Nucleotide sequences and mapping of novel heterogeneous 5'-termini of adenovirus 2 early region 4 mRNA

Shuichi Hashimoto, Michael H. Pursley and Maurice Green

Institute for Molecular Virology, St. Louis University School of Medicine, 3681 Park Avenue, St. Louis, MO 63110, USA

Received 24 November 1980

ABSTRACT

The major 5'-termini of human adenovirus type 2 early gene block 4 mRNA were sequenced. Poly(A⁺) polyribosomal RNA was isolated from Ad2 early infected cells, the 5'-terminal m⁷Gppp removed and the 5'-OH of the penultimate 2'-O-methylated nucleotide labeled with [γ -³²P]ATP using polynucleotide kinase. Ad2 E4 mRNA was purified by hybridization to the Ad2 EcoRI-C fragment and was digested with RNase T1. The resulting oligonucleotides were resolved by two dimensional paper electrophoresis-homochromatography. Four major and 3-4 minor 5'-terminal sequences were identified and characterized. The sequence of the 5'-terminal structures of the major four termini are: (1) m⁷GpppU^mUACACUGp, (2) m⁷GpppU^mUACACUGp, (3) m⁷GpppU^mACACUGp, and (4) m⁷Gppp(m⁶)A^mC^mACUGp. These major 5'-terminal sequences were aligned with nucleotide 325, 326, 327, and 329 from the righthand end of the known Ad2 DNA sequence (1) in the region mapped as the 5'-terminus of E4 mRNA by electron microscopy (2,3) and S1 nuclease-gel (4) mapping. Two potential ribosomal binding sites and an initiator codon were found at 40 to 65 nucleotides and about 80 nucleotides, respectively, from these heterogeneous 5'-termini. Ad2 E4 major mRNA species appear to be unique since mRNA molecules initiate at a pyrimidine, perhaps by RNA polymerase stuttering, or they are products of an unusual type of RNA processing.

INTRODUCTION

Human adenovirus type 2 (Ad2) early genes are models of eukaryotic gene organization, expression, and regulation (5-7). The early genes are clustered in four major regions, E1 to E4, located at map positions 1.5 to 11 (E1) and 76 to 86 (E3) on the r-strand, and 68 to 62 (including a leader at 75) (E2) and 99 to 91.5 (E4) on the l-strand (the map contains 100 units; r and l refer to rightward and leftward transcription). Studies of promoter mapping, DNA sequence of the promoter regions, and mapping of 5'-termini of mRNA to Ad2 DNA of the four Ad2 early regions are important for understanding transcription initiation and regulation. Two approaches, UV mapping (8) and "nascent RNA mapping" (9), have identified four major promoters, one at the 3'-side of map positions E1 to

E4; E1 may contain additional weaker promoters (10,11). Electron microscopic and S1 nuclease-gel mapping procedures have shown that each early region codes families of partially overlapping spliced RNAs which appear to have common 5'-termini (2-4).

Biochemical analysis has shown that Ad2 early mRNAs have heterogeneous cap structures, including penultimate pyrimidine nucleosides (U^m), at the 5'-termini (12,13), and that the number of termini is greater than the number of known promoters (14). Structural studies of the RNase T1 oligonucleotides derived from the 5'-termini of early mRNA have suggested that each gene specific mRNA has heterogeneous sequences (15,16). The 5'-terminus of E4 mRNA molecules appears to be particularly heterogeneous, and contains 5'-terminal pyrimidine nucleotides (pU^m) (15), a feature not previously observed with purified capped eukaryotic mRNAs. Recently, the nucleotide sequence of one of the heterogeneous 5'-terminal sequences was reported by Baker and Ziff (17).

In this paper, we report further detailed analysis of E4 mRNA. We have isolated four major and 3-4 minor 5'-terminal sequences by RNase T1 digestion of *in vitro* 5'- ^{32}P -labeled E4 mRNA followed by two dimensional electrophoresis-homochromatography. The major four terminal RNase T1 fragments were characterized, sequenced, and aligned with a stretch of 5 nucleotides at map position 325 to 329 from the righthand end of the Ad2 genome.

MATERIALS AND METHODS

Cell culture and virus infection - Exponentially growing KB cells in suspension were infected with Ad2 as described, and labeled from 2 to 10 h postinfection with [5,6- 3H]Urd to monitor Ad2 mRNA purification (15).

