Ise, N., & Okubo, T. (1980) *Acc. Chem. Res. 13*, 303–309.
Ladbrooke, B. D., & Chapman, D. (1969) *Chem. Phys. Lipids 3*, 304–356.
Lentz, B. R., Freire, E., & Biltonen, R. L. (1978) *Biochemistry 17*, 4475–4480.
Luna, E. J., & McConnell, H. M. (1977) *Biochim. Biophys. Acta 470*, 303–316.
Mabrey, S., Mateo, P. L., & Sturtevant, J. M. (1978) *Biochemistry 17*, 2464–2468.
Melchior, D. L., & Steim, J. M. (1979) *Prog. Surf. Membr. Sci. 13*, 211–296.
Miller, R. G. (1980) *Nature (London) 287*, 166–167.
Oldfield, E., & Chapman, D. (1972) *FEBS Lett. 23*, 285–297.
Owicki, J. C., & McConnell, H. M. (1980) *Biophys. J. 30*, 383–397.
Recktenwald, D., & McConnell, H. M. (1981) *Biochemistry 20*, 4505–4510.
Rubenstein, J. L. R., Smith, B. A., & McConnell, H. M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 15–18.
Rubenstein, J. L. R., Owicki, J. C., & McConnell, H. M. (1980) *Biochemistry 19*, 569–573.
Shimshick, E. J., & McConnell, H. M. (1973) *Biochemistry 12*, 2351–2360.
Smith, L. M., Parce, J. W., Smith, B. A., & McConnell, H. M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4177–4179.
Verkleij, A. J., Mombers, C., Leunissen-Bijuelt, J., & Ververgaert, P. H. J. Th. (1979) *Nature (London) 279*, 162–163.
Wu, S. H. W., & McConnell, H. M. (1975) *Biochemistry 14*, 847–854.

Nucleotide Sequences of 5′ Termini of Adenovirus 2 Early Transforming Region E1a and E1b Messenger Ribonucleic Acids†

Shuichi Hashimoto,* William S. M. Wold, Karl H. Brackmann, and Maurice Green

ABSTRACT: The major 5′-terminal RNase T1 oligonucleotides derived from human adenovirus type 2 (Ad2) early regions E1a [map position (mp) 1.3–4.6] and E1b (mp 4.6–11.2) were characterized and sequenced. Poly(A+) polyribosomal RNA was isolated from Ad2-infected KB cell cultured 10 h in the presence of cycloheximide, the 5′-terminal m^7Gppp was removed, and the 5′-OH of the penultimate 2′-O-methylated nucleoside was labeled with [γ-^{32}P]ATP by using polynucleotide kinase. 5′-^{32}P-labeled poly(A+) RNA was hybridized to *Eco*RI-A (mp 0–58.5), *Hpa*I-E (mp 0–4.5), and *Hpa*I-C (mp 4.5–25) and to cloned fragments representing mp 0–4.5, 4.5–8, 8–10.7, 8–17, and 17–31.5. Fragment-specific mRNA was digested with RNase T1, the resulting oligonucleotides were resolved by two-dimensional paper electrophoresis–homochromatography, and the 5′-terminal methylated oligonucleotides were characterized by nuclease P1 digestion and two-dimensional thin-layer chromatography. Two major 5′-terminal RNase T1 oligonucleotides were obtained from *Eco*RI-A-specific mRNA; these were characterized by RNase T2 digestion and two-dimensional thin-layer chromatography and were sequenced by partial nuclease P1 digestion and electrophoresis–homochromatography. These two oligonucleotides were mapped to E1a and E1b, respectively, by analysis of E1a- and E1b-specific mRNA. Their sequences are as follows: E1a, m^7Gppp(m^6)A^{m}C$^{(m)}$UCUUGp; E1b, m^7Gppp(m^6)A^{m}C$^{(m)}$AUCUGp, in which the methylations shown in parentheses are partial. The E1a 5′-terminal oligonucleotide was found as a minor component of E1b mRNA, implying that mRNAs exist with both E1a and E1b sequences. Five other minor 5′-terminal oligonucleotides were observed in E1 mRNA, implying that transcription initiation is not always precise. Sequences corresponding to these two 5′ termini were found at nucleotide positions 498–504 (E1a) and 1699–1705 (E1b) in the Ad2 DNA sequence (T. Gingeras and R. Roberts, unpublished experiments), positions where the 5′-terminal region of E1a and E1b mRNAs has been mapped by nuclease gel analysis [Berk, A. J., & Sharp, P. A. (1978) *Cell (Cambridge, Mass.) 14*, 695–711] and electron microscopic visualization [Chow, L. T., Broker, T. R., & Lewis, J. (1979) *J. Mol. Biol. 134*, 265–303; Kitchingman, G. R., & Westphal, H. (1980) *J. Mol. Biol. 137*, 23–48]. E1a contains a TATA box at −24 to −31, a potential ribosome binding site at +22 to +32, and an ATG at +62 to +64. E1b contains a TATA box at −24 to −30, potential ribosome binding sites at +3 to +10 and +34 to +47, and ATGs at +13 to +15 and +317 to +319.

A major goal of eukaryotic molecular biologists is to understand the mechanism of transcription initiation, because this is central to the problem of gene regulation. The human adenovirus (Ads) are good models for such studies: many of the genes and their products have been well characterized, and much of the genome has been sequenced. As a prerequisite to understanding transcription initiation, it s necessary to localize, at the nucleotide level, exactly where transcription starts. Studies on the Ad2 major late transcription unit have indicated that transcription in vivo probably initiates at the cap site (Ziff & Evans, 1978). This also seems to occur in in vitro transcription systems (Weil et al., 1979; Manley et al., 1979). Therefore, the transcription initiation site of any transcription unit can be identified by sequencing the capped 5′ termini of RNA and then by locating the sequence within the DNA sequence.

In this paper, we have sequenced the major 5′ termini of mRNAs synthesized during early stages of Ad2 infection from Ad2 early region E1. E1 is located at map position (mp) 1.3–11.2, and E1 mRNAs are transcribed off the r the strand (Berk & Sharp, 1978; Chow et al., 1979; Kitchingman &

† From the Institute for Molecular Virology, St. Louis University Medical Center, St. Louis, Missouri 63110. *Received February 24, 1981.* This work was supported by U.S. Public Health Service Grant 9 R01 CA-29561. W.S.M.W. is the recipient of National Institutes of Health Research Career Development Award 1 K04 CA-00650. M.G. is the recipient of National Institutes of Health Research Career Award 5 K06 AI-04739.

