MOLECULAR AND CELLULAR BIOLOGY, Apr. 1987, p. 1518–1529
0270-7306/87/041518-12$02.00/0

Vol. 7, No. 4

Products of In Vitro Cleavage and Polyadenylation of Simian Virus 40 Late Pre-mRNAs

MICHAEL D. SHEETS, PAULINE STEPHENSON, AND MARVIN P. WICKENS*

Department of Biochemistry, College of Agriculture and Life Sciences, University of Wisconsin–Madison, Madison, Wisconsin 53706

Received 15 October 1986/Accepted 29 December 1986

Formation of mRNA 3′ termini involves cleavage of an mRNA precursor and polyadenylation of the newly formed end. Cleavage of simian virus 40 late pre-mRNA in a crude nuclear extract generated two RNAs, 5′ and 3′ half-molecules. These RNAs were unmodified and linear. The 5′ half-molecule contained sequences upstream but not downstream of the poly(A) site and ended in a 3′-terminal hydroxyl. The 3′ half-molecules comprised a family of RNAs, each of which contains only sequences downstream of the poly(A) site, and ends in a 5′-terminal phosphate. These RNAs differed only in the locations of their 5′ terminus. The 3′ terminus of the 5′ half-molecule was the adenosine 10 nucleotides downstream of AAUAAA, at the +1 position. The 5′ terminus of the longest 3′ half-molecule was at +2. Thus, these two RNAs contain every nucleoside and phosphate of the precursor. The existence of these half-molecules demonstrates that endonucleolytic cleavage occurs near the poly(A) site. 5′ half-molecules generated in the presence of EDTA (which blocks polyadenylation, but not cleavage) ended at the adenosine at position +1 of the precursor. When incubated in the extract under suitable conditions, they became polyadenylated. 5′ half-molecules formed in 3′-dATP-containing reactions contained a single 3′-deoxyadenosine (cordycepin) residue added onto the +1 adenosine and were poor polyadenylation substrates. We infer that the +1 adenosine of the precursor becomes the first A of the poly(A) tract and provides a 3′ hydroxyl group to which poly(A) is added posttranscriptionally.

The 3′ terminus of most eucaryotic mRNAs is generated by two sequential reactions, cleavage and polyadenylation. Primary transcripts extend beyond the polyadenylation site by as much as 20 kilobases (40, 55). Endonucleolytic cleavage of these long mRNA precursors generates new 3′ termini to which approximately 250 adenosine residues [poly(A)] are added (10, 35, 40, 55). The mechanisms of cleavage and polyadenylation are of interest because both steps are required for the maturation of nearly all mRNAs and because both steps can be regulated (1, 2, 5, 8, 10, 45).

Sequence information sufficient for both cleavage and polyadenylation is confined to the immediate vicinity of the polyadenylation site (37, 38, 51, 53). At least two sequence elements are essential: the AAUAAA sequence (15, 21, 36, 43, 51) and a sequence downstream of the poly(A) site (9, 17, 32, 33, 47, 48, 52).

By analyzing mRNA processing in vivo, Nevins and Darnell (40) established that formation of mRNA 3′ termini requires at least one endonucleolytic scission. Two important questions could not be resolved owing to the inherent limitations of in vivo analysis. First, the site at which endonucleolysis occurs could not be determined, since the portion of the precursor which lies downstream of the poly(A) site is degraded very rapidly. Thus, cleavage might occur either at the poly(A) site or downstream (in which case an exonuclease would also be required). Second, the nucleotide to which poly(A) is added could not be identified. In most genes (4, 33), the poly(A) site coincides with an NA dinucleotide in the precursor; as a result, inspection of the sequence cannot resolve whether the first A of the poly(A) tail is derived from the primary transcript or is added posttranscriptionally. In vivo cleaved RNAs that are not yet polyadenylated cannot be detected, and so data that bear on the question cannot be obtained.

* Corresponding author.

The limitations of in vivo studies can be circumvented in vitro by examining cleavage and polyadenylation of synthetic pre-mRNAs (34) in a crude extract of HeLa cell nuclei (12, 38). Under suitable conditions, degradation of the RNA fragment downstream of the poly(A) site can be prevented, and polyadenylation can be inhibited without affecting cleavage (38, 53). Consequently, the two half-molecules that result from cleavage of the precursor can be detected and analyzed (16, 39, 53).

In this report, we examine cleavage and polyadenylation of simian virus 40 (SV40) late pre-mRNAs in vitro. We characterized the RNA products of the cleavage reaction to identify the nucleotide to which poly(A) is added and the phosphodiester bond that is hydrolyzed during mRNA 3′-end formation. In addition, we determined whether cleavage, like mRNA and tRNA splicing (18, 23, 25, 26, 42, 46), is accompanied by covalent modification of the pre-mRNA.

MATERIALS AND METHODS

Construction of pSPSV−141/+31. pSV−141/+31 contains a 172-base-pair fragment of SV40 spanning the polyadenylation site of virion protein (late) mRNAs (nucleotides [nt] 2533 to 2720 of the SV40 genome; numbering as in Tooze [50]). It was constructed by digesting *Hin*dIII-cleaved pSV−141/+70 (51) with *Bal* 31 exonuclease. *Xba*I linkers (CTCTAGAG) were added to the +31 position of SV40, and the DNA was resealed with T4 DNA ligase. The *Bam*HI-*Pst*I fragment of this clone containing the −141/+31 (*Bam*HI-*Xba*I) region of SV40 was then cloned into *Bam*HI-*Pst*I-cut pSP65 (34), yielding pSPSV−141/+31.

Preparation of −141/+31 RNA. pSPSV−141/+31 was cleaved with *Xba*I at the +31 position of SV40 and was then transcribed in vitro as described previously (34) except that the reactions contained 0.1 mM GTP, 1.2 mM GpppG (27; P-L Biochemicals, Inc., Milwaukee, Wis.), and 50 to 500 μCi of an [α-^{32}P]nucleoside triphosphate (Amersham Corp.,

Arlington Heights, Ill.). Full-length RNAs were purified by elution from urea-containing acrylamide gels (31).

Preparation of nuclear extract and cleavage in vitro. Nuclear extract was prepared by the method of Dignam et al. (12), except that $MgCl_2$ was omitted from buffers A and D. RNA (5 ng or less) was incubated at 30°C in a 25-μl reaction mixture containing 11 μl of extract (equivalent to approximately 10^6 cells) and additional components. The final concentrations of the various components (including the contributions of the extract) were as follows: 84 mM KCl, 8.8 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) (pH 7.6), 0.44 mM EDTA, 2.2 mM dithiothreitol, 8.8% glycerol, 20 mM phosphocreatine, 10 μM ATP, and 2.8% polyvinyl alcohol. Either 2 mM EDTA or 1 mM 3'-dATP (P-L Biochemicals) was included to inhibit polyadenylation (38, 53). After the incubation, RNA was purified as described previously (28). Reactions were incubated for 1 to 2 h; shorter incubation times (as little as 30 s) did not alter the pattern of RNAs observed, but reduced their yield.

RNA structural analyses. (i) RNase T_1. RNAs were mixed with 20 μg of tRNA and digested with 10 U of RNase T_1 (Calbiochem-Behring, La Jolla, Calif.) for 30 min at 37°C in 10 mM Tris hydrochloride (pH 8.0) in a 3-μl reaction mixture. The resulting oligonucleotides were separated in two dimensions (3, 6): first by high-voltage electrophoresis on cellulose-acetate at pH 3.5 and then by homochromatography on Cel 300 polyethyleneimine plates (Brinkmann Instruments, Inc., Westbury, N.Y.).

(ii) RNase T_2. RNAs were mixed with 20 μg of tRNA and digested with 10 U of RNase T_2 (Calbiochem) for 8 to 12 h at 37°C in 10 mM Tris hydrochloride (pH 8.0) in a 3-μl reaction mixure. The products were analyzed by two-dimensional thin-layer chromatography with a mixture of isobutyric acid-concentrated NH_4OH-H_2O (45:30:25, vol/vol/vol) in the first dimension and isopropanol-concentrated HCl-H_2O (75:15:15, vol/vol/vol) in the second (41). For the chromatography, unlabeled nucleoside diphosphates (pNp) were mixed with samples as markers and visualized with short-wave UV light.

(iii) Alkaline phosphatase. RNA was treated with 5 to 10 U of calf intestinal alkaline phosphatase (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) for 30 min at 37°C in 10 mM Tris hydrochloride (pH 9.0) in a 3-μl reaction mixture. The enzyme was then removed by repeated phenol-chloroform extractions in the presence of 10 mM EDTA.

(iv) Oligonucleotide-RNase T_1 assays. RNA was annealed to a 10- to 50-fold molar excess of a DNA oligonucleotide complementary to positions −22 through −8 of −141/+31 RNA (DNA sequence, 3'-GACGTTATTTGTTCAA-5'). Hybridization reaction mixtures (3-μl total volume) contained 0.5 M KCl and were incubated at 42°C for 20 min. Then, 20 μg of yeast RNA and 5 U of RNase T_1 were added, and the samples were incubated at 20°C for another 30 min. Digestion products were purified by phenol-$CHCl_3$ extraction and ethanol precipitation. The fraction of RNA that is protected from T_1 digestion varies between experiments from 10% (see Fig. 5) to 80% (data not shown).

