Unusual heterogeneity of the 5'-termini of human adenovirus type 2 early region E2 mRNA

Shuichi Hashimoto* and Maurice Green

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

Received 17 April 1984; Revised and Accepted 12 November 1984

ABSTRACT

The 5'-terminal structures of human adenovirus type 2 (Ad2) early region 2 (E2) mRNA were investigated. The E2 transcription unit has several interesting properties, including the presence of a TATA-like box that matches the consensus sequence poorly, delayed transcription during early stages of infection, and a switch in promoter recognition late after infection. E2-specific RNA, 5'-labeled *in vitro* to high specific activity was analyzed. Purified E2 mRNA was digested with RNase A or RNase T1 and the resulting oligonucleotides were resolved by two dimensional paper electrophoresis-homochromatography. Remarkably, as many as sixteen 5'-terminal RNase A oligonucleotides were identified and their sequences were deduced. The most common 5'-termini in the RNase A digest were $p(m^6)A^mCp$, $p(m^6)A^m(m)Cp$, $pG^m(m)Cp$, and $p(m^6)A^mG(m)Cp$. Two RNase A oligonucleotides originated from the E4 promoter region, consistent with electron microscopic observations (1). The sequence encoding these potential initiation sites covered about 90 nucleotides. Eleven of the sequences of the 5'-terminal RNase A oligonucleotides were aligned with the Ad2 DNA sequence in the Ad2 E2 promoter region (2). If the heterogeneous termini in the E2 promoter region were generated by a process of transcription initiation, their existence cannot be explained by stuttering of RNA polymerase II. This suggests that the transcription of Ad early region 2 has features which differ from those of other Ad2 early gene transcription units. Perhaps this is due to the absence of a conventional TATA box which is believed to position the initiation site. Alternatively, it is conceivable that the E2 promoter represents an alternate class of RNA polymerase II promoters containing different signals with different requirements for activation and/or that an E1A gene product modifies transcription initiation.

INTRODUCTION

Human adenovirus type 2 (Ad2) early genes provide useful models for the study of eukaryotic gene organization, expression, and regulation (3,4). The Ad2 genome is a linear duplex DNA molecule of Mr 23 X 10⁶ (5). Ad2 genes are expressed in two major stages during productive infection of human (KB or Hela) cells, early and late (3). During early stages, six regions of the viral genome are expressed. These include E1A (map position (mp) 1.3-4.5), E1B (mp 4.6-11.1), E2 (mp 75-62), E3 (mp 76-86), and E4 (mp 99-91.5) and late region L1 (6). A major promoter has been localized at the

5'-boundary of each early region (7-10). Electron microscopic and S1 nuclease-gel mapping procedures have shown that each early region encodes families of partially overlapping spliced RNAs which appear to have common 5'-termini (1,11-12), although heterogeneous termini were reported for E2 mRNA by Kitchingman and Westphal (12).

Direct structural analysis of the 5'-termini of Ad2 mRNA is important for understanding transcription initiation and regulation. In previous studies, we have purified mRNA specific to each Ad2 early gene region and characterized the 5'-termini of the purified mRNA. We observed heterogeneous 5'-termini in E1A, E1B, and E4, and we were able to identify "major" termini from each region which we sequenced and mapped on the Ad2 DNA genome (13-15). Similar results were reported by Baker and Ziff (16). However, we failed to obtain sufficient amounts of the 5'-termini of E2 mRNA by digestion of purified 5'-³²P-labeled mRNA with RNase T1 for definitive sequence analysis (13). The most common 5'-terminal sequences of E2 mRNA appeared to be p(m⁶)A^mGp and p(m⁶)A^mG^(m)NGp; these were present in much smaller quantities than the major terminal structures of E1A, E1B and E4 mRNAs (13). Baker *et al.* (17) detected two 5'-terminal oligonucleotides in E2 mRNA by RNase A digestion of RNA hybridized to the E2 promoter region. They sequenced and mapped the two 5'-termini, AGAGAGC (major) and GAGAGC (minor) which are encoded at adjacent positions in Ad2 DNA (17). These sequences appear to be transcription initiation sites. Subsequently, Fire *et al.* (18) and Mathis *et al.* (19) transcribed E2 RNA using an *in vitro* transcription system. Analysis of the RNA runoff products indicated that the 5'-termini were heterogeneous. Furthermore, Mathis *et al.* (19) observed two sets of heterogeneous transcription initiation sites by S1 nuclease analysis of *in vivo* Ad2 E2 mRNA; the first set was mapped at the same region as the major 5'-terminus reported by Baker and Ziff (16) and the second set was mapped at about 25-30 nucleotides upstream of the first site. They identified a potential new TATA-like box at about 26 nucleotides upstream of the second set of initiation sites.

In our previous analysis of the 5'-terminus of E2 mRNA, we could not rule out the possibility that the 5'-terminus was heterogeneous because of the low levels of 5'-³²P-labeled RNA that were obtained (13). In this report, we describe the isolation and characterization of the 5'-termini of Ad2 E2 mRNA labeled *in vitro* with a 10- to 40-fold higher ³²P-radioactivity than obtained previously (13). We isolated as many as sixteen 5'-terminal

RNase A oligonucleotides and aligned these with the DNA sequences within the Ad2 E2 early promoter region (1).

