Volume 14 Number 3 1986 Nucleic Acids Research

In vitro splicing of simian virus 40 early pre mRNA

Jonathan C.S.Noble, Carol Prives and James L.Manley

Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Received 20 November 1985; Accepted 20 December 1985

ABSTRACT

The products of splicing of simian virus 40 early pre mRNA in HeLa cell nuclear extracts have been characterized. Of the two alternative splicing patterns exhibited by this precursor in vivo, which involve the use of alternative large T and small t 5' splice sites and a single shared 3' splice site, only one, producing large T mRNA, was found to occur in vitro. A number of possible intermediates and byproducts of splicing of large T mRNA were observed, including free large T 5' exon, lariat form intron joined to 3' exon and free lariat and linear forms of large T intron. The formation of these products argues strongly for a basic similarity in the mechanism underlying large T and other, non-alternative splices. A collection of RNAs resulting from protection of early pre mRNA at specific points from an endogenous 5' to 3' exonuclease activity in vitro have also been observed. The regions of the precursor RNA protected map to positions immediately upstream of the 5' splice sites of large T and small t and the lariat branchpoint, and may represent interaction of these regions with components of the splicing machinery.

INTRODUCTION

Considerable progress has been made in recent years in regard to our understanding of the mechanism and biochemistry of eukaryotic mRNA splicing. This has been due in large part to the development of cell-free systems able to splice pre mRNAs in vitro (1-4). On the basis of kinetic studies in vitro (5,6) as well as structural analyses of pre mRNA processing products in vitro (5-7) and in vivo (8-11), a general pathway for pre mRNA splicing has been deduced (reviewed in 12). This pathway includes as a first step cleavage at the 5' splice junction and formation of a lariat involving the 5' end of the intron joined by a 2'-5' phosphodiester linkage to an A residue upstream of the 3' splice junction. This is then thought to be followed by cleavage at the 3' splice junction and ligation of the 5' and 3' exons. Although both steps in this process involve what would appear to be several sub reactions, these are believed to occur in a concerted fashion (12).

The sequences defining 5' and 3' splice sites within the pre mRNAs of higher eukaryotes are known to comprise a number of elements. Among the most obvious of these are sequences located at the 5' and 3' splice junctions themselves, which conform to a consensus

at the 5' splice site of CAG/GTAAGT and at the 3' splice site of $(Py)_nNCAG/G$ (13). In addition a rather liberal consensus sequence of PyNPyTPuAPy has been found to occur 18-37 nucleotides (nt) upstream of 3' splice junctions, the penultimate residue of which is involved in lariat branch formation (6,7,9-11,14,15). The involvement of U1 and U2 snRNPs in recognition of these elements has been hypothesized (16-18), and strengthened in the case of U1 snRNP by the demonstration of a requirement for this particle for splicing in vivo (19) and in vitro (20,21).

Whatever the identity of factors involved in splice site recognition, the mechanism underlying selection of appropriate splice site pairings must satisfy a number of criteria. Among the most exacting of these is the ability of some pre mRNAs to undergo alternative patterns of splicing. Alternative splicing pathways are exhibited by a number of viral genes including early region E1a of adenovirus (22), and early and late regions of SV40 (23-25), and also cellular genes including myosin light chain (26), αA crystallin (27) and troponin T (28). The early region of SV40, which is perhaps the simplest example of this class of RNAs, contains two 5' splice sites which in vivo may be spliced to a single shared 3' splice site yielding either large T or small t mRNAs (23). The ability to form these alternative pairings of 5' and 3' splice sites is difficult to reconcile with the notion of a simple directional scanning mechanism of splice site selection (29).

In order to define more clearly the mechanism involved in determining alternative large T and small t splicing patterns in the SV40 early region we have undertaken an analysis of the splicing of this pre mRNA in vitro. Surprisingly, we have observed that although the large T 5' splice was utilised with moderate efficiency, utilisation of the small t 5' splice site could not be detected in vitro. In addition to the spliced product, a number of probable intermediates and byproducts of splicing of this precursor to large T mRNA were observed. Finally, a collection of RNAs were produced that may represent binding of splicing factors to early pre mRNA at the 5' splice sites of large T and small t and a putative large T lariat branchpoint.

MATERIALS AND METHODS

Template construction

A 45 bp EcoRI-HindIII fragment containing a synthetic promoter with the consensus sequence for E. coli RNA polymerase (the gift of J. Rossi, 30) cloned into plasmid vector pEA300 (obtained from J. Brosius, 31) was fused to the 1169 bp HindIII B fragment of SV40 early region (nt 5171-4002) to create pYSVHdB.

Preparation of precursor RNA

Precursor RNA was prepared by run-off transcription of template pYSVHdB, and contained 2 nt and 15 nt at the 5' and 3' ends, respectively, of bacterial sequences. Transcription reactions contained in 50 µl, 0.25 pmol template and 2 pmol E. coli RNA polymerase (the gift of S. Beychok) in a buffer comprising, 20 mM Tris.HCl (pH 7.9), 0.4 M KCl, 10 mM $MgCl_2$,

0.1 mM EDTA, 2.5 mM dithiothreitol, 1 mM ATP, CTP and UTP, 500 μM GTP and 25 μCi α-^{32}P GTP. After 1 hr at 30°C transcription was terminated by adding rifampicin at 500 μg/ml and 10 min later DNase I at 20 μg/ml and incubating for 10 min at 37°C. RNA was then extracted with phenol:chloroform (1:1), purified over a column of Sephadex G100, and precipitated with ethanol. Transcriptions performed using this protocol yielded 2-4 pmol of precursor RNA.

In vitro splicing reaction

Nuclear extracts were prepared from HeLa cells using a modification of the method of Dignam et al. (32). Standard splicing reactions contained in 25 μl, 0.15 pmol precursor RNA and 5 μl nuclear extract in a buffer comprising 4 mM Tris.HCl (pH 7.9), 4% (v/v) glycerol, 20 mM KCl, 2 mM $MgCl_2$, 40 μM EDTA, 0.2 mM dithiothreitol, 4 mM creatine phosphate (disodium salt) and 500 μM ATP. After 3 hr at 30°C processing was terminated and RNA extracted using the method of Manley & Gefter (33).

S1 nuclease analysis

S1 nuclease mapping of processed RNAs was carried out by the method of Berk and Sharp (24) using as a probe the 1169 bp HindIII B fragment of SV40 early region (SV40 nt 5171-4002) labeled at either 5' or 3' ends. The 3' end labeled probe was purified to single strands by hybridization to M13 (34). Typically, 10-25 ng of probe was used for one-fifth of the products of a splicing reaction. Hybridization was carried out for 4-5 hr at 45°C, and S1 nuclease digestion performed using 4 x 10^3 units/ml S1 for 1 hr at 40°C. Following digestion, products were analyzed either under denaturing conditions on sequencing gels (35), or under native conditions using the method of Manley & Gefter (33).