Cell fractionation and isolation of poly(A^+) RNA - Polyribosomal RNA (2 to 5 mg) was isolated from 1 or 2 liters of Ad2 infected cells (3 to 5 $\times 10^5$ cells/ml) and poly(A^+) RNA was selected on poly(U)-Sephadex (12).

Ad2 DNA and restriction fragment - Ad2 DNA was isolated and EcoRI restriction fragments were prepared as described (18,19,15).

In vitro labeling of 5'-termini of Ad2 mRNA - The 5'-terminal m^7Gppp in poly(A^+) RNA was removed by periodate oxidation, β -elimination, phosphatase treatment, and the new 5'-terminal N^m residues were labeled with [γ - ^{32}P]ATP (100-500 Ci/mmol) using polynucleotide kinase (New England BioLabs) (14,15).

Isolation of EcoRI-C specific Ad2 mRNA - About 50 μg of *in vitro* 5'- ^{32}P -labeled poly(A^+) RNA from Ad2 infected cells were hybridized to 10

to 20 µg of Ad2 EcoRI-C fragment immobilized on a filter (15).

RNAse T1 digestion and fingerprinting - Hybridization purified viral RNA plus 100 µg of carrier yeast tRNA were digested with 5 µg of RNAse T1 (Sankyo Co. Ltd.) in 10 µl of 20 mM Tris-HCl, pH 7.5 at 37° for 1 h. The RNAse T1 digest was fractionated by electrophoresis on cellulose acetate strips in the first dimension and by homochromatography on DEAE-cellulose thin layer chromatography (TLC) plates in the second dimension (20). The TLC plates were autoradiographed with X-ray film (Kodak, XR-5) and two Lightning-plus (Dupont) intensifying screens at -70° for 1 to 2 h.

Analysis of 5'-³²P-labeled nucleotides of RNAse T1 oligonucleotides - RNAse T1 oligonucleotide spots observed on DEAE-cellulose TLC plates by autoradiography were eluted with 2 M triethylamine bicarbonate. After removal of triethylamine bicarbonate by repeated evaporation, each sample was dissolved in 20 µl of 20 mM sodium acetate, pH 5.3. A portion (5 µl) of each spot was digested with 2 µg of nuclease P1 (Boehringer Mannheim) at 37° for 2 h, mixed with pN^m markers (pA^m, pm⁶A^m, pC^m, pU^m, pG^m, 0.5 A_{260nm} units each, P-L Biochemicals), and applied to cellulose TLC plates (EM Laboratories, Inc.). Two dimensional (2D)-TLC was performed with isobutyric acid/0.5 N NH₄OH (5:3) in the first dimension and isopropyl alcohol/concentrated HCl/H₂O (70:15:15) in the second dimension (21); this provides excellent resolution of pNs and pN^ms (14). TLC plates were autoradiographed for 2 to 7 days; marker spots were detected with a UV-lamp.

RNAse A or RNAse T2 digestion of RNAse T1 oligonucleotides - RNAse T1 oligonucleotides identified as 5'-terminal fragments (i.e., containing ³²P-labeled pN^m) were dissolved in 10-20 µl of 20 mM Tris-HCl, pH 7.6, 2 mM EDTA, or 20 mM sodium acetate, pH 4.5, containing 100 µg tRNA, and digested with 5 µg of RNAse A (Sigma Co.) or 1 unit of RNAse T2 (Sankyo Co. Ltd.), respectively. The mixtures were incubated at 37° for 2 h, separated by 2D-homochromatography as described above, and autoradiographed. In some experiments, spots were eluted and further analyzed by nuclease P1 digestion and 2D-TLC. 5'-³²P-labeled oligonucleotide markers, ³²pACp and ³²pAAAUp, were prepared by digestion of 5'-³²P-labeled RNA with RNAse A and RNAse T1, and purification by 2D-homochromatography.

