1981

Promoters and Heterogeneous 5' Termini of the Messenger RNAs of Adenovirus 2

Carl Chester Baker

PROMOTERS AND HETEROGENEOUS 5' TERMINI OF THE MESSENGER RNAs OF ADENOVIRUS 2

A thesis submitted to the Faculty of The Rockefeller University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
by

Carl Chester Baker

March 20, 1981
The Rockefeller University
New York

ACKNOWLEDGEMENTS

I would like to thank Ed Ziff for teaching me the rudiments of RNA and DNA sequencing and for his advice, encouragement and support. I would also like to thank Bob Lamb for his help and advice on reverse transcription and S1 mapping, Greg Gasic for his gift of RNAse-free calf intestinal alkaline phosphatase, Mary Moore for excellent technical assistance, and all the past and present members of the Darnell Lab who have taught me techniques, given me useful advice and criticism, and occassionally supplied me with cloned Ad2 DNA fragments. Special thanks go to my colleagues elsewhere in the scientific community who have helped to make this work possible: Andy Fire for collaborating on *in vitro* transcription, Tom Broker and Louis Chow for their helpful discussions and for communicating their electron microscopy mapping data prior to publication, David Lee for keeping me informed about his Ad2 *in vitro* transcription studies, and Francis Galibert and Ulf Petterson for communicating the Ad2 Eco RI F fragment and Protein IX gene sequences, respectively, prior to publication. And last, but not least, I would like to thank my wife, Susan R. Haynes, for her helpful advice, critical reading of this thesis and her constant tolerance, patience and understanding during this work.

Portions of this work have previously appeared in print in Cell *18*, 569-580 (1979); Cold Spring Harbor Symposia on Quantitative Biology *44*, 415-428 (1980); and the Journal of Molecular Biology *149*, 189-221 (1981).

SUMMARY

Adenovirus-2 messenger RNAs are transcribed from nine separate transcription units which function during early, intermediate and late stages of infection. I have sequenced capped 5' termini of messenger RNAs from each of these transcription units, and located the templates for each capped terminus in the sequence of the Ad2 genome. These results define the putative initiation sites and promoters for all major Ad2 transcripts.

By comparing the termini and promoter DNA sequences I conclude that seven of nine initiation sites including EIa, EIb, PIX, major late, EIII, EIV and the late form of region IIa mRNA have clearly defined TATA box homologies. With early region IIa and IVa2, this homology is absent. For most messengers, the mRNA terminus is microheterogeneous, and encoded within a 2-7 base cap site region located approximately 30 nucleotides downstream from the TATA box. However, the two transcription units which lack TATA boxes display equivalent microheterogeneity. Purine termini are greatly favored over pyrimidine termini. I propose a model for Ad2 cap site recognition in which RNA polymerase scans short DNA segments for purine starts, and discuss the model in the context of EIV mRNAs of Ad2 and Ad5 which differ in their promoter region DNA sequence.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

Ad2	Adenovirus 2
BAP	bacterial alkaline phosphatase
CPE	cytopathic effects
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
EDTA	ethylene diamine tetraacetic acid
FCS	fetal calf serum
γ	micrograms
γ-equivalents	molar equivalent of 1 γ of viral DNA
HAP	hydroxyapatite
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HSB	high-salt buffer
λ	microliters
moi	multiplicity of infection
PBS	phosphate buffered saline
pfu	plaque forming units
p.i.	post-infection
PIPES	Piperazine-N,N'-bis[2-ethanesulfonic acid]
PNK	T4 polynucleotide kinase
RSB	reticulocyte standard buffer
RT	room temperature
SDS	sodium dodecyl sulfate
TEA	triethylamine
TES	N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid
Tris	Tris(hydroxymethyl)aminomethane
UV	ultraviolet light

INTRODUCTION

Cells exist in an environment which is in a constant state of flux. The cell functions best when a fairly constant intracellular environment is maintained, so it is beneficial to the cell to be able to react and adapt to its external environment. The regulation of the metabolism of the cell could exist solely at the level of feedback inhibition or stimulation of the metabolic enzymes themselves. However it is more economical if regulation can be at the level of gene expression also, and organisms capable of this will be more "fit" from an evolutionary point of view (Zamenhof and Eichhorn, 1967).

Jacob and Monod (1961) first proposed the idea that there are both structural and regulatory genes. The regulatory gene codes for a repressor molecule (trans acting) capable of binding to an operator element (cis acting) present in the structural gene and of blocking its transcription. A group of genes (known as an operon) can be coordinately expressed if they are transcribed, under the control of a single operator element, into a polycistronic mRNA. Jacob, Ullman and Monod (1964) showed that a genetic element, the promoter, is necessary for polymerase binding and expression of the operon. Genetic regulation in prokaryotes is highly responsive because the mRNA is very unstable. Thus a change in the transcription rate can have an immediate effect on the translation rate and therefore on the level of a protein product.

Prokaryotic transcription and genetic regulation have now been elucidated at the molecular level (for a review see Goldberger, 1979). The structural elements necessary for RNA polymerase binding at the promoter, as well as the interactions with activator and repressor proteins, are known (for a review see Siebenlist, Simpson and Gilbert, 1980).

RNA biogenesis in eukaryotes is much more complex than in prokaryotes. The DNA template for transcription is bound up in a nucleoprotein complex called chromatin (for reviews see Chambon, 1978; Felsenfeld, 1978). Chromatin is not uniform, but appears to differ structurally in active versus inactive trancription units. For example,

embryonic globin genes are more sensitive to DNAse I digestion than adult-specific globin genes in embryonic red cells while adult-specific globin genes are more DNAse I sensitive than embryonic globin genes in adult red cells. Neither embryonic nor adult globin genes are DNAse I sensitive in fibroblasts. Ovalbumin genes are not DNAse I sensitive in any of these cell types (Weintraub and Groudine, 1976; Stalder et al., 1980). However the ovalbumin gene is preferentially DNAse I sensitive in oviduct cells of the laying hen (Garel and Axel, 1976). The actively transcribed adenovirus genes in adenovirus transformed cells are also DNAse I sensitive (Flint and Weintraub, 1977). The DNAse I sensitivity of actively transcribed chromatin may be due, at least in part, to the presence of two nonhistone chromosomal proteins, HMG 14 and 17. Reconstitution of HMG depleted chromatin with HMG 14 or 17 specifically confers DNAse I sensitivity on the adult β-globin gene and other transcriptionally active genes (Weisbrod et al., 1980).

Eukaryotic RNA is transcribed by three distinct RNA polymerases: RNA polymerase I transcribes 18S and 28S rRNA; RNA polymerase II transcribes mRNA, hnRNA and some small nuclear RNAs (e.g. U1 RNA); and RNA polymerase III transcribes 5S, tRNAs and adenovirus VA RNA (Chambon, 1975; Roeder, 1976; Roop et al., 1981). Thus transcriptional control of these three sets of genes may differ. In the rest of this thesis I will deal exclusively with transcription of mRNA by RNA polymerase II.

RNA polymerase II produces a long primary transcript (called hnRNA) which is colinear with the DNA and whose 5' end is modified by capping (Perry et al., 1975; Shatkin, 1976; Salditt-Georgieff et al., 1976; Salditt-Georgieff et al., 1980a). This primary transcript is further modified by endonucleolytic cleavage and polyadenylation to produce a unique 3' end, splicing, and 6-methylation of a few adenines (Desrosiers, Friderici and Rottman, 1974). The resulting product, mRNA, is then transported from the nucleus to the cytoplasm (for a review of mRNA biogenesis, see Darnell, 1979). Gene expression is potentially regulated at all the steps of mRNA biogenesis.

Transcriptional Control

Transcriptional control is an important regulatory mechanism for at least some eukaryotic genes. For example, in Bombyx mori, fibroin mRNA accumulates in the posterior silk gland but not in the middle silk gland. Tsuda et al. (1979), using the S1 technique, showed that the fibroin mRNA precursor is found only in the posterior silk gland. This indicates that the tissue specific control of fibroin gene expression is at the transcriptional level.

Expression of the chick ovalbumin gene family (ovalbumin, X and Y genes) is hormonally controlled. By comparing the rates of transcription in in vitro nuclei isolated from oviducts of hormone withdrawn and estrogen or progesterone stimulated chicks, McKnight and Palmiter (1979) and LeMeur et al. (1981) have shown that regulation by steroid hormones is at least partially at the transcriptional level.

Several genes in Dictiostelium discoideum appear to be developmentally regulated at the transcriptional level by cAMP. Williams, Tsang and Mahbubani (1980) showed that addition of cAMP produced a rapid and specific reduction in the rate of transcription of the discoidin I gene and the acceleration of appearance of 3 mRNAs which normally appear at later stages of development when the concentration of cAMP is high. The discoidin I gene is normally not expressed at these later stages of development.

To understand transcriptional control, it is first necessary to understand the structure of a promoter and the process of transcription initiation.

Promoters, Transcription Initiation, and Capping

The 5' termini of mRNAs are modified by capping (for a review see Shatkin, 1976). The general structure of a cap is $^{7m}GpppX_mY_{(m)}Z$---. Caps are formed by two mechanisms. In the first, represented by VSV capping, the GTP donor contributes both an α and β phosphate and the 5' terminus of the mRNA contributes an α phosphate only (Abraham, Rhodes and Banerjee, 1975). It has been suggested that the 5' terminal

phosphate of the transcript is formed by cleavage of a precursor molecule prior to capping (Abraham and Banerjee, 1976), although recent work may challenge this (Testa, Chanda and Banerjee, 1980).

The other mechanism is used by vaccinia virus (Ensinger et al., 1975; Moss et al., 1976), reovirus (Furuichi and Shatkin, 1976; Furuichi et al., 1976; Furuichi and Shatkin, 1977), HeLa cell nuclei (Wei and Moss, 1977; Keith, Ensinger and Moss, 1978; Groner, Gilboa and Aviv, 1978), rat liver nuclei (Mizumoto and Lipmann, 1979) and mouse L-cell nuclei (Winicov and Perry, 1976). By labeling with α-, β-, or γ-^{32}P-nucleoside triphosphates, it can be shown that only an α-phosphate is contributed by GTP, the guanyl donor, and that both α- and β-phosphates are contributed from the RNA 5' terminus, the guanyl acceptor. Guanylation precedes methylation, since ^{7m}GTP is not a substrate for the HeLa cell guanylyltransferase (Venkatesan and Moss, 1980). 7-Methylation precedes 2'-O-methylation in reovirus (Furuichi et al., 1976). There may be no rigid order of methylation in permeabilized CV1 cells since 2'-O-methylation was present on some caps without 7-methylation (Contreras and Fiers, 1981). Kinetic studies by Perry and Kelly (1976) in mouse L-cells indicate that the 2'-O-methylation at the Y position of the cap is probably a cytoplasmic rather than a nuclear modification event and therefore occurs last.

Most evidence indicates that capped termini formed by the second mechanism are derived from triphosphate termini formed by transcription initiation of the primary transcripts. However, it has been hypothesized that pyrimidine caps derive from 5' termini formed by cleavage (Shatkin, 1976; Schibler and Perry, 1976). In cytoplasmic polyhedrosis virus, transcription is dependent on capping, since incubation with p(NH)ppA (β-γ imido ATP) which blocks capping, but not elongation, inhibits <u>in vitro</u> mRNA synthesis (Furuichi, 1978). However capping is not required for transcription in either reovirus or CV1 cells. Inhibition of reovirus capping by either pyrophosphate (Furuichi and Shatkin, 1976) or p(NH)ppG (Furuichi and Shatkin, 1977) did not decrease transcription and produced transcripts terminated by ppGC--- or p(NH)ppGC---, respectively. Incubation of SV40 infected permeabilized

CV1 cells with p(CH_2)ppA did not decrease the relative amount of SV40 specific RNA (late SV40 RNAs start predominantly with A) (Contreras and Fiers, 1981). However in the SV40 case, transcription may have compensated by initiating at alternate start sites encoding a G. The HeLa cell mRNA guanylyltransferase requires at least a dinucleotide with a 5' di- or tri- phosphate (ppXpY or pppXpY) as the guanyl acceptor (Venkatesan and Moss, 1980). Therefore initiation cannot take place with an already capped nucleotide and capping must occur after initiation.

Capping is an early event in transcription. In reovirus, short molecules synthesized during short pulse labels or in the absence of ATP and UTP are capped (Furuichi et al., 1976). In isolated nuclei of mouse L-cells, synthesis of caps in >12S hnRNA is dependent on transcription (shown by inhibiting transcription with α-amanitin) and all caps labeled in a 5' pulse are in small molecules (Winicov and Perry, 1976). Salditt-Georgieff et al. (1980a) have shown that after a 3 minute label of CHO cells with ^{3}H-methyl-L-methionine, the majority of labeled caps are in molecules shorter than 750 nucleotides. No triphosphate termini were found in hnRNA. In addition, they showed that the composition of the X position of the cap was the same in small molecules and large hnRNA. This is inconsistent with the hypothesis that large hnRNA is cleaved to form pyrimidine caps, because then hnRNA should contain more pyrimidine caps than short (nascent) molecules. Babich, Nevins and Darnell (1980) showed that transcripts from the major late promoter of Ad2 are essentially all capped before they are 100 nucleotides long. No triphosphate termini could be found, even in short molecules.

These experiments show that capping occurs very early in transcription, but do not rule out the possibility of cleavage and subsequent capping occurring after the polymerase has transcribed a short distance. More detailed analysis of the structures of the 5' termini of both the primary transcript and the mRNA was needed to answer the question of whether the capped terminus of the mRNA is derived from the initiated terminus of the primary transcript.

Roop, Tsai and O'Malley (1980) analyzed nuclear RNA from chicken

oviducts by the S1 technique (Berk and Sharp, 1977a) and showed that the 7.8 kb ovalbumin precursor has the same 5' end (±10-20 nucleotides) as ovalbumin mRNA. They were unable to detect any ovalbumin precursors larger than 7.8 kb by Northern blotting. Weaver and Weissmann (1979), also using the S1 technique, showed that the 5' terminus of the 15S mouse β-globin mRNA precursor mapped to within 5 nucleotides of the cap site encoding the 10S mRNA cap. Curtis, Mantei and Weissmann (1978) showed by direct RNA analysis that the 15S precursor had the same cap structure as the mRNA. A potential 27S β-globin precursor reported by Bastos and Aviv (1977) but not seen by other workers, has not been analyzed in this way.

Tsuda, Ohshima and Suzuki (1979) analyzed >47S precursors to the 47S fibroin mRNA by S1 and reverse transcription. The precursor 5' end was mapped by S1 to within 6 nucleotides of the mRNA cap site. Since the DNA probe was 5' end labeled within the first intervening sequence (IVS), only unspliced molecules were analyzed. The possibility of an additional 5' leader on the precursor coming from far upstream was ruled out by reverse transcription of the precursor using a DNA primer labeled within the IVS. The size of the reverse transcript was that expected for a precursor beginning at the mRNA cap site.

Perhaps the best studied system is that of the Ad2 major late transcription unit. The promoter was first mapped by hybridization analysis of nascent RNA molecules pulse labeled *in vivo* and *in vitro*. The shortest nascent chains hybridized to a DNA fragment near coordinate 16 and indicated that the promoter was at 16.3 m.u. (Weber, Jelinek and Darnell, 1977; Evans et al., 1977). Similar mapping of short transcripts prematurely terminated by UV (Goldberg, Weber and Darnell, 1977) and DRB (Fraser, Sehgal and Darnell, 1978) supported this conclusion. Ziff and Evans (1978) showed that late mRNAs had the same 5' cap undecanucleotide as late *in vivo* labeled nuclear RNA and that this cap oligonucleotide was encoded at the site to which the promoter had been mapped by pulse labeled nascent chains. In addition, Manley, Sharp and Gefter (1979a, 1979b) and Baker and Ziff (1980) showed that late Ad2 RNA prepared in pulse labeled *in vitro* nuclei had the same cap structure as

late mRNAs. No RNA could be detected which was encoded upstream from the cap site. Thus for the Ad2 major late transcription unit, the large unspliced mRNA precursor has the same 5' terminus as mRNA and this 5' terminus appears to be derived from the initiated terminus encoded at the promoter.

Two cell-free in vitro transcription systems have been developed which appear to initiate transcription accurately at the Ad2 major late promoter (Weil et al., 1979; Manley et al., 1980). In vitro transcripts have the same 5' cap structures as are found in late mRNAs and mRNA precursors. A similar result has been obtained using a human β-globin gene (Proudfoot et al., 1980) and a mouse β-globin gene (Luse and Roeder, 1980). Proudfoot et al. (1980) also showed that the 5' termini of in vitro transcripts of human β-, ε-, and α- globin genes are coterminal with their mRNAs. Wasylyk et al. (1980b) have shown that ovalbumin, conalbumin and the Ad2 EIa genes also produce transcripts with the correct 5' termini.