Primer extension analysis

Primer extension was performed by the method of Treisman et al. (36) using 5' end labeled 28 mer and 32 mer synthetic oligonucleotide primers complementary to large T intron (SV40 nt 4843-4816) or 3' exon (SV40 nt 4546-4517) regions of early pre mRNA, respectively. Generally, 0.1-0.2 ng primer was used for one-fifth of the products of a splicing reaction, although when sequencing of primer extended products was to be performed 5 ng of primer and 25 times the amount of processed RNA were used. Hybridizations were performed for 16 hr at 30°C and reverse transcription carried out using 120-400 units/ml reverse transcriptase (Life Sciences Inc.) for 1 hr at 40°C. Primer extended products were then analyzed on sequencing gels. For DNA sequencing, preparative scale primer extensions were fractionated on denaturing buffer step gradient gels and individual cDNAs eluted and sequenced (35).

RESULTS

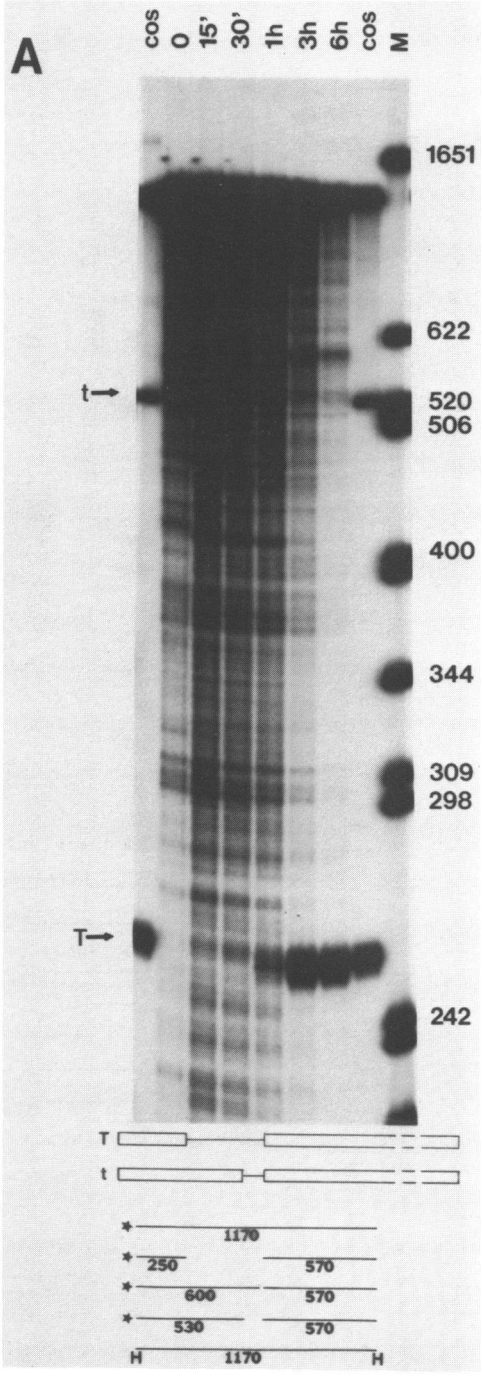
Kinetics of appearance of products of the splicing reaction

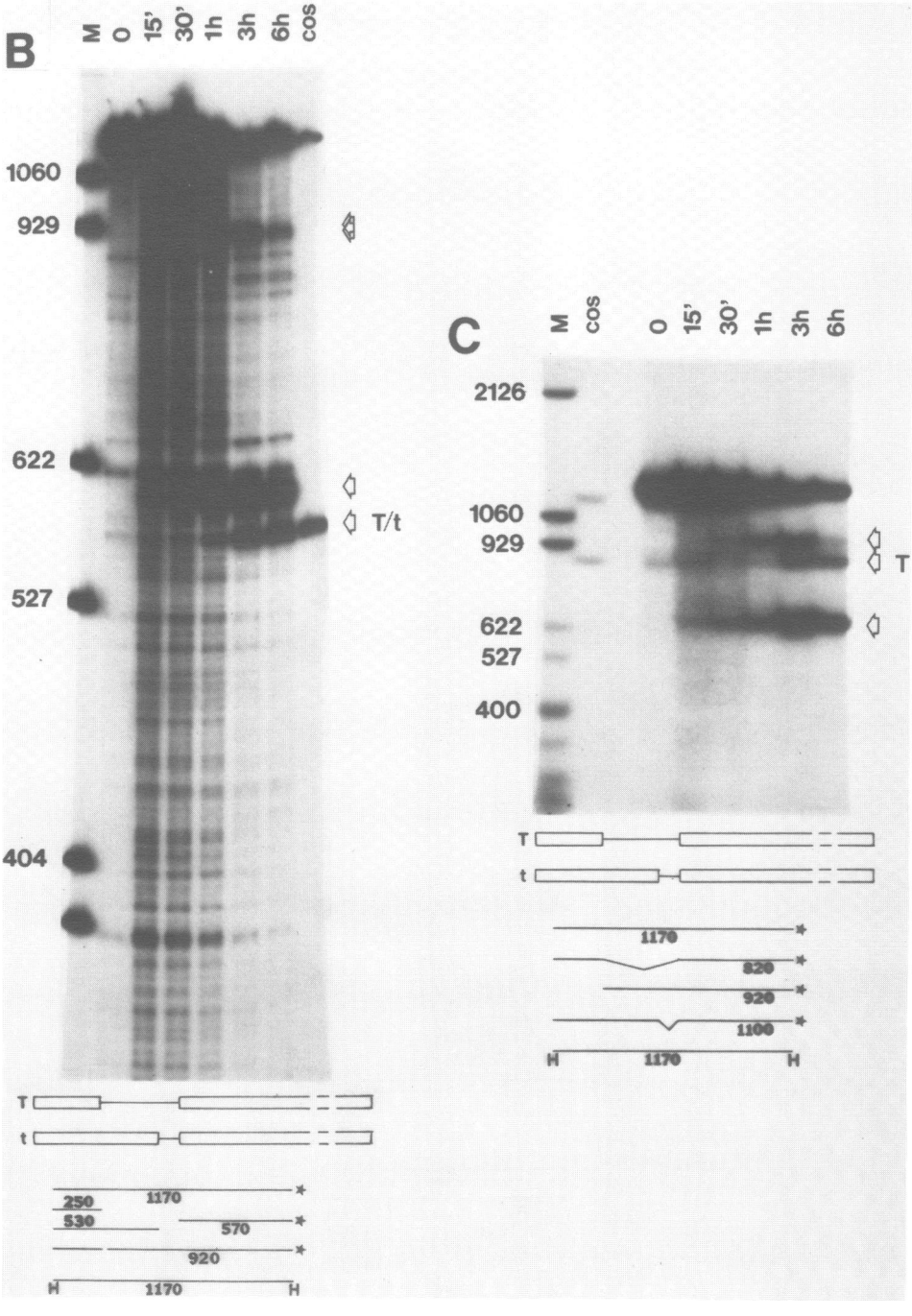
Uncapped precursor RNA spanning SV40 nt 5171 to 4002 was synthesized in vitro using E. coli RNA polymerase and pYSVHdB as template. This RNA was then incubated in HeLa cell