Sequencing of RNAse T1 oligonucleotides - The 5'-terminal RNAse T1 oligonucleotides eluted from TLC plates were dissolved in 20 µl of 20 mM sodium acetate, pH 5.3 containing 100 µg carrier yeast tRNA, and digested with 100 ng of nuclease P1 at room temperature. Aliquots were removed at

various times, and the enzyme was inactivated by heating at 100° for 5 min in 5 mM EDTA (22). The aliquots were pooled, and mixed with a complete digest (enzyme inactivated) and with an undigested sample. The mixture of oligonucleotide products was resolved by 2D-homochromatography, and the sequence was deduced by the migration position of the products detected by autoradiography (22). As a control for sequence analysis, yeast tRNA^{Phe} of known sequence (23) was analyzed. The 5'-terminus of yeast tRNA^{Phe} (Boehringer Mannheim) was labeled with ³²P as described (14) and intact [5'-³²P]tRNA^{Phe} was isolated by electrophoresis on a 10% acrylamide gel in 7 M urea. The RNA was eluted, partially digested with nuclease P1, and analyzed by 2D-homochromatography as described above.

RESULTS

5' Terminal ³²P-Labeling of Poly(A⁺) RNA From Ad2 Infected KB Cells

The recovery of poly(A⁺) RNA from Ad2 infected KB cells and the efficiency of 5'-terminal ³²P-labeling in vitro were previously described (14). Similar results were obtained in the experiments described here. Two to four pmol of ³²P was incorporated per µg of RNA, into both bonafide 5'-termini of mRNA and 5'-termini resulting from nicks in the RNA. Labeled RNA was assayed routinely to estimate how much ³²P was incorporated into each 5'-terminus ("bonafide" and "nicked") by analysis of nuclease P1 digested labeled RNA. The ³²P-radioactivity in bonafide termini are in 2'-O-methyl nucleotides, whereas that in termini resulting from internal nicks is in unmethylated nucleotides.

Purification of EcoRI-C Fragment Specific mRNA

About 50 µg of in vitro labelled [5'-³²P] poly(A⁺) RNA from Ad2 infected KB cells were annealed to 10-20 µg of Ad2 EcoRI-C fragment on filters. From 2-3% of input RNA, representing 1 to 1.5 µg of viral RNA (based on [³H]Urd radioactivity), were eluted from the filter. Under these conditions, contamination of viral RNA with cellular RNA was negligible (14).

Fractionation of 5'-terminal ³²P-labeled RNase T1 Oligonucleotides derived from EcoRI-C Specific mRNA

A typical RNase T1 fingerprint of 5'-³²P-labeled EcoRI-C specific mRNA is shown in Fig. 1. The spots were eluted, digested with nuclease P1, and the 5'-terminal nucleotides identified (see below). The oligonucleotides designated by numbers or letters contained 2'-O-methylnucleotides at the 5'-terminus. Oligonucleotides 1, 2, 3, and 4 are major

Figure 1. Fingerprint of an RNase T1 digest of *in vitro* 5'-³²P-labeled Ad2 E4 specific mRNA. About 50 µg of 5'-³²P-labeled poly(A⁺) RNA (2.9×10^7 cpm) were annealed to 10-20 µg of Ad2 EcoRI-C fragment on a filter. Hybridized RNA was eluted (4.5×10^6 cpm), precipitated with alcohol together with 100 µg of carrier tRNA, digested with 5 µg of RNase T1, and fingerprinted. The first dimension (left to right) is electrophoresis on cellulose acetate paper, pH 3.5. The second dimension (bottom to top) is homochromatography (20). Spots designated by numbers or letters were derived from the 5'-termini of E4 mRNA. B and Y are positions of xylene cyanol FF and methyl orange, respectively.

5'-terminal oligonucleotides and oligonucleotides a, b, c, and d are minor. Although a is a strong spot, the content of 2'-methylnucleotides (pm⁶A^m and pA^m) was low and variable in different preparations (10 to 60%). The fingerprint in Fig. 1 is that of RNA prepared at 10 h postinfection in the presence of cycloheximide. Viral RNA prepared at 6 and 8 h postinfection in the presence of cycloheximide also contained oligonucleotides 1, 2, 3, and 4 as major 5'-termini (about 90% of the total 5'-termini of E4 mRNA, i.e. oligonucleotides 1, 2, and 3, 10-15% each; oligonucleotide 4, ~50%).