MATERIALS AND METHODS

Cell culture and virus infection

Exponentially growing KB cells in suspension were infected with Ad2 and labeled in the presence of cycloheximide with [5,6-³H]Urd from 2 to 8 h postinfection in order to monitor Ad2 mRNA purification (21).

Cell fractionation and isolation of poly(A⁺) RNA

Polyribosomal RNA was isolated from 1 or 2 liters of Ad2 infected cells or uninfected cells (3 to 5 x 10⁵ cells per ml) and poly(A⁺) RNA was selected on poly (U)-Sephadex (21).

Ad2 DNA and restriction fragments

Ad2 DNA was isolated and EcoRI restriction fragments were prepared as described (22).

In vitro labeling of 5'-termini of poly(A⁺) RNA

The 5'-terminal m⁷Gppp in poly(A⁺) RNA from uninfected or Ad2 early infected cells was removed by periodate oxidation, β -elimination, and phosphatase treatment (15). To remove m⁷Gppp efficiently from poly(A⁺) RNA, these treatments were repeated twice. The mRNA molecules with m⁷Gppp removed were labeled at the new 5'-terminal N^m residues with [γ -³²P]ATP (1000 Ci/mmol) and polynucleotide kinase (15).

Isolation of Ad2 E2 specific RNA

Ad2 E2 RNA was purified by hybridization to the EcoRI-B fragment using two different protocols: by method (A), 5'-³²P-labeled poly(A⁺) RNA (50 μ g) was hybridized to Ad2 EcoRI-B (15-20 μ g) immobilized on nitrocellulose filters; by method (B), 5'-³²P-labeled poly(A⁺) RNA (50 μ g) was sequentially hybridized to excess amounts of EcoRI-C (10-20 μ g), EcoRI-D (10-15 μ g), and EcoRI-A (30-40 μ g), and the unhybridized RNA which should contain minimal quantities of E1, E3, and E4 RNA was hybridized to EcoRI-B (10-20 μ g).

RNase A or RNase T1 digestion and two dimensional fingerprinting

5'-³²P-labeled viral RNA and 100 μ g of carrier yeast tRNA were digested with 5 μ g of RNase A (Worthington) or RNase T1 (Sankyo Co., Ltd.) in 10 μ l of 20 mM Tris-HCl pH 7.5 at 37° for 2 h. The RNase digests were resolved by paper electrophoresis on cellulose acetate strips in the first dimension and homochromatography on DEAE-cellulose TLC plates in the second dimension (23). The TLC plates were exposed to X-ray film (Kodak, XR-5) with two Lightning-plus (Dupont) intensifying screens at -70° for 1 to 10 h.

Nuclease P1 analysis of the 5'-³²P-labeled nucleotides of hybridization purified RNA and RNase A oligonucleotides

The 5'-terminal nucleotides of hybridization purified, Ad2 early region mRNA were identified by nuclease P1 digestion of 5'-³²P-labeled RNA and two dimensional thin layer chromatography (2D-TLC) as described (15). Purified mRNA was digested in 10 μ l of 20 mM sodium acetate, pH 5.3, containing 50 μ g of carrier tRNA, with 2 μ g of nuclease P1 (Boehringer Mannheim) at 37° for 2h, mixed with pN^m markers (pA^m, p^m6A^m, pC^m, pU^m, pG^m, 0.2 A_{260nm} units each, P-L Biochemicals), and applied to cellulose TLC plates (10 x 10 cm, EM Laboratories, Inc.). 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 (15). TLC plates were autoradiographed for 1 to 7 days; marker spots were detected with a UV-lamp.

5'-terminal nucleotides of RNase digests were also identified by nuclease P1 digestion and 2D-TLC. RNase A or RNase T1 oligonucleotide spots observed on DEAE-cellulose TLC plates by autoradiography were eluted with 2M triethylamine bicarbonate. After removal of triethylamine bicarbonate by repeated evaporation, each sample was dissolved in 20 μ l of 20 mM sodium acetate, pH 5.3, and digested with 2 μ g of nuclease P1 at 37° for 2 h. A portion of each sample (5-10 μ l) was mixed with pN^m markers and applied to cellulose TLC plates (6 x 6 cm). Nucleotides were resolved and identified as described above.