nuclear extracts. Processing reactions were terminated over a range of times and the products analyzed by S1 nuclease mapping. The probe used in these analyses was the SV40 HindIII B fragment which spans the entire length of the precursor, including both large T and small t 5' splice sites and the 3' splice site. S1 mapping using this probe labeled at the 3' end coupled to denaturing gel analysis was employed to distinguish possible utilization of large T and small t 5' splice sites. From the results of this analysis, shown in Figure 1A, it is apparent that a substantial fraction of the input RNA, perhaps 75%, was degraded during the course of the incubation. In addition, at later stages of the reaction, beginning at 1hr, cleavage of the precursor RNA at the 5' splice site of large T, represented by the appearance of a protected probe fragment of 258 nt, was seen to occur. In contrast, over the same period no detectable cleavage at the small t 5' splice site, which would be represented by a probe fragment of 538 nt, was observed. This lack of small t 5' splice site cleavage has been consistently observed in vitro in all incubation conditions tested, using both uncapped and uncapped precursor RNAs, and contrasts with the ability of the same capped precursor to be cleaved at both large T and small t 5' splice sites when microinjected into Xenopus laevis oocytes (37).

In order to determine whether cleavage of the early precursor RNA at the 3' splice site and subsequent ligation of large T 5' and 3' exons was occurring during the course of the incubation, S1 nuclease mapping using the same probe labeled this time at the 5' end and coupled to denaturing and native gel analysis was performed. From the results of these analyses, shown in Figure 1B and C, it is apparent that cleavage of the precursor at the 3' splice site did occur, as evidenced by the appearance of a protected probe fragment of 570 nt (Figure 1B), and with kinetics closely paralleling those of large T 5' splice site cleavage. That this cleavage product was ligated to the large T 5' exon is shown in Figure 1C by the simultaneous appearance of an RNA:DNA hybrid of length approximately 820 bp corresponding to spliced product. In addition, another product, represented by a protected probe fragment of length 920nt, is apparent in Figure 1B and C. The length of this protected fragment maps the 5' end of this product to the 5' end of the intron, suggesting that it probably reflects cleavage at the 5' splice site of large T without cleavage at the 3' splice site.

Together with these expected products, three unexpected products, represented by probe fragments of 930, 650 and 600 nt (Figure 1B and C), corresponding to cleavages close to and upstream of large T and small t 5' splice sites and the 3' splice site respectively, accumulated as very early products during the the reaction. That these products were due to the activity of a 5' to 3' exonuclease endogenous to the nuclear extract is suggested by the native and denaturing gel analyses shown in Figures 1C and A. The results of Figure 1C demonstrate on the one hand that these RNAs are cleavage products, while Figure 1A shows that products corresponding to the 5' portions of the precursor RNA cleaved at each of these points could not be detected. This notion is strengthened further by the observation that when capped precursor RNA was used as a substrate in the processing reaction, these species were greatly


A
cos
0
15'
30'
1h
3h
6h
cos
M
1651
622
t
520
506
400
344
309
298
T
242
T
t
1170
250
570
600
570
530
570
H
1170
H


B
M
0
15'
30'
1h
3h
6h
cos
1060
929
622
527
404
T/t
T
t
250
530
1170
570
920
H
1170
H
C
M
cos
0
15'
30'
1h
3h
6h
2126
1060
929
622
527
400
T
t
1170
820
920
1100
H
1170
H

reduced in amount (M. Chaudhuri, personal communication).

Although these unexpected products appear to differ significantly in their kinetics of appearance from spliced large T mRNA, two (920 and 600 nt) increased considerably over 3hr while the third (650 nt) remained at a constant and relatively low level throughout the incubation, these RNAs have the same requirements (including ATP and Mg) and exhibit the same optimum conditions for production as spliced large T mRNA (unpublished data).

Analysis of processing products in polyacrylamide-urea gels

In order to define more clearly the products of the in vitro splicing reaction, RNA isolated after a 3hr incubation in the nuclear extract was fractionated on a 5% polyacrylamide-urea gel. The pattern of products revealed by this method, shown in Figure 2A, comprises eight components (I-VIII) including unprocessed precursor RNA (III). Two of these products migrate in this gel with apparent sizes greater than precursor (I,II), while the remainder migrate with sizes approximately 930 nt (IV), 845 nt (V), 600 nt (II), 340 (VII), and 255nt (VIII). From a comparison of the sizes of these RNAs and the size of protected probe fragments determined in the S1 analysis shown in Figure 1B and C, two of the components may be identified tentatively as products of cleavage upstream of the large T 5' (IV) and 3' splice sites (VI). In addition, it may be noted that products VII (340 nt) and VIII (255 nt) are of sizes close to those expected for free large T intron (346 nt) and free large T 5' exon (261 nt) respectively, while product V (845 nt) is of a size close to that expected for large T spliced RNA (840 nt).

When each of these RNAs was eluted and analyzed on a 3.5% polyacylamide-urea gel, shown in Figure 2B, all but two migrated with a consistent apparent size. The two exceptions (I and II), which on the 5% polyacrylamide gel had both migrated with extremely large apparent sizes, ran on the 3.5% polyacrylamide with greatly reduced sizes, the smaller of the two (II) migrating with an apparent size of only 490 nt. One interpretation of the anomalous behavior of these latter two species on different percentage polyacrylamide gels would be that they possess a non-linear structure. Similar anomalous electrophoretic mobilities have been observed by others for splicing intermediates and byproducts containing lariat form intron sequences (5,6,9-11).