Analysis of 5'-terminal Nucleotides of RNase T1 Oligonucleotides

The oligonucleotides shown in Fig. 1 were eluted, digested with nuclease P1, and analyzed on 2D-TLC, as described in Materials and Methods. Oligonucleotides that contain 2'-O-methyl nucleotide(s) should represent 5'-terminal structures of mRNA. Figure 2 shows the analysis of the 5'-terminal nucleotides of oligonucleotides 1 to 4 and a to d. Oligonucleotides 1, 2, and 3 contained pU^m and pN^{*m} . As previously described, pN^{*m} is probably a 2'-O-methyl uridine derivative (15). Recently, Harada *et al.* (24) isolated a 2'-O-methyluridine derivative (pU^{*m}), from tRNA^{Pro} present in Moloney murine leukemia virions; pU^{*m} migrated at the same position as our pN^{*m} on 2D-TLC using the solvent system reported by these authors (24). Thus E4 mRNA has three sequences that may initiate with U and modified U. Oligonucleotide 4 contained mainly pm^6A^m , with minor quantities of pA^m and pA ; therefore, base methylation of A to m^6A is partial in this terminus.

Oligonucleotides a, b, and c contained low quantities of pm^6A^m and pA^m . We were unable to sequence these oligonucleotides because of the relatively higher contamination of these oligonucleotides with unmethylated oligonucleotides. Oligonucleotide d contained very low quantities of a 5'-terminal nucleotide which did not co-migrate with any marker nucleotide (plate D in Fig. 2), and thus was not identified.

Further characterization of the 5'-termini of the Major RNase T1 Fragments

The 5'-termini of oligonucleotides 1 to 4 were further analyzed by RNase T2 or RNase A digestion. Plates 1 and 2 in Fig. 3 show 2D-homochromatography of the RNase T2 digests of oligonucleotides 1 and 2, respectively. Our results show that they have the same 5'-terminal three nucleotides. Two identical ^{32}P -labeled spots were produced from oligonucleotides 1 and 2. The 5'-terminal nucleotides of these spots are pU^m and pN^{*m} as shown by nuclease P1 digestion and 2D-TLC (see Fig. 2) and by direct analysis of eluted spots 1 and 2 (not shown). The two spots in plates 1 and 2 must represent cap 1 ($pN^m pNp$) and cap 2 ($pN^m pN^m pNp$) structures; we have shown previously that there are cap 1 ($m^7GpppN^m pNp$) and cap 2 ($m^7GpppN^m pN^m pNp$) structures at the 5'-terminus of Ad2 early mRNA (12). From the migration orientation of $pU^m U^m Up$ (cap 2) and $pU^m Up$ (cap 1) (the general rules for the interpretation of mobility shifts is described in the next section), the third base of cap 2 is deduced to be U, i.e., there are only 3 possible dinucleotides which start with U: $pUpUp$, $pUpCp$, $pUpAp$ ($pUpGp$ is ruled out by the results of RNA T1 diges-

Figure 2. Analysis of ^{32}P -labeled 5'-terminal nucleotides of RNase T1 oligonucleotides derived from Ad2 E4 mRNA. Oligonucleotides 1 to 4 (major 5'-terminal sequences) and a to d (minor 5'-terminal sequences) shown in Fig. 1 were eluted, digested with nuclease P1, and 5'-terminal nucleotides were identified by 2D-TLC using pN^{m} and pN markers as described under Materials and Methods. Variable amounts of unmethylated nucleotides, mainly pA, were detected in nuclease P1 digests, reflecting background contamination with ^{32}P -labeled internal nicks in RNA. Two unusual spots (pN^{m}) in plates 1, 2, and 3 are most likely a 2'-O-methyl uridine derivative previously described (14,15). The unusual spot (pN^{m}) in plate d is not identified.

Figure 3. Two-dimensional homochromatography of the 5'-terminal RNase T1 oligonucleotides 1 to 4 of Fig. 1 after digestion with RNase T2 or RNase A. Plate 1: RNase T2 digest of oligonucleotide 1. The 5'-terminal nucleotides of spots 1 and 2 were eluted and shown to be pU^m and pU^mU^m by nuclease P1 digestion and 2D-TLC. Plate 2: RNase T2 digest of oligonucleotide 2. Plate 3: RNase T2 digest of oligonucleotide 4. The 5'-terminal nucleotides of the two spots were shown to be $p(n^6A^m)$ and pA^m by nuclease P1 digestion and 2D-TLC. Plate 4: RNase A digest of oligonucleotide 4. The 5'-terminal nucleotides of both spots (1 and 3) were shown to be $p(n^6A^m)$ and pA^m as described above. Plate 5: Spot 1 of plate 3 and marker $^{32}pA^mCp$ (about 100 cpm each) were mixed and analyzed by 2D-homochromatography. Plate 6: An RNase A digest (500 cpm) of oligonucleotide 4 was mixed with $^{32}pAAAUp$ (200 cpm) and analyzed by 2D-homochromatography. 2D-Homochromatography was performed as described in Figure 1. A different batch of the homomixture was used for plate 4, accounting for the difference in the relative migration of spots to dye markers in the second dimension.

