structure as judged by computer analysis (59). Indeed it is particularly devoid of structure in the poly A tract, leaving this region free to hybridize. The finding of this repetitive tract, albeit with different sequences, in all these repeat families (43) lends support to this suggestion.

The finding of an active Alu transcription unit within 400 bp of the initiation site for AFP transcription, raises several interesting possibilities regarding the expression and regulation of these genes in vivo. Can both genes be transcribed from the same template simultaneously? In the experiments presented here, multiple copies of the template precludes a definitive answer to this question. In mouse cells that are actively transcribing the AFP gene, however, we do not detect transcripts of the Alu gene in the relative quantities observed here (see Fig. 6). Either the 220 nucleotide RNA is not present in these cells or its synthesis and/or stabilization relative to AFP mRNA is greatly depressed. One could envision that transcription of the AFP gene might interfere with the transcription of Alu sequence via some physical and topological interference between the two transcriptional complexes or as some indirect effect of the alteration of adjacent chromatin structure.

In any case, the AFP minigene seems well suited to addressing the role of the transcription of the repeat element on AFP expression in mouse cells using DNA mediated gene transfer.

ACKNOWLEDGEMENTS

This work was supported by NCI grant CA-28050, a Basil O'Connor Starter Grant, NIH grants CA-06927 and RR-05539 awarded to the Institute for Cancer Research and an appropriation from the Commonwealth of Pennsylvania. S.M.T. is the recipient of an American Cancer Society Junior Faculty Award, P.R.Y. and R.W.S. are U.S.P.H.S. postdoctoral fellows.

REFERENCES

* Abbreviations used: kilobase pairs, Kb; base pairs, bp; 0.15M NaCl, 0.015 M Na citrate, SSC; α-fetoprotein, AFP.

1. Wigler, M., Silverstein, S., Lee, L.-S., Pellicer, A., Cheng, Y.-C., and Axel, R. (1977) Cell 11, 223-232.
2. Shortle, D. and Nathans, P. (1978) Proc. Natl. Acad. Sci., U.S.A. 75, 2170-2176.
3. Shortle, D., Koshland, D., Wienstock, G. M., and Botstein, D. (1980) Proc. Natl. Acad. Sci., U.S.A. 77, 5375-5379.
4. Sakonju, S., Bogenhagen, D. F., and Brown, D. D. (1980) Cell 19, 13-25.
5. Bogenhagen, D. F., Sakonju, S., and Brown, D. D. (1980) Cell 19, 27-35.
6. McKnight, S. L., Gavis, R., Kingsbury, R., and Axel, R. (1981) Cell 25, 385-398.
7. Galli, G., Hofstetter, H., and Birnsteil, M. L. (1981) Nature 294, 626-631.
8. Fowlkes, D. M. and Shenk, T. (1980) Cell 22, 405-413.
9. Lee, F., Mulligan, R., Berg, P., and Ringold, G. (1981) Nature 294, 228-232.
10. Huang, A. L., Ostrowski, M. C., Berard, D., and Hager, G. L. (1981) Cell 27, 245-255.
11. Kioussis, D., Eiferman, F., van de Rijn, Gorin, M. B., Ingram, R. S., and Tilghman, S. M. (1981) J. Biol. Chem. 256, 1960-1967.
12. Dziadek, M. and Adamson, E. (1978) J. Embryol. Exp. Morphol. 43, 289-313.
13. Hogan, B. L. M., Taylor, A., and Adamson E. (1981) Nature 291, 235-237.
14. Eiferman, F. A., Young, P. R., Scott, R. W., and Tilghman, S. M. (1981) Nature 294, 713-717.
15. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L., and Gilbert, W. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 3727-3731.
16. Grunstein, M. and Hogness, D. C. (1975) Proc. Natl. Acad. Sci., U.S.A. 72, 3961-3965.
17. Maniatis, T., Jeffrey, A., and Kleid, D. G. (1975) Proc. Natl. Acad. Sci., U.S.A. 72, 1184-1188.
18. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Fischer, J., Rutter, W. J., and Goodman, H. M. (1977) Science 196, 1313-1319.
19. Pavlakis, G. N., Hizuka, N., Gorden, P., Seeburg, P., and Hamer, D. H. (1981) Proc. Natl. Acad. Sci., U.S.A. 78, 7398-7402.
20. Seigh, M. J., Topp, W. C., Hanich, R., and Sambrook, J. F. (1978) Cell 14, 79-88.
21. Hamer, D. H., Kaehler, M., and Leder, P. (1980) Cell 21, 697-708.
22. Hirt, B. (1967) J. Mol. Biol. 26, 365-369.
23. Scherrer, K. (1969) In Fundamental Techniques in Virology, (Habel, K. and Saltzman, N. P., eds.) Academic Press, New York, pp.413-432.
24. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci., U.S.A. 69, 1408-1412.
25. Wahl, G. M., Stern, M., and Stark, G. R. (1979) Proc. Natl. Acad. Sci., U.S.A. 76, 3683-3687.
26. Thomas, P. S. (1980) Proc. Natl. Acad. Sci., U.S.A. 77, 5201-5205.
27. Berk, A. J. and Sharp, P. A. (1977) Cell 12, 721-732.
28. Southern, E. M. (1975) J. Mol. Biol. 38, 503-517.
29. Gluzman, Y., Sambrook, J., and Frisque, R. (1980) Proc. Natl. Acad. Sci., U.S.A. 77, 3848-3902.
30. Benoist, C. and Chambon, P. (1981) Nature 290, 304-310.
31. Grosschedl, R. and Birnsteil, M. (1980) Proc. Natl. Acad. Sci., U.S.A. 77, 7102-7106.