While all these experiments support the hypothesis that the 5' end of the mRNA derives from the initiated terminus of the primary transcript, they do not rule out a mechanism in which initiation occurs a short distance upstream from the cap site and the capped terminus is produced by cleavage when the polymerase reaches the cap site, followed by capping at the cleavage site and immediate degradation of the RNA transcribed upstream (or 5' exonucleolytic degradation as the RNA is transcribed). The best evidence that the cap site is the site of initiation comes from experiments labeling with β-^{32}P-nucleoside triphosphates. According to the favored model for capping an initiated triphosphate terminus, the β-phosphate comes from the triphosphate RNA terminus. Hagenbuchle and Schibler (1981) have shown that the caps produced by Ad2 major late and β-globin (both have A caps only) promoter fragments in the Manley in vitro transcription system are preferentially labeled over internal nucleotides by β-^{32}P-ATP only. SV40 late mRNAs have multiple A and G caps and 5' termini encoded over a 200 nucleotide region (see references in the legend to Figure 46). Contreras and Fiers (1981) have shown that by labeling permeabilized SV40 infected CV1 cells

with β-^{32}P-XTPs, all of the caps could be labeled. Little β-^{32}P was exchanged with the α-phosphates and the caps were labeled only in the β position. Only A caps were labeled with β-^{32}P-ATP and G caps with β-^{32}P-GTP. Thus the multiple SV40 late cap sites represent separate initiation sites and not cleavage sites. Similar experiments have also been done with early SV40 infected permeabilized cells (Gidoni et al., 1981). In this case the C caps (the major SV40 early caps are CU (Kahana et al., 1981)) were labeled by β-^{32}P-CTP in the β position of the cap. Therefore it is likely that all caps, including pyrimidine caps, derive from the initiated triphosphate terminus of the primary transcript.

From these experiments it can be concluded that the cap site is the site of initiation and that the cap region will probably contain the promoter for RNA polymerase II. When I began this work nothing was known about what constitutes an RNA polymerase II promoter. In the meantime many putative promoter regions have been sequenced and analyzed by in vivo and in vitro genetic studies which have helped to elucidate the functional elements of the promoter (for current reviews, see Breathnach and Chambon, 1981; Shenk, 1981). This work will be discussed in detail in the Discussion section.

Splicing

Intramolecular RNA splicing was first discovered for Adenovirus-2 late mRNAs by Berget et al. (1977), Chow et al. (1977a) and Klessig (1977). Almost all protein coding genes analyzed to date have been shown to have at least one intervening sequence, or intron. Their mRNA precursors therefore must be spliced to form a functional mRNA (for reviews see Abelson, 1979; Breathnach and Chambon, 1981). Exceptions are the Ad2 Protein IX gene (Alestrom et al., 1980), the human interferon-α gene (Nagata, Mantei and Weissmann, 1980), the histone genes (Kedes, 1979) and Herpes simplex virus type 1 thymidine kinase gene (McKnight, 1980; Wagner, Sharp and Summers, 1981). Since splicing allows several different proteins to be made from one transcription unit, it is a potential control point in genetic expression. There is

also some evidence that splicing may be required for transport of an mRNA from the nucleus to the cytoplasm (Hamer and Leder, 1979).

Many splice junctions have been compared and found to have sequences in common (Lerner et al., 1980; Rogers and Wall, 1980). All introns begin with GU and end with AG. The consensus sequence at the splice junctions is:

5'-(exon)-$\begin{matrix}A\\C\end{matrix}$AG/GUAAGUA-(intron)-UUUUPyUUUUUUCUUNCAG/G-(exon)-3'

In addition the consensus sequence is complementary to the 5' terminal sequence of U-1 snRNA, the most abundant snRNA (10^6 copies/nucleus). U-1 snRNA could align the splice junctions for splicing (Lerner et al., 1980; Rogers and Wall, 1980). Therefore it is hypothesized that U-1 snRNA is part of a splicing enzyme.

Polyadenylation and Transcription Termination

Eukaryotic mRNAs are modified at their 3' ends by variable lengths of poly(A) which is not template encoded (for a review see Brawerman, 1976). Histone mRNAs from HeLa cells are an exception in that they are not polyadenylated (Adesnik and Darnell, 1972). Comparison of the 3' noncoding regions of several polyadenylated mRNAs reveals the common sequence AAUAAA approximately 20 nucleotides upstream from the poly(A) (Proudfoot and Brownlee, 1976). An alternative sequence AUUAAA may be used by Ad2 Early Region III mRNA (Herisse, Courtois and Galibert, 1980; Fraser et al., 1981) and the mouse pancreatic α-amylase mRNA (Hagenbuchle, Bovey and Young, 1980).

In most systems, transcription does not terminate at the poly(A) addition site, but continues for at least an additional 1000 nucleotides. The Ad2 major late transcription unit, which has 5 different poly(A) addition sites (Fraser and Ziff, 1978; Nevins and Darnell, 1978a, 1978b), exhibits equimolar transcription throughout its length, even 2500-3000 bp past the last poly(A) site (Nevins and Darnell, 1978b; Fraser et al., 1979). In addition, Fraser et al. (1979) have shown by UV sensitivity that the transcripts extending beyond the most promoter-distal poly(A) site belong to the major late transcription unit. In similar experiments, transcription has been shown to continue past the

poly(A) sites of SV40 late mRNAs (Ford and Hsu, 1978), Adenovirus-2 early regions IIa and IV mRNAs (Nevins, Blanchard and Darnell, 1980) and mouse β-globin mRNA (Hofer and Darnell, 1981). In addition, since polyoma late mRNAs often have multiple copies of the 5' leader sequence (Legon et al., 1979), transcription must continue past the poly(A) site several times. Chicken ovalbumin may be an exception, since transcription appears to terminate at the 3' end of the structural gene (Roop, Tsai and O'Malley, 1980). Since for most genes transcription proceeds beyond the poly(A) site, the primary transcript must be cleaved endonucleolytically prior to poly(A) addition. Poly(A) addition must occur immediately after transcription, since in Chinese hamster ovary cells and HeLa cells, poly(A) is added within one minute after RNA chain synthesis (Salditt-Georgieff et al., 1980b). For the major late Ad2 transcription unit, cleavage and polyadenylation occur soon after transcription of the poly(A) site and before the polymerase terminates (Nevins, 1979). In addition, polyadenylation occurs before splicing in Ad2 (Nevins and Darnell, 1978b; Weber et al., 1980).

The existence of alternate poly(A) sites in a single transcription unit is a potential control point for gene regulation. For example, in B-lymphocyte development the switch from a membrane bound IgM heavy chain (μ_m) to a secretory IgM heavy chain (μ_s) is due to the use of a different poly(A) site which leads to different 3' coding regions on the mRNA and different C termini of the proteins (Rogers et al., 1980; Early et al., 1980).

Adenoviruses as a Model System for Eukaryotic mRNA Biogenesis

Adenoviruses were first isolated from human adenoid tissue cultures (Rowe et al., 1953) as an agent which induced CPE (cytopathic effects) in human epithelial cells in culture but could not be cultured in cell free medium alone. Subsequently the same agent was isolated from throat washings of men suffering from acute respiratory illness (Hilleman and Werner, 1954). Since then, adenoviruses have also been linked with epidemic keratoconjunctivitis (Ad8) (Jawetz et al., 1958). Adenoviruses were first shown to be oncogenic when Trentin, Yabe and

Taylor (1962) produced tumors in newborn Syrian hamsters with Ad12. Adenoviruses have been divided into 3 groups by their oncogenicity (for a review see Green, 1970): Group A (Ad 12, 18, 31) are highly oncogenic in newborn hamsters; Group B (Ad 3, 7, 11, 14, 16, 21) are weakly oncogenic; and Group C (Ad 1, 2, 5, 6) are nononcogenic. All groups, however, transform rodent cells in vitro (McBride and Weiner, 1964; Freeman et al., 1967; MacAllister et al., 1979).

Adenoviruses are nonenveloped icosahedral viruses with a diameter of 65-80 nm and a linear double-stranded DNA genome 35kb long (for a review of adenovirus structure, see Flint, 1980). Each end of the DNA genome contains a covalently linked protein which may prime DNA synthesis (Rekosh et al., 1977). The DNA in the virion may be organized into a "chromatin-like" structure by core proteins V and VII (Corden, Engelking and Pearson, 1976).

The lytic cycle of adenoviruses lasts roughly 36 hours and, except for protein synthesis, is confined to the nucleus (for a review of lytic infection see Flint and Broker, 1980). The cycle is divided into two phases: early events precede the onset of DNA replication and late events follow it. Adenovirus virions do not carry an RNA polymerase (Chardonnet, Gazzolo and Pogo, 1972) and in addition early transcription is not inhibited by the protein synthesis inhibitor cycloheximide (Parsons and Green, 1971). Also pure adenovirus DNA is infectious (Nicolson and MacAllister, 1972). Therefore early transcription must use preexisting host enzymes. Determination of the divalent cation and ionic strength requirements and α-amanitin sensitivity of the RNA polymerase responsible for adenovirus transcription has led to the conclusion that the host nuclear RNA polymerase II is used both early and late in infection (Ledinko, 1971; Chardonnet, Gazzolo and Pogo, 1972; Price and Penman, 1972; Wallace and Kates, 1972; Weinmann, Raskas and Roeder, 1974). The nuclear levels of RNA polymerase II remain constant throughout the infective cycle (Weinmann et al., 1976).

The Ad2 (and closely related Ad5) genome contains at least 9 transcription units (for a review of Ad2 transcription see Ziff, 1980).

Each of these transcription units has its own promoter, as shown by UV target size (Berk and Sharp, 1977b) and nascent chain mapping (Evans et al., 1977; Ziff and Evans, 1978). Additional evidence for separate promoters comes from mapping nascent molecules prematurely terminated by UV (Goldberg, Weber and Darnell, 1977; Wilson, Fraser and Darnell, 1979) and DRB (Fraser et al., 1978; Sehgal, Fraser and Darnell, 1979). The 9 transcription units are temporally regulated by a complex cascade mechanism which is still being elucidated. Infection in the presence of stringent protein synthesis inhibition by anisomycin produce only the mRNAs for a 13.5k protein encoded between 17 and 21.5 m.u. and the mRNA for the late 52,55k protein encoded between 29 and 34 m.u. (Lewis and Mathews, 1980). These genes have been called immediate early genes because they require no viral products for expression. These mRNAs probably come from the major late transcription unit which at early times produces mRNAs from the first 3' coterminal mRNA family (L1) only (Chow, Broker and Lewis, 1979; Thomas and Mathews, 1980; Lewis and Mathews, 1980; Shaw and Ziff, 1980; Kitchingman and Westphal, 1980). The pre-early genes, early region Ia, are dependent on immediate-early gene products for expression (Lewis and Mathews, 1980) and are required for expression of the other early genes. Neither Ad5 dl312 (a deletion mutant lacking EIa) nor Ad5 hr1 (a host range mutant of EIa) produce mRNA from the other early regions (Jones and Shenk, 1979; Berk et al., 1979). The early genes produce products necessary for DNA replication and are expressed in the order EIV, EIb, EIII, EIIa and EIIb (Nevins et al., 1979). Late in the early phase, the Protein IX (PIX) and IVa2 mRNAs are produced from two transcription units whose expression is also independent of DNA replication (Pettersson and Mathews, 1977; Persson, Petterson and Mathews, 1978; Galos et al., 1979; Chow, Lewis and Broker, 1980). At this time there appears to be a general depression of adenovirus specific transcription, especially of regions EII, EIII and EIV (Nevins et al., 1979; Blanton and Carter, 1979; Shaw and Ziff, 1980). Nevins and Winkler (1980) have shown that the repression of EIV, but not EII, transcription is due to DNA binding protein mediated inhibition of transcription initiation. After the onset of DNA replication

(approximately 6-8 hours p.i.), the major late transcription unit expresses 3' coterminal families L2-L5 as well as L1 (Thomas and Mathews, 1980; Shaw and Ziff, 1980). Also, during the late phase, the DNA binding protein mRNA is transcribed from a new late specific cap site (promoter) at 72 m.u. (Chow et al., 1979).

In summary, adenovirus-2 is a good model system for the study of mRNA biogenesis for the following reasons:

1) Transcription is nuclear and uses the host RNA polymerase II. All the processing events and mRNA modifications are the same as for cellular mRNAs (for a review see Nevins and Chen-Kiang, 1981). Also transcription of integrated virus genes in transformed cells can be studied.

2) The Ad2 genome contains at least 9 transcription units. Probably most mechanisms of transcriptional and post-transcriptional control are utilized in the complex expression of these transcription units. Very accurate maps of all the mRNAs are available (Berk and Sharp, 1978; Chow et al., 1979; Chow et al., 1980; Stillman et al., 1981).

3) The Ad2 genome can be cleaved with restriction enzymes to produce very specific DNA probes. Accurate restriction maps are available for many enzymes (Tooze, 1980).

4) Most mRNAs are synthesized in sufficient quantities for hybridization analyses and also direct sequence analysis. During the early phase, viral specific nuclear RNA synthesis comprises 0.5-1% of the total RNA synthesis (J. Nevins, pers. comm.) and viral specific polysomal mRNA may be as much as 18% of the total (Lindberg, Persson and Philipson, 1972). At late times, 40% of newly synthesized nuclear RNA is viral (Philipson et al, 1975) and as much as 95% of cytoplasmic polysomal mRNA is viral (Lindberg et al., 1972).

In this thesis I use adenovirus-2 as a model system to study transcription initiation and the structural features of eukaryotic RNA polymerase II promoters. The capped 5' termini of the mRNAs from 8 out of 9 transcription units were sequenced (the major late cap was sequenced by Ziff and Evans (1978)), along with the DNA encoding them. The cap sites were then identified by scanning the DNA sequences for DNA matching the cap sequences. I show that all Ad2 mRNAs, except possibly

those from the major late transcription unit and early region Ia, exhibit microheterogeneity at their 5' ends which are encoded at multiple cap sites clustered within a 2 to 7 nucleotide region. Purine caps are greatly favored over pyrimidine caps. The latter are generally neighbors of major purine starts. In addition I show that 7 out of 9 promoters have an AT rich sequence, the TATA box, in common. When this sequence is present, the 2-7 nucleotide cap region is located a fixed distance downstream. Region IV, which encodes a set of 5 staggered minor U termini and a major A terminus, has been analyzed in the related virus, Ad5, where there is an insertion of an additional T in the cap region. Ad5 EIV produces mRNA with 6 U termini instead of 5. In addition, Ad5 EIV has been analyzed in the in vitro transcription system of Manley et al. (1980). RNAs transcribed in vitro have the same proportions of A and U 5' termini as found in EIV mRNA. Taken together, the data suggest a model for RNA initiation in which the RNA polymerase measures a fixed distance from the TATA box and then scans a 2-7 nucleotide region for available purine starts.

RESULTS

Figure 1 shows a simplified mRNA map of the Ad2 genome. As mentioned above, each early region has its own promoter, demonstrated by UV and nascent chain analyses. These experiments, however, had poor resolution at the nucleotide level. Since most, if not all, capped mRNA 5' termini are derived from the initiating nucleotide of the primary transcript, sequence analysis of capped termini can be used to locate transcription initiation sites in the DNA. Likewise, the identity or nonidentity of the capped termini of two mRNAs can be used to establish that the two mRNAs are transcribed from the same or different promoters. For example, all of the mRNAs from the major late transcription unit were shown to have the same capped undecanucleotide, even though their message bodies map in different regions of the genome (Gelinas and Roberts, 1977; Ziff and Evans, 1978). Cap analysis of total Ad2 early mRNA showed that they have at least 4 (and probably many more) cap structures (Hashimoto and Green, 1976), consistent with multiple early transcription units and promoters.

In this thesis I set out to analyze the cap structures of mRNA from each Ad2 transcription unit. At first the purpose was to determine more accurately the number of transcription units within early region I (EI) and also to determine if EI mRNAs are expressed from the same promoters at early and late times after infection. I then proceeded to sequence the capped 5' termini of all Ad2 mRNAs except those from the major late transcription unit and used these RNA sequences to locate the cap sites for these mRNAs in the genomic DNA sequences. It was hoped that comparison of many cap regions would elucidate the sequence of a promoter, and also possible control regions.