tion), and only pUpUp migrates to the right of the methyl orange marker (see Fig. 3). Therefore, the nucleotide sequence of the RNase T2 digest of both spots 1 and 2 should be $pU^m pU^{(m)} pUp$ in which the second 2'-O-methylation is partial.

By similar analysis of oligonucleotide 3, $pU^m pUp$ was obtained (not shown). The spot representing $pU^m pU^m pNp$ was not detected; probably the second 2'-O-methylation is incomplete.

The fingerprints of RNase T2 and RNase A digests of oligonucleotide 4 are shown in plates 3 and 4 (Fig. 3), respectively. The 5'-terminal nucleotide of each spot was shown to be pm^6A^m and pA^m by elution and nuclease P1 digestion followed by 2D-TLC (not shown). Spot 1 of plates 3 and 4 were mixed with a $^{32}pACp$ marker and analyzed by 2D-homochromatography. They were not separable (plate 5 in Fig. 3). Therefore, spot 1 most likely is $p(m^6)A^m pCp$. The third base of the cap 2 structure (spot 2 in plate 3) could be an A or C based upon the relative mobility of spot 1 and spot 2. However, spot 2 was not produced by RNase A digestion, and therefore the third base must be a purine, i.e. A. Co-chromatography of the RNase A digest with $^{32}pAAAUp$ marker indicated that the nucleotide length of the second spot (spot 3 in plate 4) should be four (plate 6 in Fig. 3), and that the fourth base should be C, because the base was cleaved by RNase A (C or U) and the relative migration of spots 1, 2, and 3 disclosed no U in spot 3. Therefore, the 5'-terminal sequence of oligonucleotide 4 is deduced to be $p(m^6)A^m C^{(m)} ACp$, in which the methylation in parenthesis is partial. The sequence of oligonucleotides 1 to 4 deduced from RNase T2 or RNase A digestion were confirmed by partial nuclease P1 digestion and 2D-homochromatography as described below.

Sequencing the major Ad2 E4 mRNA 5'-terminal RNase T1 Fragments

The major Ad2 E4 mRNA 5'-terminal RNase T1 fragments were sequenced by the partial nuclease P1 digestion and 2D-homochromatography method which has been developed for sequencing RNA (22) and DNA (25) oligonucleotides. The mobility of the partial digestion products of oligonucleotides 1 to 4 (Fig. 1) is shown in Fig. 4 (plates 1 to 4 respectively). As the partial digests were mixed with the complete digest and the undigested sample prior to the 2D-homochromatography, the smallest product (top) should be pN^m and the largest product (bottom) should be the intact 5'-terminal oligonucleotide. Since the 3'-phosphate of the oligonucleotide is removed by nuclease P1 (26), the second spot from the bottom should be the 5'-RNase T1 oligonucleotide lacking its 3'-phosphate (22).

Figure 4. Sequence analysis of the major 5'-terminal RNase T1 oligonucleotides 1 to 4 of Ad2 E4 mRNA by partial nuclease P1 digestion and 2D-homochromatography. Oligonucleotides 1 to 4 (plates 1 to 4) in Fig. 1 were eluted, digested partially with nuclease P1, and mixed with a complete digest and with the undigested oligonucleotide, as described in Materials and Methods. The mixture was analyzed by 2D-homochromatography. The spot of pU^m in plates 1, 2, and 3 was eluted and analyzed on 2D-TLC, as described in Figure 2. This spot contained pU^m and pN^m . Therefore, pU^m and pN^m are not distinguishable by electrophoresis and homochromatography. Several unusual spots of which the 5'-terminal nucleotide was pC were observed in plate 3. These are derived from background contamination. As shown in plate 4, pm^m and pA^m were also not distinguishable by 2D-homochromatography.