Preparation of RNA markers. The markers used in Fig. 6 were prepared as follows. For −22/−1 (CUGCAA UAAACAAGUUAACAAG_{OH}), DNA oligonucleotide-T_1 protection (see above) was performed with a mutant −141/+31 RNA in which the −1 position was a G, not a C, thereby generating a −22/−1 RNA ending with a 3'-terminal Gp. The 3'-terminal phosphate was removed with phosphatase, and the oligonucleotide was purified by gel electrophoresis.

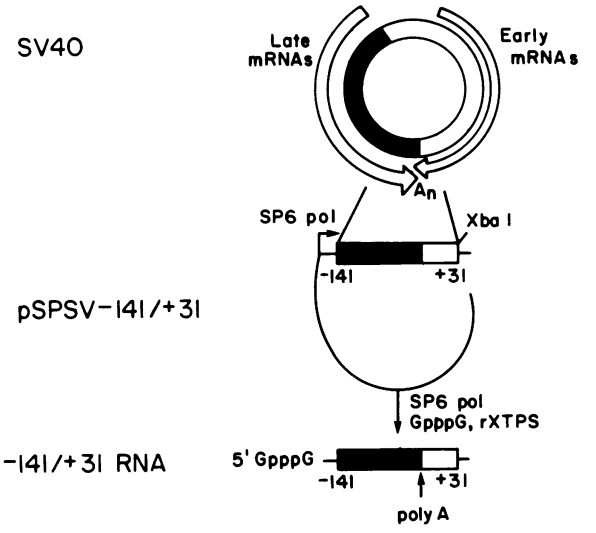


FIG. 1. Structure of the RNA substrates. The region of SV40 from 141 nt before the late polyadenylation site (the natural *Bam*HI site) to 31 nt beyond (a synthetic *Xba*I site) was cloned into pSP65 (34), generating pSPSV−141/+31. Transcription of *Xba*I-cleaved pSOSV−141/31 in the presence of GpppG (28, 34) and ribonucleoside triphosphates (rXTPs) produces −141/+31 RNA. The sequence of the transcript is given in Fig. 3.

−22/+1 (CUGCAAUAAACAAGUUAACAAGA$_H$) was prepared by incubating −22/−1 RNA with *Escherichia coli* poly(A) polymerase in the presence of 3'-dATP. A total of 10% of the RNA receives a single cordycepin residue, becoming −22/+1 RNA; the remainder does not react. Thus, this marker is a mixture of −22/−1 and −22/+1 oligonucleotides.

−22/+3 (CUGCAAUAAACAAGUUAACAACAAG$_{OH}$) was prepared exactly like −22/−1, but with a mutation in which the +3 position was a G, not a C.

An alkaline hydrolysate of uniformly labeled −141/+31 RNA was prepared by the method of Donis-Keller et al. (13).

Polyacrylamide gel electrophoresis. Electrophoresis through 6 or 15% polyacrylamide gels containing 7 M urea was performed as described previously (49). Lengths of RNAs were estimated by comparison with known T_1 oligonucleotides and with a partial alkaline hydrolysate of labeled RNA. For quantitation, films were scanned with a laser microdensitometer; several exposures of the same gel were used.

RESULTS

RNA substrates. SV40, a tumor virus with a circular double-stranded DNA genome, directs the synthesis of virion protein (late) mRNAs late in infection of monkey cells (Fig. 1). We constructed a plasmid, pSPSV−141/+31, in which a bacteriophage (SP6) promoter lies upstream of a 172-base-pair SV40 fragment containing the polyadenylation site of these mRNAs. Transcription of *Xba*I-cleaved pSPSV−141/+31 by SP6 polymerase generated a 207-nt transcript. This RNA contained 170 nt before the poly(A) site (141 nt of SV40 preceded by 29 nt of vector) and 37 nt after (31 nt of SV40 followed by 6 nt of vector). The transcript, called −141/+31 RNA to indicate the SV40 sequence it contains, possesses a 5'-terminal cap (27) and includes both the AAUAAA sequence and the downstream element (9, 51).

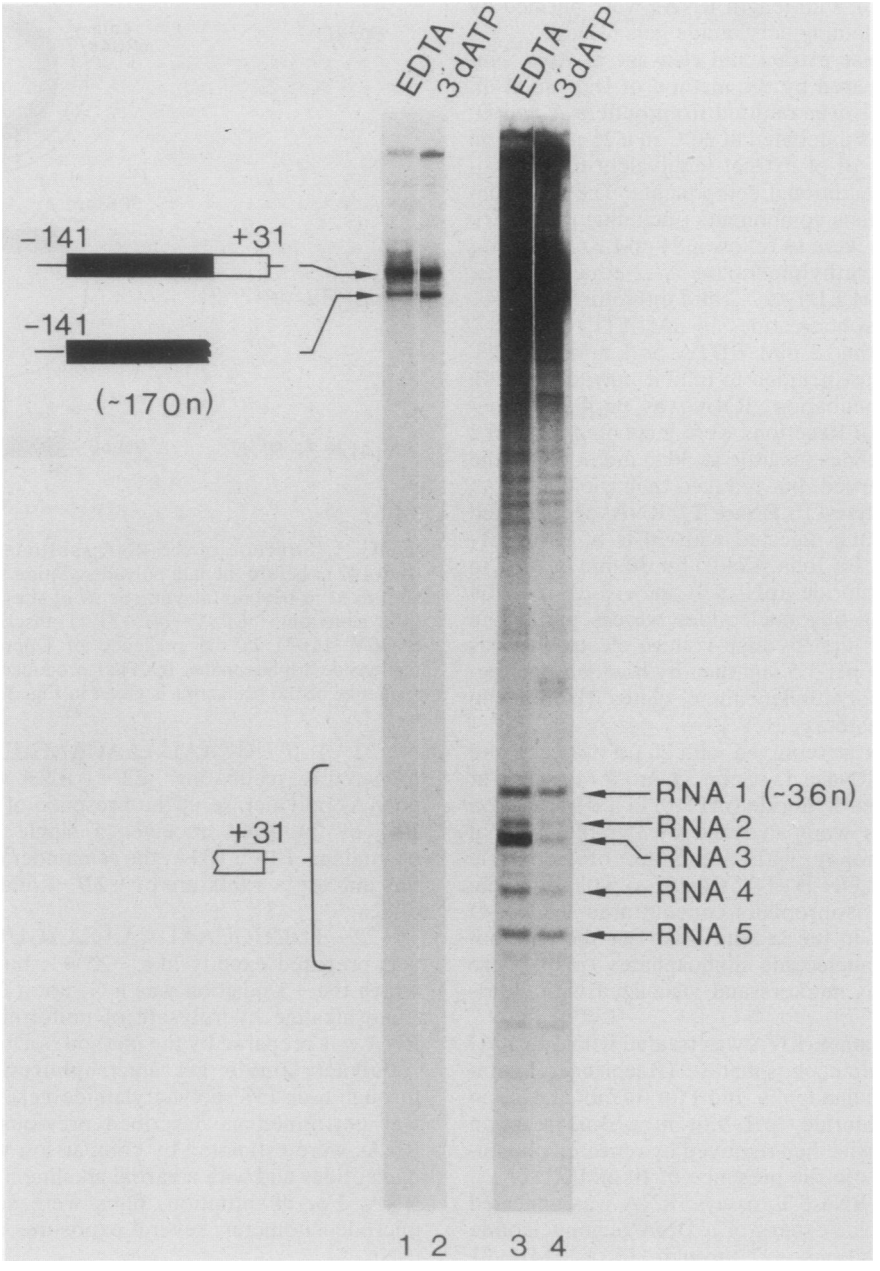


FIG. 2. Two types of cleavage products. Precursor containing 170 nt (n) 5′ of the polyadenylation site (141 nt of SV40) and 36 nt 3′ to the polyadenylation site (31 nt of SV40) was incubated in nuclear extract reaction mixtures containing either EDTA (lane 1) or 3′-dATP (lane 2). The products were analyzed by electrophoresis through a 15% acrylamide gel (49). Lanes 3 and 4 are a longer autoradiographic exposure of the same gel.

Two types of cleavage products observed. Uniformly labeled −141/+31 RNA was incubated in a crude extract of HeLa cell nuclei (12, 38). Polyadenylation was inhibited with either EDTA or 3′-dATP (cordycepin triphosphate) (38, 53). After 1 h, RNA was recovered and analyzed by polyacrylamide gel electrophoresis.

With either inhibitor, two types of RNA products were generated (Fig. 2). The first is a 170-nt RNA, evident even in a brief autoradiographic exposure (Fig. 2, lanes 1 and 2). The second, detected in a longer exposure, is a group of five short RNAs (RNAs 1 through 5), the longest of which contains approximately 36 nt (Fig. 2, lanes 3 and 4). The molar yield of RNAs 1 to 5 relative to the 170-nt RNA varies between experiments from 50 to 75%. The 170-nt RNA and RNAs 1 to 5 are generated by cleavage, not by nonspecific nucleases, since their appearance is prevented by mutations in AAUAAA or in the downstream element (D. Zarkower, data not shown).