The Cap Structure

A knowledge of the molecular structure of a cap, shown in Figure 2, is important for the understanding of the techniques used for isolation and sequencing of cap oligonucleotides. The cap is composed of an inverted ^{7m}G linked via a 5'-5' triphosphate linkage to the penultimate

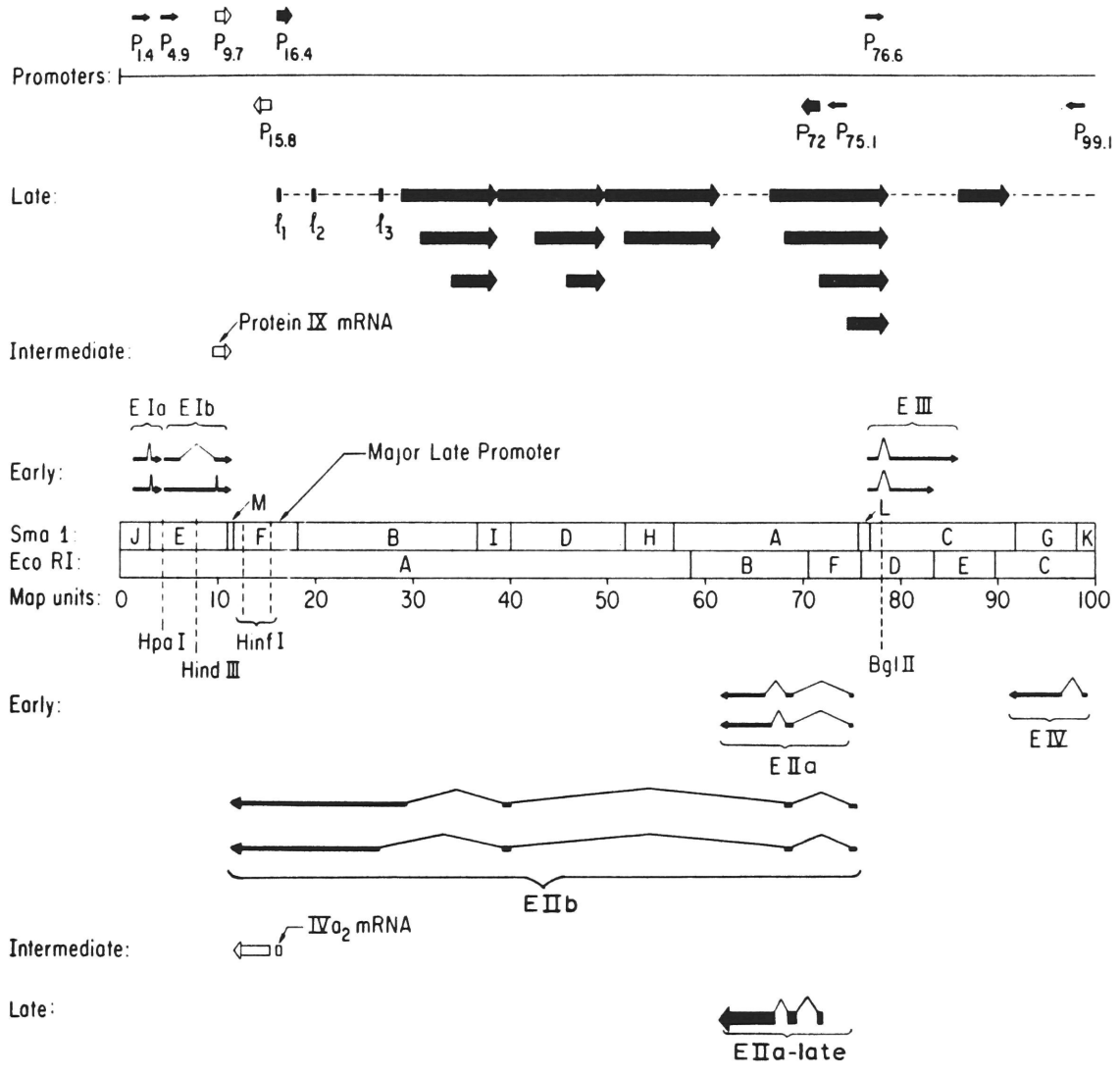


Figure 1. Promoters and Major mRNA Species of Adenovirus-2

The physical map of Ad2 is shown and is subdivided into 100 map units. Full restriction maps for Eco RI and Sma I are indicated and their respective DNA fragments identified. Selected Hpa I, Hind III, Hinf I and Bgl II cleavage sites are also shown. Ad2 RNA polymerase II promoters (P) for early (fine arrow), intermediate (open arrow) and late (thick arrow) transcription units are shown, together with the major mRNA species produced by each transcription unit. Promoter subscripts indicate the cap site coordinate. Detailed mRNA maps can be found in Berk and Sharp (1978), Chow et al. (1979), Chow et al. (1980) and Stillman et al. (1981).

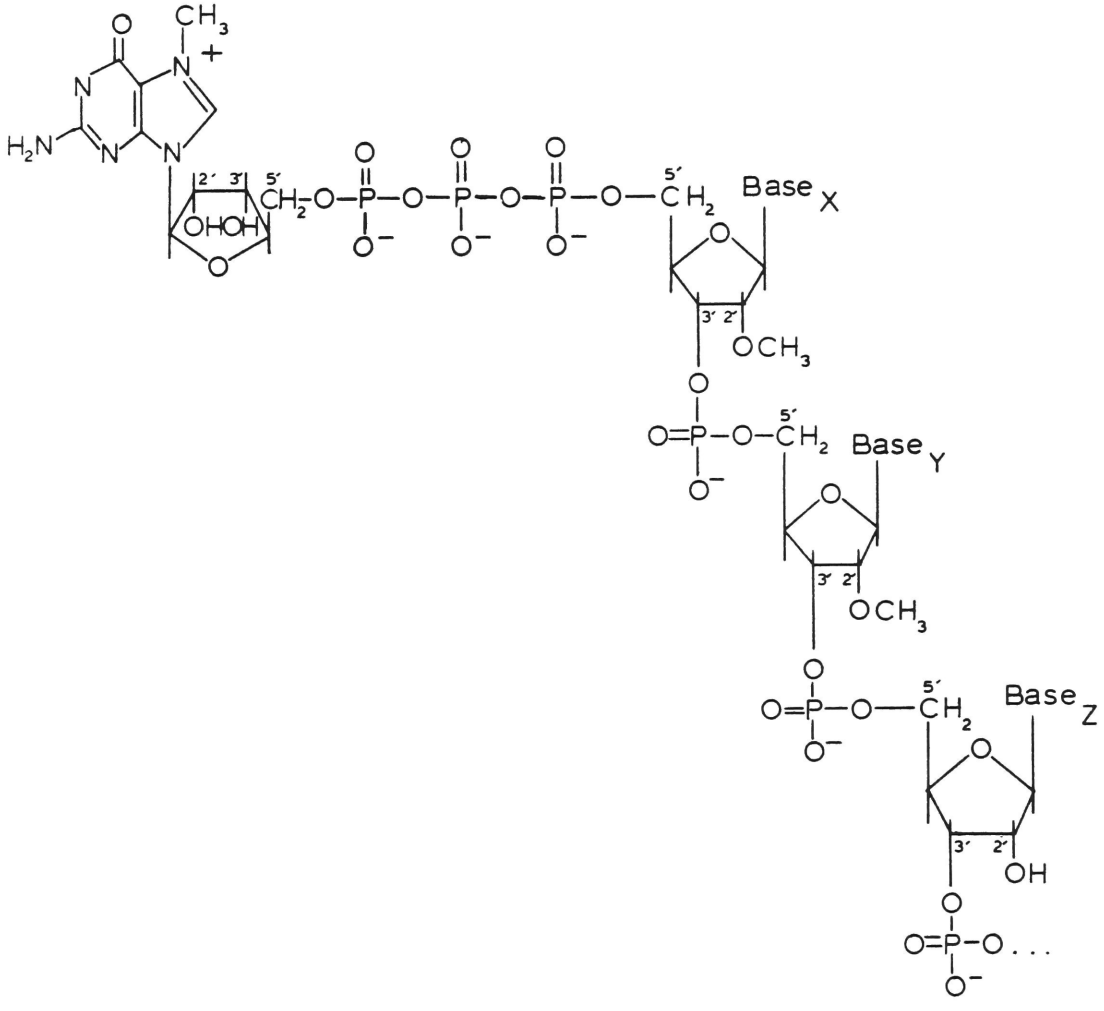


Figure 2. The Structure of an mRNA Cap

^{7m}G is linked via a 5'-5' triphosphate linkage to the 5' terminus of the mRNA. 7-Methylation of the guanine ring introduces a positive charge to the ring. 2'-O-Methylation at the Y position is optional, being present on only some of the mRNA molecules.

nucleotide (X position) of the mRNA 5' terminus. The methyl group at the 7 position of the guanine ring introduces a positive charge to the ring structure. The caps from most animal cells contain a methyl on the 2'-OH of the ribose of the penultimate nucleotide. A portion of the caps found in the cytoplasm also have a methyl on the 2'-OH of the ribose of the pen-penultimate nucleotide (Y). Those caps with only the first 2'-O-methylation are called type I caps; those with both 2'-O-methyl groups are type II caps. The peculiar nature of a cap allows it to be distinguished from the rest of the mRNA molecule by various chemical and enzymatic methods. Figure 3 shows the cleavage specificities of the RNAses and nucleases which are useful for cap oligonucleotide sequence determination. Several chemical and enzymatic techniques were considered for the isolation and analysis of cap oligonucleotides and are discussed below.

Bovine Brain Exoribonuclease

Guha (1975) isolated a 5' exoribonuclease from bovine brain which digests only RNA with a free 5'-OH. Thus 5' termini blocked by caps or phosphates are resistant to degradation. To isolate cap oligonucleotides, one would first digest the RNA with an endonuclease which leaves 5'-hydroxyls and then treat with bovine brain exoribonuclease to digest all internal oligonucleotides to mononucleotides. The cap oligonucleotide could then be separated from the mononucleotides on the basis of charge. I tried this technique on Adenovirus mRNAs and obtained an enrichment for cap oligonucleotides, but found that the level of contaminating endonuclease was too high for this technique to be used to isolate large cap oligonucleotides.

Dihydroxyboryl Columns

Dihydroxyboryl-substituted methacrylic acid polymer (Schott et al., 1973) or cellulose (Rosenberg, Wiebers and Gilham, 1972; Rosenberg, 1974) selectively and reversibly binds any molecules with cis-diols. Thus both caps and poly(A) from mRNA can be bound because they each have free 2',3'-OHs. I tried using these columns to isolate cap

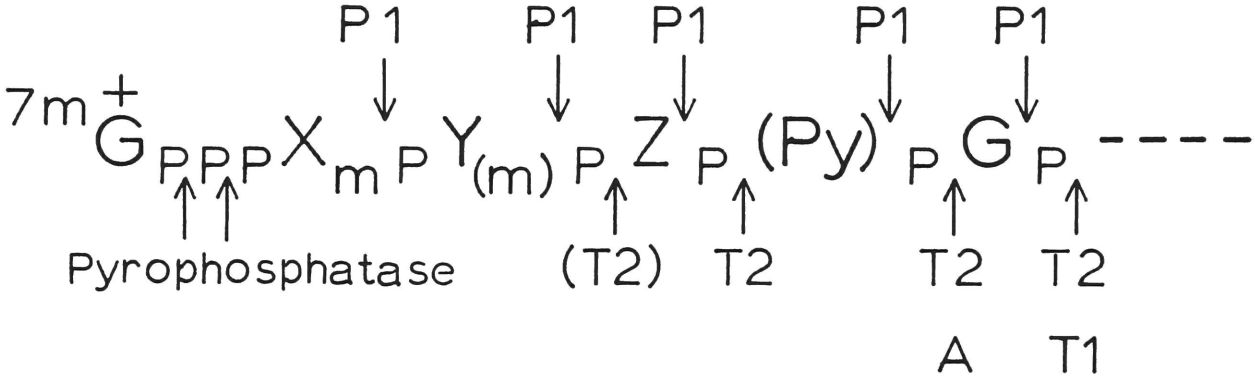


Figure 3. Cleavage Specificities of Enzymes Useful in Sequence Analysis of mRNA 5' Termini

Of the enzymes used in this study, only nucleotide pyrophosphatase cleaves pyrophosphate bonds. Nuclease P1 cleaves all phosphodiester bonds, producing 5' mononucleotides, but has difficulty cleaving phosphodiester bonds with a 2'-O-methylation (Appendix A). RNAse T1 cleaves only after G, RNAse A cleaves only after pyrimidines (C and U), and RNAse T2 has no base specificity and cleaves after all bases. These three enzymes can not cleave phosphodiester bonds with 2'-O-methyl groups, because cleavage proceeds via a 2'-3' cyclic intermediate, producing 3' mononucleotides.

oligonucleotides, but was unable even to enrich for caps. A possible reason for this will be discussed below.

Two-Dimensional RNAse T2 Fingerprints

Caps can be isolated and separated as RNAse T2 resistant cap cores by two dimensional chromatography on cellulose acetate and DEAE paper, both at pH 3.5 (Adams and Cory, 1975; Cory and Adams, 1975). Early and late Ad2 mRNA labeled with ^{32}P *in vivo* was first selected by hybridization to DNA restriction fragments, each encoding only a single mRNA 5' terminus, digested with RNAse T2, and fingerprinted (data not shown). Each fingerprint contained more than five cap spots; late mRNA from region Ib gave more than 8 spots and early mRNA from region IV gave more than 13 spots. Most spots, however, did not contain sufficient radioactivity for adequate secondary analyses. If each mRNA had only a single capped terminus, each fingerprint should have only 2 spots, the type I and type II cap cores. Part of the reason for the large number of spots is an artifactual charge heterogeneity in the cap structure. As mentioned above, 7-methylation of the guanine ring of the cap structure introduces a positive charge in the imidazole ring. This destabilizes the ring, allowing ring fission and loss of the positive charge under alkaline conditions and hydrolysis of the glycosidic linkage at neutral and acidic pHs (Hendler, Furer and Srinivasan, 1970). mRNA for this experiment had been alkali broken (pH 13.3, 0°C) to remove polyadenylated 3' termini on poly(U)-sepharose for separate analysis. These conditions would have been sufficient to ring-open at least some of the caps. The existence of 2 charge species for each cap core would have resulted in at least 4 spots for each fingerprint. As we shall see, however, part of the reason for the large number of spots is explained by the existence of more than one capped terminus for each mRNA.

Even if I had been able to unambiguously determine the sequences of all the RNAse T2 cap cores, 2-3 nucleotide sequences are insufficient to either establish the identity of two mRNA 5' termini, or to map these 5' termini within the DNA sequences. I therefore settled on the

experimental protocol shown in Figure 4 for the analysis of mRNA caps labeled with ^{32}P *in vivo*. For the preparation of early mRNA, a high moi was used to give a large number of templates for transcription. In addition, cycloheximide was present during the labeling period to: 1) inhibit DNA replication and thereby indirectly prevent late transcription, and 2) stabilize early mRNAs so that much larger amounts could be accumulated (Parsons and Green, 1971; Eggerding and Raskas, 1978). No drugs were used in the preparation of late mRNA. Before hybridization, early or late mRNA was first fragmented to 350-500 nucleotide pieces with RNAse T1 rather than NaOH, to prevent alkali induced imidazole ring fission. The mRNA was fragmented for two reasons: 1) The polyadenylated fragments were reisolated on poly(U)-sepharose for mapping of the poly(A) sites and 2) since the mRNA was hybridized to several different filter-bound DNA fragments in the same vial, I did not want to lose mRNA 5' termini by hybridization to DNA complementary to only the 3' end of the mRNA. The DNA fragments which were chosen were those which would hybridize the 5' termini of only a single family of mRNAs. However, as in the case of PIX, the DNA fragment is complementary to the message bodies of both EIb and PIX mRNAs and would have selected both 5' termini if the mRNAs had been intact. Figure 5 shows the restriction enzyme cleavage maps of Ad2 used for preparation of DNA fragments for hybridization and DNA sequence analysis. After hybridization, the filters were treated with RNAse to remove noncomplementary RNA. This was necessary to ensure that a cap oligonucleotide was encoded within the fragment used for hybridization. The eluted hybrids were then digested with RNAse T1 or A and fingerprinted. Cap oligonucleotides were identified by the assay shown in Figure 6. All spots were eluted and redigested with RNAse T2. Internal oligonucleotides release only 3' mononucleotides. Cap oligonucleotides release mononucleotides plus type I and II cap cores. All redigestions were electrophoresed at pH 3.5 on DEAE paper. The cap cores, because of their high charge, have very low mobilities and are easily distinguished from the mononucleotides. Cap cores were eluted from the DEAE paper and digested with nuclease P1 or sequentially digested with both nuclease P1 and

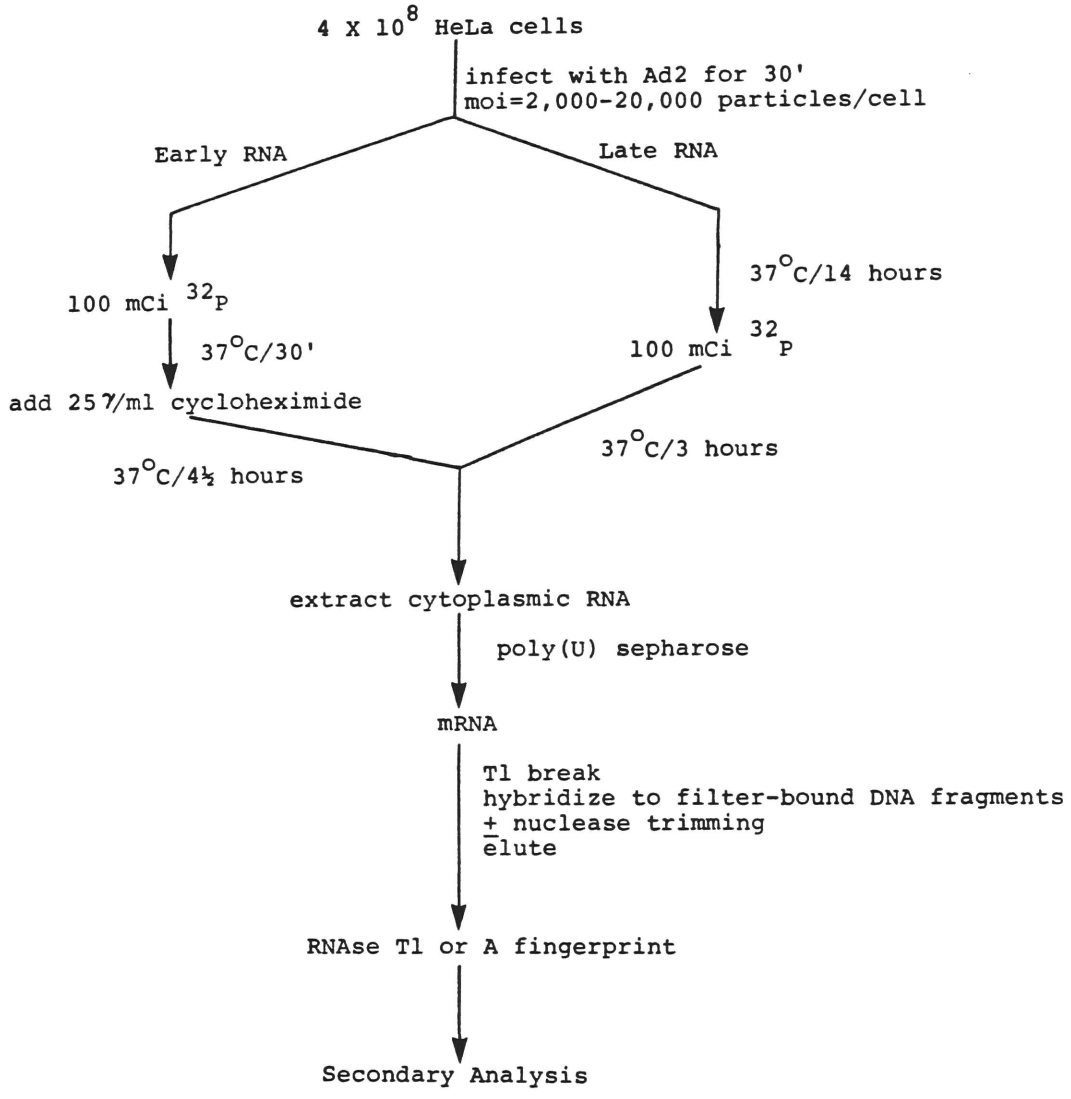


Figure 4. Preparation and Analysis of Ad2 mRNA Labeled *in Vivo* with ^{32}P

See Experimental Procedures for details. Secondary analyses are described in Figures 6 and 7 and in the legend to Table II.