RNase T1 or RNase T2 digestion of RNase A oligonucleotides

RNase A oligonucleotides identified on DEAE-cellulose plates by autoradiography were dissolved in 10-20 μ l of 20 mM Tris-HCl, pH 7.6, or 20 mM sodium acetate, pH 4.5, containing 100 μ g of tRNA, and digested with 5 μ g of RNase T1 (Sankyo Co. Ltd.) or 1 unit of RNase T2 (Sankyo Co. Ltd.), respectively. The reaction mixtures were incubated at 37° for 2 h, mixed with 5'-³²P-labeled marker oligonucleotides, resolved by 2D-electrophoresis-homochromatography as described above, and autoradiographed. The 5'-³²P-labeled oligonucleotide markers, ³²pACp, ³²pAACp, ³²pAAACp, ³²pAUp, ³²pAAUp and ³²pAAAUp were prepared by digestion of 5'-³²P-labeled RNA with RNase A and RNase T1 and purification by 2D-homochromatography. ³²pGCp and ³²pGUp were prepared by labeling of GCp and GUp, respectively, with [γ -³²P]ATP and polynucleotide kinase, and purification by 2D-homochromatography (24). ³²pU^mUp and ³²pU^mU^mUp were prepared by digestion of ³²pU^mU^(m)UUACACUGp (prepared from E4 mRNA) with RNase T2, and purification by 2D-homochromatography (14). ³²pAGp was prepared by digestion of 5'-³²P-labeled RNA with RNase T1 and purification by 2D-homochromatography (24).

RESULTS

Purity of 5'-³²P-labeled Ad2 E2-specific mRNA

Ad2 early mRNA was purified by hybridization to restriction DNA fragments after *in vitro* labeling of the 5'-termini of poly(A⁺) RNA with [γ -³²P]ATP using polynucleotide kinase (15). Since poly(A⁺) RNA had been labeled with [³H]Urd *in vivo*, ³H-radioactivity was used to monitor the purification of Ad2 RNA. Ad2 E2 mRNA was purified by two different hybridization protocols: by method (A), 5'-³²P-labeled RNA was hybridized to EcoRI-B (encodes E2); by method (B), 5'-³²P-labeled RNA was hybridized sequentially to excess EcoRI-C (encodes E4), then to excess EcoRI-D (encodes E3), then to excess EcoRI-A (encodes E1), and then finally to EcoRI-B. E2 mRNA purified by method (B) should not be significantly contaminated with E1, E3 or E4 mRNAs. The E2 mRNA purified by method (A) was 1.64% and that by method (B) was 1.4% of the input poly(A⁺) RNA. The total Ad2 early RNA that hybridized to Ad2 DNA was 7-7.5% of the input poly(A⁺) RNA. The Ad2 early RNA purified by successive hybridization to EcoRI-C, -D, -A and -B was 3.80, 0.81, 0.62, and 1.4% respectively, at each step. Each EcoRI fragment specific RNA preparation was digested with RNase T1, fingerprinted, and the 5'-terminal nucleotide of each RNase T1 oligonucleotide was determined (see below). Our analysis showed that each fragment-specific RNA contained the same 5'-terminal RNase T1 oligonucleotides as reported previously, except for EcoRI-B specific RNA (13). In the experiments reported here, we observed many more 5'-terminal RNase T1 oligonucleotides in EcoRI-B specific RNA. In control experiments, 5'-³²P-labeled poly(A⁺) RNA from uninfected KB cells was successively hybridized to the EcoRI fragments as described above. Only about 0.05% of the input RNA was retained, presumably nonspecifically, to the EcoRI-B fragment. Since 1.4-1.6% of the poly(A⁺) RNA from Ad2 infected cells was EcoRI-B specific, contamination of EcoRI-B specific RNA with cellular RNA was only about 3%. Furthermore, as described below, we did not detect 5'-terminal 2'-O-methyl nucleotides (³²pN^m) in the RNA retained nonspecifically.

Identification of 5'-terminal nucleotides (³²pN^m) in Ad2 E1, E2, E3, and E4 specific RNA

5'-³²P-labeled RNA preparations were hybridized separately to EcoRI-A, EcoRI-B, EcoRI-C and EcoRI-D fragments, digested with nuclease P1, and the 5'-terminal nucleotides were identified. The 5'-terminal nucleotide (³²pN^m) identified in each RNA fraction was: EcoRI-A (pA^m, pm⁶A^m), EcoRI-B (pA^m, pm⁶A^m, pG^m, pC^m, pU^m), EcoRI-C (pA^m, pm⁶A^m, pU^m), EcoRI-D (pA^m, pm⁶A^m, pG^m,

Figure 1: Fingerprints (1 and 2) and diagrams (3 and 4) of RNase A digests of 5'-³²P-labeled Ad2 E2 specific mRNA purified by two different protocols.

Fingerprints: about 50 µg of 5'-³²P-labeled poly (A⁺) RNA (5.4 X 10⁷ cpm) were annealed to 15-20 µg of Ad2 EcoRI-B fragment on a filter by two different protocols, (A) and (B) (see Materials and Methods). Hybridized RNA was eluted [4.5 X 10⁵ cpm by method (A) and 2.2 X 10⁵ cpm by method (B), respectively], precipitated by alcohol together with 100 µg of carrier tRNA, digested with 5 µg of RNase A, 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 (23). Spots designated by numbers were eluted and 5'-terminal nucleotides were identified by nuclease P1 digestion and 2D-TLC. B and Y are the positions of xylene cyanol FF and methyl orange, respectively. To

visualize spots 25 and 26, plate 2 in this figure was exposed longer, and thus some spots in the plate were overexposed. Exposure to X-ray film was 1 h for plate 1 and 2.5 h for plate 2. Diagrams: the diagrams of plate 1 (method A) and plate 2 (method B) are shown in plate 3 and plate 4, respectively. The numbers assigned to the spots correspond to those in the fingerprints and in Table I. The RNase A oligonucleotides derived from 5'-termini of mRNA (pN^m is present at the 5'-terminus) are denoted by hatches or dots.

and very slight amounts of pC^m) (data not shown). In addition to these 2'-O-methyl nucleotides, unmethylated nucleotides (^{32}pA , ^{32}pC , ^{32}pU , ^{32}pG) derived from internal nicks were identified in each RNA. Only unmethylated nucleotides (pA , pC , pU , pG) were identified in the analysis of uninfected cell RNA retained nonspecifically to EcoRI-B.