0006-2960/81/0420-6640$01.25/0

Westphal, 1980). E1 is divided into E1a (mp 1.3–4.6) and E1b (mp 4.6–11.2). E1a has one promoter (Berk & Sharp, 1977; Evans et al., 1977; Wilson et al., 1979) and encodes three overlapping mRNAs, 13 S, 12 S, and 9 S (Spector et al., 1978; Esche et al., 1980), whose 5′ termini map at 1.3 and whose 3′ termini map at 4.6, but which have different internal regions spliced out. Polypeptides of approximately M_r 53 000 (53K) and 41K are coded by the 13S mRNA, those of 47K and 35K by the 12S mRNA, and those of 28K by the 9S mRNA (Spector et al., 1978; Esche et al., 1980; Halbert et al., 1979). E1b probably contains two promoters in the region from mp 4.6–11, one at mp 4.6, and the other at mp 10.2 (Wilson et al., 1979; Sehgal et al., 1979). The promoter at mp 4.6 initiates at least two mRNAs, 22 S and 13 S (Spector et al., 1978; Esche et al., 1980), whose 5′ and 3′ termini map at 4.6 and 11.2, respectively, but which have different section spliced out (Chow et al., 1979; Kitchingman & Westphal, 1980). The promoter at mp 10.2 initiates an unspliced 9S mRNA (Spector et al., 1978; 1978; Esche et al., 1980) whose 3′ terminus maps at 11.2 (Chow et al., 1978; Kitchingman & Westphal, 1980; Aleström et al., 1980). Polypeptides of M_r 53K, 19K, and 15K (protein IX) are coded by the 22S, 13S, and 9S mRNAs, respectively (Esche et al., 1980; Halbert et al., 1979; Aleström et al., 1980). Thus, E1 contains three promoters, and two of the promoters control the transcription of overlapping genes with different exons and introns. Another point of interest is that the E1a 9S mRNA, and the E1b 13S and 9S mRNAs, continues to be synthesized during late stages of infection (Spector et al., 1978; Esche et al., 1980), pointing to a differential regulation of these mRNAs with respect to the other mRNAs in the same transcription units. E1 is of further interest because it contains the sequences responsible for Ad2 cell transformation (Van der Eb et al., 1977; Gallimore et al., 1974).

In our previous studies on the 5′ termini of E1 mRNAs, we identified two major and one minor 5′-terminal RNase T1 oligonucleotides, all of which terminate in pA^m and pm^6A^m (Hashimoto & Green, 1980). In our present study, we have sequenced the two major 5′-terminal RNase T1 oligonucleotides, and we have assigned them to either E1a or E1b, by means of hybridizing E1 mRNAs to subfragments of E1, and by identifying the sequences of these termini within the DNA sequence. The DNA sequences proximal to the E1a and E1b capping sites contain "TATA boxes", possible "CCAAT boxes", possible ribosome binding sites, and possible translation initiation codons. The same conclusions on the 5′-terminal sequence and map positions of the major E1a and E1b mRNAs were recently published by Baker & Ziff (1980), although complete analysis of the methylated constituents was not provided. Our detailed description of our findings presented here provides independent confirmation of the genome location and 5′-terminal sequences of the Ad2 transforming region mRNAs. The methodology that we have independently developed for these analyses should be of general application for the sequence of 5′ termini of any capped eukaryotic mRNA present at low concentration in cellular RNA.

Materials and Methods

Cell Culture and Virus Infection. Exponentially growing KB cells in suspension were infected with 100–200 PFU/cell (plaque-forming units per cell) of Ad2 (Hashimoto & Green, 1979). After 1-h adsorption, cells were diluted 10-fold in medium containing 5% horse serum, and 25 μg/mL cycloheximide and labeled from 2- to 10-h postinfection with 10–20 μCi of [^{3}H]Urd ([5,6-$^3H_2$]uridine, 40–50 Ci/mmol, New England Nuclear) per mL. The [^{3}H]Urd label was used to monitor the purification of Ad2 mRNA.

Cell Fractionation and Isolation of Poly(A+) RNA. Polyribosomal RNA (2–5 mg) was isolated from 1 or 2 *l* of Ad2-infected cells [(3–5) × 10^5 cells/mL)], and poly(A+) RNA was selected on poly(U)–Sepharose (Pharmacia) (Hashimoto & Green, 1979).

Ad2 DNA and Restriction Enzyme Fragments. Ad2 DNA was isolated (Landgraf-Leurs & Green, 1971) and *Eco*RI restriction fragments were prepared on agarose gels as described (Büttner et al., 1976). *Eco*RI-A fragment (mp 0–58.5) was eluted from the agarose gel, redigested with *Eco*RI, reelectrophoresed, eluted, and immobilized to nitrocellulose filters (Schleicher & Schuell, Inc.) (Hashimoto & Green, 1980). *Eco*RI-A was digested with *Hpa*I, and the fragments were separated by agarose gel electrophoresis, eluted, and immobilized on nitrocellulose filters (Büttner et al., 1976).

In some experiments, Ad2 DNA restriction fragments cloned in pBR322 and amplified in *Escherichia coli* (strain HB101) were used to purify E1 mRNAs. The fragment descriptions and map positions are F1 (mp 0–4.5), F2 (mp 4.5–8.0), F3 (mp 8.0–10.7), F4 (mp 8.0–17.0), and F5 (mp 17.0–31.5). The cloned plasmids for F1, F2, and F3 were obtained from M. Wilson and J. Darnell (Rockefeller University), and those for F4 and F5 were obtained from S.-L. Hsu and J. Sambrook (Cold Spring Harbor Laboratories). Plasmids containing inserted Ad2 DNA fragments were grown in *E. coli* HB101 in M9 media in the presence of 50 μg/mL ampicillin. After growth of bacteria to A_{560nm} = 0.5, plasmids were amplified by the addition of 250 μg/mL chloramphenicol with incubation continued overnight. Plasmid DNA was isolated as described by Clewell & Helinski (1969). All plasmid DNA was checked for correct viral DNA insert by excision followed by agarose gel elelctrophoresis with a *Hin*dIII digest of Ad2 DNA as molecular weight markers. Purified plasmid DNAs were digested with *Eco*RI, denatured with alkali, and immobilized on nitrocellulose filters.

Labeling of 5′ Termini of Poly(A+) RNA. The 5′-terminal m^7Gppp in poly(A+) RNA from Ad2-infected KB cells was removed by periodate oxidation and β elimination and treatment with calf intestine alkaline phosphate (Boehringer Mannheim) (Hashimoto & Green, 1980; Hashimoto et al., 1980). The decapped mRNAs were labeled at their new 5′-terminal N^m residues, using [γ-^{32}P]ATP (100-500 Ci/mmol) and polynucleotide kinase (New England Biolabs) (Hashimoto & Green, 1980; Hashimoto et al., 1980). In the experiments reported here, the treatments for removal of m^7Gppp were done twice. This produced more internal nicks than previously reported (Hashimoto & Green, 1980), and therefore, the ratio of $^{32}pN^m$ to total ^{32}P incorporated decreased, although total incorporation was 2–3-fold higher.

Isolation of Ad2 DNA Fragment Specific mRNA. About 50 μg of in vitro 5′-^{32}P-labeled poly(A+) RNA from Ad2-infected KB cells was hybridized to Ad2 DNA fragments immobilized on a nitrocellulose filter, as described previously (Hashimoto & Green, 1980). The following amounts of DNA were immobilized to each filter: *Eco*RI-A, 30–40 μg; *Hpa*I-C, 20 μg; *Hpa*I-E, 10 μg; plasmid DNA, 40–80 μg.