Since the precursor contains 170 nt before the poly(A) site and 36 nt beyond, these results suggest that the two types of RNA are 5′ and 3′ half-molecules generated by cleavage near the poly(A) site (16, 38, 39, 53). In the remainder of the results, we first focus on characterization of the 170-nt RNA and then on RNAs 1 through 5.

The 170-nt RNA is a 5′ half-molecule. The 170-nt RNA purified from reactions containing either EDTA (Fig. 3C) or

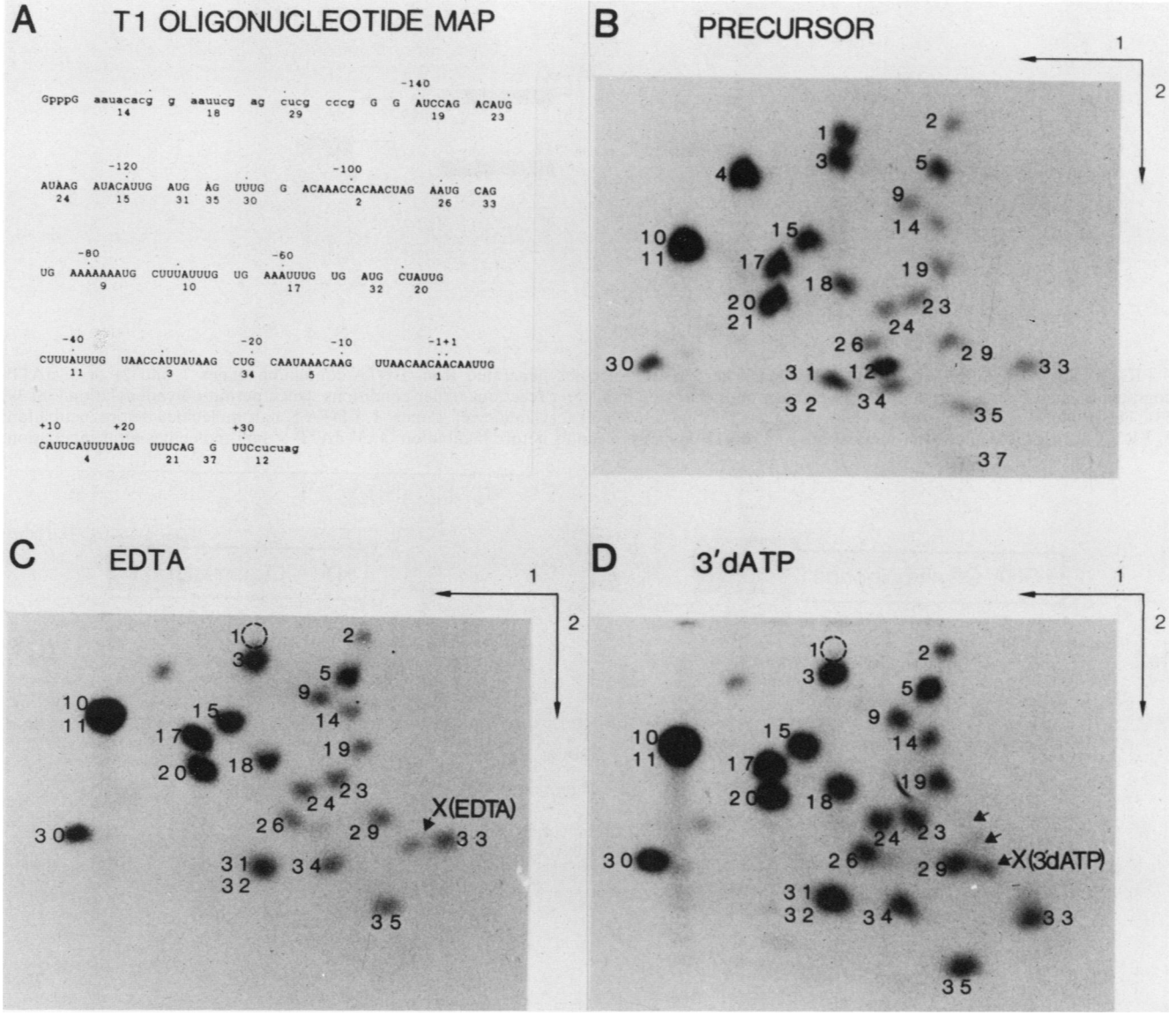


FIG. 3. The 170-nt RNA is a 5′ half-molecule. [α-^{32}P]UTP-labeled −141/+31 RNA was incubated in nuclear extract containing either EDTA or 3′-dATP. The resulting 170-nt RNAs were purified by gel electrophoresis. Each RNA was digested with RNase T_1, and the products were fractionated by two-dimensional chromatography (see Materials and Methods). (A) Sequence of −141/+31 RNA. The sequences of the precursor and the predicted T_1 oligonucleotides are provided, including numbering of the oligonucleotides (below the sequence) and positions relative to the polyadenylation site (above the sequence). (B) Oligonucleotides generated from −141/+31 precursor RNA. (C) Oligonucleotides generated from the 170-nt RNA formed in EDTA-containing reactions. The new oligonucleotide, oligonucleotide X(EDTA), is indicated by an arrow. (D) Oligonucleotides generated from the 170-nt RNA formed in 3′-dATP-containing reactions. The prominent new oligonucleotide, designated oligonucleotide X(3′-dATP), is indicated by an arrow. Minor, less abundant, new oligonucleotides are also observed, and are indicated by arrows above oligonucleotide X(3′-dATP). The dotted circle in C and D indicates the position at which oligonucleotide 1 would appear were it present.

3′-dATP (Fig. 3D) was digested with T_1 RNase. The resulting T_1 oligonucleotides were separated by electrophoresis and chromatography (3, 6). To provide a standard, −141/+31 RNA was analyzed in parallel (Fig. 3B).

Both 170-nt RNAs contained all oligonucleotides upstream of the polyadenylation site, but none downstream (Fig. 3). They lacked the oligonucleotide which contains the polyadenylation site (oligonucleotide 1) but instead contained a new oligonucleotide not found in the precursor [oligonucleotide X(EDTA) or oligonucleotide X(3′-dATP)]. As we shall demonstrate below, the new oligonucleotides include the 3′ end of each RNA. From these data we conclude that the 170-nt RNAs are 5′ half-molecules.

The 3′ end of half-molecules cleaved in the presence of EDTA is the polyadenylation site. Oligonucleotide X (see above) is one residue longer in 5′ half-molecules prepared from reactions containing 3′-dATP rather than EDTA [compare X(3′-dATP) (Fig. 3C) with X(EDTA) (Fig. 3D)]. Oligonucleotides two and three residues longer were also detected (Fig. 3D, arrows), although in much smaller quantities. These differences in oligonucleotide X are the only differences between the two 170-nt RNAs and are confirmed by

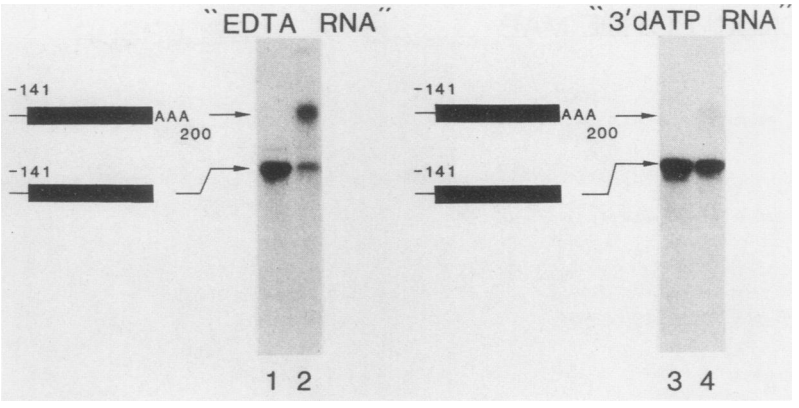


FIG. 4. Polyadenylation of 5′ half-molecules. The 5′ half-molecules generated from EDTA-containing (lanes 1 and 2) or 3′-dATP-containing (lanes 3 and 4) reactions were isolated and incubated in nuclear reactions under conditions which permit polyadenylation (38, 39, 53). The products were analyzed by electrophoresis through a 15% polyacrylamide gel. Lanes: 1, EDTA 5′ half-molecules before incubation; 2, EDTA 5′ half-molecules after incubation; 3, 3′-dATP 5′ half-molecules before incubation; 4, 3′-dATP 5′ half-molecules after incubation.