In the photograph of sequencing data (Fig. 4), some spots are faint. But they are clearly observed on the original films, and are reproducibly detected. The nucleotide sequences were deduced by: (1) the general rules for the interpretation of mobility shifts of oligonucleotides differing by one nucleotide (22, 25, 27); (2) the characterization of nuclease P1, RNase T2, and RNase A digests (Fig. 3); and (3) the specificity of RNase T1. The general rules for mobility shifts are: (a) the loss of pA or pC from an oligonucleotide produces a shift of the oligonucleotide spot to the right with the angle of the shift being greater for pC than for pA by 2D homochromatography; (b) the distance between two spots in the second dimension (homochromatography) produced by the loss of pA is larger than that for the loss of pC; (c) a loss of pU or pG from an oligonucleotide produces a shift of the spot to the left with the angle of the shift being greater for pU than for pG; (d) if the 5'-terminal two or three nucleotides in an oligonucleotide are composed of only A or C, such as spot 4 in Fig. 4, spot shifts by loss of A or C are slightly to the left by 2D homochromatography because of the slower migration of pA or pC in the first dimension (paper electrophoresis) (see 22), and therefore the spot shifts of ACAC in oligonucleotide 4 and oligonucleotides 1, 2, and 3 are different during sequencing in Fig. 4; (e) the loss of 3'-phosphate in an oligonucleotide produces a shift to the left. In control experiments, the 5'-terminal 15 nucleotides of yeast tRNA^{Phe} was analyzed by this method and the mobility changes of the oligonucleotides produced by nuclease P1 digestion corresponded to the known sequence (23).

The purity of oligonucleotides 1 and 2 is 95-100%, but oligonucleotides 3 and 4 are contaminated (~10%) with oligonucleotides containing pC and pA termini, respectively. Since contaminated oligonucleotides are generally mixtures of sequences, they are usually undetectable in 2D-homochromatographs. In some cases, such as plate 3 of Fig. 4, a contaminated spot with pC as 5'-terminal nucleotide was visible. However, since contaminating spots are usually not present in the major sequencing line, it is possible to distinguish them from the spots representing the bona-fide sequence. The spots in plate 4 of Fig. 4 were eluted, digested with nuclease P1, and the 5'-terminal nucleotide in these spots was shown to be pm⁶A^m and pA^m. From these sequencing experiments, and from the analysis of the RNase A or RNase T2 digests described above, we conclude that oligonucleotides 1, 2, 3, and 4 have the following sequences:

oligonucleotide 1, $pU^{m(m)}UUACACUGp$; oligonucleotide 2, $pU^{m(m)}UACACUG$; oligonucleotide 3, $pU^{m(m)}ACACUGp$; oligonucleotide 4, $p(m^6)A^mC(m)ACUGp$ (the methylation in the parenthesis is partial).

DISCUSSION

Four major distinct 5'-terminal sequences were obtained by RNase T1 digestion of *in vitro* 5'- ^{32}P -labeled Ad2 E4-specific mRNA. These 5'-terminal RNase T1 fragments were sequenced by 2D-homochromatography of RNase T2 and RNase A digests and of partial nuclease P1 digests. The deduced T1 sequences can be aligned with the nucleotide sequence of the *r*-strand of Ad2 E4 DNA (1). Fig. 5 shows the DNA sequence at nucleotide 271 to 430 from the righthand end and the position of the 5'-terminal RNase T1 fragments that correspond to the DNA sequence. The 5'-terminal nucleotide of these four RNase T1 oligonucleotides map at nucleotide 325, 326, 327, and 329, which are at mp 99.1. The entire E4 region (mp 91.5-99) has been sequenced by Sciaky and Roberts (personal communication), and the sequences of oligonucleotides 1 to 4 do not occur anywhere but at mp 99.1 on the *r*-strand. As shown in Fig. 5, two sequences that could form a hybrid with 3'-terminal sequence of 18S ribosomal RNA (28), i.e. potential ribosome binding sites, are present at 40 to 50 and 55 to 65 residues from these heterogeneous 5'-termini. The first translation initiation codon (AUG) is observed at about 80 residues from the terminus. Several 5'-terminal leader sequences of gene block 4 mRNAs have been mapped at about map position 99.2, and the length of the leader sequence was estimated as approximately 0.1 to 0.3 map units (2). Chow *et al.* (2) also observed one mRNA species lacking a leader and mapping at position 99.2 to 94.4.