RNase T1 Digestion and Fingerprinting. Hybridization-purified viral RNA and 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 °C for 1 h (Hashimoto & Green, 1980). The RNase T1 digest was separated by paper electrophoresis on cellulose acetate strips in the first dimension and homochromatography on DEAE-cellulose thin-layer chromatography (TLC) plates in the second di-

mension (Brownlee & Sanger, 1969). The TLC plates were exposed to X-ray film (Kodak, XR-5) with Lightning-plus (Du Pont) intensifying screens for 2–24 h.

Analysis of 5′-^{32}P-Labeled Nucleotides of RNase T1 Oligonucleotides. *RNase* T1 oligonucleotide spots on DEAE-cellulose TLC plates were identified by autoradiography, scraped, and eluted with 2 M triethylamine bicarbonate. The 5′-^{32}P-labeled RNase T1 fragments were dissolved in 30 μL of 20 mM sodium acetate (pH 5.3) and digested with 4 μg of nuclease P1 (Boehringer Mannheim); 5–15-μL aliquots were mixed with pN^m markers and applied to cellulose TLC plates (Hashimoto & Green, 1980). Two-dimensional TLC (2D-TLC) was performed with isobutyric acid, 0.5 N NH_4OH (5:3) in the first dimension, and isopropyl alcohol/concentrated HCl/H_2O (70:15:15) in the second dimension (Nishimura, 1972). Most pN and pN^m residues are separable in the TLC system (Hashimoto et al., 1980). TLC plates were exposed to X-ray film for 2–7 days with intensifying screens; marker spots were detected with UV lamp.

RNase T2 Digestion of RNase T1 Oligonucleotides. The RNase T1 oligonucleotides that derived from the 5′ termini of E1 mRNAs were dissolved in 10–20 μL of 20 mM sodium acetate buffer (pH 4.5) containing 100 μg of tRNA. One unit of RNase T2 (Sankyo Co., Ltd.) was added, and the reaction mixture was incubated at 37 °C for 2 h (Lockard & RajBhandary, 1976). The products were separated by two-dimensional homochromatography as described above. The spots, detected by autoradiography, were eluted and further analyzed by nuclease P1 digestion and 2D-TLC. The 5′-^{32}P-labeled oligonucleotide marker, ^{32}pACp, was prepared by digestion of 5′-^{32}P-labeled RNA with RNase A and purified by two-dimensional homochromatography.

Partial Digestion of RNase T1 Oligonucleotides with Nuclease P1 and Sequence Analysis. The 5′-terminal RNase T1 oligonucleotides eluted from TLC plates were dissolved in 20 μL of 20 mM sodium acetate buffer (pH 5.3) containing 100 μg of carrier yeast tRNA, mixed with 100 ng of nuclease P1 (Boehringer Mannheim), and incubated at room temperature. Aliquots were removed at various times, and the enzyme was inactivated by heating at 100 °C 5 min in 5 mM EDTA (Lockard & RajBhandary, 1976). The aliquots were pooled and mixed with a completely digested sample and with an undigested sample. The digests were separated by two-dimensional homochromatography, and the sequence was deduced by the migration of the spots detected by exposure to X-ray film (Lockard & RajBhandary, 1976). As a control for sequence analysis, yeast tRNA^{Phe} of known sequences (RajBhandary & Chang, 1968) was analyzed. The 5′ terminus of yeast tRNA^{phe} (Boehringer Mannheim) was labeled with ^{32}P as described (Hashimoto et al., 1980), and intact [5′-^{32}P]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 two-dimensional homochromatography as described above.

Results

Preparation of 5′-^{32}P-Labeled 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 labeling in vitro with ^{32}P were previously described (Hashimoto & Green, 1980). Basically, similar results were obtained in the present study except that the RNA-specific radioactivity was about 2–3-fold higher. Since ^{32}P was incorporated into both bonafide 5′ termini of mRNA and 5′ termini resulting from nicks in the RNA, labeled RNA was assayed routinely with nuclease P1 as described (Hashimoto et al., 1980) to estimate how much ^{32}P was incorporated into each 5′ terminus. Bonafide termini were easily distinguished from termini resulting from nicks, because in the former the ^{32}P radioactivity was in 2′-*O*-methyl nucleotides, whereas in the latter the radioactivity was in unmethylated nucleotides.

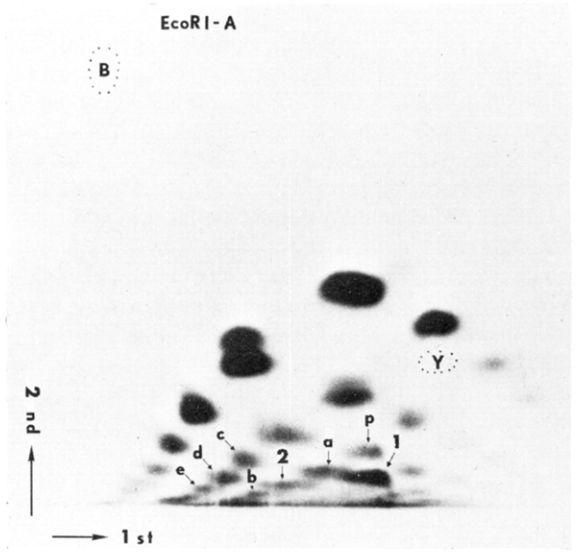

FIGURE 1: Fingerprint of RNase T1 digest of 5′-^{32}P-labeled *Eco*RI-A-specific Ad2 early mRNA. About 50 μg of 5′-^{32}P-labeled poly(A+) RNA from Ad2-infected KB cells was annealed to 30–40 μg of Ad2 *Eco*RI-A fragment immobilized on a nitrocellulose filter. The hybridized RNA was eluted, mixed with 100 μg of carrier tRNA, and precipitated with alcohol. The RNA was digested with 5 μg of RNase T1 and fingerprinted. The first dimension (from left to right) is electrophoresis on cellulose acetate paper, pH 3.5, and the second dimension (from bottom to top) is homochromatoggraphy. the spots were eluted, and the 5′-terminal nucleotide was analyzed. The spots indicated by arrows contain 5′-terminal sequences of the mRNA, and the 5′-terminal analysis of these spots is shown in Figure 2. B and Y are positions of xylene cyanol FF and methyl orange, respectively.

Purification of Total E1, E1a, and E1b mRNAs. About 50 μg of in vitro 5′-^{32}P-labeled poly(A+) RNA from Ad2-infected KB cells was hybridized to Ad2 restriction fragments immobilized on nitrocellulose filters. *Eco*RI-A was used to select total E1 mRNA, and *Hpa*I-E and *Hpa*I-C were used to select E1a and E1b mRNAs, respectively. Cloned fragments (F1, F2, and F3) were also used to select E1a and E1b mRNA. About 2% of total poly(A+) RNA hybridized to *Eco*RI-A; of this, 70–80% and 20–30% were E1a and E1b mRNAs, respectively. As described below, protein IX mRNA was not detected. Less than 0.2% of total poly(A+) RNA hybridized to F5 (mp 17.0–31.5). About 0.1% of input 5′-^{32}P-, [^{3}H]Urd-labeled RNA from uninfected cells was retained, presumably nonspecifically, to filters. As described previously (Hashimoto & Green, 1980), specific 5′-terminal RNase T1 oligonucleotides could not be detected in the cellular RNA that was retained on filters, supporting the idea that these are nonspecific sequences, and eliminating the posibility that these sequences compromised our analysis of the E1 5′ termini.