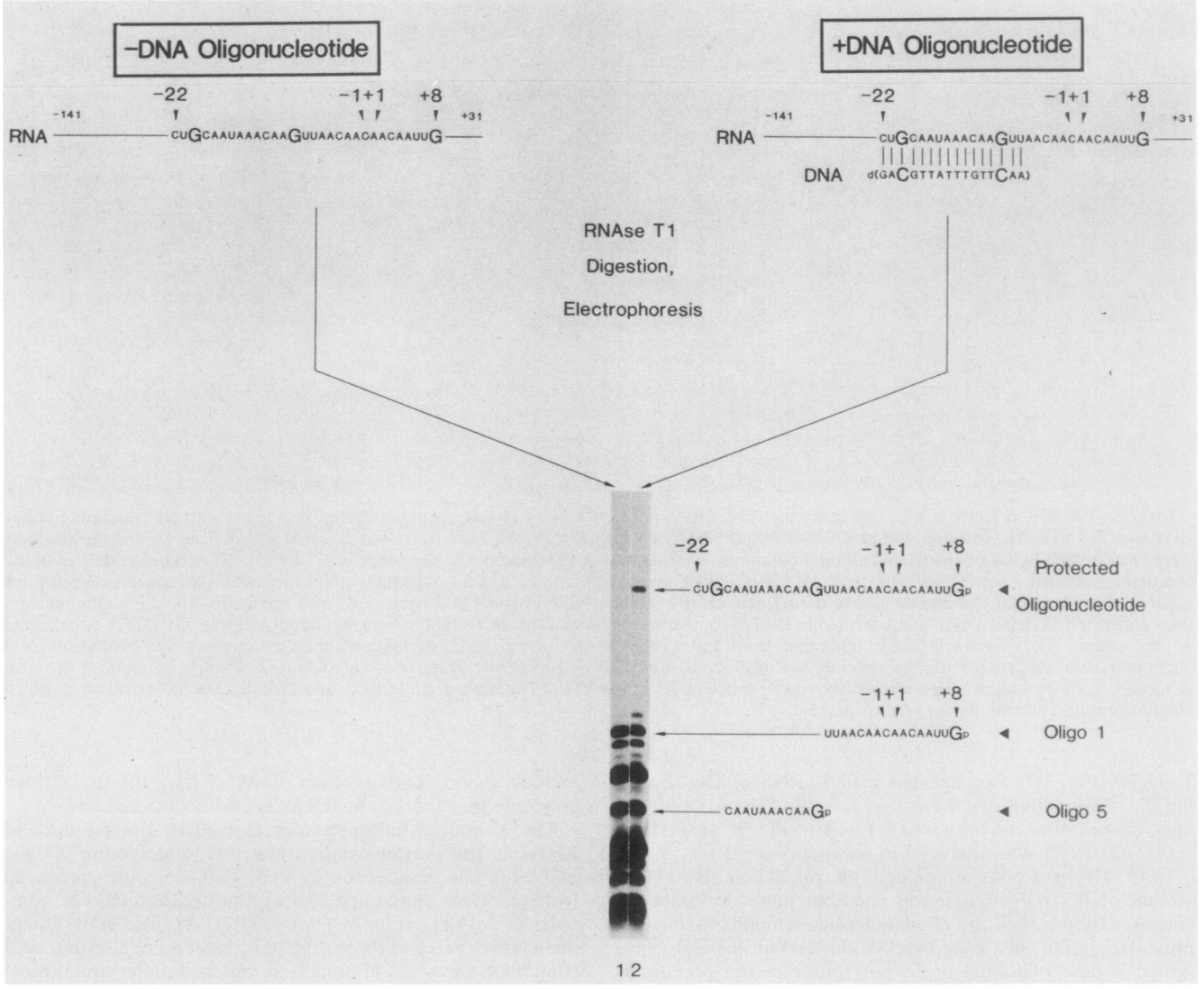


FIG. 5. Oligonucleotide protection assay. For details of the method see text. The products of RNase T_1 digestion of −141/+31 RNA were analyzed on a 15% acrylamide–7 M urea gel. Lanes: 1, RNase T_1 digestion in the absence the DNA oligonucleotide; 2, RNase T_1 in the presence of the DNA oligonucleotide. The −22/+8 protected oligonucleotide contains the poly(A) site.

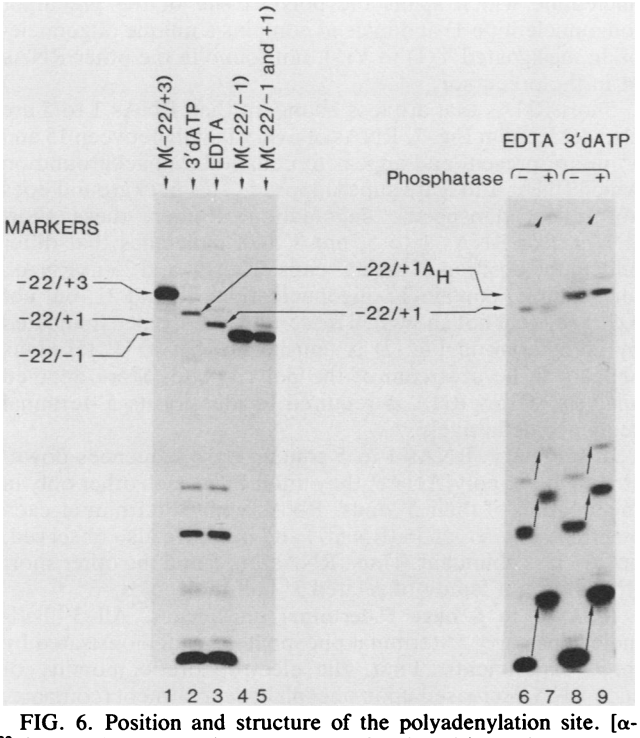


FIG. 6. Position and structure of the polyadenylation site. [α-^{32}P]UTP-labeled −141/+31 RNA was incubated in nuclear extract reaction mixtures containing either EDTA or 3'-dATP. The resulting 5' half-molecules were purified by gel electrophoresis. The 5' half-molecules were analyzed by the oligonucleotide protection assay followed by electrophoresis through a 15% acrylamide–7 M urea sequencing gel. Lanes 1, 4, and 5 contain marker (M) oligonucleotides. Lane 1, −22/+3 (CUGCAAUAAACAAGUUAACA ACAAG$_{OH}$); lane 4, −22/−1 (CUGCAAUAAACAAGUUAACA AG$_{OH}$); lane 5, −22/−1 and −22/+1 (CUGCAAUAAACAAGU UAACAAGA$_H$); lane 2, oligonucleotide-T_1 protection assay of the 170-nt RNA purified from a 3'-dATP-containing reaction; lane 3, oligonucleotide-T_1 protection assay of the 170-nt RNA purified from an EDTA-containing reaction. The 3'-terminal oligonucleotides (−22/+1:A$_H$ and −22/+1) are indicated with arrows, as is each marker. Lanes 6 to 9 demonstrate that 170-nt RNAs lack a 3' phosphate. Purified 170-nt RNAs were digested with RNase T_1 in the presence of DNA oligonucleotide (as in lanes 2 and 3). After T_1 digestion, half of the sample was treated with phosphatase (+ lanes); the other half was not (− lanes). Lane 6, 170-nt RNA purified from an EDTA-containing reaction, no phosphatase; lane 7, 170-nt RNA purified from an EDTA-containing reaction, phosphatase treated; lane 8, 170-nt RNA purified from a 3'-dATP-containing reaction, no phosphatase; lane 9, 170-nt RNA purified from a 3'-dATP-containing reaction, phosphatase treated. The arrows indicate the altered mobility of internal T_1 oligonucleotides upon phosphatase treatment. The 3'-terminal oligonucleotides (−22/+1:A$_H$ and −22/+1) do not shift.

gel electrophoresis of the oligonucleotides (data not shown). We infer that in the presence of 3'-dATP, a cordycepin molecule is added to the polyadenylation site, usually immediately after cleavage, but sometimes after one or two adenosines have already been polymerized.

To test this inference, we compared 170-nt RNAs generated in EDTA and 3'-dATP as polyadenylation substrates; a teminal cordycepin (3'-deoxyadenosine) should block polyadenylation since it lacks a 3' hydroxyl group. Purified 170-nt RNAs were incubated in nuclear extract under conditions which permit polyadenylation (38, 39, 51) (Fig. 4). The EDTA RNA was efficiently polyadenylated, as judged by its increase in length (lanes 1 and 2) and its retention by oligo(dT)-cellulose (data not shown). In contrast, the 3'-dATP RNA was a very poor substrate for polyadenylation (lanes 3 and 4). Since the two RNAs differ only at their 3' termini (Fig. 3), we conclude that in reactions containing 3'-dATP, a single cordycepin residue is added which then blocks elongation of the poly(A) tract. The residue to which that cordycepin is added, the 3' end of the EDTA RNA, is the polyadenylation site.

The polyadenylation site is the adenosine at +1. To identify the polyadenylation site precisely, we determined the length of a 3'-terminal oligonucleotide by gel electrophoresis. Oligonucleotide X is sufficiently short that its length cannot be determined unambiguously by comparison with markers owing to differences in base composition. To circumvent this problem, longer 3'-terminal oligonucleotides were generated by a DNA oligonucleotide-RNase T_1 protection assay. The method is outlined in Fig. 5.