Figure 1. S1 nuclease analysis of processed RNAs. (A) S1 mapping performed using a single-stranded 3' end labeled HindIII B fragment probe (SV40 nt 5175-4003) analyzed under denaturing conditions. Cos corresponds to total cytoplasmic RNA from an SV40-transformed monkey cell line. M represents size markers. A diagram of expected S1 nuclease-resistant probe fragments is shown below. (B) and (C) S1 mapping carried out using a double-stranded 5' end labeled HindIII B fragment probe analyzed under denaturing (B) or native conditions (C). Note that RNA:DNA hybrids migrate slightly above DNA markers of equivalent size in the native analysis. Diagrams of expected S1 nuclease resistant probe and RNA:DNA hybrid fragments are shown below.

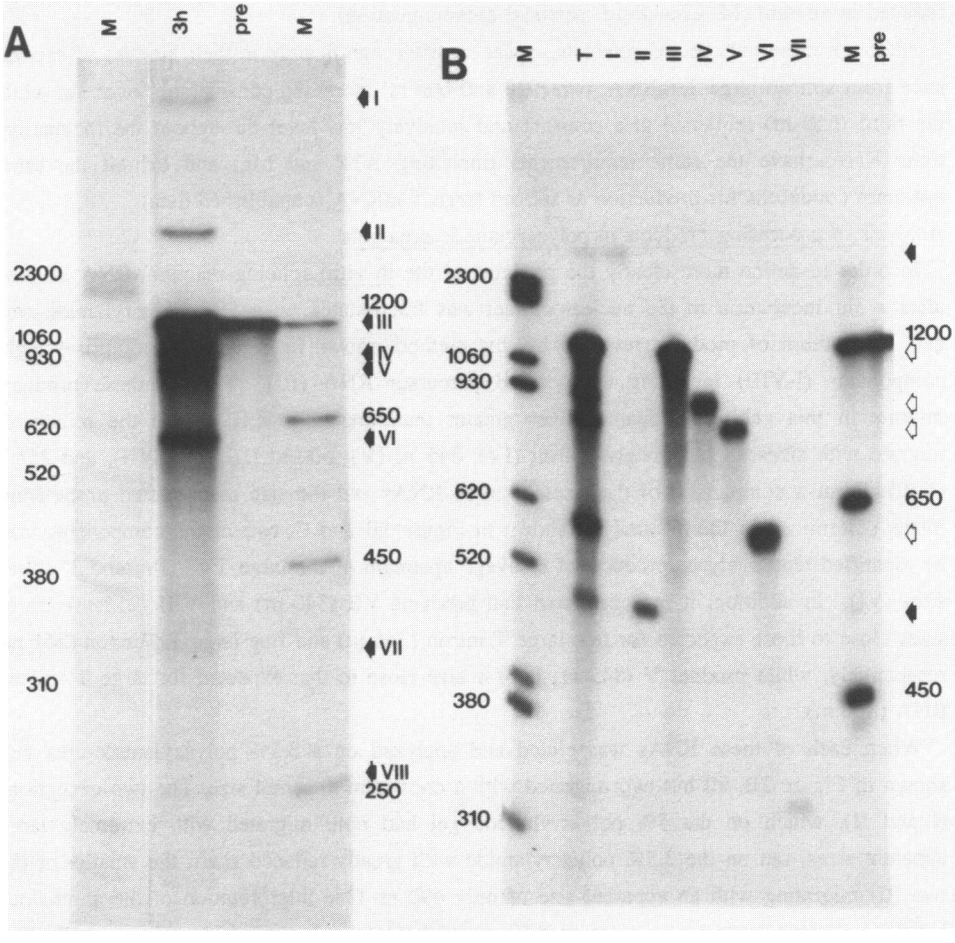


Figure 2. Polyacrylamide-urea gel analysis of processed RNAs. (A) Processed RNA run on a preparative 5% polyacrylamide-8 M urea gel. Pre corresponds to precursor RNA. Products of the splicing reaction are designated in descending order of apparent size by the roman numerals I through VIII. (b) Individual processed RNAs purified from the gel in part (A) run on a 3.5% polyacrylamide-8 M urea gel. Purified RNAs are indicated by their designation in part (A). T corresponds to the total products of a splicing reaction unfractionated.

S1 nuclease and primer extension analysis of processing products

Further characterization of each of the products of the in vitro splicing reaction was achieved by S1 nuclease and primer extension analysis of individual RNAs purified from the 5% polyacrylamide-urea gel shown in Figure 2a. S1 nuclease analysis of component VIII (255 nt), using as a probe the HindIII B fragment labeled at the 3' end, yielded a protected probe fragment of 255 nt, identifying this species as the free 5' exon of large T (unpublished data).

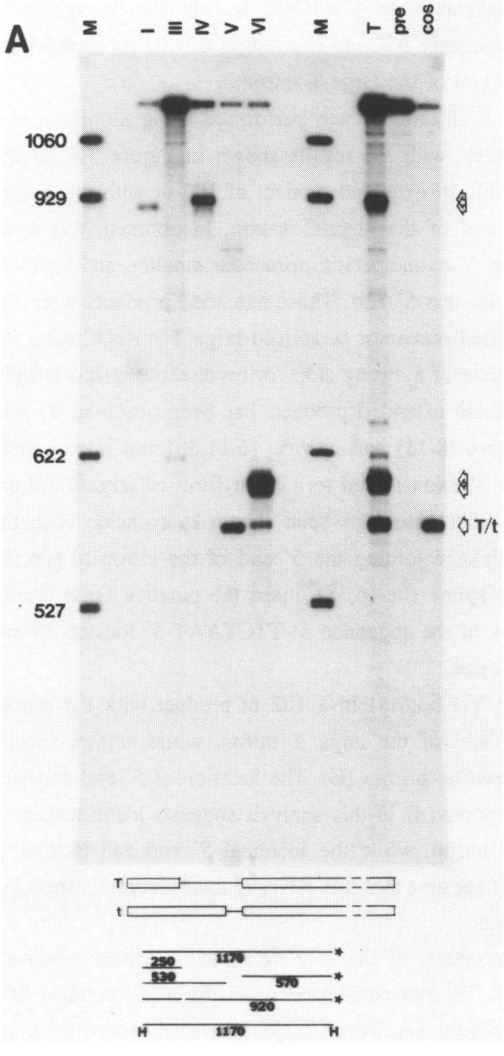
S1 analysis of components I,IV,V and VI using the same probe labeled at the 5' end was performed with the result shown in Figure 3A. The size of the protected probe fragment observed for component IV (930 nt) suggests that this RNA corresponds to the product of precursor cleavage upstream of the large T 5' splice site. The pattern of protected probe fragments evident for component VI (600 nt doublet and 560 nt) suggests that this RNA represent the product of precursor cleavage upstream of the 3' splice site. The protection by this RNA of a probe fragment of 560 nt might indicate that a small degree of uncoupling of cleavage at the 3' splice junction from T exon ligation has occurred. The protection of a 560 nt probe fragment by component V establishes that this RNA is a product of cleavage at the 3' splice junction, consistent with its identification as spliced large T mRNA. In addition, component I yields a protected probe fragment of approximately 920 nt. The length of this protected fragment maps the 5' end of this RNA to the 5' end of the large T intron.