Identification of 5′-Terminal RNase T1 Oligonucleotides Derived from E1-Specific mRNA. A typical RNase T1 fingerprint of 5′-^{32}P-labeled *Eco*RI-A-specific mRNA is shown in Figure 1. The spots were eluted and digested with nuclease P1, and the 5′-terminal nucleotides were identified by TLC. As shown in Figure 2, spots 1, 2, and a–e contained pA^m as well as pm^{6}A^m. The presence of both pA^m and pm^{6}A^m in the same spot can be explained by assuming that not all mRNA molecules were methylated at the N^6 position. Spots 1 and 2 contained at least 90% of the pA^m and pm^{6}A^m residues and are designated by numbers. Spots a–e contained less than 10%

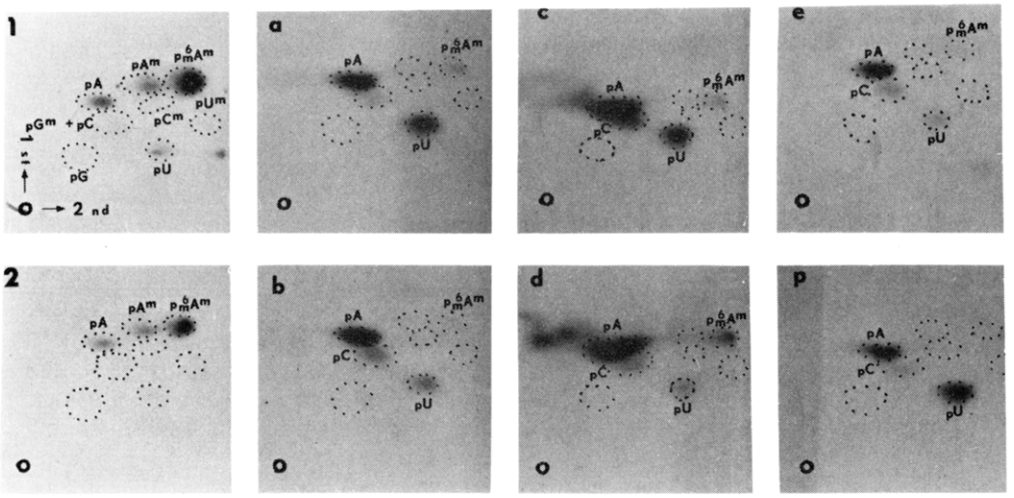

FIGURE 2: Identification of 2′-*O*-methyl nucleotides in 5′-terminal RNase T1 oligonucleotides derived from *Eco*RI-A-specific mRNA. Spots 1 and 2 (major 5′-terminal oligonucleotides) and a–e (minor 5′-terminal oligonucleotides), and P in Figure 1, were eluted and digested with nuclease P1, and the 5′-terminal nucleotides were identified by 2D-TLC using pN^m and pN markers. Variable amounts of unmethylated nucleotides, mainly pA, were observed in the nuclease P1 digest of spots a–e, reflecting background contamination with ^{32}P-labeled internal nicks in RNA. Spot P, which is at the location predicted from the known structure of the 5′ termini of protein IX mRNA (Alestöm et al., 1980), did not contain methylated nucleotides.

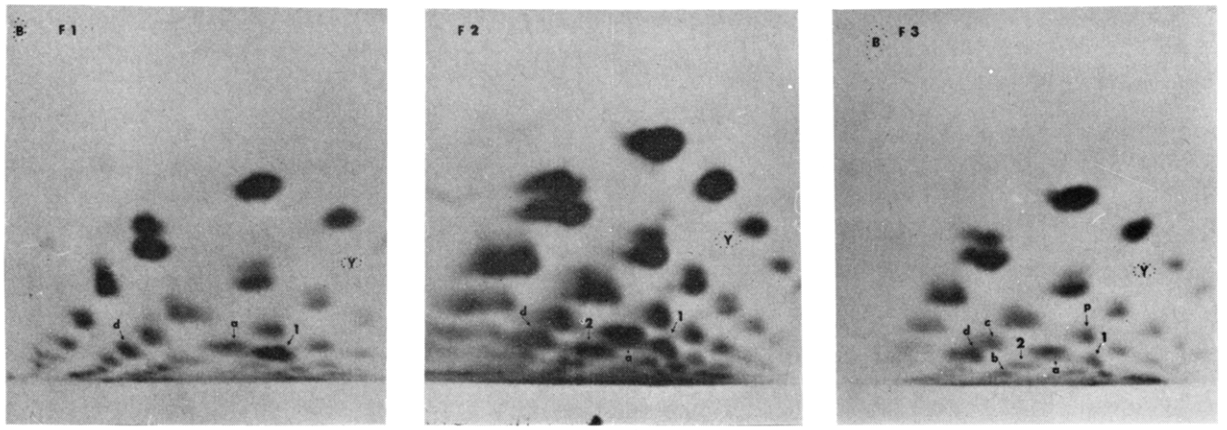

FIGURE 3: Fingerprint of RNase T1 digest of 5′-^{32}P-labeled E1a- and E1b-specific Ad2 early mRNA. About 50 μg of 5′-^{32}P-labeled poly(A+) RNA from Ad2-infected KB cells was annealed to 40–80 μg of plasmid DNA F1 (plate 1), F2 (plate 2), and F3 (plate 3) immobilized on nitrocellulose filters. F1, F2, and F3 contain map positions 0–4.5, 4.5–8.0, and 8.0–10.7 of Ad2 DNA, respectively. Hybridized RNA was eluted and precipitated with alcohol in the presence of 100 μg of carrier tRNA. The RNA was digested with 5 μg of RNase T1 and fingerprinted as described in Figure 1. The spots were eluted, and the 5′-terminal nucleotide was analyzed as described in Figure 2. With the exception of spot P, the spots indicated by arrows contained 2′-*O*-methyl nucleotides.

of the 2′-*O*-methyl nucleotides and are designated by letters. Although pm^6A^m in plates b and e is faint in this analysis (Figure 2), it was more pronounced in the analysis of different RNA preparations. The spot of pA was sometimes smeared when too much sample was loaded onto the TLC plate, as shown in plates c and d.