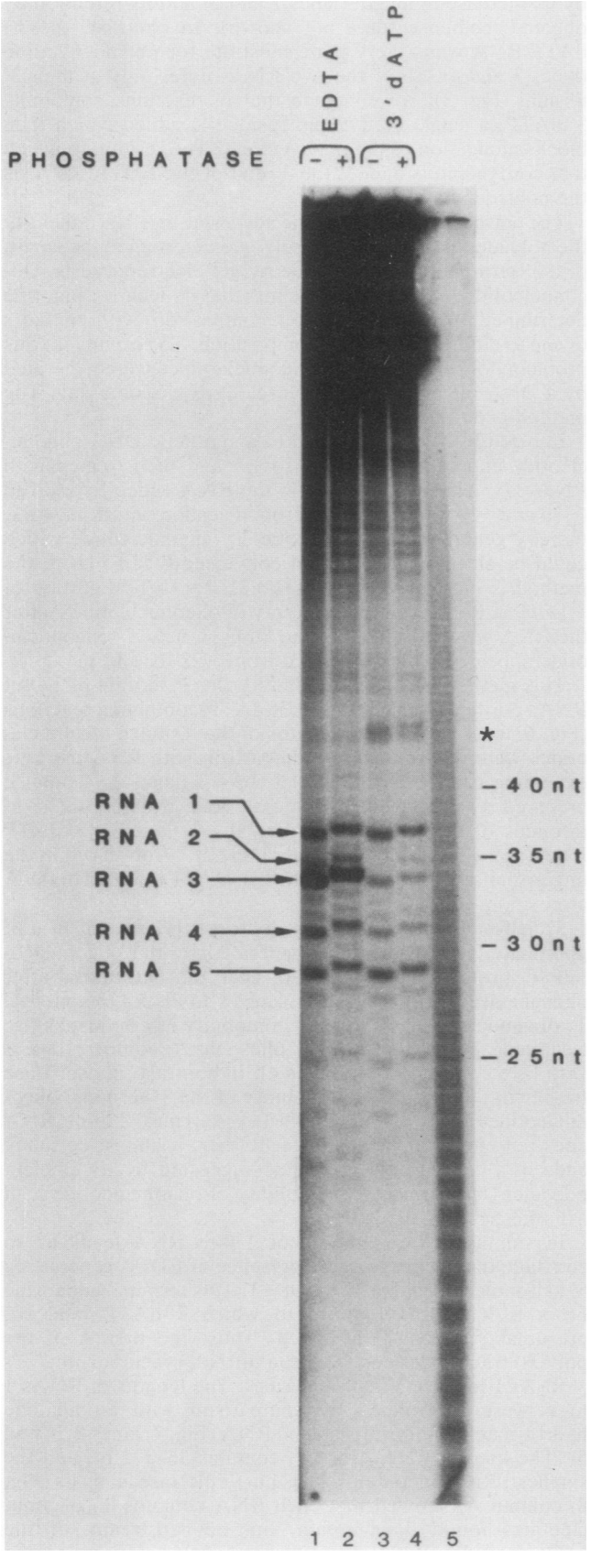
Labeled RNA was annealed to a synthetic DNA oligonucleotide extending from −22 to −7 and then treated with RNase T_1. Those G residues in the RNA which hybridized (−20 and −9) were protected from digestion with RNase T_1, thereby generating a novel, large T_1 oligonucleotide which could be sized accurately and conveniently. In Fig. 5, the method is illustrated with −141/+31 RNA. T_1 digestion of −141/+31 RNA produced an array of oligonucleotides; after the RNA was hybridized to the DNA, a new T_1 oligonucleotide appeared which extended from −22 to +8.

This method was used to identify the 3' termini of 170-nt RNAs purified from EDTA- or 3'-dATP-containing reactions (Fig. 6, lanes 1 to 5). The lengths of the terminal oligonucleotides were determined by comparison with RNA markers extending from −22 to −1, +1, or +3 (lanes 1, 4, and 5; prepared as described in Materials and Methods). The 3' terminus of the EDTA RNA lies at +1 (lane 3). 3'-dATP RNAs were 1, 2, and 3 nt longer (lane 2), as observed in the fingerprint analyses (Fig. 3). Thus, 3'-dAPT is added to the A at position +1.

5' half-molecules end in a 3' hydroxyl group, not in a 3' phosphate. Oligonucleotides generated by the oligonucleotide T_1 protection method were analyzed before and after phosphatase treatment (Fig. 6, lanes 6 to 9). All the internal T_1 oligonucleotides decreased in mobility after phosphatase treatment, since they bear 3' phosphates (compare lane 1 with lane 2, or lane 3 with lane 4). In contrast, phosphatase treatment did not alter the mobility of the 3'-terminal oligonucleotides (−22/+1 or −22/+1:A$_H$). Thus, 170-nt RNA does not end in a 3' phosphate. Rather, it almost certainly ends in a 3' hydroxyl group, as suggested by its efficient polyadenylation and by the mobility of oligonucleotide X in a T_1 fingerprint (Fig. 3).

In summary, our analysis of 170-nt RNA leads us to conclude that cleavage in the presence of EDTA generates a 5' half-molecule with a 3' end at +1. This terminal adenosine bears a 3' hydroxyl group to which 3'-dATP [and so, presumably, poly(A)] is added. Thus, the first A of the poly(A) tract is derived from the mRNA precursor.

RNAs 1 to 5 are 3' half-molecules. The lengths of RNAs 1 to 5 were determined by comparison with an alkaline hydrolysate of uniformly labeled RNA (Fig. 7, lanes 1, 3, and 5). The longest RNA, RNA 1, contains 36 ± 1 nt, and the smallest, RNA 5, contains 28 ± 1 nt. T_1 RNase analysis (Fig. 8, column A) showed that each RNA contains T_1 oligonucleotides found downstream, but not upstream, of the poly(A) site, and possesses the same 3' terminus as the precursor (oligonucleotide 12). Each RNA lacks the oligo-



nucleotide which spans the poly(A) site of the precursor (oligonucleotide 1) and instead contains a unique oligonucleotide, designated Y(1) to Y(5), not found in the other RNAs or in the precursor.

Short RNAs that are less abundant than RNAs 1 to 5 are also detected in Fig. 7. RNAs of every length between 15 and 45 nt are present and appear to constitute a background on which RNAs 1 to 5 are superimposed. This background does not reflect nonspecific degradation. Rather, these short RNAs, like RNAs 1 to 5, are 3′ half-molecules that differ only in the position of their 5′ ends. The 20- and 25-nt RNAs, for example, contain T_1 oligonucleotides 12 and 21, but not 4 (Fig. 8; data not shown). The 45-nt RNA species (indicated by an asterisk in Fig. 7) is unusual in that its 5′ terminus appears to lie upstream of the poly(A) site. More detailed analysis of this RNA is required to identify its 5′-terminal sequence definitively.

In summary, RNAs 1 to 5 contain only sequences downstream of the poly(A) site; they differ from each other only in the location of their 5′ ends. RNAs with 5′ termini at each position between nt 36 (RNA 1) and nt 15 are also observed, but are less abundant. Thus, RNAs 1 to 5 and the other short RNAs form a family of related 3′ half-molecules.

RNAs 1 to 5 have 5′-terminal phosphates. All 3′ half-molecules carry 5′-terminal phosphates, as demonstrated by three experiments. First, the electrophoretic mobility of each RNA decreased upon phosphatase treatment (compare, in Fig. 7, lanes 1 and 2, or lanes 3 and 4). Second, phosphatase treatment before T_1 digestion caused each Y oligonucleotide to shift in the fingerprinting analysis, without affecting the mobility of the other oligonucleotides (Fig. 8, column B versus column A). Third, as we show below (see Fig. 9), T_2 RNase digestion generated a pNp product from the 5′ end of each RNA.

Structure of RNAs 1 to 5. From the lengths of each RNA (Fig. 7) and the fingerprinting analysis (Fig. 8), we could position the 5′ termini with an uncertainty of 1 nt (Fig. 9). To eliminate the uncertainty and confirm each structure, we performed T_2 RNase analyses.

T_2 digestion of each RNA should generate a characteristic pNp product, in which N is the 5′-terminal residue of the RNA. The pNp product should be eliminated if the RNA is treated with phosphatase before T_2 digestion. Using a precursor with [α-^{32}P]nucleoside triphosphates, the pNp will be detected if N or the adjacent, downstream nucleotide is labeled. Thus, by varying the labeled nucleoside triphosphate used to synthesize the precursor, we can deduce the terminal dinucleotides of RNAs 1 to 5.

An analysis of this type (Fig. 10) demonstrated that the sequences proposed in Fig. 9 are correct and that each RNA bears a 5′-terminal phosphate. For example, RNA 1 derived from [α-^{32}P]CTP-labeled −141/+31 RNA generated pAp after T_2 RNase digestion, as well as each of the four 3′

FIG. 7. Approximate length of RNAs 1 to 5. [α-^{32}P]UTP-labeled −141/+31 RNA was incubated in nuclear extract reaction mixtures containing either EDTA or 3′-dATP. Half of each sample was then treated with phosphatase (+ lanes); the other half was not treated (− lanes). The products were analyzed by electrophoresis through a 15% acrylamide–7 M urea sequencing gel. Lengths were estimated by comparison with a partial alkaline hydrolysate (13) of −141/+31 RNA and with known purified oligonucleotides (not shown). Lanes: 1, EDTA reaction, no phosphatase; 2, EDTA reaction, phosphatase added; 3, 3′-dATP reaction, no phosphatase; 4, 3′-dATP reaction, phosphatase added; 5, partial alkaline hydrolysate. The asterisk indicates a 45-nt RNA species referred to in the text.