Fractionation of 5'-terminal ^{32}P -labeled RNase A oligonucleotides derived from EcoRI-B specific mRNA and identification of 5'-terminal pN^m

5'- ^{32}P -labeled EcoRI-B specific RNAs were purified by two different protocols, digested with RNase A, and fingerprinted by 2D-electrophoresis-homochromatography, as shown in Fig. 1. Plates 1 and 2 are fingerprints of EcoRI-B specific RNA purified by methods (A) and (B), respectively; plates 3 and 4 show the diagrams of plates 1 and 2, respectively. The presence of 2'-O-methyl nucleotides is deduced from Fig. 2 (see below). Basically the same fingerprint pattern was obtained for the two RNA preparations.

Oligonucleotides in Fig. 1 could derive from authentic 5'-termini in mRNA or from internal nicks. Authentic 5'-termini should be capped, i.e. contain 2'-O-methyl nucleotides, whereas internal nicks should not. To identify authentic 5'-termini, each 5'-terminal nucleotide of each spot visible in Fig. 1 was analyzed by nuclease P1 digestion and 2D-TLC. Fig. 2 shows that 2D-TLC analysis of the spots that had 2'-O-methyl nucleotides (pN^m) at the 5'-termini (although some 2'-O-methyl nucleotides are faint in the photograph, these were clearly and reproducibly detected in the original film). Many of the spots in Fig. 1 were mixtures of oligonucleotides derived from authentic 5'-termini which have 5'-terminal 2'-O-methyl nucleotides (pN^m) and from internal nicks which have 5'-terminal unmethylated nucleotide (pN). As we have previously shown (13-15), oligonucleotides containing methylated components (both 2'-O-methyl and base methylated oligonucleotides) do not migrate differently from unmethylated oligonucleotides by 2D-electrophoresis-homochromatography when the nucleotide sequence or composition is the same. We did not obtain major 5'-terminal oligonucleotides in our analysis of E2 mRNA purified by two different protocols, i.e. all were at about the same concentration. In contrast, we previously

Figure 2: Nuclease P1 analysis of ^{32}P -labeled 5'-terminal nucleotides of RNase A oligonucleotides derived from Ad2 E2 mRNA. Spots 1 to 25 in Fig. 1 were eluted, digested with nuclease P1, and 5'-terminal nucleotides identified by 2D-TLC using pN^{m} and pN markers. The analysis of the spots contain pN^{m} in plate 1 of Fig. 1 is shown. The plate numbers correspond to the spot numbers shown in Fig. 1. Similar results were obtained in the analysis of the spots in plate 2 except that the contents of 2'-O-methyl nucleotides of spots 25 and 26 decreased, and those of spots 8 and 9 were undetectable. Resolution of spot 27 from spot 22 in plate 1 was not adequate. Therefore, the 5'-terminal analysis of spot 27 shown here is that of plate 2 in Fig. 1. Although resolution of pG^{m} from pC in these plates is poor (designated by $\text{pC} + \text{pG}^{\text{m}}$), the nucleotides detected in spots 8, 10, 16, and 22 are concluded to be pG^{m} because of their RNase A specificity.

detected major 5'-terminal oligonucleotides in Ad2 E1A, E1B, and E4 mRNA, although some heterogeneous termini were also present (13-15). This suggests that the heterogeneity of the 5'-termini of E2 mRNA is much higher and of a different nature than that of other Ad2 early region mRNAs.

As many as sixteen RNase A oligonucleotides contained 5'-terminal 2'-O-methylnucleotides, although several of the sequences were reduced in amount in mRNA purified by method (B). The nucleotide sequence of most oligonucleotides could be deduced from 5'-terminal nucleotide analysis, co-chromatography with marker oligonucleotides, enzyme specificity, and relative mobility of the spots (25, 26) (see legend to Fig. 2 and Table 1). Several oligonucleotides (spots 2, 3, and 10) appeared to be most common in both mRNA preparations. The sequences of these spots listed in Table I are: spot 2, $p(m^6)A^mCp$; spot 3, $p(m^6)A^mA^{(m)}Cp$; and spot 10, $p(m^6)A^mG^{(m)}Cp$ and $pG^mA^{(m)}Cp$. In addition, spot 27, pC^mU^mUp and $p(m^6)A^mU^mUp$, appeared to be the next most common sequences. The sequences pU^mUp (spot 25) and pU^mU^mUp (spot 26) that possess sequences common to the 5'-termini of E4 mRNA were reduced in amount in the mRNA purified by method (B). Therefore, at least some of these sequences appear to be derived from the E4 promoter region. In addition, pG^mCp (spot 8) and $p(m^6)A^mA^{(m)}Up$ (spot 9) are found in mRNA prepared by method (A) but not by method (B), a more stringent hybridization procedure. These sequences are not found in E4 mRNA 5'-termini and are probably derived from another promoter region. The E2 mRNA 5'-terminal sequences that were reported previously (17), AGAGAGC and GAGAGC, were detected in only small amounts in our fingerprints. Our spot 20 may contain $p(m^6)A^mG^{(m)}AGAGCp$.