Figure 3 shows RNase T1 fingerprints of mRNA specific to F1 (mp 0–4.5), F2 (mp 4.5–8.0), and F3 (mp 8.0–10.7). Each spot was eluted, digested with nuclease P1, and analyzed by TLC; the spots shown by arrows contained pA^m and pm^6A^m. Spots 1, a, and d were found in F1 mRNA, spots 1, 2, a, and d in F2 mRNA, and spots 1, 2, and a–d in F3 mRNA. The correspondence between these spots and those in Figure 1 is indicated by their diagnostic position in the fingerprint and by the presence of pA^m and pm^6A^m. Spot 1 in F1 mRNA and spot 2 in F2 and F3 mRNA contain about 90% of the 2′-*O*-methyl nucleotides detected. Spot 1 represented 10–20% of the pA^m and pm^6A^m residues in F2 and F3 mRNA, and the lettered spots represented less than 10% in the F1, F2, or F3 mRNA. Similar results were obtained from analysis of the *Hpa*I-E- and *Hpa*I-C-specific mRNA. Analysis of F4 (mp 8.0–17.0) mRNA gave results that were similar to those from F3 mRNA. The radioactivity of F5 (mp 17.0–31.5) mRNA was too low to analyze the RNase T1 fragments. We conclude that spot 1 is the major 5′ terminus in E1a mRNA and that spot 2 is the major 5′ terminus in E1b mRNA. The presence of spot 1 in F3 mRNA probably reflects read-through from the E1a promoter into E1b. Since cloned restriction fragments were used to purify the RNA, contamination of the E1b fragment by E1a DNA sequences cannot be a problem.

It was difficult to map unequivocally the minor lettered 5′ termini because of low radioactivity and high contamination by unmethylated oligonucleotides. Spots a and d were observed in F1, F2, and F3 mRNA, and spots b and c were found in F3 mRNA. The possible origins of these spots will be discussed below.

The 5′-terminal sequence of the E1b 9S mRNA for protein IX (15K) has been shown to be pA^mU^mCUGp (Aleström et al., 1980). This oligonucleotide should be located at position P in our fingerprints (Hashimoto & Green, 1980; Jay et al., 1974; Domdey & Gross, 1979). As shown in Figure 2, spot P did not contain pA^m or pm^6A^m. Therefore, the protein IX 9S mRNA must be synthesized at levels too low to be detected under our infection conditions.

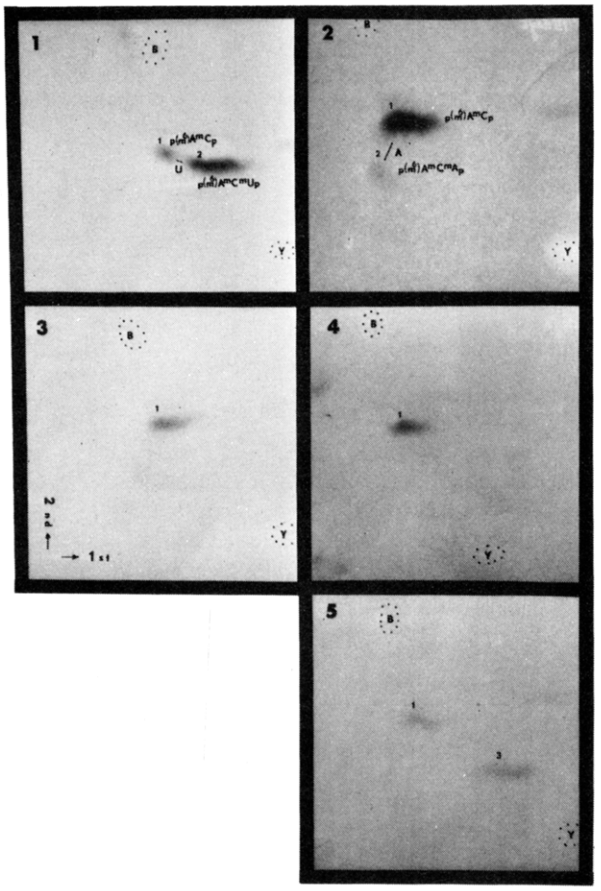

FIGURE 4: Characterization of spots 1 and 2 by RNase T2 and RNase A digestion and two-dimensional homochromatography. Plates 1 and 2: Oligonucleotide spots 1 (plate 1) and 2 (plate 2) in Figure 1 were eluted, digested with RNase T2, and resolved by two-dimensional homochromatography. The 5′-terminal nucleotides of the separated spots were shown to be pm^6A^m and pA^m by nuclease P1 digestion and 2D-TLC. Plates 3 and 4: Spot 1 [$p(m^6)A^mpCp$] of plates 1 and 2 was eluted, mixed with about the same number for counts of ^{32}pApCp marker, and resolved by two-dimensional homochromatography in plates 3 and 4, respectively. Plate 5: Oligonucleotide spot 2 in Figure 1 was digested with RNase A and resolved by two-dimensional homochromatography. The spot 2 used in this experiment is a different preparation from that of plate 2.

Characterization of the RNase T2 and RNase A Resistant Oligonucleotides of Spots 1 and 2. The 5′-terminal structures of RNase T1 spots 1 and 2 were characterized by RNase T2 digestion and two-dimensional homochromatography. As shown in Figure 4, two RNase T2 resistant oligonucleotides were obtained from both spot 1 (plate 1) and spot 2 (plate 2). We have previously shown (Hashimoto & Green, 1979; Hashimoto et al., 1980) that Ad2 early mRNA contains cap 1 (m^7GpppN^mpNp) and cap 2 (m^7GpppN^mpNp) structures at their 5′ termini; therefore, these two oligonucleotides must be pN^mpNp and pN^mpN^mpNp. These oligonucleotides were further characterized to determine their sequence. First, the 5′-terminal nucletides were shown to be pm^6A^m and pA^m by nuclease P1 digestion and 2D-TLC (not shown). Second, the spots labeled $p(m^6)A^mpCp$ in plates 1 and 2 were shown to cochromatograph with a ^{32}pApCp marker on two-dimensional homochromatography, as shown in plates 3 and 4, respectively; approximately the same amounts of radioactivity of ^{32}p-$(m^6)A^mpCp$ and the ^{32}pApCp marker were mixed prior to the two-dimensional homochromatography. These results indicate that the first two nucleotides of both spots are AC. The third base of cap 2 of spots 1 and 2 can be deduced by the relative migration of cap 1 (pN^mpNp) and cap 2 (pN^mpN^mpNp) structures on the two-dimensional homochromatogram shown in Figure 4 (the general rules for the interpretation of mobility shifts are described in the next section). The third base in the cap 2 structure could be an A, C, or U (G is ruled out by the results of RNase T1 digestion). Since only the loss of Up produces a shirt of the spot to the left, the third base of spot 1 must be U. The third base of spot 2 could be an A or C, based upon the relative mobility of cap 1 and cap 2. However, RNase A digestion did not produce spot 2 of plate 2 (plate 5 in Figure 4); rather, it produced a new spot (spot 3) which is possibly $p(m^6)A^mC^mAUp$. Therefore, the third base must be a purine, i.e., A. We conclude that the third base is U for spot 1 and A for spot 2. Therefore, the 5′-terminal sequences of spots 1 and 2 must be $p(m^6)A^mpC^{(m)}pUp$ and $p(m^6)A^mpC^{(m)}pAp$, respectively, in which the methylations shown in parentheses are partial. These sequences were confirmed by partial nuclease P1 digestion and two-dimensional homochromatography as described below.