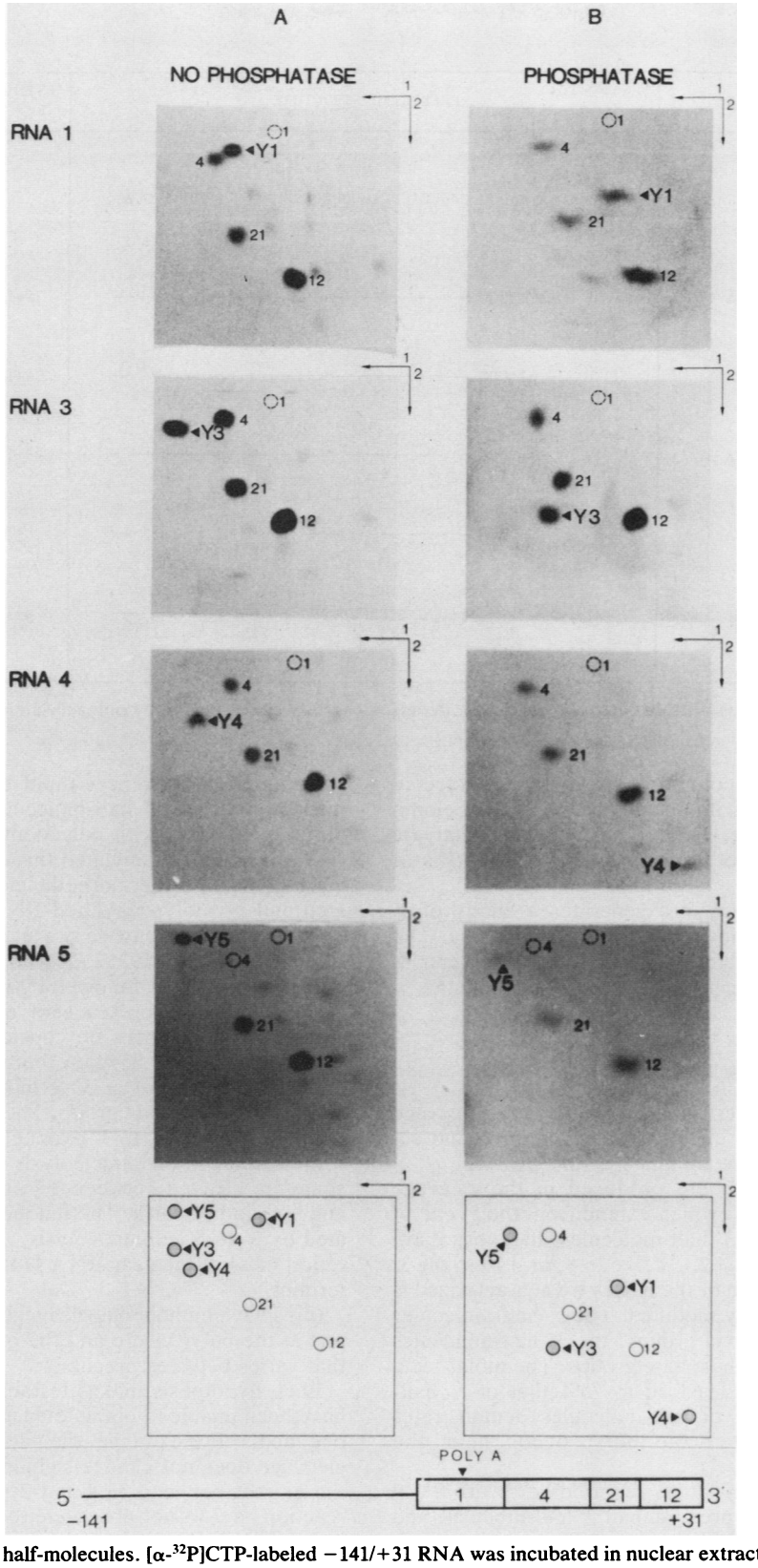


FIG. 8. RNAs 1 to 5 are 3′ half-molecules. [α-^{32}P]CTP-labeled −141/+31 RNA was incubated in nuclear extract in the presence of EDTA. RNAs 1, 3, 4, and 5 were purified by gel electrophoresis. Half of each was treated with phosphatase (column B); the other half was not (column A). RNAs were then digested with RNase T_1, and the products were fractionated by two-dimensional chromatography. The RNA species being analyzed in each pair of fingerprints is indicated to the left. Oligonucleotides 1, 4, 12, and 21 are numbered in each fingerprint, and their positions in the precursor are diagrammed at the bottom of the figure. New oligonucleotides (not found in the precursor) are designated Y and are indicated by an arrowhead. The positions at which oligonucleotides 1 and 4 would appear, were they present, are indicated by dotted circles. The drawings below the autoradiograms are a schematic compilation of the data. The Y oligonucleotides are shown as shaded circles, and oligonucleotides 4, 21, and 12 are shown as open circles.

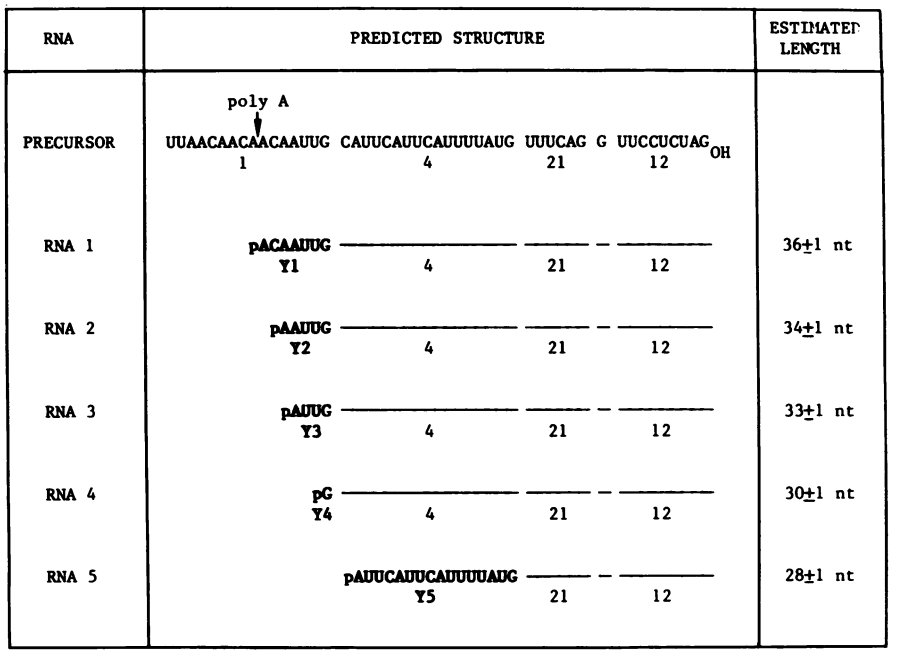


FIG. 9. Predicted structures of RNAs 1 to 5. See text for details. Lengths were estimated by polyacrylamide gel electrophoresis (Fig. 7).

monophosphates (Fig. 10, column A). pAp was not detected if RNA 1 was first treated with phosphatase (Fig. 10, column B). The structures in Fig. 11 are the only ones that are consistent with both these T_2 analyses and the lengths of each RNA.

In summary, cleavage in vitro generates a family of 3′ half-molecules, each bearing a 5′-terminal phosphate. The longest member of this family, RNA 1, has a 5′ end at position +2. Thus, it directly abuts the 3′ end of the 5′ half-molecule.

DISCUSSION

In this report we analyzed RNAs formed during cleavage of SV40 late pre-mRNAs in vitro. The structures and sequences of these RNAs are summarized in Fig. 11.

Two types of molecules are produced in the cleavage reaction: 5′ half-molecules (which extend from the 5′ end of the precursor to +1), and 3′ half-molecules (the most abundant of which extend from +2, +4, +5, +8, or +10 to the 3′ end of the precursor). Each of the RNAs we characterized is linear and is not detectably modified. The 5′ half-molecules possess 3′-terminal hydroxyl groups; the 3′ half-molecules each possess 5′-terminal phosphate groups. The molar yield of 3′ half-molecules is at least 50 to 75% that of 5′ half-molecules. The existence of half-molecules demonstrates that endonucleolytic cleavage must occur near the polyadenylation site.

The general pattern of products reported here with SV40 late pre-mRNAs, a single, predominant 5′ half-molecule and multiple 3′ half-molecules with staggered 5′ termini, recently has been observed with adenovirus L3 (39) and histone H4 (16) pre-mRNAs. With histone mRNA (16), as with SV40 late mRNA, the largest 3′ half-molecule directly abuts the unique 5′ half-molecule. Thus, in both cases, these two half-molecules contain every nucleoside and phosphate of the pre-mRNA. In contrast, L3 pre-mRNA generates 5′ and 3′ half-molecules that are separated by at least 1 nt (39). This gap might simply reflect rapid exonucleolysis of the first nucleotide of the 3′ half-molecule. Regardless, the general patterns obtained with polyadenylated (SV40 and adenovirus) and nonpolyadenylated (histone H4) mRNAs are strikingly similar. Furthermore, the cleavage of both polyadenylated and nonpolyadenylated RNAs requires sequences on either side of the cleavage site and, probably, a small nuclear ribonucleoprotein (19, 37, 38). Thus, cleavage may be similar whether or not it is followed by polyadenylation.

To explain the presence of multiple 3′ half-molecules, three hypotheses may be considered. (i) Endonucleolysis occurs only at one position, between the adenosines at +1 and +2, and leaves a 3′-terminal hydroxyl group and a 5′-terminal phosphate group. The multiple 3′ half-molecules are generated by 5′-to-3′ exonucleolysis.