Fractionation and Identification of 5'-terminal ^{32}P -labeled RNase T1 Oligonucleotides derived from EcoRI-B specific mRNA

The extreme heterogeneity of the 5'-termini of E2 mRNA observed in the RNase A digests was confirmed by analysis of RNase T1 oligonucleotides derived from 5'- ^{32}P -labeled E2 mRNA. Ad2 E2 mRNA purified by method (B) was digested with RNase T1 and resolved by 2D-homochromatography. The RNase T1 oligonucleotides were eluted, digested with nuclease P1, and the 5'-terminal nucleotides were identified by 2D-TLC (data not shown). At least fifteen RNase T1 oligonucleotides contained 2'-O-methylnucleotides at the 5'-termini (data not shown). At least eleven oligonucleotides contained $p(m^6)A^m$, two of which contained pU^m and pC^m as well (see 27). The relative migration by 2D-homochromatography of the two spots that contained pU^m appeared to be almost the same as $pU^mU^{(m)}UUACACUGp$ and $pU^mU^{(m)}UACACUGp$ which are major

Table I Identification of 5'-terminal nucleotides and sequence deduction of RNase A oligonucleotides derived from Ad2 E2 mRNA

Spot No. ^a	5'-Terminal Nucleotide Identified by 2D-TLC ^b	Nucleotide Length ^d	Deduced Sequence	Deduced Bonafide 5'-Terminal Sequence ^e
1	pC	1	pCp	
2	p(m ⁶)A ^m , pA	2	pACp	p(m ⁶)A ^m Cp
3	p(m ⁶)A ^m , pA	3	pAACp	p(m ⁶)A ^m A ^(m) Cp
4	p(m ⁶)A ^m , pA	4	pAAACp	p(m ⁶)A ^m A ^(m) ACp
5	pA	5	pAAAACp	
6 ^c	pN* ^m	2		(pN* ^m Cp)
7	p(m ⁶)A ^m , pA	2	pAUp	p(m ⁶)A ^m Up
8	pG ^m , pG	2	pGCp	pG ^m Cp
9	p(m ⁶)A ^m , pA, pG	3	pAAUp	p(m ⁶)A ^m A ^(m) Up
10	p(m ⁶)A ^m , pG ^m , pA, pG	3	p(AG)Cp	p(m ⁶)A ^m G ^(m) Cp + pG ^m A ^(m) Cp
11	pA	4	pAAAUp	
12	pA, pG	4	p(AAG)Cp	
13	pA	5		
14	pA	6		
15	p(m ⁶)A ^m , pA, pG	3	p(AG)Up	p(m ⁶)A ^m G ^(m) Up
16	pG ^m , pA, pG	4	p(AAG)Up	pG ^m A ^(m) AUp
17	pA	5		
18	pA	5 or 6		
19	pA	6		
20	p(m ⁶)A ^m , pA	6 or 7		
21	pU	1	pUp	
22	pG ^m , pG	2	pGUp	pG ^m Up
23	pG ^m , pG	3	pGGUp	pG ^m G ^(m) Up
24	pA, pG	4		
25	pU ^m	2		pU ^m Up
26	pU ^m	3		pU ^m U ^m Up
27 ^f	pC ^m , pm ⁶ A ^m	3		pC ^m U ^m Up + pm ⁶ A ^m U ^m Up

a. Spot numbers correspond to those of Fig. 1. Spots 2, 3, and 10 which are underlined contain the most common 5'-terminal sequences of E2 mRNA.

b. The oligonucleotide spots from Fig. 1 were eluted, digested with nuclease P1, and analyzed by 2D-TLC as described (20). The analysis is shown in Fig. 2.

c. pN^{*m} (spot 6) did not migrate with pN^m and pN markers on 2D-TLC. As the migration of pN^m was close to that of pG^m , it could be a 2'-O-methyl modified guanylic acid (pG^{*m}). This sequence was not included in the total number of 5'-termini of E2 mRNA.

d. Nucleotide length of each oligonucleotide was estimated by the relative mobility of oligonucleotides and by co-chromatography with marker oligonucleotide using 2D-electrophoresis-homochromatography.

e. The nucleotide sequences of most spots observed in Fig. 1 were deduced by 5'-terminal analysis, RNase A specificity, co-chromatography with marker oligonucleotides, and RNase T1 digestion. The nucleotide sequences of bonafide 5'-termini were deduced by 5'-terminal analysis and from the nucleotide sequence of the spot which included the bonafide 5'-terminal oligonucleotide. The sequences of spots 25 and 26 were deduced by the 5'-terminal analysis and co-chromatography with $^{32}pU^mUp$ and $^{32}pU^mU^mUp$ markers, respectively.

f. Spot 27 is a mixture of pC^mU^mUp and $p(m^6)A^mU^mUp$, because (i) the 5'-terminal nucleotides are pC^m and $p(m^6)A^m$; (ii) the migration of pC^m and $p(m^6)A^m$ is almost the same in the first dimension (14); and (iii) the spot migrated faster than $pAUp$ or $pAGUp$ in the first dimension, thus the sequences must be more U-rich than $p(m^6)A^mUp$, $p(m^6)A^mG^mUp$, pC^mUp , pC^mG^mUp , which suggests that this spot contains trinucleotides.