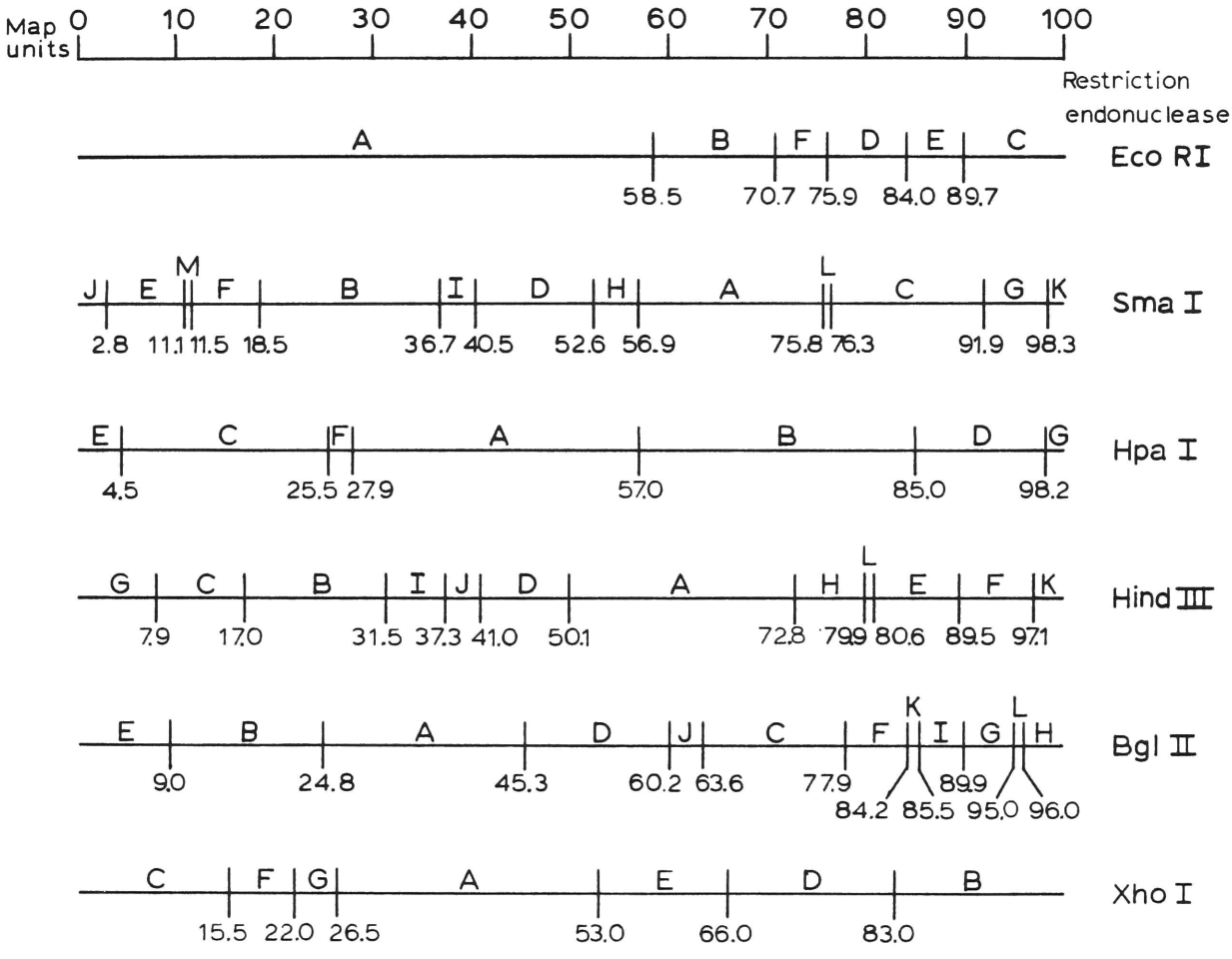


Figure 5. DNA Restriction Endonuclease Cleavage Maps of Ad2

The cleavage sites and fragment names are shown for the restriction endonucleases used in this thesis. These maps were taken from Tooze (1980).

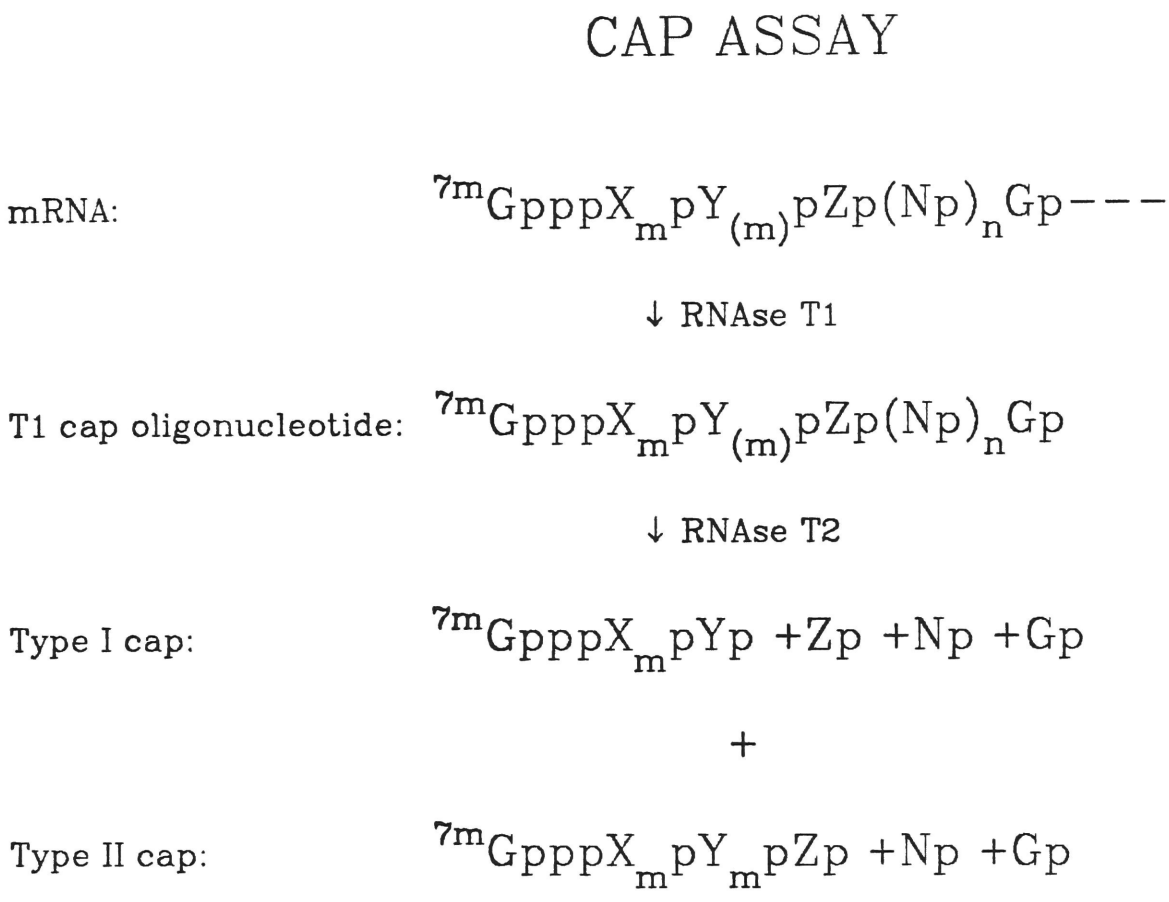


Figure 6. Use of RNAse T2 to Distinguish 5' Cap Oligonucleotides from Internal Oligonucleotides

All oligonucleotides from RNAse T1 or A fingerprints of in vivo labeled mRNA were eluted, redigested with RNAse T2, and electrophoresed on DEAE paper at pH 3.5. Internal oligonucleotides will be completely digested by RNAse T2 to mononucleotides, while cap oligonucleotides will yield highly charged RNAse T2 resistant type I and II cap cores which remain near the origin (see Figure 37G for an example). Further analysis of the cap cores is described in Figure 7.

pyrophosphatase as shown in Figure 7. Analysis of the type I cap core identifies nucleotides X and Y; analysis of the type II cap core identifies nucleotide Z.

Direct sequence analysis of *in vivo* labeled cap oligonucleotides was frequently complicated by the instability of the ^{7m}G (described above) or made impossible by insufficient radioactivity. Thus most 5' terminal sequences were obtained using a modification of the 5' terminal labeling technique of Lockard and RajBhandary (1976) outlined in Figure 8. Total unlabeled mRNA from Ad2 infected cells was oxidized with $NaIO_4$ (Whitfeld and Markham, 1953), a procedure which affects only the cis-diols of the inverted ^{7m}G of the cap and the terminal A_{OH} of poly(A). Oxidized 7-methyl guanosine was removed by β-elimination induced by the primary amine, lysine (Neu and Heppel, 1964), leaving a 5' terminal triphosphate. The phosphates were removed with RNAse-free alkaline phosphatase and the 5'-OH was labeled by the polynucleotide kinase catalyzed transfer of ^{32}P from γ-^{32}P-ATP. The 5' labeled RNA was then selected by hybridization and fingerprinted as already described for *in vivo* labeled RNA, except that hybrids were not nuclease treated. In theory, only the 5' terminal oligonucleotides will be labeled, but in practice there is some internal nicking and kinasing of the RNA which results in a background of spots. Since the 5' labeling procedure does not remove 2'-O-methyl groups, true 5' terminal oligonucleotides can be distinguished from internal oligonucleotides by resistance to RNAse T2. Enzymatic sequence analysis of 5' terminal oligonucleotides will be described below.

The application of these sequencing techniques to 9 Ad2 transcription units will now be described. Each transcription unit will be discussed separately.

The Transforming Region

The left 11% of the Ad2 genome contains the transforming region and is composed of 3 transcription units, EIA, EIb and PIX, each with its own RNA initiation site (Sehgal et al., 1979; Wilson et al., 1979). The mRNAs from these transcription units are shown in Figure 9. Two of

ANALYSIS OF RNASE T2 CAPS

Type I cap:

$$^{7m}GpppX_mpYp \xrightarrow{P1} {}^{7m}GpppX_m + pY + p_i$$

$$^{7m}GpppX_mpYp \xrightarrow{P1+pyro.} {}^{7m}ppG + pX_m + pY + p_i$$

Type II cap:

$$^{7m}GpppX_mpY_mpZp \xrightarrow{P1} {}^{7m}GpppX_m + pY_m + pZ + p_i$$

$$^{7m}GpppX_mpY_mpZp \xrightarrow{P1+pyro.} {}^{7m}ppG + pX_m + pY_m + pZ + p_i$$

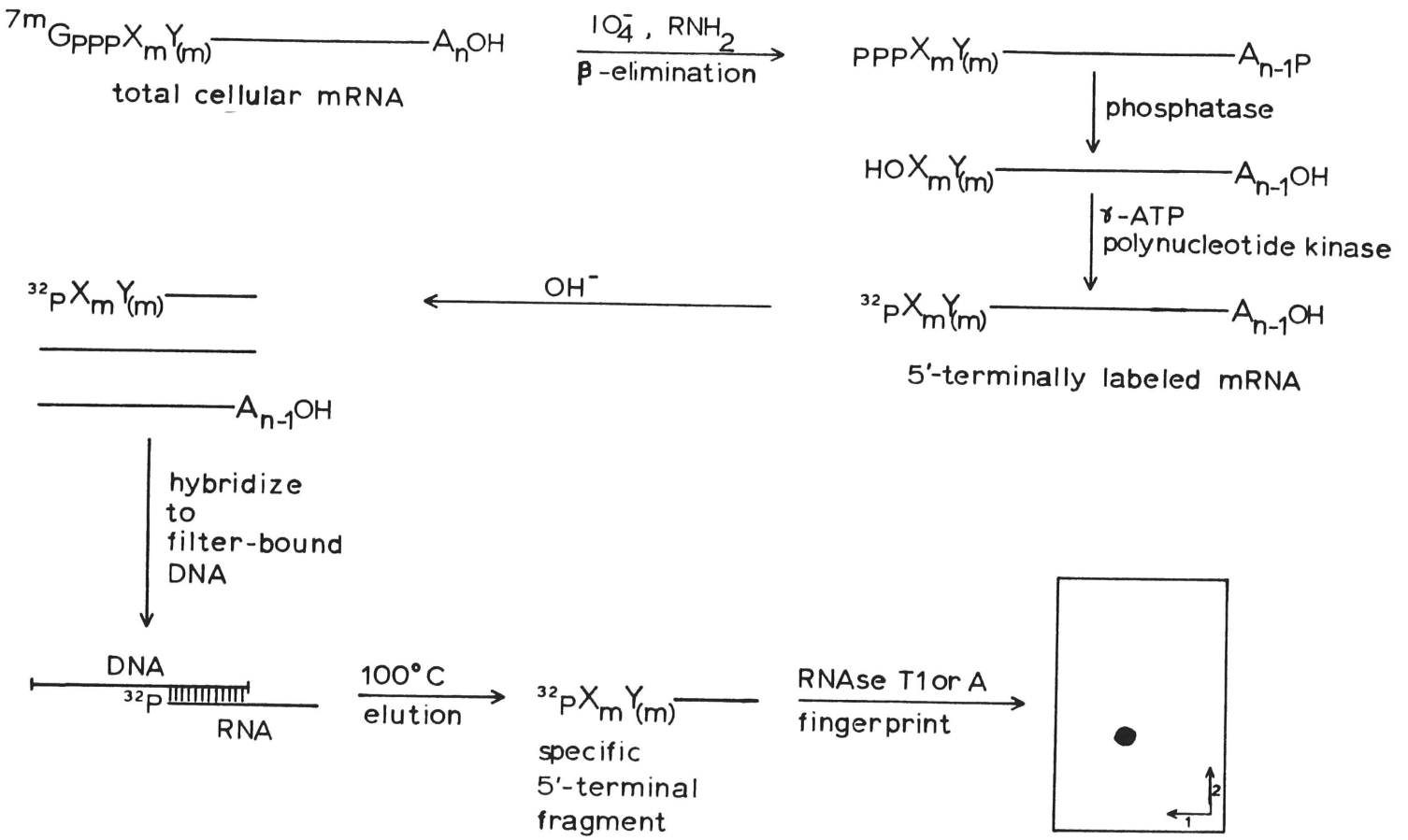
Figure 7. Sequence Analysis of RNAse T2 Resistant Cap Cores

The RNAse T2 resistant type I and II cap cores separated by electrophoresis at pH 3.5 on DEAE paper were redigested with nuclease P1 or nuclease P1 plus pyrophosphatase and electrophoresed on Whatman 540 paper at pH 3.5. Nuclease P1 digestion of the type I cap core identifies Y and combined digestion with P1 and pyrophosphatase identifies X. Analysis of the type II cap core can then be used to identify Z.

Figure 8. *In Vitro* 5' Labeling of mRNA

This procedure is used to produce the end labeled 5' terminal oligonucleotides which are analyzed in Table I and Figure 11. Details are presented in Experimental Procedures.