(ii) The site of endonucleolysis varies between precursors: some are cleaved between +1 and +2, others between +3 and +4, and so forth. The unique 5′ half-molecule is generated by 3′-to-5′ exonucleolysis, as occurs during the maturation of small nuclear RNA (14, 30, 54) and tRNA (20) 3′ termini.

(iii) Two simultaneous cleavages occur in each precursor: one at the poly(A) site and the other downstream, at a site that varies between precursors.

Each hypothesis makes testable predictions. As yet, we have been unable to obtain evidence decisively favoring any one model. For example, the pattern of RNAs generated by cleavage does not change significantly during incubation in the extract between 30 s and 2 h. Thus, even early in the reaction, we do not observe either intermediates extending beyond the poly(A) site or the conversion of RNA 1 to RNAs 2 to 5 (data not shown). Similarly, purified RNA 1 is not shortened during a second incubation in extract (data not shown). A priori, the most attractive model to us is that endonucleolytic cleavage occurs only at the polyadenylation site, since all pre-mRNA molecules would be cleaved identically.

In each of these models, endonucleolytic cleavage must

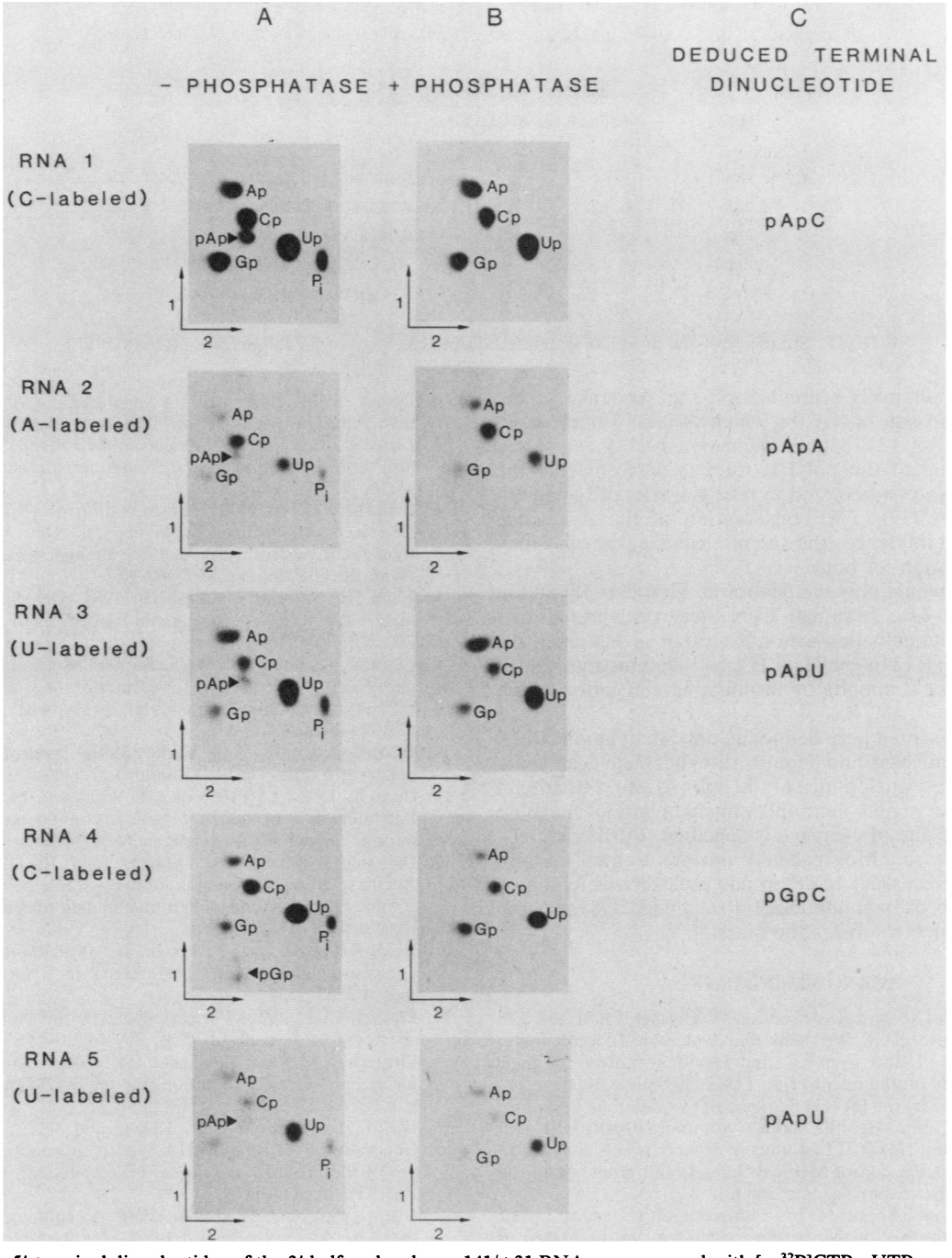


FIG. 10. The 5′-terminal dinucleotides of the 3′ half-molecules. −141/+31 RNA was prepared with [α-^{32}P]CTP, -UTP, or -ATP (indicated in the left column). RNAs 1 to 5 were purified with phosphatase (column B) or left untreated (column A). The RNAs were digested with RNase T_2, and the products were separated by two-dimensional chromatography. (A) RNase T_2 digestion products generated without prior phosphatase treatment. The pNp residue observed in each case is indicated with an arrow. Inorganic phosphate (P_i) is liberated during RNase T_2 digestion but is due to contaminating nuclease activities, since it also is liberated from −141/+31 RNA that has never been in the extract. (B) RNase T_2 digestion products generated after each RNA was treated with phosphatase. (C) The 5′-terminal dinucleotides deduced through this analysis.

generate a 5′ phosphate and a 3′ hydroxyl group. These same termini are left by the cleavages which occur during nuclear mRNA splicing (22) and self-splicing (7). Whether this reflects a fundamental similarity in the biochemical mechanism of 3′-end formation and splicing remains to be determined.

Poly(A) is added to the adenosine residue at +1. We draw this conclusion because, after cleavage in the presence of 3′-dATP, a single cordycepin residue was added to +1 (Fig. 6). We assume that 3′-dATP is added to the same residue as is ATP normally. This is suggested by the fact that one or two adenosines sometimes are added to +1 before a cordycepin causes chain termination (Fig. 3 and 6).

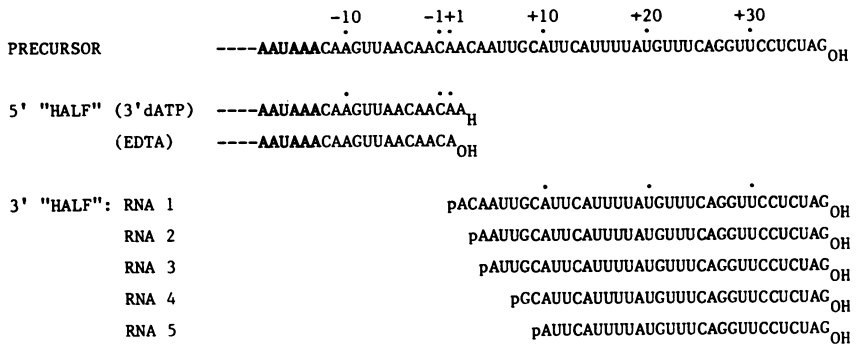


FIG. 11. Summary of the products generated during cleavage of SV40 late pre-mRNAs in vitro.

In virtually all mRNA precursors, the sequence at the polyadenylation site is NA (in which N is any nucleotide, most commonly C) (4, 33). Presumably, in all these cases, the first residue of the poly(A) tract is derived from the precursor, as shown here and in recent work with adenovirus L3 mRNA (39). The conservation of the A residue suggests that it influences the site of cleavage, the efficiency of polyadenylation, or both.

3'-dATP prevents polyadenylation in vivo (11, 35) and in vitro (24, 29, 44). Although it has been suggested that 3'-dATP inhibits polyadenylation by acting as a competitive inhibitor of ATP (24), our data (Fig. 4) demonstrate that it does not; rather it inhibits by terminating elongation of the poly(A) tract.

The data presented here define the polyadenylation site of SV40 late pre-mRNAs and demonstrate that cleavage occurs at or within a few nucleotides of the poly(A) site. Mutational analysis of the region near the polyadenylation site may reveal how the site of cleavage is specified. Further experiments, perhaps requiring partially purified extract components, will be necessary to determine precisely which phosphate bond is broken endonucleolytically and to identify the component which catalyzes the reaction.

ACKNOWLEDGMENTS

The advice and technical instruction of Elsebet Lund and Jim Dahlberg were invaluable. We thank Phil Anderson, Judith Kimble, Vivian Bardwell, Laura Conway, and David Zarkower for their helpful comments on the manuscript. DNA oligonucleotide synthesis was performed by the Protein Sequence-DNA Synthesis Facility, supported by Public Health Service grants (S10-RR01684 and CA07175) from the National Institutes of Health, a G.R.S. grant to the University of Wisconsin Medical School, and grants from the University of Wisconsin Graduate School.

This work was supported by Public Health Service grant GM31892 from the National Institutes of Health, by the March of Dimes (Basil O'Connor Starter research grant), and by a Steenbock Career Development award to M.W.