Sequencing of Spots 1 and 2 by Partial Nuclease P1 Digestion. The nucleotide sequences of spots 1 and 2 were determined by the partial nuclease P1 digestion and two-dimensional homochromatography method (Lockard & RajBhandary, 1976). The general rules for the interpretation of mobility shifts of oligonucleotides differing by one nucleotide are the following (Jay et al., 1974; Galibert et al., 1974; Lockard & RajBhandary, 1976): (i) in two-dimensional homochromatography, 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 pA; (ii) 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; (iii) a loss of pU or pG from an oligonucleotide produces a shift of the spot to the left with the angle of shift being greater for pU than for pG; (iv) if the 5′-terminal two or three nucleotides in an oligonucleotide are composed of only A or C, spot shifts by loss of A or C are slightly to the left by two-dimensional homochromatography because of the slower migration of pA or pC in the first dimension; (v) the loss of 3′-phosphate in an oligonucleotide produces a shift to the left.

The mobilities of the partial digests of spots 1 and 2 are shown in plates 1 and 2, respectively, of Figure 5. The smallest nucleotide (top) should be pN^m, and the largest nucleotide (bottom) should be the undigested sample. The 3′-phosphate of oligonucleotides is removed by nuclease P1 (Fujimoto et al., 1974), and therefore, the second spot from the bottom should be the intact oligonucleotide lacking 3′-phosphate (Lockard & RajBhandary, 1976). The smallest nucleotide (top) in both plates was eluted and shown to be a mixture of pm^6A^m and pA^m by 2D-TLC. This shows that pm^6A^m and pA^m are not separable on two-dimensional homochromatography. The spot of pm^6A^m and pA^m in plate 1 is faint. This problem is probably due to incomplete transfer of mononucleotides from the cellulose acetate strip to the DEAE-cellulose plate prior to homochromatographic analysis (Lockard & RajBhandary, 1976). From this sequencing experiment and the analysis of the nuclease P1 and RNase T2 digests, we conclude that the nucleotide sequences are $p(m^6)A^mC^{(m)}UCUUGp$ for spots 1 and $p(m^6)A^mC^{(m)}AUCUGp$ for spot 2, in which the methylations shown in parentheses are partial.

Discussion

The 5′-terminal RNase T1 oligonucleotides of Ad2 early region E1 mRNAs were studied. Two major 5′-terminal oligonucleotides, spots 1 and 2, obtained from mRNA hybridized

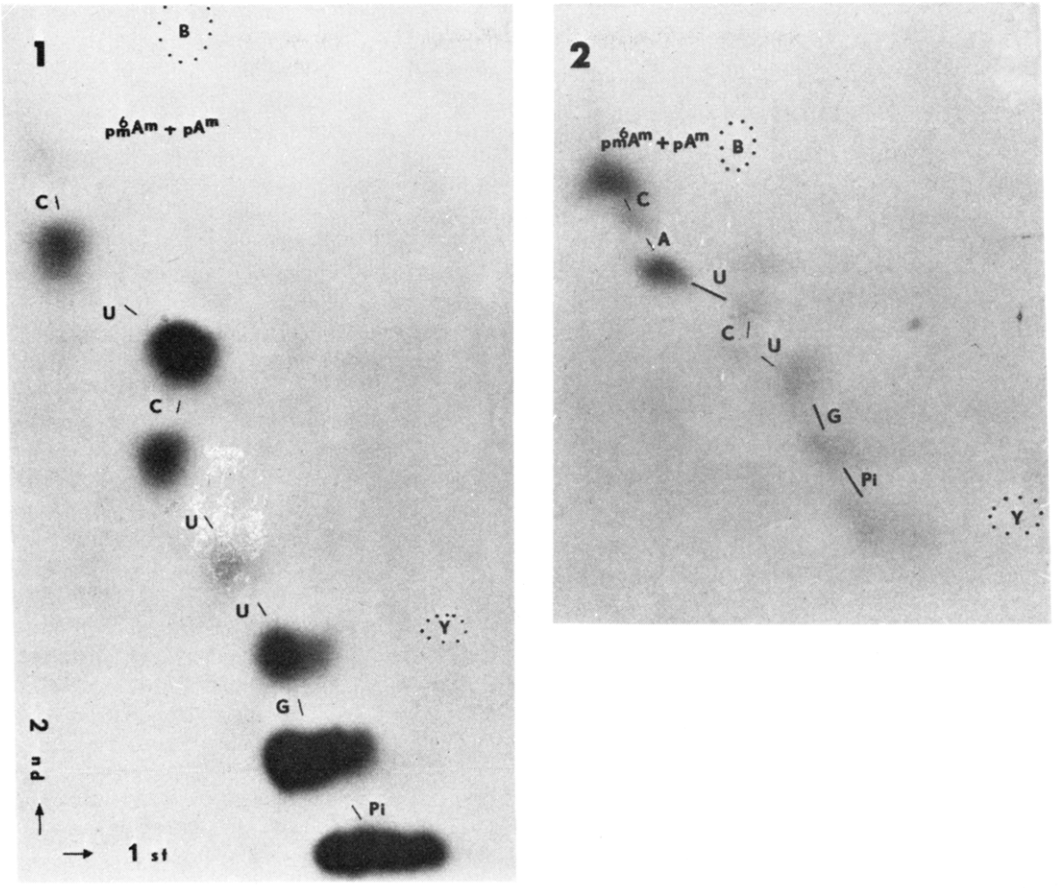

FIGURE 5: Nuclease P1 partial digestion and sequencing of RNase T1 spots 1 and 2. Oligonucleotide spots 1 and 2 from Figure 1 were partially digested with nuclease P1, mixed with completely digested sample and an undigested sample, and resolved by two-dimensional homochromatography. The nucleotide sequences of the spots were deduced by migration of the partial digest as described (Lockard & RajBhandary, 1976). The homomixture used in plate 1 was a different batch from that in plate 2, and therefore, the migration of the spots relative to the dye markers is different. the additional faint spot in plate 2 (near the upper U) contains only pA and was derived from slight contamination of spot 2 with an unmethylated oligonucleotide (see Figure 2).

to *Eco*RI-A fragment, were characterized by nuclease P1 and RNase T2 digestion and were sequenced by partial nuclease P1 digestion and two-dimensional homochromatography. Spots 1 and 2 were mapped to E1a (mp 1.3–4.6) and E1b (mp 4.6–11.2), respectively, by analysis of mRNA hybridized to subfragments of the E1 DNA region. Spot 1 was also found as a major component of the E1b 5′ termini, which suggests that the 5′ termini of some E1b mRNAs are derived from the E1a transcription initiation site. Large cytoplasmic mRNAs (about 24 S) that contain both E1a and E1b sequences have been observed by polyacrylamide gel electrophoresis of E1 mRNA (Büttner et al., 1976; Chinnadurai et al., 1976). ExoVII gel electrophoresis analysis indicates that the 5′ termini of E1b mRNAs were derived from both mp 4.6 and 1.3, although the 1.3-terminated molecules were minor (Berk & Sharp, 1978). The spot 1 oligonucleotides that we have found in E1b mRNA may be derived from these large RNA molecules, which represent a minor fraction of the total E1 mRNA. On the other hand, although we obtained our mRNA from polyribosomal pellets, we cannot rule out the possibility of minor contamination of the mRNA fraction by the unspliced 3340-nucleotide viral nuclear RNA (Kitchingman & Westphal, 1980).