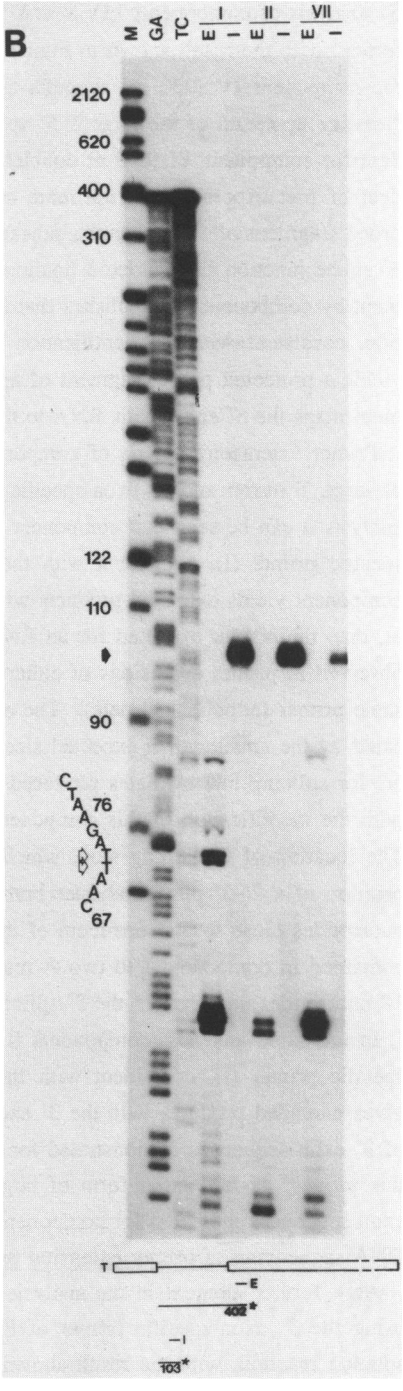
Primer extension analysis of components I, II and VII was performed using a combination of large T intron and 3' exon-specific primers, with the results shown in Figure 3B. In this analysis it can be seen that component I yields an extended product of 102 nt with the intron-specific primer (I), coincident with the 5' end of the large T intron. In contrast, this same component yields extended products with the 3' exon-specific primer far smaller, at 73 and 74 nt, than the 401 nt expected for an RNA with this 5' end. These extended products were not observed in primer extensions of either purified precursor or spliced large T mRNA using the same primer (unpublished data). The existence of a strong stop for reverse transcription indicated by the smaller than expected size of these extended products has been observed by others for splicing intermediates produced in vivo (8-11) and in vitro (6,11,38) and is consistent with the identification of this component as 3' exon linked to a lariat form of large T intron. The location of the strong stop, which in other cases has been shown to coincide with the position of a 2'-5' phosphodiester branch linkage joining the 5' end of the intron to specific nucleotides close to and upstream of the 3' splice site (6,38), maps the putative lariat branch contained in component I to two A residues of the sequence 5'-TTCTAAT-3' located 18 and 19 nucleotides upstream of the 3' splice junction.

In the same analysis, components II and VII both yield a 102 nt product with the intron-specific primer (I), coincident with the 5' end of the large T intron, while neither species gives extended products with the 3' exon specific primer (E). The location of 5' end and lack of 3' exon sequences demonstrated for component II in this analysis suggests identification of this product as free lariat form of large T intron, while the identical 5' end and lack of 3' exon sequences indicated for component VII suggest that this RNA is linear large T intron.

DNA sequencing of primer extension products

As a further element in the analysis of products of the splicing reaction, primer extension using the 3' exon-specific primer of Figure 3B was performed upon the total products of a splicing reaction, with the result shown in Figure 4A. Seven major extended products includ-


A
M
I
III
IV
V
VI
M
T
pre
cos
1060
929
622
527
T/t
T
t
290
1170
530
570
920
H
1170
H


B
M
GA
TC
I
E
I
II
E
I
VII
E
I
2120
620
400
310
122
110
90
76
67
T
E
402*
I
103*

ing that corresponding to precursor (662nt, III) were observed. These products are of approximate sizes 400 nt (IV), 310 nt (V), 130 nt (VIa), 86 nt (VIb), 80 nt (VI c), and 72 nt (I). Each of these species was eluted from the gel and sequenced using the Maxam-Gilbert technique. The sequences of products IV, VIa, and VIb,c (not shown) were found to correspond, as expected from their sizes, to RNAs cleaved upstream of large T and small t 5' splice sites and the 3' splice site, respectively. The exact positions of these cleavages were mapped, as shown in Figure 4C, to 8-10 nucleotides upstream of both 5' splice sites (IV and VIa), and 9-10 (VIc) and 15-16 (VIb) nucleotides upstream of the putative lariat branchpoint. Similar sequence analysis of species I (not shown) established that it represents a primer extension product terminating at the putative lariat branchpoint.