DECAPPING AND KINASING OF mRNA



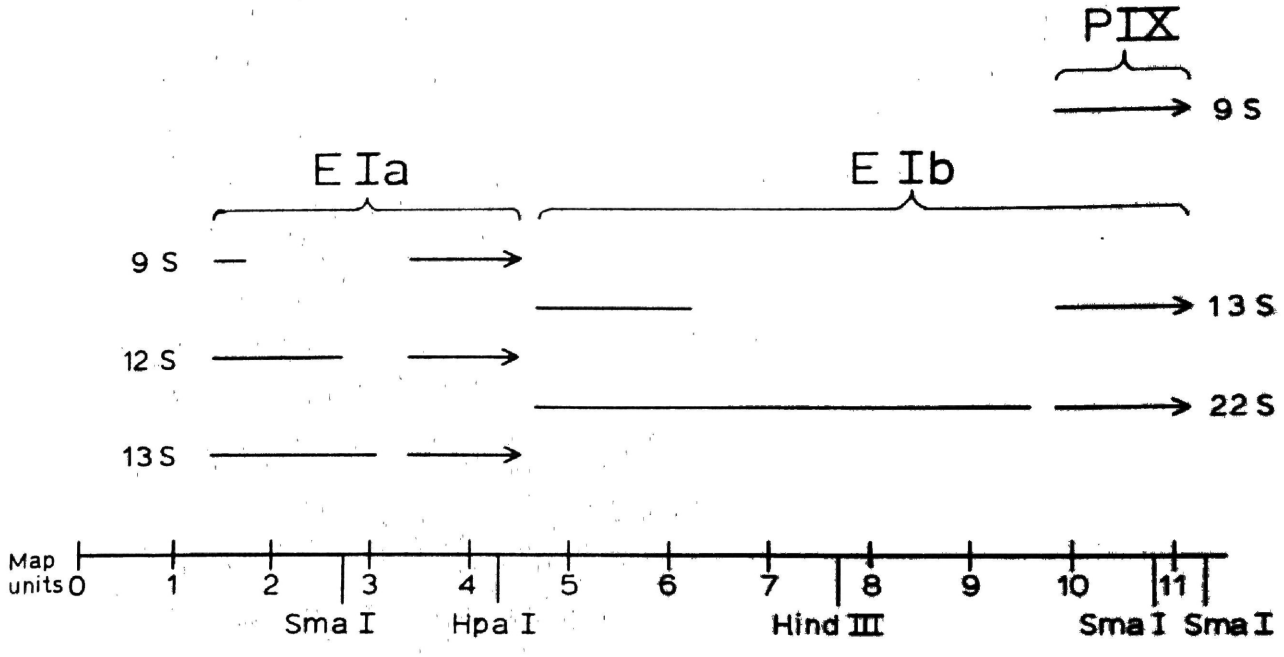


Figure 9. mRNAs of the Transforming Region

The mRNAs of the transforming region were mapped by S1 (Berk and Sharp, 1978) and electron microscopy (Chow et al., 1979, 1980). The exact splice points and 3' termini were determined from the cDNA sequences of the EIa 12S and 13S mRNAs (Perricaudet et al., 1979) and the EIb 13S and 22S mRNAs (Perricaudet et al., 1980). Nucleotide positions for these points and for restriction enzyme cleavage sites were converted to map units using the Ad5 DNA sequence (van Ormondt et al., 1980) as a guide and the conversion of 1 map unit=365 nucleotides (Chow et al., 1980).

these transcription units, EIa and EIb, function throughout infection (Spector, McGrogan and Raskas, 1978), while the third, for the protein IX mRNA, begins expression at intermediate times, approximately 6-8 hours post-infection (Persson et al., 1978). The splicing patterns of the steady state EIa and EIb mRNAs also change as infection proceeds (Spector et al., 1978; Chow et al., 1980; Wilson et al., 1980).

The EIa 12S mRNA is the predominant species present at early times, while the 9S is predominant late (Chow et al., 1980). The 12S and 13S mRNAs have been cloned and sequenced, so the splice junctions are known (Perricaudet et al., 1979). However, the clones did not extend all the way to the cap. Region Ia is both necessary and sufficient for cell transformation (Van der Eb et al., 1977). A product from EIa may also be required to activate transcription from the other early regions (Jones and Shenk, 1979; Berk et al., 1979).

The EIb 22S mRNA is the predominant species early in the presence of cycloheximide; however at late times and in the absence of cycloheximide the 13S mRNA is the predominant species. The 22S and 13S mRNAs have been cloned and partially sequenced by Perricaudet, Le Moullec and Pettersson (1980). The 22S and 13S mRNAs encode the 67K and 20K T antigens which are important for complete transformation of cells (van der Eb et al., 1977; van der Eb et al., 1980).

The 9S mRNA which encodes the minor virion component IX maps between 9.9 and 11.2 m.u. (Chow et al, 1977b). Both the genomic DNA encoding PIX mRNA and a cDNA clone have been sequenced (Alestrom et al., 1980), proving that the PIX mRNA is unspliced. Although PIX mRNA is not expressed until very late in the early phase (6-8 hours), it is still expressed in the presence of cycloheximide and therefore in the absence of DNA replication (Persson et al., 1978), in contrast to the late RNAs which require the onset of DNA replication for expression.

The entire transforming region of Ad5 has been sequenced (van Ormondt et al., 1978; Maat and van Ormondt, 1979; Maat, van Beveren and van Ormondt, 1980; Van Ormondt, Maat and Van Beveren, 1980).

EIa mRNA

Figure 10A shows the result when early mRNA labeled *in vivo* with ^{32}P was hybridized to coordinate 0-3 DNA (Sma I J) to select the EIa 5' end. Some preparations of late mRNA yielded a very similar fingerprint (not shown), while others yielded a simpler fingerprint, probably due to varying abundances of the different spliced species. All spots were eluted, redigested with RNAse T2, and electrophoresed on DEAE paper. The spot indicated by the arrow yielded type I and II cap cores (identified by their low electrophoretic mobility) plus Cp, 2 Up and Gp with RNAse T2 digestion. The type II core released pC and pU plus P1 resistant core with nuclease P1 and pA, pC and pU with combined P1 and pyrophosphatase digestion. The type I core was contaminated with the type II core. These data are consistent with the structure $^{7m}GpppA_m(CU)p$. Late mRNA also gave this capped oligonucleotide, with identical digestion products.

When early mRNA was 5' labeled *in vitro* and selected by 0-3 DNA, only a single 5' terminal oligonucleotide, shown in Figure 10B, was obtained. Nuclease P1 digestion yielded pA (Table I and Appendix A), and the type I and type II 5' labeled oligonucleotides released by RNAse T2 were the same as those released by RNAse A, as seen in Figure 11, lanes 1-2. Therefore the 5' sequence was deduced to be pAPyPy . From the DEAE mobilities in Figure 11, the 5' sequence is $pA_mC_{(m)}U-$. This was confirmed by the nuclease P1 partial digestion products given in Table I. [A discussion of the incomplete digestion by nuclease P1 and its usefulness in sequencing is described in Appendix A.] The same 5' terminal sequence was obtained from late mRNA, except that the methylation pattern was mostly type II. A partial RNAse T2 digest of the late 5' terminal oligonucleotide, shown in Figure 10C, which was analyzed by the mobility shift fingerprinting rules (Galibert et al., 1974) given in Figure 10D, identified the sequence -CUUGp at the oligonucleotide's 3' end. Partial RNAse T2 digestion (not shown) of the early 5' terminal oligonucleotide yields -UCUUGp because a significant proportion of the molecules are type I. The 5' sequence of EIa mRNA was thus determined to be $^{7m}GpppA_mC_{(m)}UCUUG-$.

Figure 10. The 5' Terminus of Region Ia mRNA

(A) RNAse T1 fingerprint of early mRNA hybrid to 0-3 m.u. DNA. Polyadenylated cytoplasmic RNA was isolated from Ad2 infected cells labeled with ^{32}P in the presence of cycloheximide 0-5 hr post-infection, cleaved with RNAse T1 to approximately 350-500 nucleotides and hybridized to 0-3 m.u. DNA (Sma I J) . After RNAse T1 trimming and elution, the RNA was fingerprinted. The arrow indicates the only cap oligonucleotide, as determined by RNAse T2 digestion.

(B) 5' terminus of EIa mRNA labeled by kinasing. Cold polyadenylated cytoplasmic RNA, isolated at 5 hrs p.i. from Ad2 infected cells incubated in the presence of cycloheximide, was decapped, kinased, hybridized to Sma I J, and RNAse T1 fingerprinted. RNAse T2 resistance of spot 1 confirmed that it was from a 5' terminal cap structure.

(C) Sequence of EIa 5' terminus. The late 5' terminal oligonucleotide, equivalent to spot 1 of panel (B), was partially digested with RNAse T2 and fingerprinted to give a series of labeled 5' coterminal oligonucleotides. Loss of a single nucleotide from the 3' end of an oligonucleotide produces a mobility shift which may be interpreted with the aid of the diagram in panel (D). The sequence of the 3' portion of spot 1 deduced from the fingerprint in panel (C) is thus -CUUGp.

In all fingerprints the first dimension is electrophoresis at pH 3.5 from right to left and the second dimension is homochromatography from bottom to top. The xylene cyanol F.F. dye is indicated by b and the Orange G by y.


A
b
y
B
b
1
C
y
C
U
U
G
D
A
G
C
U

Table I. Nuclease P1 Analyses of 5' Terminal Oligonucleotides.

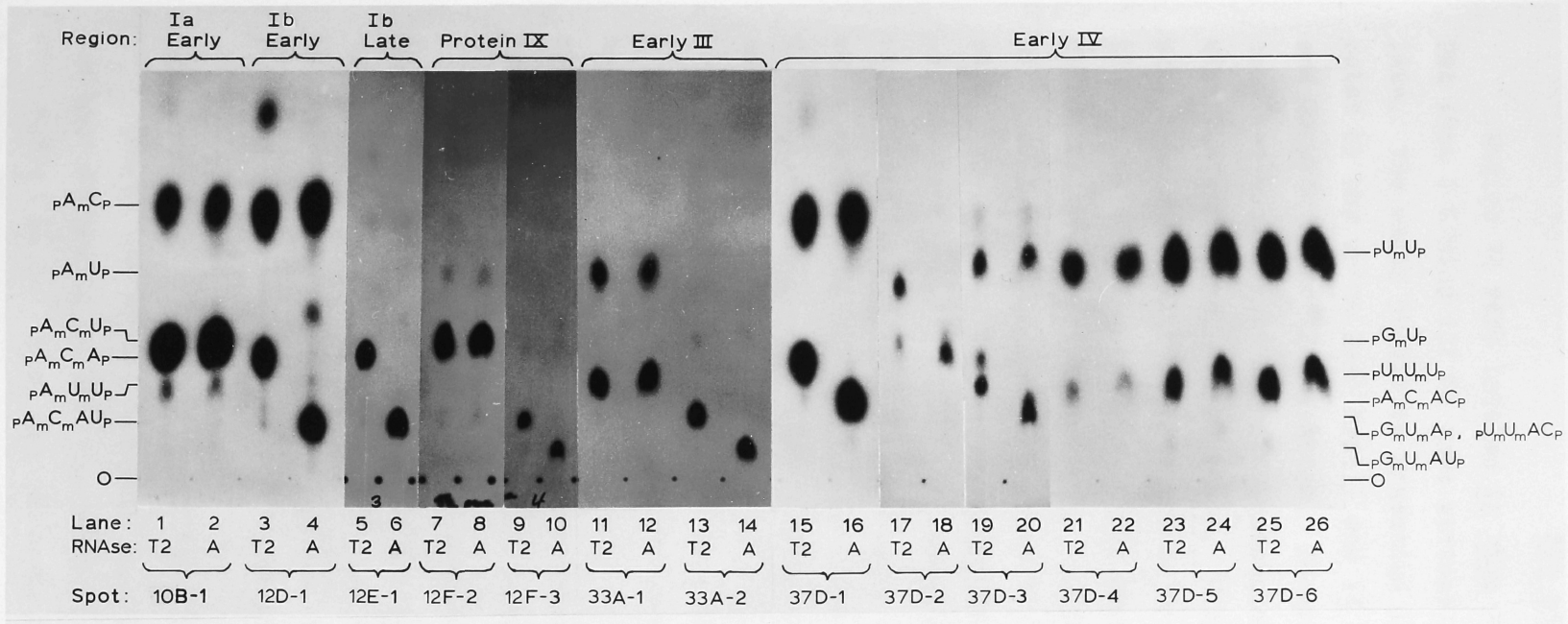
Figure	Spot Number[a]	Complete P1 Product[b]	Partial P1 Products[c]
10B	1	pA	pAC,pACU
12D	1	pA	pAC,pACA
12E	1	pA	pAC,pACA
12E	2	pA	pAU
12F	1	pA	pAC,pACA
12F	2	pA	pAU
12F	3	pG	-
18A	1	pA	pAG
18A	2	pU	pUC
18A	3	pA	pAG
18A	4	pA	pAC,pACU
18B	1	pU(>90%),pC(<10%)	p(UC)
18B	2	pA	pAC
18B	3	pA	pAC,pACU
18B	4	pA	pAG
18B	5	pU	pUC
28A	1	pG	-
28A	2	pA	-
30A	1	pA	-
30A	2	pA	pAC,pACA
30A	3	pA	-
30B	1	pA	pAC
30B	2	pA	-
30B	3	pA	-
30B	4	pA	-
33A	1	pA	pAUU
33A	2	pG	-
37D	1	pA	pAC,pACA
37D	2	pU	pUAC
37D	3	pU	pUU,pUUA
37D	4	pU	pUU
37D	5	pU	pUU,pUUU
37D	6	pU	pUU,pUUU

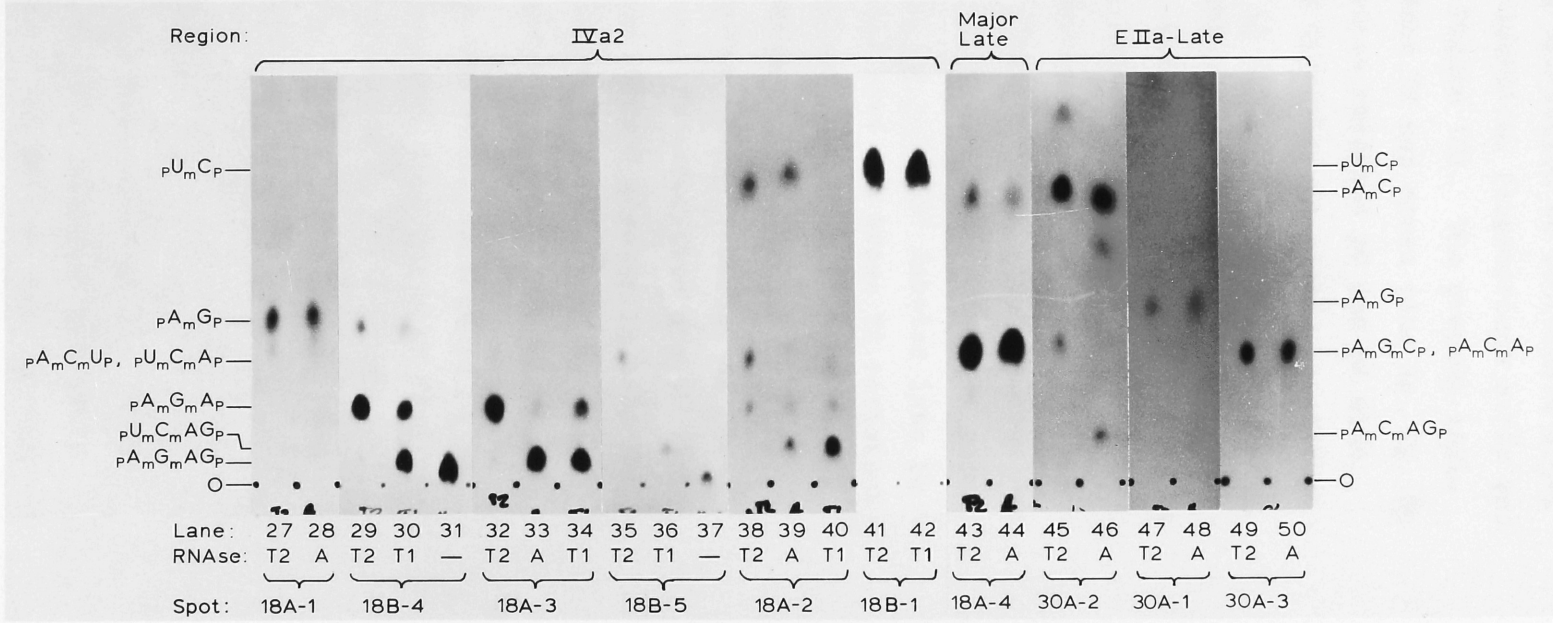
a. Spot number refers to the corresponding spot in the fingerprint in the given figure.
b. This is the product produced when nuclease P1 digestion is complete and represents the 5' terminal nucleotide. Products were separated by electrophoresis at pH 3.5 on Whatman 540 paper.
c. These are the additional products produced when nuclease P1 digestion is not complete. Cleavage by nuclease P1 is greatly inhibited by 2'-O-methylations, especially on pyrimidine nucleosides (Yamada and Ishikura, 1975). For details see Appendix A.

Figure 11. Sequencing of 5' Residues of Ad 2 mRNAs by Secondary Analysis of *in Vitro* Labeled 5' Terminal Oligonucleotides

Messenger RNA from Ad 2 infected cells was decapped and labeled with ^{32}P at the 5' terminus by polynucleotide kinase as described in Experimental Procedures. This RNA was partially degraded with alkali and 5' fragments of particular messengers were selected by hybridization to appropriate DNA fragments and fingerprinted with RNAse T1 or RNAse A in Figures 10, 12, 18, 28, 30, 33 and 37. All spots derived from cap structures as identified by RNAse T2 resistance were subjected to secondary analysis. Shown here are the pH 3.5 DEAE paper electrophoresis fractionations of RNAse T2 and RNAse T1 or RNAse A (where appropriate) digests of the 5' terminal oligonucleotides. The spot origins and enzymes for digestion are given below the panels. For example, lane 1 is the RNAse T2 digest of spot 1 of Figure 10B. mRNA regions are given above the panels. The origin for electrophoresis is indicated by O.