32. Struhl, K. (1981) Proc. Natl. Acad. Sci., U.S.A. 78, 4461-4465.
33. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
34. Hamer, D. H., Smith, K. D., Boyer, S. H., and Leder, P. (1979) Cell 17, 725-735.
35. Gruss, P., Lai, C.-J., Dhar, R., and Khoury, G. (1979) Proc. Natl. Acad. Sci., U.S.A. 76, 4137-4321.
36. Gruss, P., Efstratiadis, A., Karathanasis, S., Konig, M., and Khoury, G. (1981) Proc. Natl. Acad. Sci., U.S.A. 78, 6091-6095.
37. Ghosh, P. K., Roy, P., Barkah, A., Mertz, J. E., Weissman, S., and Lebowitz, P. (1981) Proc. Natl. Acad. Sci., U.S.A. 78, 1386-1390.
38. Hentschel, C. C. and Birnstiel, M. L. (1981) Cell 25, 301-313.
39. Nagata, S., Mantei, N., and Weissman, C. (1980) Nature 287, 401-408.
40. Gruss, P., Dhar, R., and Khoury, G. (1981) Proc. Natl. Acad. Sci., U.S.A. 78, 943-947.
41. Schibler, U., Tosi, M., Pittet, A.-C., Fariani, L., and Wellauer, P. K. (1980) J. Mol. Biol. 142, 93-116.
42. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci., U.S.A. 74, 560-564.
43. Haynes, S. R., Toomey, T. P., Leinwand, L., and Jelinek, W. R. (1981) Mol. Cell Biol. 1, 573-583.
44. Leder, A., Miller, H. I., Hamer, D. H., Seidman, J. G., Norman, B., Sullivan, M., and Leder, P. (1978) Proc. Natl. Acad. Sci., U.S.A. 75, 6187-6191.
45. Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayev, A. A., and Georgiev, G. P. (1980) Nucl. Acids Res. 8, 1201-1215.
46. Stumph, W. E., Kristo, P., Tsai, M.-J., O'Malley, B. W. (1981) Nucl. Acids Res. 9, 5383-5397.
47. Jellinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Hovck, C. M., Deininger, P. L., and Schmid, C. W. (1980) Proc. Natl. Acad. Sci., U.S.A. 77, 1398-1402.
48. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P., and Chambon, P. (1981) Nucl. Acids Res. 9, 6047-6068.
49. Banerji, J., Rusconi, S., and Schaffner, W. (1981) Cell 27, 249-308.
50. Sharp, P. A. (1981) Cell 23, 643-646.
51. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978) Proc. Natl. Acad. Sci., U.S.A. 75, 4853-4857.
52. Trapnell, B. C., Tolstroshev, P., and Crystal, R. G. (1980) Nucl. Acids Res. 8, 3659-3672.
53. Page, G. S., Smith, S., and Goodman, H. M. (1981) Nucl. Acids Res. 9, 2087-2104.
54. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., and Chambon, P. (1979) Nature 282, 567-574.
55. Fuhrman, S. A., Deininger, P. L., LaPorte, P., Friedmann, T., and Geiduschek, E. P. (1981) Nucl. Acids Res. 9, 6439-6456.
56. DiSegni, G., Carrara, G., Tocchini-Valentini, G. R., Shoulders, C. C., and Baralle, F. E. (1981) Nucl. Acids Res. 9, 6709-6722.
57. Haynes, S. R. and Jelinek, W. R. (1981) Proc. Natl. Acad. Sci., U.S.A. 78, 6130-6134.
58. Jelinek, W. and Leinwand, L. (1978) Cell 15, 205-214.
59. Wuilmart, C., Urbain, J. and Givol, D. (1977) Proc. Natl. Acad. Sci., U.S.A. 74, 2526-2530.