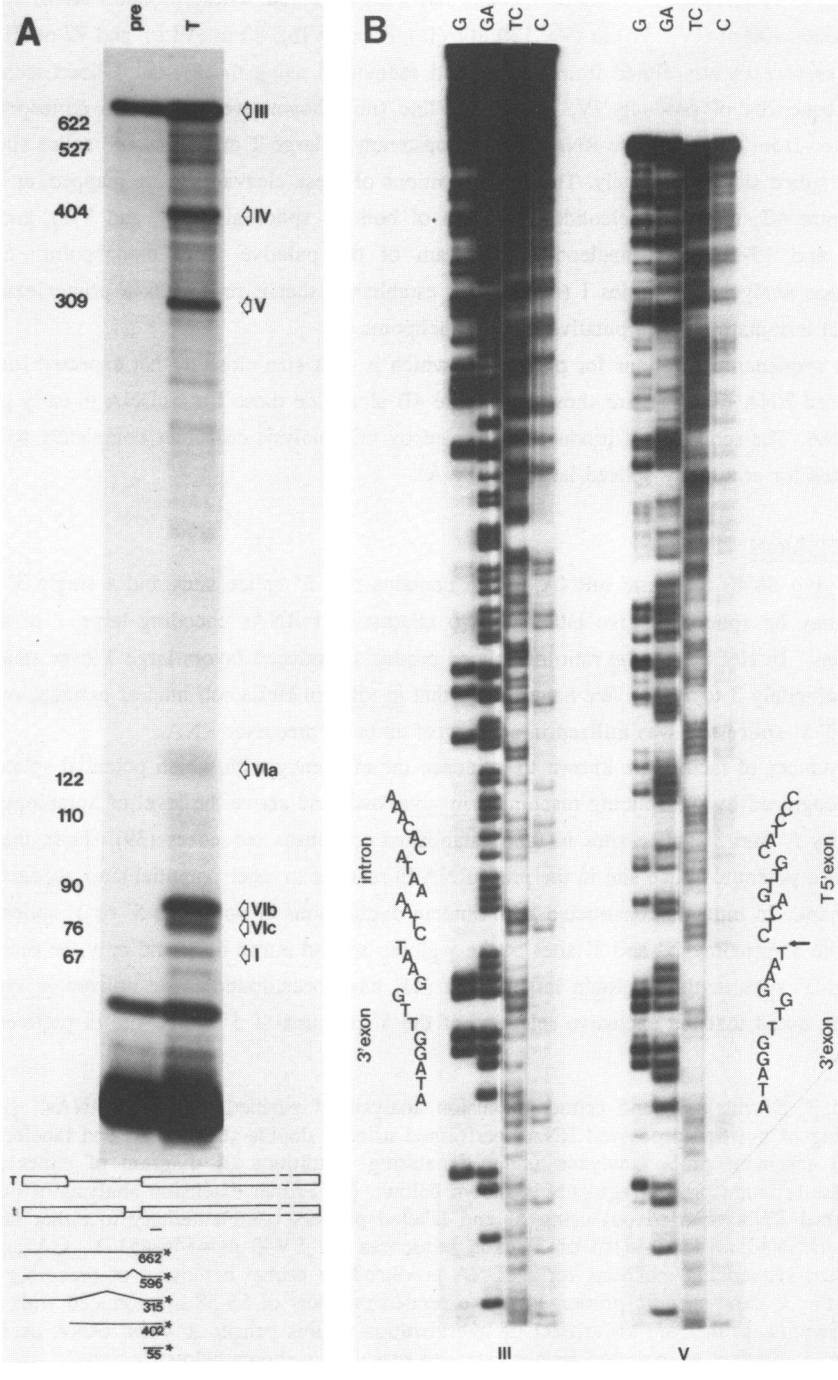
The sequencing reactions for product V, which is of a size close to that expected for large T spliced RNA (316 nt), are shown in Figure 4B alongside those for a cDNA to early precursor RNA. The sequence of product V revealed by this analysis coincides completely with that expected for accurately spliced large T mRNA.

DISCUSSION

In vivo SV40 early pre mRNA, which contains two 5' splice sites and a single 3' splice site, may be spliced to give either of two alternative mRNAs encoding large T or small t antigens. In HeLa cells the ratio of spliced products produced favors large T over small t by approximately 3 to 1 (37). We have shown that *in vitro* in HeLa cell nuclear extracts, only the large T 5' splice site was utilized in splicing of an early precursor RNA.

A number of factors are known to influence the efficiency with which potential splice sites are recognized by the splicing machinery *in vivo*, over and above the level of homology exhibited by 5' and 3' splice sites to their established consensus sequences (39). First, the position of a potential splice site in the pre mRNA in relation to other potential sites appears to be important. In mutants constructed with tandem duplications of wild-type 5' or 3' splice sites, only the 5' proximal 5' and 3' sites in the γ globin second intron (40), and only the outermost 5' and 3' sites in the β globin large intron (41) have been found to be utilized *in vivo*. It may be noted that the exclusive selection of the 5' proximal T 5' splice site in preference to

Figure 3. S1 nuclease and primer extension analysis of purified processed RNAs. (A) S1 mapping of purified processed RNAs performed using a double-stranded 5' end labeled HindIII B fragment probe analyzed under denaturing conditions. A diagram of expected S1 nuclease-resistant probe fragments is shown below. (B) Primer extension analysis of purified processed RNAs carried out using 5' end labeled primers complementary to either large T intron (I:SV40 nt 4843-4816) or 3' exon sequences (E: SV40 nt 4546-4517). GA and TC represent sequencing reactions for a cDNA produced by primer extension of precursor RNA using the 3' exon-specific primer. Primer extended products of 56-58 nt produced with the 3' exon specific primer are an artifact of hybridization of this primer to yeast tRNA used as a carrier. A diagram of expected primer extended products is shown below.


A
pre
T
622
527
404
309
◊III
◊IV
◊V
122
110
90
76
67
◊VIa
◊VIb
◊VIc
◊I
T
t
662*
596*
315*
402*
55*
B
G
GA
TC
C
intron
3'exon
5'exon
III
V

C

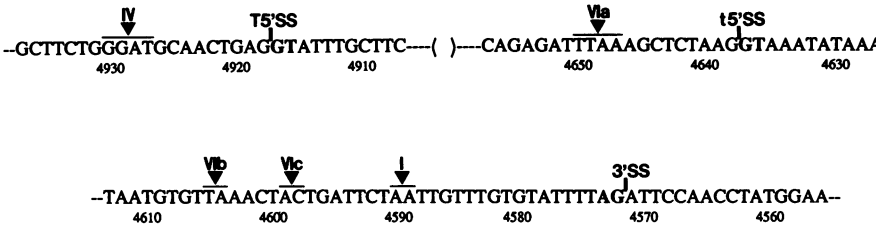


Figure 4. Primer extension and nucleotide sequence analysis of processed RNAs. (A) Large scale primer extension analysis carried out using the 5' end labeled 3' exon-specific primer described in Figure 2 part (B), fractionated on a buffer step gradient. Primer extended products are designated in accordance with the system describing RNA products in Figure 2 part (A). Pre represents precursor RNA. A diagram of expected primer extended products as shown below. (B) Maxam-Gilbert sequence analysis of primer extended products III and V corresponding to unspliced precursor and large T spliced RNAs respectively. An arrow indicates the 3' splice junction and point of divergence between the sequences of the two extended products. (C) Nucleotide sequences surrounding the large T and small t 5' splice sites and 3' splice site. Arrows indicate the positions of cleavage sites (IV, VIa,b, and c), and the lariat branchpoint (I) corresponding to primer extended products described in part (A).