5'-termini of E4 mRNA (14). These spots were present at much higher levels in the RNase T1 fingerprints of E2 mRNA prepared by method (A) than by method (B). Therefore, the sequences that initiate with pU^m in E2 mRNA may be derived from the E4 promoter regions. Since only short leader sequences are present in E2 mRNA that would hybridize to the E4 region (1), pre-hybridization to the EcoRI-C fragment may not be sufficient to eliminate these 5'-terminal sequences completely.

DISCUSSION

The 5'-termini of E2 mRNA have been mapped at mp 75.1 on the Ad2 DNA genome by electron microscopy and S1 nuclease-gel mapping procedures (1,11, 12). Minor 5'-termini of E2 mRNA were also mapped at coordinate 99.1 in the E4 promoter by electron microscopic visualization (1). Baker *et al.* (17) isolated 5'-terminal oligonucleotides by RNase A digestion of E2 selected mRNA. The sequences they identified were AGAGAGC (major) and GAGAGC (minor) at nucleotides 1458 and 1457 (nucleotide numbers refer to the Ad2 EcoRI-F fragment DNA), respectively, at about mp 75.1; the transcription initiation sites deduced from these sequences are shown in lane 1 of Fig. 3 as a circle (major nucleotide) and a dot (minor nucleotide). Mathis *et al.* (19) analyzed transcription initiation sites of the Ad2 E2 promoter region using S1 nuclease analysis. They observed two sets of heterogeneous 5' ends, one close to the same site as mapped by Baker *et al.* (17) (at nucleotides 1456, 1457, 1458, 1459, 1460, and 1461), and the second at about 25-30 nucleotides

Figure 3: Mapping of the 5'-terminal RNase A or RNase T1 oligonucleotides of Ad2 E2 mRNA in the E2 promoter region of Ad2 DNA.

DNA sequence and nucleotide number (left hand end of EcoRI-F fragment is designated as nucleotide 1) were determined by Galibert *et al.* (2). The nucleotide sequence of the sense strand of the E2 promoter region and mapping of the 5'-termini of *in vivo* (16,17) and *in vitro* (19) E2 mRNA as well as the 5'-termini of E2 mRNA that could be mapped in this report are shown (see Text). The nucleotide sequence of the sense strand of the E2 promoter region is shown. The positions of the A-terminus (major) and the G-terminus (minor) are shown with a circle and a dot, respectively. Lane (2) shows the 5'-termini of RNA transcribed from the E2 promoter region using an *in vitro* transcription system by Mathis *et al.* (19). The major terminal nucleotides are shown with circles and the minor terminal nucleotides are shown with dots.

upstream of the first set (at nucleotides 1484 and 1485). A candidate TATA-like sequence for the second set of initiation sites was predicted at nucleotide 1516 (19). The putative transcription initiation sites detected by Mathis *et al.* (19) are shown as circles and dots in lane 2 of Fig. 3. In addition, Elkaim *et al.* (28) found heterogeneous transcription initiation sites for both *in vivo* and *in vitro* generated E2 mRNA as analyzed by S1 nuclease. The exact positions of these heterogeneous initiation sites observed by these workers are uncertain, because the S1 nuclease procedure does not permit a precise identification of 5'-terminal nucleotides, nor does it show that the nucleotide is capped (29,30).

In the present study, we have directly analyzed the terminal RNase A and RNase T1 oligonucleotides of chemically decapped Ad2 E2 mRNA. By the analysis of increased amounts of 5'-³²P-labeled RNA, we have detected reproducibly a large number of additional 5'-termini. To eliminate the possibility that our results might reflect contamination with other viral early mRNA sequences, we have used a subtractive hybridization procedure to

further purify E2 mRNA prior to analysis. Our data suggest that the transcription of Ad early region 2 has features which differ from those of other Ad2 early gene transcription units.

We have isolated as many as sixteen RNase A 5'-terminal oligonucleotides and at least fifteen RNase T1 structures. Among these heterogeneous termini, $p(m^6)A^mCp$, $p(m^6)A^mA^{(m)}Cp$, $p(m^6)A^mG^{(m)}Cp$, and $pG^mA^{(m)}Cp$ in the RNase A digest and $p(m^6)A^mGp$, $p(m^6)A^mU^{(m)}Gp$, and two $p(m^6)A^mN^{(m)}NGps$ in the RNase T1 digest appeared to be the most common in the mRNA purified by a hybridization procedure designed to remove possible contamination with RNA from other early regions. When E2 mRNA was purified without pre-hybridization to the *EcoRI*-C fragment which encodes E4, pU^mUp , pU^mU^mUp , and pG^mCp were also among the common sequences in the RNase A digest. This probably reflects the minor 5'-termini of E2 mRNA mapped in E4 at coordinate 99.1 (1).