The sequences of spots 1 and 2 were shown to be ACUCUUG and ACAUCUG, respectively. These sequences can be found at positions 498–504 (spot 1) and 1699–1705 (spot 2) within the Ad2 DNA sequence (T. Gingeras and R. Roberts, unpublished experiments) and at no other location within E1. Therefore, E1a mRNAs must initiate at position 498, and E1b mRNAs at position 1699. Identical results have been obtained by Baker & Ziff (1981).

Figure 6 gives the DNA sequences (sense strand) of the region flanking the E1a and E1b capping sites. Each site has a TATA box located at positions −24 to −30 or −31. TATA boxes have been found at this position in most eukaryotic transcription units and are believed to function somehow in the correct initiation of transcription. As with most TATA boxes, the E1a and E1b boxes are flank by GC-rich segments. Efstratiadis et al. (1980) have noted that most TATA boxes contain the trinucleotide ATA, located at approximately −28 to −30. ATA is found at −27 to −29 in E1b, but only ATT is found in E1a at this position.

A second region of homology has been found at approximate positions −70 to −90 in variety of eukaryotic transcription units (Benoist et al., 1980) and has been referred to as a "CCAAT box" (Efstratiadis et al., 1980). The only sequences in E1a and E1b that resemble this sequence are TCAAA at −65 to −73 in E1a and ATAAT at −72 to −79 in E1b. Another possible common sequence, a so-called "capping signal", PyPyATTCPu, which has been found immediately after the cap site of the sea urchin histone (Sures et al.,, 1978) and *Drosophila* 70K protein (Ignolia et al., 1980) genes, may be marginally represented at the Ela and E1b cap sites.

The first translation initiation codons (AUG) are found at +62 in E1a and +13 in E1b. These codons may initiate translation, if E1a and E1b mRNAs follow the general rule that translation of eukaryotic mRNA is initiated at the first AUG downstream from the cap site (Kozak, 1978). Indeed,

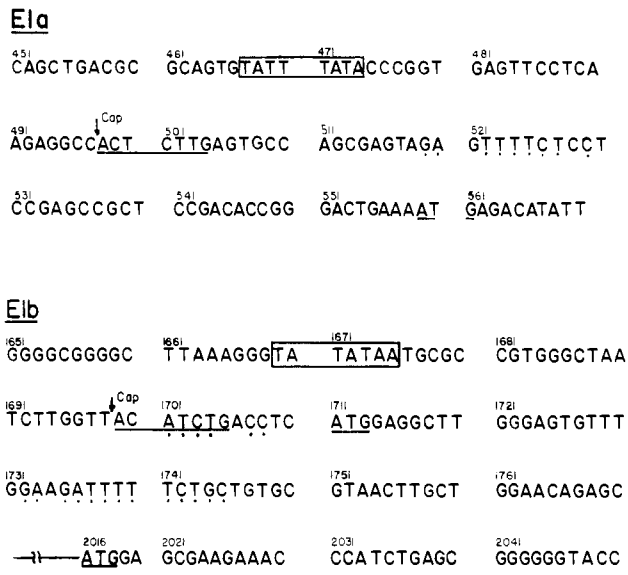

FIGURE 6: Alignment of 5′-terminal sequences of E1a and E1b mRNAs on Ad2 DNA. The Ad2 DNA sequences shown are from 451 to 570 (E1a) and from 1651 to 1770, and from 2016 to 2041 (E1b) on the 1 strand (the sense strand). The numbers refer to the distance, in nucleotides, from the left end of the genome. The DNA sequence was determined by T. Gineras and R. Roberts (unpublished experiments). The sequences corresponding to the deduced 5′-terminal sequence of E1a and E1b mRNAs are underscored. The TATA box is boxed. Nucleotides that are capable of forming a hybrid with the 3′-terminal nucleotides of 18S rRNA are shown with dots. Two separate ribosome homology sequences were found in E1b. The first possible translation initiation codon (AUG) in E1a and E1b and the second possible initiation codon in E1b are underscored.

the +13 AUG in E1b has been shown to code the N terminus of the E1b 15K polypeptide (Anderson & Lewis, 1980).

In many eukaryotic mRNAs, – pyrimidine-rich sequence that is complementary to the purine-rich 3′-terminal sequence of 18S rRNA has been observed between the cap and translation intiation sites (Kozak, 1978; Kozak & Shatkin, 1977; Hagenbüchle et al., 1978). This sequence could possibly be involved in the binding of mRNAs to ribosomes (Hagenbüchle et al., 1978), although at present there is no evidence for this. Both E1a and E1b contain sequences (Figure 6, indicated by dots), located between the cap site and the initiation codon, that can form hybrids with the 3′ end of 18S rRNA. Interestingly, E1b also contains a longer sequence (13 nucleotides) located about 20 nucleotides downstream from the ATG for E1b 15K that also is complementary to the 3′ end of 18S rRNA. An intriguing speculation is that this sequence is involved in the interaction of the E1b 22S mRNA with ribosomes and that translation of the E1b 53K polypeptide is initiated at the AUG coded at position 2016 (Figure 6). We have translated E1b mRNA in the reticulocyte lysate system and characterized the translation products by polyacrylamide gel electrophoresis and by peptide mapping (T. Matsuo, W. Wold, S. Hashimoto, A. Rankin, J. Symington, and M. Green, unpublished experiments). We have observed three E1b polypeptides, 53K, 20K, and 19K, with 53K and 20K being related to each other but not to 19K. This suggests that there may be more than one translation initiation site and/or transcription initiation site in E1b.

In addition to the two major 5′ termini, we observed several minor 5′-terminal RNase T1 oligonucleotides, designated spots a–e. Spots a and d were found in F1, F2, and F3 mRNA, and b and c in F3 mRNA. This suggests that there may be multiple minor transcription initiation sites for both E1a and E1b mRNAs. In this sense, E1a and E1b seem to be typical eukaryotic transcription units, because heterogeneous 5′ termini have been found for the other Ad2 early region (Hashimoto et al., 1981; Baker & Ziff, 1981), and for SV40 (Canaani et al., 1979; Kahana et al., 1981; Haegeman & Fiers, 1980; Ghosh et al., 1978; Reddy et al., 1979) and polyoma virus (Flavell et al., 1979).

The technology we described here is a simple and direct method for 5′-terminal nucleotide sequencing of a capped mRNA which exits in very low abundance in cellular RNA. For example, Ad2 E1b mRNA is less than 0.5% of the total poly(A+) RNA under our infection conditions. This method can be used to determine the 5′-terminal sequence (generally five to ten nucleotides) of any mRNA which exists at this level, if the mRNA can be purified by hybridization. Our methodology has several advantages over that used by other workers: (1) the penultimate nucleoside is labeled in vitro with [γ-^{32}P]ATP, permitting the identification of nucleosides in which the base may be partially methylated, the ribose may be 2′-O-methylated, and the nucleoside may be modified in an unusual manner (Shatkin, 1976; Hashimoto & Green, 1980); (2) the purity of each 5′-terminal oligonucleotide is demonstrated prior to sequencing; and (3) RNA sequencing is done by partial nuclease P1 digestion which produced 3′-OH nucleotides (Fujimoto et al., 1974), rather than partial RNase T2 digestion which produces a mixture of 3′-phosphate and 2′,3′-cyclic phosphate nucleotides(Sato et al., 1966).