the downstream small t 5' splice site in vitro agrees at least superficially with this pattern of utilisation. Second, in a number of cases, including early and late regions of SV40 (42,43), sequences distal to splice sites may play a role in determining splice site strength. For the early region, deletion of sequences toward the 5' end of the large T intron has been shown to reduce accumulation small t mRNA relative to large T mRNA in vivo (42). Third, the distance separating potential 5' and 3' splice sites appears to effect splice site utilisation. In mutants containing extensive deletions of intron sequence studied in vivo in the β globin large intron (44), and in vitro in the β globin first intron (38) and E1a 13S intron (45), the efficiency of splice site utilization has been found to be dramatically reduced for introns smaller than approximately 80 nucleotides. It may be noted that the small t intron at 66 nucleotides is considerably smaller than the suggested threshold length. One explanation for the discrepancy observed between in vivo and in vitro patterns of splicing of the SV40 early precursor RNA would be that differences may exist between the heirarchy of these influences as they affect splice site utilisation in vivo and in vitro. Previous observations of differences between in vivo and in vitro efficiences of splicing of mutant pre mRNAs would be consistent with such an explanation (4,38,14,46). We are currently attempting to test this notion by analyzing the effects upon small t 5' splice site utilization in vivo and in vitro of deletion of the large T 5' splice site, and expansion of the small t intron.

In addition to large T spliced mRNA, we have demonstrated the production in vitro of a

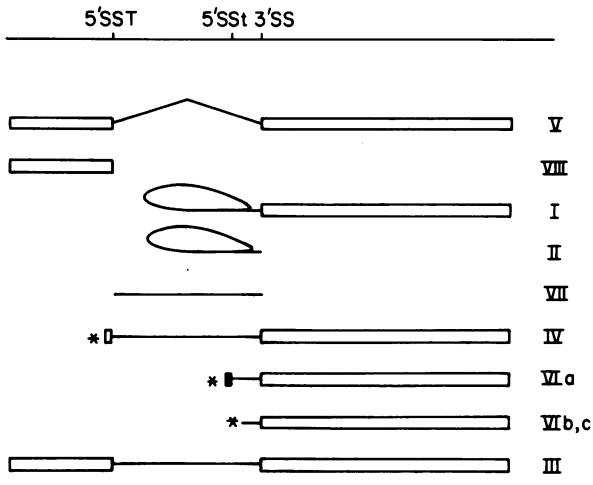


Figure 5. Schematic summary of RNAs produced in the processing of SV40 early precursor RNA. Each RNA is identified according to designation in Figures 2(A) and 4(A). Asterisks identify those products resulting from the activity of a 5' to 3' exonuclease.

number of potential intermediates and byproducts of the splicing of early precursor RNA to large T mature mRNA which bear on the mechanism of the large T splice (Figure 5). Two possible intermediates we have observed are the free 5' exon of large T and an RNA containing the 3' exon joined to what we believe to be a lariat form of the large T intron. Similar intermediates have been observed by others in the splicing of adenovirus leaders 1 and 2 and β globin first intron in vitro (5,6), and β globin second intron in vivo (11). Their production in the course of splicing of SV40 early precursor to large T mature mRNA indicates that the mechanism underlying the large T splice is fundamentally similar to that underlying other, non-alternative splices. We have mapped the putative branch formed as an intermediate in splicing of SV40 large T mRNA to the sequence TTCTAAT located close to and upstream of the 3' splice junction. This sequence conforms closely to the branchpoint consensus sequence (6,15), and lies within the expected distance of the 3' splice site. The precise position of the putative branchpoint, determined from the location of a primer extension block characteristic of such branched nucleotides, corresponds to the fifth and sixth residues of this consensus sequence, both As, located respectively 19 and 18 nucleotides upstream of the 3' splice junction. The ambiguity in mapping the position of this branch may be due to the technique employed or reflect a true flexibility in location of the branch noted in other instances in which the fifth nucleotide of the branchpoint consensus sequences is an A (14).

Among the products of splicing of SV40 early precursor RNA, we have observed two possible byproducts of the reaction, namely free lariat and linear forms of the large T intron. Free

lariat introns have been observed by others as byproducts of splicing in vitro (5,6,11) and in vivo (11). Free linear introns have also been observed in splicing of β globin first and second introns in vivo (11), although in vitro splicing products of the same and other pre mRNAs have been found to include only free introns with non-linear structures (6,11). Our characterization of the putative linear large T intron would appear to exclude the possibility that this RNA might have a non-linear structure. Instead its production may be explained by the presence in the nuclear extract we have used of an activity similar to that described in an S100 fraction by Ruskin et al. (38,47), which catalyzes debranching of free lariat large T intron.

In the course of our analysis of the products of splicing of early precursor, we have identified a number of species that appear to result from protection of precursor RNA in specific regions from the activity of a 5' to 3' exonuclease endogenous to the nuclear extract. These protection products were observed in the earliest stages of the splicing reaction well in advance of the appearance of splicing intermediates. Their relevance to the splicing process is suggested both by the locations of the protected regions described and by the observation that they are produced only in the same conditions as bona fide products of the splicing process. Our interpretation of their appearance is that they represent binding to the precursor RNA of components of the splicing machinery as a prelude to its participation in events of the splicing process.

Of the four major protection sites we have mapped in the SV40 early pre mRNA, two are located 8-10 nucleotides upstream of large T and small t 5' splice sites, respectively. It has been known for some time that the 5' end of U1 snRNA exhibits complementarity to the 5' splice site consensus sequence (16,17), and, more recently, partially purified U1 snRNP particles have been shown to bind in vitro specifically to a region of the mouse β globin first intron encompassing the 5' splice site (48). The region protected by U1 snRNP particles in this analysis was found to extend from 10-11 nucleotides upstream of the 5' splice junction to as much as 32 nucleotides downstream of it, and thus corresponds closely in its 5' boundary to the region of SV40 early precursor protection observed for both 5' splice sites in this in vitro system. The fact that we observe protection of the small t 5' splice site under conditions in which we fail to see cleavage at this site may imply that non-utilization of this splice junction in vitro is not a reflection of its inability to be recognized as such.

Two other protection sites we have mapped in early pre mRNA correspond to locations 27-28 and 33-34 nucleotides upstream of the 3' splice junction. Although complementarity between a region of U1 snRNA and the 3' splice site consensus has also been noted (16,17), attempts to demonstrate binding of purified U1 snRNP particles to 3' splice site sequences in vitro have been unsuccessful (48,49). However, recently, binding of U2 snRNP during the splicing reaction to a region encompassing the lariat branchpoint has been demonstrated (50). The region protected by U2 snRNP particles in this analysis was found to extend from 19-25 nucleotides upstream of the branchpoint to 9-16 nucleotides downstream of it, which

corresponds well to the 5' limit of sequences at the 3' splice site protected in the SV40 early precursor in this in vitro splicing system.