Recently, Kitchingman and Westphal (3) observed E4 mRNAs with 5'-termini at mp 96.8 as well as 99.3, by electron microscopy. Possibly, the minor 5'-terminal sequences that we observe are located at mp 96.8. The four terminal sequences that begin at 325, 326, 327, and 329 from the right end of Ad2 DNA could represent multiple transcription initiation sites due to "stuttering" by RNA polymerase II. The usual eukaryotic RNA polymerase recognition sequence, termed the TATA box, exists at -25 to -30 upstream from oligonucleotide 4, ACACUG. This is typical of several eukaryotic mRNAs (29).

It is of interest that none of the well characterized mRNAs have a pyrimidine nucleotide at the 5'-terminus. Furthermore, we did not observe

REGION E4 CAP SITE (RNase T1 fragments 1 to 4)

Figure 5. Alignment of 5'-terminal sequences of E4 mRNA to Ad2 DNA. The sequences of oligonucleotides 1 to 4 were aligned in arrows 1 to 4, respectively, to the Ad2 DNA sequence. The DNA sequence shown is nucleotide 271 to 430 from the righthand end of the α -strand (1). The position of the predicted RNA polymerase II recognition site, the TATA box, is shown within the boxed area. Two series of nucleotides that are capable of forming a hybrid with 3'-terminal nucleotides of 18S ribosomal RNA are shown with dots. The first potential translation initiation codon (AUG) is observed at nucleotide 406 to 408; however, we have no information on whether this is the true initiation codon, or whether it is spliced out, i.e. whether it is present within the intron.

5'-terminal pyrimidine nucleotides in mRNAs of Ad2 E1, E2, and E3 (15). Therefore, these U-termini could represent a sequence produced by mechanisms different from those used to generate other Ad2 early mRNAs. The nucleotide sequence of one of these 5'-terminal sequences, ACACUG, was recently reported by Baker and Ziff (17). Basically the same results in the sequences and mapping of the U-terminal T1 oligonucleotides have been obtained by these authors (Baker and Ziff, manuscript in preparation).

We previously reported that the 5'-terminal U was partially modified (15). Similar results were obtained by 2D-TLC analysis of the 5'-terminal nucleotides of RNase T1 oligonucleotides (see Fig. 2). The modified nucleotide (pN^{*m}) is similar if not identical to pU^{*m} (U^* is modified U) reported in Moloney murine leukemia virus primer tRNA (24), on the basis of similar Rf values by 2D-TLC.

We have obtained at least six polypeptides by cell-free translation of E4 mRNA (30). Three to six mRNAs that have different splicing patterns have been mapped in the E4 region (2-4). Important questions are (i) how the U-terminus is produced; (ii) whether mRNA with a U terminus is translatable, and (iii) how do the heterogeneous 5'-termini correspond to the heterogeneous mRNA species?

Recently, heterogeneous 5'-terminal sequences have been identified in both early and late SV40 mRNA (31-34) and in polyomavirus late mRNA (35). Thus, there may be general features of transcription initiation or processing of precursor mRNA which may produce heterogeneous 5'-terminal sequences of adenovirus, SV40, and polyomavirus mRNA. The multiple 5'-ends of SV40 late mRNA were shown to be dispersed over large areas on SV40 DNA, and thus cannot be simply explained by a stuttering of RNA polymerase II.

ACKNOWLEDGEMENTS

We thank Dr. H. Thornton for generous assistance with cell culture and virus infection, Drs. D. Sciaky and R. Roberts for making their adenovirus 2 DNA sequence available to us prior to publication, and Carolyn E. Mulhall for editorial assistance.

This work was supported by Public Health Service Grant 5 R01 AI-01725 to MG, and by Core Grant 1 P30 CA-25843 from the NIH. MG is the recipient of Research Career Award 5 K06 AI-04739 from the NIH.