We were able to align eleven of the sixteen 5'-terminal RNase A oligonucleotides [isolated by method (B)] on the DNA sequence of the Ad2 E2 promoter region from nucleotides 1434 to 1518, as shown in Fig. 3 (see lane 3). Three possible RNA polymerase recognition sites, i.e. modified TATA boxes, are indicated by the dashed boxes (nucleotides 1486 to 1483, 1468 to 1463, and 1516 to 1511). Three sequences (GAC, GGU, and GAAU) detected in E2 mRNA isolated by both hybridization procedures were not found in the E2 promoter region near mp 75.1. We did not attempt to align two sequences (GC and AAU) observed only in E2 mRNA prepared by method (A) in the E2 promoter region; these two sequences possibly may be derived from other promoter regions.

The sequences that we deduced for these eleven RNase A 5'-terminal structures could be aligned at 29 sites in the promoter region (solid arrows in Fig. 3), since many 5'-terminal oligonucleotides were represented more than once. Each of these 29 sites is a potential transcription initiation site. Many of the potential sites of these termini are located in regions contiguous to the two clusters of initiation sites reported by Mathis et al. (19) (see lane 3 of Fig. 3). Six of these 29 sites have a high probability of being true transcription initiation sites, since their sequences were detected in both RNase A and RNase T1 oligonucleotides. These are shown as dashed arrows in Fig. 3. These six sequences, starting at nucleotide 1438, 1454, 1470, 1477, 1481, and 1491, most likely are bonafide transcription initiation sites, and appear to represent the more common RNase T1 oligonucleotide sequences $p(m^6)A^mGp$ and $p(m^6)A^mN^{(m)}NGp$. In addition, three sequences (AC, AAC, and AAAC) that were aligned at nucleotide 1494, 1495, and 1496, respectively, also appear to be bonafide transcription initiation

sites, because these sequences are unique to the DNA sequence of this promoter region. The exact map position of the six additional putative transcription initiation sites could not be defined, because of the short length of the RNase A oligonucleotides and the occurrence of some of these sequences at more than one site in the promoter region.

The finding of heterogeneous 5'-terminal cap structures several regions of the E2 promoter as described here appears to differ from the early reports of Baker *et al.* (16,17) which described one major and one minor cap structure at the same location. This disparity may be due in part to the use of different experimental conditions. For example, we analyzed 8 h infected cell RNA whereas they analyzed 5 h infected cell RNA. In addition, we analyzed RNA from polyribosomes whereas they (16,17) analyzed RNA from a post-mitochondrial cytoplasmic fraction. Moreover, larger amounts of radioactivity were analyzed in our present studies; this may facilitate the detection of multiple 5'-termini. Consistent with our findings, several laboratories have recently reported evidence for 2 or 3 sets of heterogeneous termini in E2 mRNA (19,28,31,32).

Heterogeneity of the 5'-termini of adenovirus early mRNAs has been demonstrated in the analysis of E1A, E1B, E3, and E4 mRNAs (13-16). Only two to five terminal sequences were obtained from each of these mRNA regions, and these could be aligned within the span of several nucleotides of each transcription initiation site. However, as shown here, the heterogeneity of the 5'-termini of E2 mRNA is much higher and basically different from those of transcripts from other adenovirus early genes. The E2 mRNA termini could be mapped within a large stretch of nucleotides, 90 nucleotides within the E2 promoter region. Genetic studies suggest that the TATA box functions to position the transcription initiation site ~25-30 nucleotides downstream. Perhaps the presence of a TATA-like box in E2 that matches the consensus sequence poorly accounts for the extreme heterogeneity in initiation. The situation with the Ad2 E2 region may be analogous to the late promoter of SV40 and polyoma, which has no TATA box and which also exhibits extreme heterogeneity of transcription initiation sites (33,34,35). On the other hand, the E2 promoter may represent an alternate class of RNA polymerase II promoters containing different signals with different requirements for activation.

ACKNOWLEDGEMENTS

This work was supported by grant CA29561 from the National Cancer

Institute. M.G. holds Research Career Award 5 K06 AI04739 from the National Institute for Allergy and Infectious Diseases. We thank Dr. H. Thornton for generous assistance with cell culture, Drs. W.S.M. Wold and G. Chinnadurai for critical reading of the manuscript, M.H. Pursley for technical support, and C.E. Mulhall for editorial assistance.