Each RNAse T2 product found above will also be obtained with either RNAse T1 (implying 3' G), with RNAse A (implying 3' C or U) or with none of these (implying 3' A). When RNAse T2 products from type I cap structures (high mobility) and from type II cap structures (low mobility) are compared with their respective RNAse T1 or RNAse A products and the result of P1 nuclease digestion is considered, short sequences at the 5' ends may be deduced. These are given on the sides of the panels. Simultaneous analysis of many products, and consideration of the M values for mobility changes (Sanger et al., 1965) and the nuclease P1 partial digestion products given in Table I also contribute to sequence deduction.


Region: Ia Early
Ib Early
Ib Late
Protein IX
Early III
Early IV
pAmCp
pAmUp
pAmCmUp
pAmCmAp
pAmUmUp
pAmCmAUp
O
pUmUp
pGmUp
pUmUmUp
pAmCmACp
pGmUmAp, pUmUmACp
pGmUmAUp
O
Lane: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
RNAse: T2 A T2 A T2 A T2 A T2 A T2 A T2 A T2 A T2 A T2 A T2 A T2 A T2 A T2 A
Spot: 10B-1 12D-1 12E-1 12F-2 12F-3 33A-1 33A-2 37D-1 37D-2 37D-3 37D-4 37D-5 37D-6


Region: IVa2
Major Late
EIIa-Late
pUmCp
pAmGp
pAmCmUp, pUmCmAp
pAmGmAp
pUmCmAGp
pAmGmAGp
O
pUmCp
pAmCp
pAmGp
pAmGmCp, pAmCmAp
pAmCmAGp
O
Lane: 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
RNAse: T2 A T2 T1 — T2 A T1 T2 T1 — T2 A T1 T2 T1 T2 A T2 A T2 A T2 A
Spot: 18A-1 18B-4 18A-3 18B-5 18A-2 18B-1 18A-4 30A-2 30A-1 30A-3

EIb mRNA

Region Ib mRNA labeled *in vivo* and hybridized to coordinate 4.5-8 DNA (Hpa I E/Hind III G) gave virtually identical fingerprints early and late. The early result is presented in Figure 12A. The product indicated by the arrow gave type I and II RNAse T2 cap cores plus Ap, Cp, Up and Gp with RNAse T2 digestion. These cores released pC and pA with combined nuclease P1 and pyrophosphatase digestion in yields consistent with the structure $^{7m}GpppA_m(CA)p$. With *in vitro* 5' labeling, the early mRNA hybrid to 3-11 DNA gave one major RNAse T2 resistant oligonucleotide, spot 1 of Figure 12D. Minor spots were not investigated. From P1 digestion, presented in Table I, and the RNAse T2 and RNAse A analyses in lanes 3 and 4 of Figure 11, the 5' sequence was deduced to be $pA_mC_{(m)}AU-$. RNAse T2 partial digestion fingerprinting gave the sequence -AUCUGp (not shown). Thus the sequence of the EIb 5' terminal oligonucleotide is $pA_mC_{(m)}AUCUGp$. In Figure 12E, *in vitro* 5' labeled late Ib mRNA, selected by hybridizing to 4.5-8 DNA, gave two RNAse T2 resistant oligonucleotides. The major oligonucleotide, spot 1, was shown by the data of Table I, Figure 11, lanes 5-6, and the partial RNAse T2 fingerprint of Figure 12G to have the same sequence, $pA_mC_{(m)}AUCUGp$, as the major early Ib mRNA 5' terminal oligonucleotide identified above. Note however that the proportion of type II cap is much greater in late RNA than early RNA. The minor oligonucleotide, spot 2, analyzed by nuclease P1, RNAse T2 and RNAse A had the 5' sequence $pA_mU_{(m)}C-$ (data not shown).

Early mRNA was always isolated from cycloheximide treated cells, as discussed above. It has been shown, however, that the mRNA from EIa and EIb have the same 5' termini when synthesized in the presence or absence of cycloheximide (Fred Cross, pers. comm.). The proportions of the type I and II caps may differ slightly, however.

Protein IX mRNA

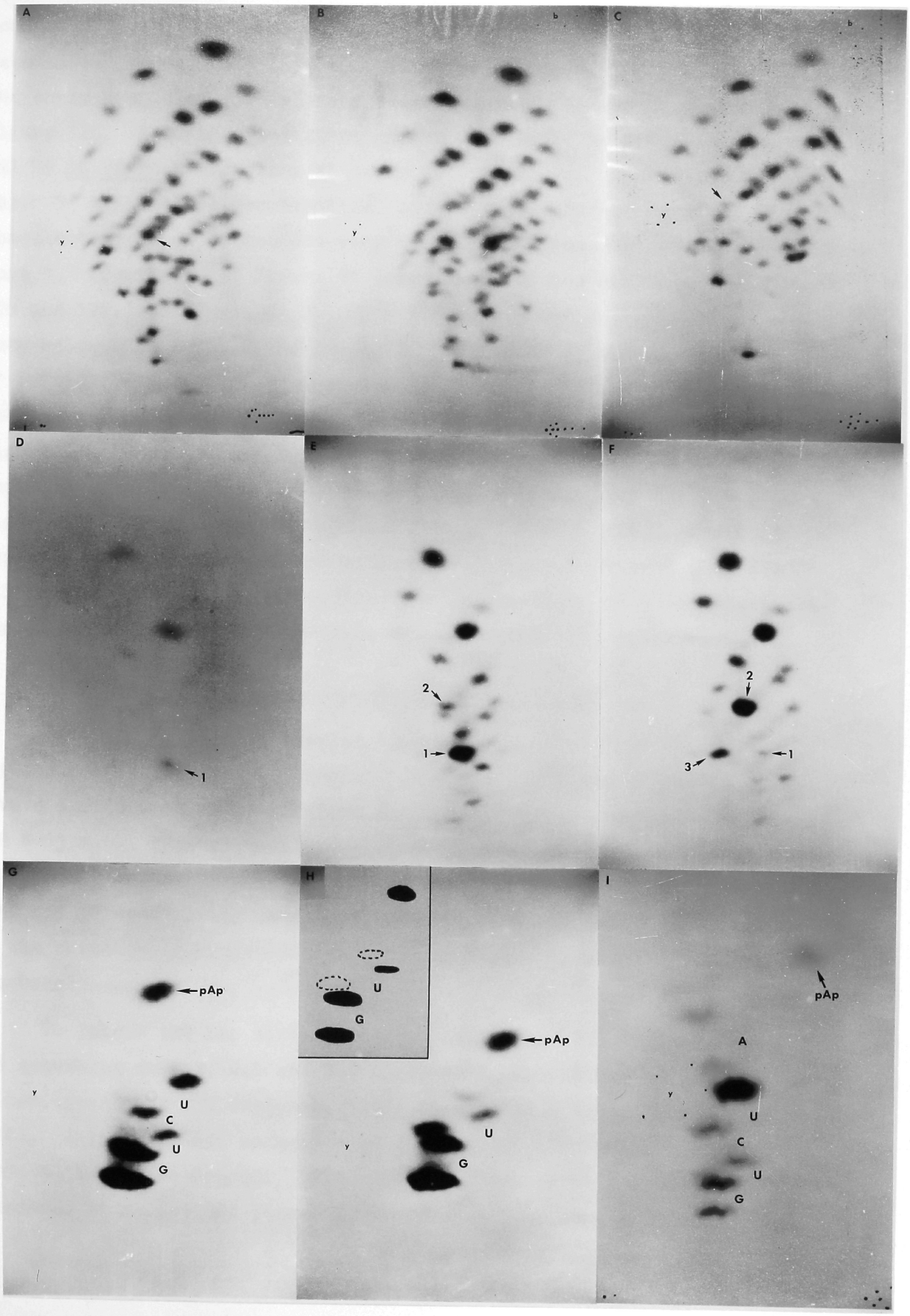
The 5' end of the PIX mRNA maps at 9.9 m.u. (Chow et al., 1979) and therefore will be selected by 8-11 m.u. DNA. Although the early mRNA hybrid to 8-11 m.u. DNA shown in Figure 12B gave no cap, as assayed by RNAse T2 resistance, a new RNAse T2 resistant oligonucleotide,

Figure 12. The 5' Termini of Region Ib and Protein IX mRNAs

(A), (B) and (C) are RNAse T1 fingerprints of *in vivo* labeled mRNA. (A) is the 5' portion of early Ib mRNA selected by 4.5-8 m.u. DNA (Hpa I E/Hind III G). (B) is the 3' portion of early Ib mRNA selected by 8-11 DNA (Sma I E/Hind III C). (C) is the RNAse T1 fingerprint of late mRNA selected as in (B) and consists largely of PIX mRNA. All hybrids were RNAse T1 trimmed before elution. Major capped oligonucleotides were identified by RNAse T2 digestion and are indicated by arrows.

(D), (E) and (F) are RNAse T1 fingerprints of mRNA labeled at 5' termini by polynucleotide kinase. (D) is early mRNA hybridized to 3-11 DNA (Sma I E) and shows the major early Ib 5' terminal oligonucleotide, spot 1. (E) is late mRNA hybridized to 4.5-8 DNA (Hpa I E/Hind III G) and shows spot 1, the major, and spot 2, the minor 5' terminal oligonucleotide of late region Ib mRNA. (F) is late mRNA hybridized to 8-11 DNA (Sma I E/Hind III C) and shows spots 2 and 3, the major and minor PIX 5' terminal oligonucleotides, and spot 1, a contaminant from Ib mRNA (see Results).

(G), (H) and (I) are fingerprints of partial RNAse T2 digests of 5' labeled oligonucleotides from mRNA termini. (G) is the major late Ib 5' terminal oligonucleotide, spot 1 of panel (E). (H) is the major PIX 5' terminal oligonucleotide, spot 2 of panel (F). The inset distinguishes the major product trail (solid spots) from paired minor products (open spots) (which potentially have 2'-3' cyclic phosphates). Major and minor products of a pair were identical on DEAE electrophoresis at pH 3.5 (not shown). (I) is the minor PIX 5' terminal oligonucleotide, spot 3 of panel (F). The sequences deduced in panels (G)-(I) are respectively -UCUGp, spot 1; -UGp, spot 2; and -AUCUGp, spot 3. The pAp spot in panels (G)-(I) is from background contaminants and is not part of the sequencing trail.


A
B
C
D
E
F
G
H
I
pAp
U
C
G
A
1
2
3

indicated by the arrow in Figure 12C, was found with in vivo labeled late mRNA. We estimate that this late-specific product is present in the early hybrid of Figure 12B at less than 1% of the late yield seen in Figure 12C. I further investigated the late-specific products from 8-11 DNA by in vitro 5' labeling as shown in Figure 12F. I obtained one major 5' terminal oligonucleotide, spot 2, one minor 5' terminal oligonucleotide, spot 3, and one very minor 5' terminal oligonucleotide, spot 1. From Table I, Figure 11 lanes 7-10 and the partials of Figures 12H and 12I, spot 2 was $pA_mU_{(m)}CUGp$ and spot 3 was $pG_mU_{(m)}AUCUGp$. I show below that spots 2 and 3 derive from the major and minor PIX mRNA 5' termini, respectively. Spot 1 starts $pA_mC_{(m)}AU-$ and is probably from intact region Ib mRNA which hybridized to the 8-11 m.u. DNA. Spot 2 of Figure 12E (from a minor late Ib terminus) probably has the same sequence as the major PIX mRNA 5' terminal oligonucleotide, spot 2 of Figure 12F. To check that these were indeed from distinct termini I repeated the hybridization using cloned 4.5-8 m.u. and 8-11 m.u. DNA. All results were the same, confirming that the minor Ib 5' terminal oligonucleotide is a separate product and was not a PIX contaminant.

Three Cap Regions at the Left End

Because the in vivo labeled caps were isolated from nuclease trimmed hybrids, it can be concluded that EIa, EIb and PIX mRNAs each have a distinct major 5' terminus encoded between 0-3 m.u., 4.5-8 m.u., and 8-11 m.u., respectively. Furthermore, the EIa and EIb mRNA 5' termini do not change structure early versus late in infection. The PIX mRNA cap is essentially undetectable at early times (when infection is in the presence of cycloheximide), but is abundant late (>100X increase).

To locate the cap sites of region Ia, Ib and PIX mRNAs, I scanned DNA sequences at 1.3, 4.6 and 9.9 m.u., their respective 5' coordinates (Kitchingman, Lai and Westphal, 1977; Chow et al., 1979; Berk and Sharp, 1978). Initially, Ad5 sequences at 1.3 and 4.6 (Van Ormondt et al., 1978; Maat and Van Ormondt, 1979) and Ad5 (Maat et al., 1980) and Ad2 (Alestrom et al., 1980) sequences around 9.9 m.u. were examined. (Ad2

and Ad5 share approximately 99% sequence homology at the left end.) In addition I sequenced the Ad2 genome in the regions of 1.3 and 4.6 m.u. by the strategies shown in Figure 13 A and B, respectively. The Ia 5' terminus, $^{7m}GpppA_mC_{(m)}UCUUG-$, is encoded at nucleotide 499 (1.37 m.u.) in the Ad5 sequence and nucleotide 138 of the Ad2 sequence presented in Figure 14. The major Ib 5' terminus, $^{7m}GpppA_mC_{(m)}AUCUG-$, is present on >95% of Ib mRNA and is encoded at nucleotide 1702 (4.66 m.u.) in the Ad5 sequence and nucleotide 361 of the Ad2 sequence presented in Figure 15. The minor Ib 5' terminus, $^{7m}GpppA_mU_{(m)}CUG-$, is present on <5% of Ib mRNA and can be accommodated 2 nucleotides downstream from the major EIb cap site in both Ad2 and Ad5. The major PIX mRNA 5' terminus, $^{7m}GpppA_mU_{(m)}CUG-$, is present on >75% of this mRNA and is found at nucleotide 3583 (9.8 m.u.) in Ad5 and in the same location and nucleotide sequence in Ad2, as shown in Figure 16. This agrees with the cap sequence and location determined by Alestrom et al. (1980) for PIX. The minor PIX mRNA 5' terminus, $^{7m}GpppG_mU_{(m)}AUCUG-$, is present on less than 25% of this mRNA and can be accommodated 2 nucleotides upstream from the major PIX cap site in both Ad2 and Ad5 (at nucleotide 3581 of Maat et al., 1980). Thus the Ia mRNAs have a single 5' terminus, while Ib and PIX mRNAs have heterogeneous 5' termini.

IVa2 mRNA and Major Late mRNAs

The region around 16 m.u. contains two diverging transcription units. The major late transcription unit produces most of the late rightward mRNAs, all of which have a short first leader encoded at 16.5 m.u. (Figure 1) (Chow and Broker, 1978), and encode most of the virion structural proteins. The major late promoter is also active at early times, and at the same level as the early promoters (Shaw and Ziff, 1980). The major late mRNAs' 5' terminus and promoter (16.45 m.u.) were the first 5' terminus and promoter in Ad2 to be sequenced (Ziff and Evans, 1978).