LITERATURE CITED

1. **Alt, F. W., A. L. M. Bothwell, M. Knapp, E. Siden, E. Mather, M. Koshland, and D. Baltimore.** 1980. Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. Cell **20**:293–302.
2. **Amara, S. G., V. Jonas, M. G. Rosenfeld, E. S. Ong, and R. M. Evans.** 1982. Alternate RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptides. Nature (London) **298**:240–244.
3. **Barrell, B. G.** 1971. Fractionation and sequence analysis of radioactive nucleotides, p. 751–759. *In* G. L. Cantonit and D. R. Davies (ed.), Proceedings in nucleic acids research, vol. 2. Harper & Row, Publishers, Inc., New York.
4. **Berget, S. M.** 1984. Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature (London) **309**:179–182.
5. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3' end formation: the end is in site! Cell **41**:349–359.
6. **Brownlee, G. G.** 1972. Determination of sequences in RNA. Elsevier/North-Holland Publishing Co., New York.
7. **Cech, T. R., and B. L. Bass.** 1986. Biological catalysis by RNA. Annu. Rev. Biochem. **55**:599–629.
8. **Colot, H. V., and M. Rosbash.** 1982. Behavior of individual maternal pA$^+$ RNAs during embryogenesis of *Xenopus laevis*. Dev. Biol. **94**:79–86.
9. **Conway, L., and M. P. Wickens.** 1985. A sequence downstream of AAUAAA is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes. Proc. Natl. Acad. Sci. USA **82**:3949–3953.
10. **Darnell, J. E., Jr.** 1982. Variety in the level of gene control in eukaryotic cells. Nature (London) **297**:365–371.
11. **Darnell, J. E., L. Philipson, R. Wall, and M. Adesnik.** 1971. Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. Science **174**:507–510.
12. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475–1488.
13. **Donis-Keller, H., A. M. Maxam, and W. Gilbert.** 1977. Mapping adenines, guanines, and pyrimidines in RNA. Nucleic Acids Res. **4**:2527–2538.
14. **Eliceiri, G. L.** 1980. Formation of low molecular weight RNA species in HeLa cells. J. Cell. Physiol. **102**:199–207.
15. **Fitzgerald, M., and T. Shenk.** 1981. The sequence AAUAAA forms part of the recognition site for polyadenylation of SV40 late RNAs. Cell **24**:251–260.
16. **Gick, O., A. Kramer, W. Keller, and M. L. Birnstiel.** 1986. Generation of histone mRNA 3' ends by endonucleolytic cleavage of the pre-mRNA in a snRNP-dependent *in vitro* reaction. EMBO J. **5**:1319–1326.
17. **Gil, A., and N. J. Proudfoot.** 1984. A sequence downstream of the AAUAAA is required for rabbit β-globin mRNA 3'-end formation. Nature (London) **312**:473–474.
18. **Grabowski, P. J., R. A. Padgett, and P. A. Sharp.** 1984. Messenger RNA splicing *in vitro*: an excised intervening sequence and a potential intermediate. Cell **37**:415–427.
19. **Hashimoto, C., and J. A. Stietz.** 1986. A small nuclear ribonucleoprotein associates with the AAUAAA polyadenylation signal *in vitro*. Cell **45**:581–591.
20. **Henryk, C., and M. P. Deutscher.** 1980. Apparent involvement of ribonuclease D in the 3' processing of tRNA precursors. Proc. Natl. Acad. Sci. USA **77**:837–841.
21. **Higgs, D. R., S. E. Y. Goodbourn, J. Lamb, J. B. Clegg, D. J. Weatherall, and N. J. Proudfoot.** 1984. Alpha thalassemia caused by a polyadenylation signal mutation. Nature (London) **306**:398–400.
22. **Keller, W.** 1984. The RNA lariat: a new ring to the splicing of mRNA precursors. Cell **39**:423–425.

23. **Knapp, G., R. C. Ogden, C. L. Peebles, and J. Abelson.** 1979. Splicing of yeast tRNA precursors: structure of the reaction intermediates. Cell **18**:37–45.
24. **Koch, S., and J. Niessing.** 1978. Mode of inhibition of nuclear poly(A) polymerase by 2'dATP and 3'dATP. FEBS Lett. **96**:354–356.
25. **Konarska, M., W. Filipowics, H. Domdey, and H. J. Gross.** 1981. Formation of a 2'-phosphomonoester 3',5'-phosphodiester linkage by a novel RNA ligase in wheat germ. Nature (London) **293**:112–116.
26. **Konarska, M., W. Filipowics, and H. J. Gross.** 1982. RNA ligation via 2'-phosphomonoester, 3',5'-phosphodiester linkage: requirement of 2',3'-cyclic phosphate termini and involvement of a 5'-hydroxyl polynucleotide kinase. Proc. Natl. Acad. Sci. USA **79**:1474–1478.
27. **Konarska, M. M., R. A. Padgett, and P. A. Sharp.** 1984. Recognition of cap structure in splicing *in vitro* of mRNA precursors. Cell **38**:731–736.
28. **Krainer, A. R., T. Maniatis, B. Ruskin, and M. R. Green.** 1984. Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. Cell **36**:993–1005.
29. **Maale, G., G. Stein, and R. Mans.** 1975. Effects of cordycepin and cordycepin triphosphate on polyadenylic and ribonucleic acid-synthesizing enzymes from eukaryotes. Nature (London) **255**:80–82.
30. **Madore, S. J., E. D. Weiben, and T. Pederson.** 1984. Intracellular site of U1 snRNA processing and ribonucleoprotein assembly. J. Cell Biol. **98**:188–192.
31. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
32. **McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins.** 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. Cell **37**:993–999.
33. **McLauchlin, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucleic Acids Res. **13**:1347–1368.
34. **Melton, D., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
35. **Mendecki, J., S. J. Lee, and G. Brawerman.** 1972. Characteristics of the polyadenylic acid segment associated with messenger ribonucleic acid in mouse sarcoma 180 ascites cells. Biochemistry **11**:792–801.
36. **Montell, C., E. F. Fisher, M. H. Caruthers, and A. J. Berk.** 1983. Inhibition of RNA cleavage but not polyadenylation by a point mutation in the 3' consensus sequence AAUAAA. Nature (London) **305**:600–605.
37. **Moore, C., and P. A. Sharp.** 1984. Site-specific polyadenylation in a cell-free reaction. Cell **36**:581–591.
38. **Moore, C., and P. A. Sharp.** 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell **41**:845–855.
39. **Moore, C. L., H. Skolnik-David, and P. A. Sharp.** 1986. Analysis of RNA cleavage at the adenovirus-2 L3 polyadenylation site. EMBO J. **5**:1929–1938.
40. **Nevins, J. R., and J. E. Darnell.** 1978. Steps in the processing of Ad2 mRNA: poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell **15**:1477–1493.
41. **Nishimura, S.** 1972. Minor components in transfer RNA: their characterization, location, and function. Prog. Nucleic Acids Res. Mol. Biol. **12**:50–85.
42. **Padgett, R. A., M. Konarska, P. J. Grabowski, S. F. Hardy, and P. A. Sharp.** 1984. Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. Science **225**:898–903.
43. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3' non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
44. **Rose, K. M., L. E. Bell, and S. T. Jocob.** 1977. Specific inhibition of chromatin-associated poly(A) synthesis *in vitro* by cordycepin 5' triphosphate. Nature (London) **267**:178–180.
45. **Rosenthal, E., T. Tansey, and J. Ruderman.** 1983. Sequence-specific adenylation and deadenylations accompany changes in the translation of maternal messenger RNA after fertilization in *Spisula* oocytes. J. Mol. Biol. **166**:309–327.
46. **Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. Cell **38**:317–331.
47. **Sadofsky, M., and J. C. Alwine.** 1984. Sequences on the 3' side of hexanucleotide AAUAAA affect efficiency of cleavage at polyadenylation site. Mol. Cell. Biol. **4**:1460–1468.
48. **Sadofsky, M., S. Connelly, J. L. Manley, and J. C. Alwine.** 1985. Identification of a sequence element on the 3' side of AAUAAA which is necessary for simian virus 40 late mRNA 3'-end processing. Mol. Cell. Biol. **5**:2713–2719.
49. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**:107–110.
50. **Tooze, J.** 1981. DNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
51. **Wickens, M. P., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent mRNA 3' end formation. Science **226**:1045–1051.
52. **Woychik, R. P., R. H. Lyons, L. Post, and F. M. Rottman.** 1984. Requirement for the 3' flanking region of the bovine growth hormone gene for accurate polyadenylation. Proc. Natl. Acad. Sci. USA **81**:3944–3948.
53. **Zarkower, D., P. Stephenson, M. Sheets, and M. Wickens.** 1986. The AAUAAA sequence is required both for cleavage and for polyadenylation of simian virus 40 pre-mRNA in vitro. Mol. Cell. Biol. **6**:2317–2323.
54. **Zieve, G., and S. Penman.** 1976. Small RNA species of the HeLa cell: metabolism and subcellular location. Cell **8**:19–31.
55. **Ziff, E. B.** 1980. Transcription and RNA processing in the DNA tumor viruses. Nature (London) **287**:491–499.