ACKNOWLEDGMENTS

We thank M. Chaudhuri for help in construction of pYSVHdB, L. Ryner for help in preparation of nuclear extracts, and R. Matthews for expert technical assistance. This work was supported by NIH grant CA 33620 and NSF grant PCM-82-16798.

REFERENCES

1. Kole, R. & Weissman, S. M. (1982) Nucleic Acids Res. 10, 5429-5445.
2. Padgett, R. A., Hardy, S. F. & Sharp, P. A. (1983) Proc. Natl. Acad. Sci. USA 80, 5230-5234.
3. Hernandez, N. & Keller, W. (1983) Cell 35, 89-99.
4. Krainer, A. R., Maniatis, T., Ruskin, B. & Green, M. R. (1984) Cell 36, 993-1005.
5. Grabowski, P. J., Padgett, R. A. & Sharp, P. A. (1984) Cell 37, 415-427.
6. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) Cell 38, 317-331.
7. Padgett, R. A., Konarska, M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. (1984) Science 225, 898-903.
8. Pikielny, C. W., Teem, J. L. & Rosbash, M. (1983) Cell 34, 395-403.
9. Rodriguez, J. R., Pikielny, C. W. & Rosbash, M. (1984) Cell 39, 603-610.
10. Domdey, H., Apostol, B., Lin, R-J., Newman, A., Brody, E. & Abelson, J. (1984) Cell 39, 611-621.
11. Zeitlin, S. & Efstratiadis, A. (1984) Cell 39, 589-602.
12. Keller, W. (1984) Cell 39, 423-425.
13. Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472.
14. Reed, R. & Maniatis, T. (1985) Cell 41, 95-105.
15. Keller, E. B. & Noon, W. A. (1984) Proc. Natl. Acad. Sci. USA 81, 7417-7420.
16. Rogers, J. & Wall, R. (1980) Proc. Natl. Acad. Sci. USA 77, 1877-1879.
17. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980) Nature 283, 220-224.
18. Ohshima, Y., Itoh, M., Okada, N. & Miyata, T. (1981) Proc. Natl. Acad. Sci. USA 78, 4471-4474.
19. Fradin, A., Jove, R., Hemenway, C., Keiser, H. D., Manley, J. L. & Prives, C. (1984) Cell 37, 927-936.
20. Padgett, R. A., Mount, S. M., Steitz, J. A. & Sharp, P. A. (1983) Cell 35, 101-107.
21. Kramer, A., Keller, W., Appel, B. & Luhrmann, R. (1984) Cell 38, 299-307.
22. Chow, L. T., Broker, T. R. & Lewis, J. B. (1979) J. Mol. Biol. 134, 265-303.
23. Berk, A. J. & Sharp, P. A. (1978) Proc. Natl. Acad. Sci. USA 75, 1274-1278.
24. Reddy, V. B., Ghosh, P. K., Lebowitz, P. & Weissman, S. M. (1978) Nucleic Acids Res. 5, 4195-4213.
25. Ghosh, P. K., Reddy, V. B., Swinscoe, J., Lebowitz, P. & Weissman, S. M. (1978) J. Mol. Biol. 126, 813-846.
26. Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. & Ogata, K. (1984) Nature 308, 333-338.
27. King, C. R. & Piatigorsky, J. (1983) Cell 32, 707-712.
28. Breitbart,R. E., Nguyen, H. T., Medford, R. M., Destree, A. T., Mahdavi, V. & Nadal-Ginard, B. (1985) Cell 41, 67-82.

29. Sharp, P. A. (1981) Cell 23, 643-646.
30. Rossi, J. J., Soberon, X., Marumoto, Y, McMahon, J. & Itakura, K. (1983) Proc. Natl. Acad. Sci. USA 80, 3203-3207.
31. Maniatis, T., Fritsch, E. F. & Sambrook, J., eds. (1982) Molecular Cloning-A Laboratory Manual. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
32. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. (1983) Nucleic Acids Res. 11, 1475 1489.
33. Manley, J. L. & Gefter, M. L. (1981) Gene Amplification and Analysis, Vol. 2, Chirikjian, J. G. & Papas, T. S., eds. (Elsevier/North Holland ,NY), pp. 369-382.
34. Wickens, M. P. & Gurdon J. B. (1983) J. Mol. Biol. 163, 1-26.
35. Maxam, A. M. & Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.
36. Treisman, R., Proudfoot, N. J., Shander, M. & Maniatis, T. (1982) Cell 29, 903-911.
37. Manley, J. L., Noble, J. C. S., Chaudhuri, M., Fu, X.-Y., Michaeli, T., Ryner, L. & Prives, C. (1985) Cancer Cells, Vol. 4 (in press).
38. Ruskin, B., Greene, J. M. & Green, M. R. (1985) Cell 41, 833-844.
39. Montell, C. & Berk, A. J. (1984) Nucleic Acids Res. 12, 3821-3827.
40. Lang, K. M. & Spritz, R. A. (1983) Science 220, 1351-1355.
41. Kuhne, T., Weiringa, B., Reiser, J. & Weissmann, C. (1983) EMBO J. 2, 727-733.
42. Khoury, G., Gruss, P., Dhar, R. & Lai, C-J. (1979) Cell 18, 85-92.
43. Piatak, M., Subramanian, K. N., Roy, P. & Weissman, S. M. (1981) J. Mol. Biol. 153, 589-618.
44. Weiringa, B., Hofer, E. & Weissmann, C. (1984) Cell 37, 915-925.
45. Ulfendahl, P. J., Pettersson, U. & Akusjarvi, G. (1985) Nucleic Acids Res. 13, 6299-6315.
46. Orkin, S. H., Kazazian Jr, H. H., Antonarakis, S. E., Ostrer, H., Goff, S. C. & Sexton, J. P. (1982) Nature 300, 768-769.
47. Ruskin, B. & Green, M. R. (1985) Science 229, 135-140.
48. Mount, S. M., Pettersson, I., Hinterberger, M. Karmas, A. & Steitz, J. A. (1983) Cell 33, 509-518.
49. Tatei, K., Takemura, K., Mayeda, A., Fujiwara, Y., Tanaka, H., Ishihama, A. & Ohshima, Y. (1984) Proc. Natl. Acad. Sci. USA 81, 6281-6285.
50. Black, D. L., Chabot, B. & Steitz, J. A. (1985) Cell 42, 737-750.