The IVa2 mRNA is expressed at intermediate to late times (Shaw and Ziff, 1980) from the l strand and has a 5' terminus mapped at 16.1 m.u. (Chow and Broker, 1978). The synthesis of this mRNA is blocked by

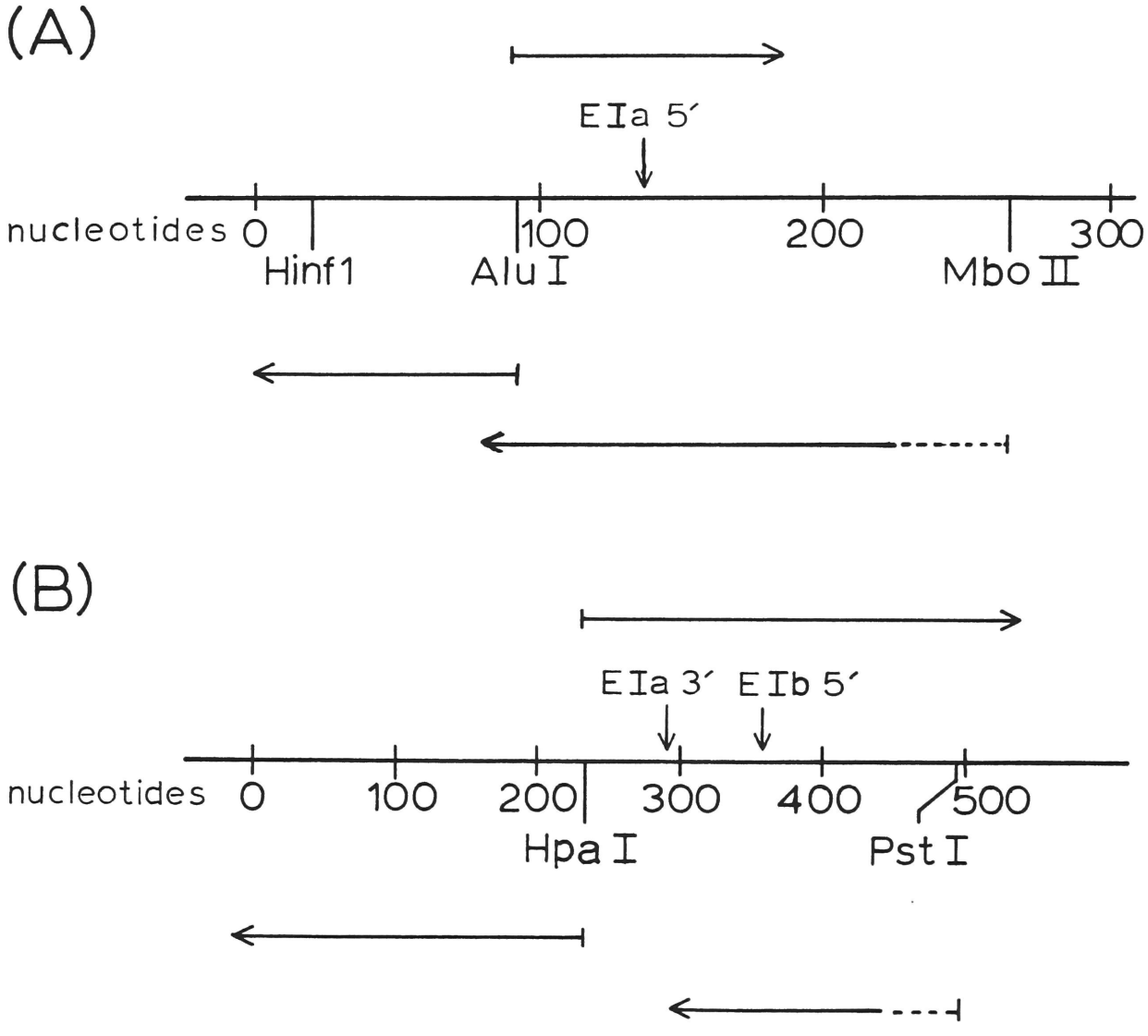


Figure 13. DNA Sequencing of the EIa and EIb Cap Regions

(A) and (B) show the strategy for sequencing the EIa and EIb cap regions. Nucleotide numbers refer to the sequences presented in Figures 14 and 15, respectively. Solid arrows above the line indicate sequence obtained of the l strand and below the line of the r strand. Dashed portions of arrows represent sequence information which was not obtained from a kinased fragment. The positions in the sequence of the EIa and EIb 5' termini and EIa 3' terminus are indicated.

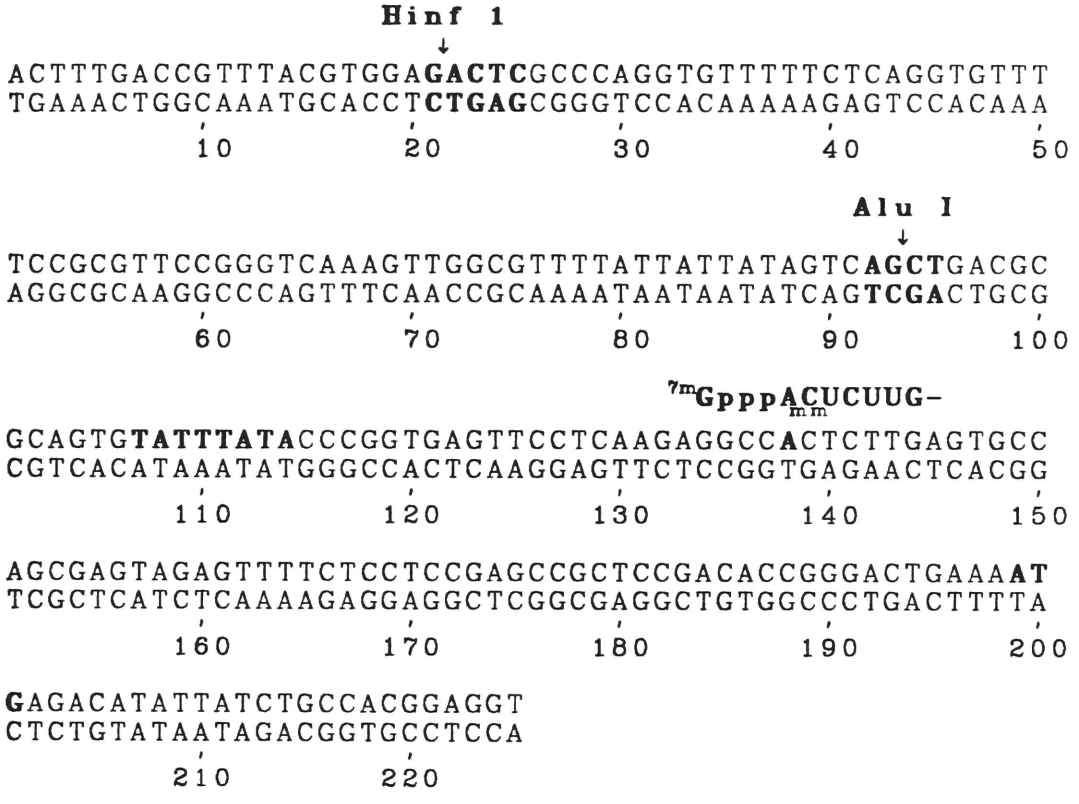


Figure 14. DNA Sequence of the Region Encoding the Ad2 EIa mRNA 5' Terminus

The alignment of the EIa 5' terminus with the DNA sequence at nucleotide 138 is shown. The TATA box and position corresponding to the 5' terminus are indicated in bold type. Restriction endonuclease recognition sequences are indicated in bold type and the cleavage sites by a vertical arrow. The first ATG is at nucleotide 199.

Figure 15. DNA Sequence of the Region Encoding the Ad2 EIa mRNA 3' Terminus and EIb mRNA 5' Termini

The EIa mRNA polyadenylation signal AAUAAA is encoded at nucleotide 270-275 and the poly A addition site is located at nucleotide 291 (Perricaudet et al., 1979, Fraser et al., 1981). The EIb TATA box is located at nucleotide 331-338 and the major and minor EIb mRNA 5' termini are encoded at nucleotide 361 and 363, respectively. The first translation initiation codon (AUG) is encoded at nucleotide 373-375.

```
l AGAGAATGCAATAGTAGTACGGATAGCTGTGACTCCGGTCCTTCTAACAC
r TCTCTTACGTTATCATCATGCCTATCGACACTGAGGCCAGGAAGATTGTG
          10        20        30        40        50

ACCTCCTGAGATACACCCGGTGGTCCCGCTGTGCCCCATTAAACCAGTTG
TGGAGGACTCTATGTGGGCCACCAGGGCGACACGGGGTAATTTGGTCAAC
          60        70        80        90        100

CCGTGAGAGTTGGTGGGCGTCGCCAGGCTGTGGAATGTATCGAGGACTTG
GGCACTCTCAACCACCCGCAGCGGTCCGACACCTTACATAGCTCCTGAAC
         110       120       130       140       150

CTTAACGAGTCTGGGCAACCTTTGGACTTGAGCTGTAAACGCCCCAGGCC
GAATTGCTCAGACCCGTTGGAAACCTGAACTCGACATTTGCGGGGTCCGG
         160       170       180       190       200

                                  Hpa I
                                    ↓
ATAAGGTGTAAACCTGTGATTGCGTGTGTG GTTAAC GCCTTTGTTTGCTG
TATTCCACATTTGGACACTAACGCACACAC CAATTG CGGAAACAAACGAC
         210       220       230       240       250

                                          Ela Poly(A)
                                               ↓
AATGAGTTGATGTAAGTTT AATAAA GGGTGAGATAATGTTTAACTTGCAT
TTACTCAACTACATTCAAATTATTTCCCACTCTATTACAAATTGAACGTA
         260       270       280       290       300

GGCGTGTTAAATGGGGCGGGGCTTAAAGGG TATATAAT GCGCCGTGGGCT
CCGCACAATTTACCCCGCCCCGAATTTCCCATATATTACGCGGCACCCGA
         310       320       330       340       350

      7mGpppAUCUG-
                mm
     7mGpppACAUCUG-
               mm
AATCTTGGTTACATCTGACCTC ATG GAGGCTTGGGAGTGTTTGGAAGATT
TTAGAACCAATGTAGACTGGAGTACCTCCGAACCCTCACAAACCTTCTAA
         360       370       380       390       400

TTTCTGCTGTGCGTAACTTGCTGGAACAGAGCTCTAACAGTACCTCTTGG
AAAGACGACACGCATTGAACGACCTTGTCTCGAGATTGTCATGGAGAACC
         410       420       430       440       450

                                                Pst I
                                                  ↓
TTTTGGAGGTTTCTGTGGGGCTCCTCCCAGGCAAAGTTAGT CTGCAG AAT
AAAACCTCCAAAGACACCCCGAGGAGGGTCCGTTTCAATCA GACGTC TTA
         460       470       480       490       500

TAAGGAGGATTACAAGTGGGAATTTGAAGAGCTTT
ATTCCTCCTAATGTTCACCCTTAAACTTCTCGAAA
         510       520       530
```

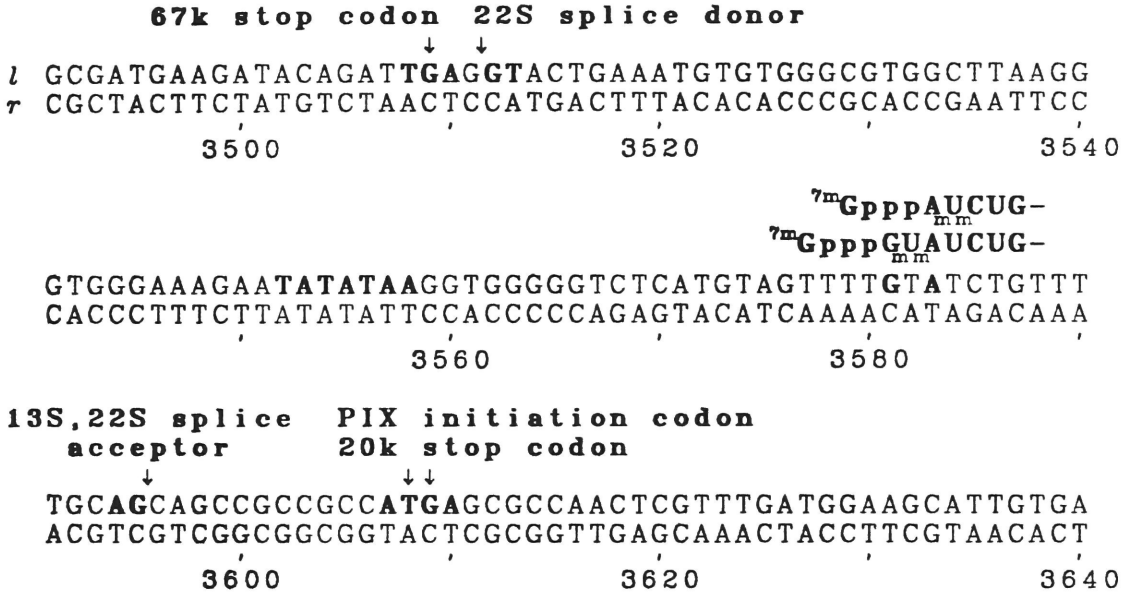


Figure 16. DNA Sequence of the Ad2 PIX Cap Region

The DNA sequence of nucleotides 3491-3544 was obtained by kinasing at the Bgl II site at 9.0 m.u. and sequencing the l strand up to nucleotide 3544. Nucleotides 3511-3640 are from Alestrom et al. (1980). Nucleotide numbering is from the Ad5 sequence of Van Ormondt et al. (1980). The PIX TATA box is located at nucleotide 3552-3558 and the major and minor PIX mRNA 5' termini are encoded at nucleotide 3583 and 3581, respectively. The sites encoding the 67K protein stop codon, the 22S mRNA splice donor, the 13S and 22S mRNA splice acceptor, the 20K protein stop codon and PIX initiation codon are taken from Perricaudet et al. (1980) and Alestrom et al. (1980).

cycloheximide but not by Ara C and therefore is transcribed independently of DNA synthesis (Chow et al., 1979). The IVa2 mRNA codes for a virion maturation protein (Persson et al., 1979).

The IVa2 5' terminus was analyzed as follows. Nonradioactive mRNA isolated 24 hrs post-infection was 5' labeled in vitro and hybridized to Sma I F/Hinf I (15-18), a fragment which encodes both the IVa2 and major late mRNA first leaders (Figure 17). The eluted RNA was RNAse T1 and RNAse A fingerprinted, as shown in Figure 18 A and B, respectively. All major oligonucleotides which were RNAse T2 resistant are numbered. RNAse T1 spot 4 (Figure 18A) has a $pA_mC_{(m)}U$- 5' terminus, identified in Table I and Figure 11 lanes 43-44, and is the 5' terminal oligonucleotide from the major late transcription unit sequenced by Ziff and Evans (1978). RNAse A spots 2 and 3 (Figure 18B) have the DEAE mobilities of the type I and type II RNAse T2 oligonucleotides, respectively, of RNAse T1 spot 4 (Figure 18A; data not shown) and are therefore pA_mCp and pA_mC_mUp, respectively. Thus RNAse T1 spot 4 (Figure 18A) and RNAse A spots 2 and 3 (Figure 18B) all represent the 5' terminus of the major late mRNAs ($^{7m}GpppA_mC_{(m)}U$-) shown in Figure 20.

RNAse A spot 4 (Figure 18B) has a pA 5' terminus (Table I), and RNAse T1 and T2 redigestion patterns and mobilities (shown in Figure 11 lanes 29-31) similar to that for the EIIa $pA_mG_{(m)}AG$- 5' terminal oligonucleotide, analyzed in greater detail below. From its position in the fingerprint, it is 5 nucleotides long. The partial digestion of the type II oligonucleotide with RNAse T1 to give an oligonucleotide with mobility equal to that of the T2 oligonucleotide is due to digestion by a contaminant in the RNAse T1 and does not indicate a G in the third position. This is corroborated by the RNAse T1 fingerprint, Figure 18A, (discussed below) which was made using a different RNAse T1 batch, and by the mobility shifts in the partial RNAse T2 fingerprint, Figure 18C, which gives the partial sequence -AGUp. RNAse T1 spot 1 (Figure 18A) has a pA terminus (Table I) and has a mobility on DEAE identical to the type I digestion products of RNAse A spot 4 (Figure 18B; compare Figure 11 lanes 27-28 with lanes 29-30). RNAse T1 spot 3 (Figure 18A) has a pA 5' terminus (Table I), and its redigestion products have DEAE mobilities