REFERENCES

1. Shinagawa, M., Padmanabhan, R.V., and Padmanabhan, R. (1980) *Gene* **9**, 99-114.
2. Chow, L.T., Broker, T.R., and Lewis, J. (1979) *J. Mol. Biol.* **134**, 265-303.
3. Kitchingman, G.R. and Westphal, H. (1980) *J. Mol. Biol.* **137**, 23-48.
4. Berk, A.J. and Sharp, P.A. (1978) *Cell* **14**, 695-711.
5. Green, M. (1962) *Cold Spring Harbor Symp. Quant. Biol.* **27**, 219-235.
6. Green, M., Parsons, J.T., Pina, M., Fujinaga, K., Caffier, H., and Landgraf-Leurs, M. (1970) *Cold Spring Harbor Symp. Quant. Biol.* **35**, 803-818.

7. Flint, S.J. and Broker, T.R. (1980) In DNA Tumor Viruses, Tooze, J., Ed., pp. 443-546, Cold Spring Harbor Laboratory.
8. Berk, A.J. and Sharp, P.A. (1977) *Cell* 12, 45-55.
9. Evans, R.M., Fraser, N., Ziff, E., Weber, J., Wilson, M., and Darnell, J.E. (1977) *Cell* 12, 733-739.
10. Wilson, M.C., Fraser, N.W., and Darnell, J.E. (1979) *Virology* 94, 175-184.
11. Sehgal, P.B., Fraser, N.W., and Darnell, J.E. (1979) *Virology* 94, 185-191.
12. Hashimoto, S. and Green, M. (1976) *J. Virol.* 20, 425-435.
13. Hashimoto, S. and Green, M. (1979) *Virology* 94, 254-272.
14. Hashimoto, S., Pursley, M.H., Wold, W.S.M., and Green, M. (1980) *Biochemistry* 19, 294-300.
15. Hashimoto, S. and Green, M. (1980) *J. Biol. Chem.* 255, 6780-6788.
16. Baker, C.C., Herisse, J., Courtois, G., Galibert, F., and Ziff, E. (1979) *Cell* 18, 569-580.
17. Baker, C.C. and Ziff, E. (1980) Cold Spring Harbor Symp. Quant. Biol. 44, 415-428 (1980).
18. Landgraf-Leurs, M. and Green, M. (1971) *J. Mol. Biol.* 60, 185-202.
19. Büttner, W., Veres-Molnar, Z., and Green, M. (1976) *J. Mol. Biol.* 107, 93-114.
20. Brownlee, G.G. and Sanger, F. (1969) *Eur. J. Biochem.* 11, 395-399.
21. Nishimura, S. (1972) In Progress in Nucleic Acid Research and Molecular Biology, Davidson, J.N. and Cohn, W.D., Eds., Vol. 12, pp. 50-85, Academic Press, New York.
22. Lockard, R.E. and RajBhandary, U.L. (1976) *Cell* 9, 746-760.
23. RajBhandary, U.L. and Chang, S.H. (1968) *J. Biol. Chem.* 243, 598-608.
24. Harada, F., Peters, G.G., and Dahlberg, J.E. (1979) *J. Biol. Chem.* 254, 10979-10985.
25. Jay, E., Bambara, R., Padmanabhan, R., and Wu, R. (1974) *Nucl. Acids Res.* 1, 331-353.
26. Fujimoto, M., Kunitaka, A., and Yoshino, H. (1974) *Agr. Biol. Chem.* 38, 785-791.
27. Galibert, F., Sedat, J., and Ziff, E. (1974) *J. Mol. Biol.* 87, 377-407.
28. Hagenbüchle, O., Santer, M., Steitz, J.A., and Mans, R.J. (1979) *Cell* 13, 551-563.
29. Rosenberg M. and Court, D. (1979) In Annual Review of Genetics, Roman, H.L., Ed., Vol. 13, pp. 319-354, Annual Reviews, Inc., Palo Alto.
30. Matsuo, T., Hashimoto, S., Wold, W.S.M., and Green, M. Submitted for publication.
31. Canaani, D., Kahana, C., Kukamel, A., and Groner, Y. (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76, 3078-3082.
32. Haegeman, G. and Fiers, W. (1980) *J. Virol.* 35, 955-961.
33. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P., and Weissman, S.W. (1978) *J. Mol. Biol.* 126, 813-846.
34. Reddy, V.B., Ghosh, P.K., Lebowitz, P., Piatak, M., and Weissman, S.W. (1979) *J. Virol.* 30, 279-296.
35. Flavell, A.J., Cowie, A., Legon, S., and Kamen, R. (1979) *Cell* 16, 357-371.