Acknowledgments

We thank Dr. H. Thornton for assistance with cell culture, M. H. Pursley for technical support, and Rose Epler for preparing cloned DNA. We also thank Drs. M. Wilson, J. E. Darnell, S.-L. Hu, and J. Sambrook for providing the clones of the F1 to F5 fragments. We are indebted to Drs. T. Gingeras and R. Roberts for making their DNA sequence data available to us prior to publication. We thank Alice Smith for typing the manuscript.

References

Aleström, P., Akusjärvi, G., Perricaudet, M., Mathews, M. B., Klessig, D. F., & Petterson, U. (1980) *Cell (Cambridge Mass.) 19*, 671–681.

Anderson, C. W., & Lewis, J. B. (1980) *Virology 104*, 27–41.

Baker, C. C., & Ziff, E. (1980) *Cold Spring Harbor Symp. Quant. Biol. 44*, 415–428.

Baker, C. C., & Ziff, E. (1981) *J. Mol. Biol. 149*, 189–221.

Benoist, C., O'Hara, K., Breathnach, R., & Chambon, P. (1980) *Nucleic Acids Res. 8*, 127–142.

Berk, A. J., & Sharp, P. A. (1977) *Cell (Cambridge, Mass.) 12*, 45–55.

Berk, A. J., & Sharp, P. A. (1978) *Cell (Cambridge, Mass.) 14*, 695–711.

Brownlee, G. G., & Sanger, F. (1969) *Eur. J. Biochem. 11*, 395–399.

Büttner, W., Veres-Molnar, Z., & Green, M. (1976) *J. Mol. Biol. 107*, 93–114.

Canaani, D., Kahana, C., Kukamel, A., & Groner, Y. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3078–3082.

Chinnadurai, G., Rho, H. M., Horton, R., & Green, M. (1976) *J. Virol. 20*, 255–263.

Chow, L. T., Broker, T. R., & Lewis, J. (1979) *J. Mol. Biol. 134*, 265–303.

Clewell, D. B., & Helinski, D. R. (1969) *Proc. Natl. Acad. Sci. U.S.A. 62*, 1159–1166.

Domdey, H., & Gross, H. J. (1979) *Anal. Biochem. 93*, 321–328.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G.,

Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., & Proudfoot, N. J. (1980) *Cell (Cambridge, Mass.) 21*, 653–668.
Esche, H., Mathews, M. B., & Lewis, J. B. (1980) *J. Mol. Biol. 142*, 399–417.
Evans, R. M., Fraser, N., Ziff, E., Weber, J., Wilson, M., & Darnell, J. E. (1977) *Cell (Cambridge, Mass.) 12*, 733–739.
Flavell, A. J., Cowie, A., Legon, S., & Kamen, R. (1979) *Cell (Cambridge, Mass.) 16*, 357–371.
Fujimoto, M., Kuninaka, A., & Yoshino, H. (1974) *Agric. Biol. Chem. 38*, 785–791.
Galibert, F., Sedat, J., & Ziff, E. (1974) *J. Mol. Biol. 87*, 377–407.
Gallimore, P. H., Sharp, P. A., & Sambrook, J. (1974) *J. Mol. Biol. 89*, 49–72.
Ghosh, P. K., Reddy, V. B., Swinscoe, J., Lebowitz, P., & Weissman, S. W. (1978) *J. Mol. Biol. 126*, 813–846.
Haegeman, G., & Fiers, W. (1980) *J. Virol. 35*, 955–961.
Hagenbüchle, O., Santer, M., Steitz, J. A., & Mans, R. J. (1978) *Cell (Cambridge, Mass.) 13*, 551–563.
Halbert, D. N., Spector, D. J., & Raskas, H. J. (1979) *J. Virol. 31*, 621–629.
Hashimoto, S., & Green, M. (1979) *Virology 94*, 254–272.
Hashimoto, S., & Green, M. (1980) *J. Biol. Chem. 255*, 6780–6788.
Hashimoto, S., Pursley, M. H., Wold, W. S. M., & Green, M. (1980) *Biochemistry 19*, 294–300.
Hashimoto, S., Pursley, M. H., & Green, M. (1981) *Nucleic Acids Res. 9*, 1675–1689.
Ignolia, T. D., Craig, E. A., & McCarthy, B. J. (1980) *Cell (Cambridge, Mass.) 21*, 669–679.
Jay, E., Bambara, R., Padmanabhan, R., & Wu, R. (1974) *Nucleic Acids Res. 1*, 331–353.
Kahana, C., Gidoni, D., Canaani, D., & Groner, Y. (1981) *J. Virol. 37*, 7–16.
Kitchingman, G. R., & Westphal, H. (1980) *J. Mol. Biol. 137*, 23–48.
Kozak, M. (1978) *Cell (Cambridge, Mass.) 15*, 1109–1123.
Kozak, M., & Shatkin, A. (1977) *J. Biol. Chem. 252*, 6895–6908.
Landgraf-Leurs, M., & Green, M. (1971) *J. Mol. Biol. 60*, 185–202.
Lockard, R. E., & RajBhandary, U. L. (1976) *Cell (Cambridge, Mass.) 9*, 746–760.
Manley, J. L., Sharp, P. A., & Gefter, M. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 160–164.
Nishimura, S. (1972) *Prog. Nucleic Acid Res. Mol. Biol. 12*, 50–85.
RajBhandary, U. L., & Chang, S. H. (1968) *J. Biol. Chem. 243*, 598–608.
Reddy, V. B., Ghosh, P. K., Lebowitz, P., Piatak, M., & Weissman, S. W. (1979) *J. Virol. 30*, 279–296.
Sato, S., Uchida, T., & Egami, F. (1966) *Arch. Biochem. Biophys. 115*, 48–52.
Sehgal, P. B., Fraser, N. W., & Darnell, J. E. (1979) *Virology 94*, 185–191.
Shatkin, A. J. (1976) *Cell (Cambridge, Mass.) 9*, 645–653.
Spector, D. J. McGrogan, M., & Raskas, H. (1978) *J. Mol. Biol. 126*, 395–414.
Sures, I., Lowry, J., & Kedes, L. (1978) *Cell (Cambridge, Mass.) 15*, 1033–1044.
Van der Eb, A. J., Mulder, C., Graham, F. L., & Houweling, A. (1977) *Gene 2*, 115–132.
Weil, P. A., Lusa, D. S., Segall, J., & Roeder, R. G. (1979) *Cell (Cambridge, Mass.) 18*, 469–484.
Wilson, M. C., Fraser, N. W., & Darnell, J. E. (1979) *Virology 94*, 175–184.
Ziff, E., & Evans, R. E. (1978) *Cell (Cambridge, Mass.) 15*, 1463–1475.