*Present address: Meiji Institute of Health Science, 540 Naruda, Odawara, 250 Japan

REFERENCES

1. Chow, L.T., Broker, T.R., and Lewis, J. (1979) *J. Mol. Biol.* **134**: 265-303.
2. Galibert, F., Hérisse, J., and Courtois, G. (1979) *Gene* **6**, 1-22.
3. Green, M., Parsons, J.T., Piña, M., Fujinaga, K., Caffier, H., and Langraf-Leurs, I. (1970) *Cold Spring Harbor Symp. Quant. Biol.* **35**, 803-818.
4. Green, M. (1978) *Perspect. Biol. Med.* **21**, 373-397.
5. Green, M. Piña, M., Kimes, R., Wensink, P.C., MacHattie, L.A., and Thomas, C.A. (1967) *Proc. Natl. Acad. Sci. U.S.A.* **57**, 1302-1309.
6. Pettersson, U. and Akusjärvi, G. (1983) in *Advances in Viral Oncology*, Klein, G., Ed., Vol. 3, pp. 83-131 Raven Press, New York.
7. Berk, A.J. and Sharp, P.A. (1977) *Cell* **12**, 45-55.
8. Evans, R.M., Fraser, N., Ziff, E., Weber, J., Wilson, M., and Darnell, J.E. (1977) *Cell* **12**, 733-739.
9. Wilson, M.C., Fraser, N.W., and Darnell, J.E. (1979) *Virology* **94**, 175-184.
10. Sehgal, P.B., Fraser, N.W., and Darnell, J.E. (1979) *Virology* **94**, 185-191.
11. Berk, A.J. and Sharp P.A. (1978) *Cell* **14**, 695-711.
12. Kitchingman, G.R. and Westphal, H. (1980) *J. Mol. Biol.* **137**, 23-48.
13. Hashimoto, S. and Green, M. (1980) *J. Biol. Chem.* **255**, 6780-6788.
14. Hashimoto, S., Pursley, M.H., and Green, M. (1981) *Nucl. Acids Res.* **9**, 1675-1689.
15. Hashimoto, S., Wold, W.S.M., Brackmann, K.H., and Green, M. (1981) *Biochemistry* **20**, 6640-6647.
16. Baker, C.C. and Ziff, E. (1981) *J. Mol. Biol.* **149**, 189-221.
17. Baker, C.C., Hérisse, J., Courtois, G., Galibert, F., and Ziff, E. (1979) *Cell* **18**, 569-580.
18. Fire, A., Baker, C.C., Manley, J.L., Ziff, E.B., and Sharp, P.A. (1981) *J. Virol.* **40**, 703-719.
19. Mathis, D.J., Klkaïm, R., Kédinger, C., Passone-Corsi, P., and Chambon, P. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7383-7387.
20. Hashimoto, S., Pursley, M.H., Wold, W.S.M., and Green, M. (1980) *Biochemistry* **19**, 294-300.
21. Hashimoto, S. and Green, M. (1976) *J. Virol.* **20**, 425-435.
22. Büttner, W., Veres-Molnar, Z., and Green, M. (1976) *J. Mol. Biol.* **107**, 93-114.
23. Brownlee, G.G. and Sanger, F. (1969) *Eur. J. Biochem.* **11**, 395-399.
24. Székely, M. and Sanger, F. (1969) *J. Mol. Biol.* **43**, 607-617.
25. Jay, E., Bambara, R., Padmanabhan, R., and Wu, R. (1974) *Nucl. Acids Res.* **1**, 331-353.
26. Domdey, H. and Gross, H.J. (1979) *Anal. Biochem.* **93**, 321-328.

27. The 5'-terminal RNase T1 oligonucleotides obtained were: $p(m^6)A^mGp$; $p(m^6)A^mU^mGp$; two $p(m^6)A^mN^mNGps$; $p(m^6)A^mN^m(N)_2Gp$; four $p(m^6)A^mN^m(N)_3Gps$; two $p(m^6)A^mN^m(N)_{5=6}Gps$; two $pC^mN^m(N)_{6=7}Gps$; two $pU^mN^m(N)_{6=7}Gps$, in which $p(m^6)A^mGp$, and $p(m)A^mU^mGp$, and two $p(m^6)A^mN^mNGps$ appear to be the most common sequences. N in the parenthesis is A, C, or U. $(N)_{5=6}$ is five or six nucleotide lengths. The relative migration by 2D-homochromatography of the two oligonucleotides initiated with pU^m were almost the same as the pU^m -terminated oligonucleotide of E4 mRNA (14).
28. Elkaim, R., Goding, C., and Kédinger, C. (1983) Nucl. Acids Res. 11, 7105-7117.
29. Green, M.R., and Roeder, R.G. (1980) Cell 22, 231-242.
30. Cowie, A., Jat, P., and Kamen, R. (1982) J. Mol. Biol. 159, 225-255.
31. Leff, T., Elkaim, R., Goding, C.R., Jalinot, P., Sassone-Corsi, P., Perricaudet, M., Kédinger, C., and Chambon, P. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 4381-4385.
32. Kingston, R.E., Kaufman, R.J., and Sharp, P.A. (1984) Mol. Cell. Biol. 4, 1970-1977.
33. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P., and Weissman, S.M. (1978) J. Mol. Biol. 126, 813-846.
34. Canaani, D., Kahana, C., Mukamel, A., and Groner, Y. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3078-3082.
35. Cowie, A., Tyndall, C., and Kamen, R. (1982) Nucl. Acids Res. 9, 6305-6322.