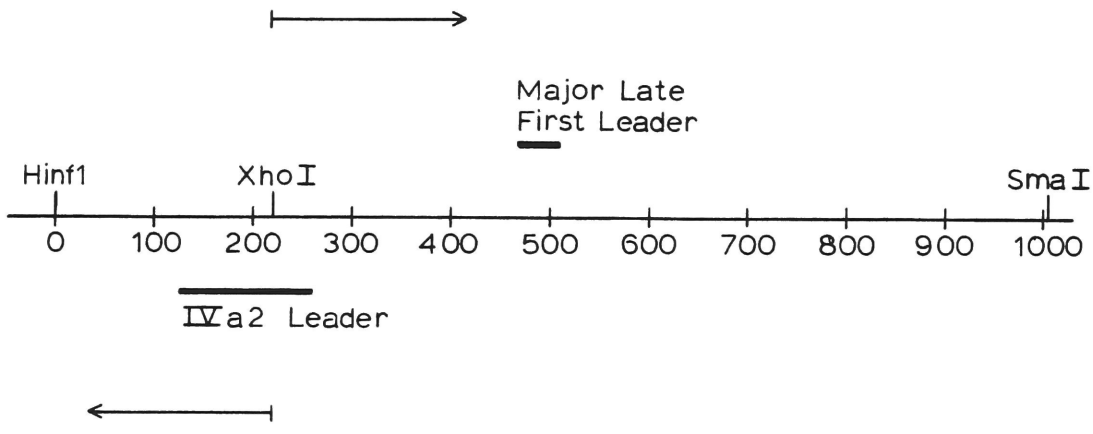


Figure 17. Map of the First Leaders of the Major Late and IVa2 mRNAs

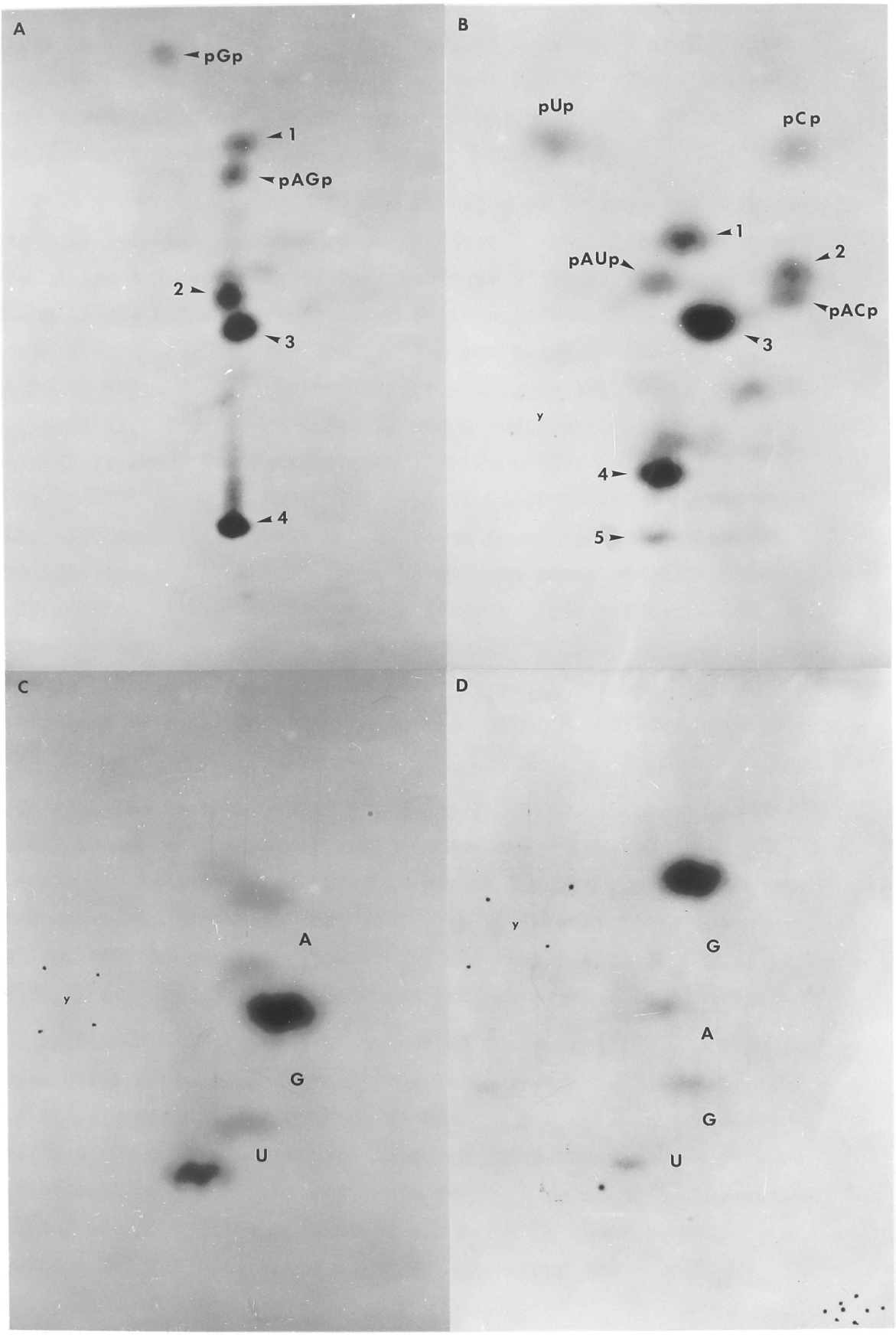
The positions of the first leaders of the major late and IVa2 mRNAs within the Sma I F/Hinf I (15-18 m.u.) fragment are indicated. Nucleotide numbers are those of the sequence presented in Figure 20. The horizontal arrows represent sequence data obtained for the region of the IVa2 leader by kinasing at the Xho I site.

Figure 18. The 5' Termini of the IVa2 and Major Late mRNAs

Late mRNA was decapped and kinased and selected by hybridization to 15-18 DNA (a Hinf I fragment of Sma I F). Because this fragment overlaps the cap sites of both the IVa2 and major late mRNAs, RNA yielding the 5' terminal oligonucleotides from both of these messengers were selected together.

(A) is the RNAse T1 fingerprint and (B) is the RNAse A fingerprint, both of the 15-18 DNA hybrid. Spots 1, 2 and 3 from panel (A) and spots 1, 4 and 5 from panel (B) are from the IVa2 mRNA. Spots 4 from panel (A) and 2 and 3 from panel (B) are from the major late mRNAs.

Panels (C) and (D) are fingerprints of partial RNAse T2 digests of spots 4 and 5 respectively from panel (B), and analyze sequences of two microheterogeneous termini of the IVa2 mRNAs. The derivation of the sequences of the spots in panels (A) and (B) is given in the text and their sites of encoding in the DNA are given in Figure 20.


A
pGp
1
pAGp
2
3
4
B
pUp
pCp
1
pAUp
2
pACp
3
y
4
5
C
A
y
G
U
D
y
G
A
G
U

identical to the type II redigestion products of RNAse A spot 4 (Figure 18B; compare Figure 11 lanes 32-33 with lanes 29-30). Therefore RNAse T1 spots 1 and 3 (Figure 18A) and RNAse A spot 4 (Figure 18B) all represent one terminus with the sequence $^{7m}GpppA_mG_{(m)}AGU-$.

RNAse A spot 1 (Figure 18B) has mostly a pU 5' terminus with some pC (Table I) and DEAE mobility of pU_mCp (type I only; Figure 11 lanes 41-42). RNAse A spot 5 (Figure 18B) has a pU 5' terminus (Table I), and the RNAse T2 and RNAse T1 redigestion pattern indicates type II methylation only with A and G in the 3rd and 4th positions, respectively, shown in Figure 11 lanes 35-37. RNAse T1 spot 2 (Figure 18A) has a pU 5' terminus (Table I). The type I RNAse T2 and A redigestion products have the mobilities found for RNAse A spot 1 (Figure 18B; compare Figure 11 lanes 38-39 with lanes 41-42). The type II RNAse T2 and A redigestion products have mobilities identical to those found for RNAse A spot 5 (Figure 18B; compare Figure 11 lanes 38-39 with lanes 35-36). Therefore RNAse T1 spot 2 (Figure 18A) and RNAse A spots 1 and 5 (Figure 18B) all come from one terminus with the 5' sequence $pU_mC_{(m)}AG-$. Combining this with the partial sequence -GAGUp of RNAse A spot 5 (Figure 18B) obtained from the RNAse T2 partial fingerprint (Fig. 18D), I conclude that this terminus begins $^{7m}GpppU_mC_{(m)}AGAGU-$.

In addition to these major 5' terminal oligonucleotides, several very minor 5' terminal oligonucleotides have been detected. Several faint spots can be seen below RNAse T1 spot 4 (Figure 18A), but were not analyzed except to establish that they have 2'-O-methylation and thus derive from mRNA 5' termini. Spots with the composition $pC_m(CU)p$, pC_mUp and $p(CU)_mUp$ were detected in the RNAse A fingerprint (Figure 18B).

Hybridization to separated strands of Sma I F/Hinf I was used to determine which 5' termini derived from which transcription unit. The RNAse A fingerprints of the hybrids to the l and r strands are shown in Figures 19 A and B, respectively. Although the separated strands were cross contaminated, each spot was preferentially selected by one strand. The l stand preferentially hybridizes RNA yielding spots 1, 4 and 5 and the r strand RNA yielding spots 2 and 3. Therefore the $^{7m}GpppA_mC_{(m)}U-$

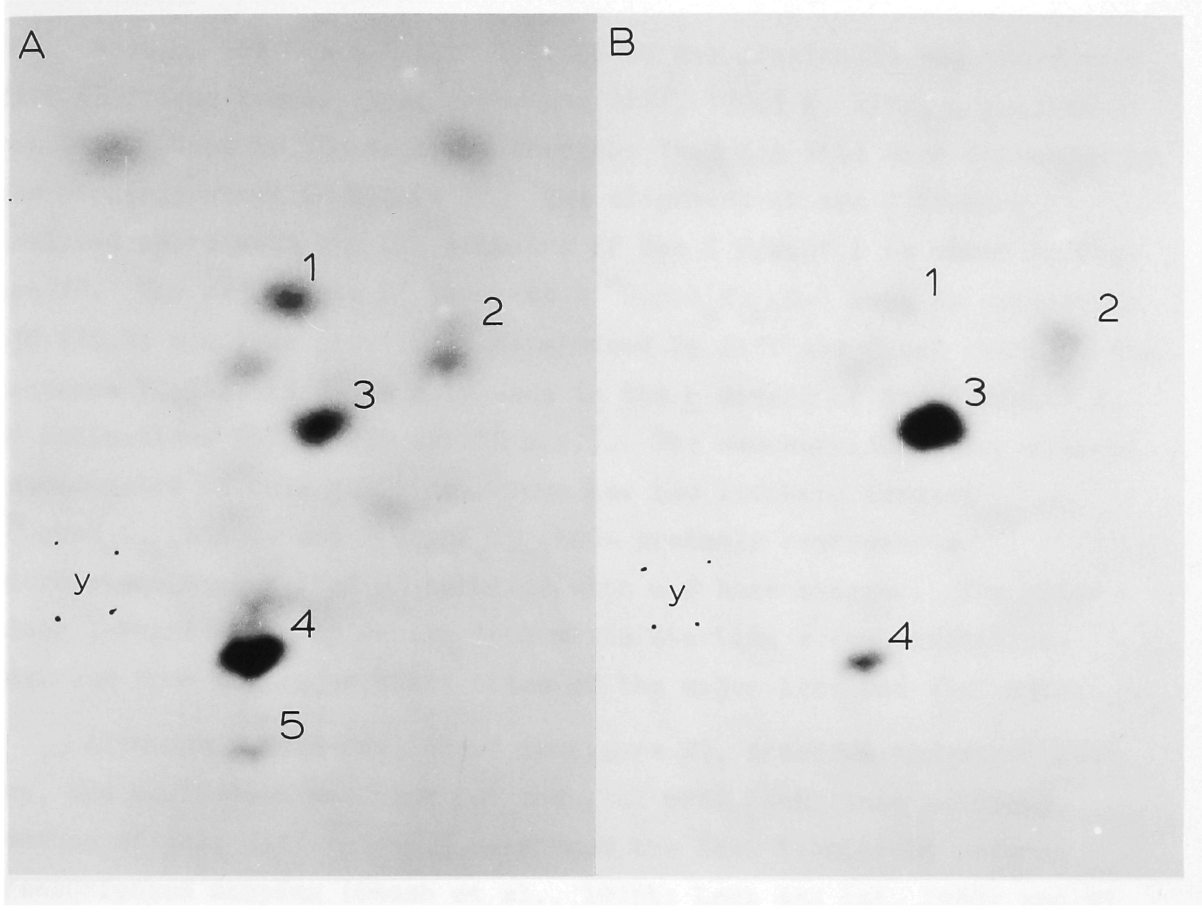


Figure 19. Separated Strand Hybridization of the Major Late and IVa2 mRNAs

To prove that the pU$_m$C$_{(m)}$AGAGUp and pA$_m$G$_{(m)}$AGUp 5' terminal oligonucleotides are from l strand encoded mRNAs and that the pA$_m$C$_{(m)}$U- 5' terminal oligonucleotide is from r strand encoded mRNA, late 5' labeled mRNA was hybridized to separated l and r strands of Sma I F/Hinf I, eluted, and RNAse A fingerprinted as shown in panels (A) and (B), respectively. Spot numbers correspond to those in Figure 18B. RNAs yielding spots 1, 4 and 5 are preferentially selected by the l strand and RNAs yielding spots 2 and 3 by the r strand.

terminus is from a rightward transcription unit (major late) and the $^{7m}GpppU_mC_{(m)}AGAGU-$ and $^{7m}GpppA_mG_{(m)}AGU-$ termini are from a leftward transcription unit (IVa2).

Most of the Sma I F/Hinf I fragment was previously sequenced by E. Ziff (Ziff and Evans, 1978; Baker and Ziff, 1980; E. Ziff, unpublished results). Gaps in the sequence near the IVa2 cap site were filled in by the strategy shown in Figure 17. The alignment of the 3 termini analyzed above with the DNA sequence of Sma I F/Hinf I is shown in Figure 20. The major late 5' terminus ($^{7m}GpppA_mC_{(m)}U-$) maps to nucleotide 470 (16.45 m.u.) as previously determined by Ziff and Evans (1978). The sequence TCAGAGT is found only once in the r strand of Sma I F/Hinf I, at nucleotides 253 to 259 (15.85 m.u.). The sequence AGAGT can also be accommodated at this position. Thus the two leftward termini $^{7m}GpppU_mC_{(m)}AGAGU-$ and $^{7m}GpppA_mG_{(m)}AGU-$ probably represent a microheterogeneous IVa2 5' terminus with a 2 base stagger. The other minor termini probably derive from mRNAs starting a few nucleotides upstream from the major start sites of the major late and IVa2 mRNAs.

Although a TATA box, shown in Figure 20, precedes the major late cap, the equivalent position for the IVa2 mRNA lacks this homology. Because of this difference, I confirmed the IVa2 termini by reverse transcription mapping (Ghosh et al., 1978b; Lamb and Lai, 1980) and S1 mapping (Berk and Sharp, 1977a; Weaver and Weissman, 1979). A schematic of this experiment is shown in Figure 21. Genomic Ad2 DNA was cut with Xho I and kinased to yield a 5' terminally labeled Xho I F fragment (15.5-22 m.u.). For reverse transcription, a 17 nucleotide primer, 5' labeled at the Xho I site at 15.5 m.u. and complementary to the IVa2 leader, was made by recutting the Xho I F fragment with Ava II (Figure 20). After hybridization to late mRNA, the 3' end of the primer was extended to the cap (approximately 30 nucleotides away) with AMV reverse transcriptase and cold dNTPs. The reaction products were electrophoresed on a 13% acrylamide-8M urea thin sequencing gel alongside a Maxam-Gilbert sequencing pattern of a fragment labeled at the same Xho I site, but extending to the Sma I site at 18 m.u., as shown in Figure 22 (left-hand and center lanes, respectively). It is important to note

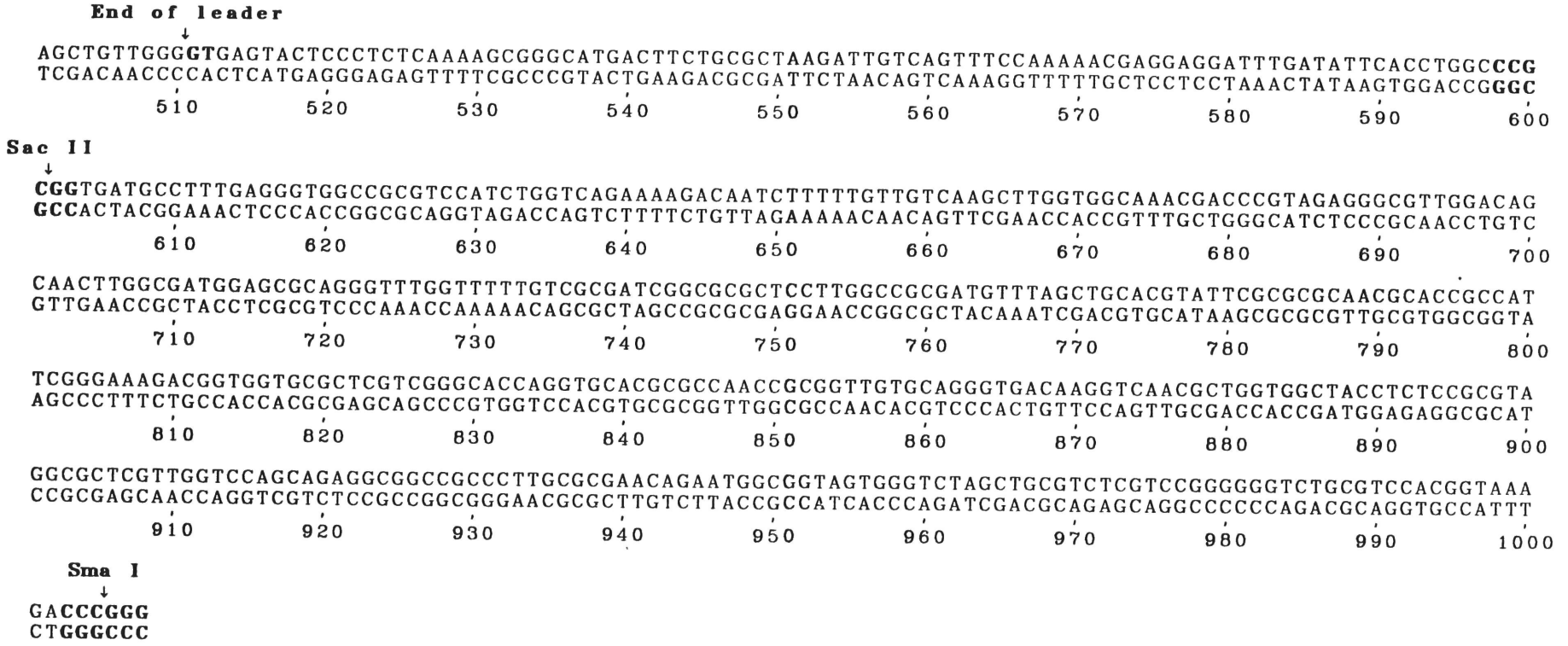
Figure 20. Alignment of the IVa2 and Major Late 5' Termini with the DNA

The sequence of the Ad2 Sma I F/Hinf I fragment (15-18 m.u.) was mostly determined by E. Ziff (Ziff and Evans, 1978; Baker and Ziff, 1980; E. Ziff, unpublished results). The sequence in the region of the IVa2 5' terminus was determined by the strategy shown in Figure 17.

The U and A termini of the IVa2 mRNA and the A terminus of the major late mRNAs are aligned with the DNA sequence. The bases corresponding to the 5' termini are indicated in bold. In addition all RNAse T1 and RNAse A 5' terminal oligonucleotides from Figure 18A and B are aligned with the mRNA sequences from which they are derived. The secondary analyses of the 5' terminal oligonucleotides are presented in Table I and Figure 11 lanes 27-44 and are discussed in detail in the text. The symbol * indicates the only phosphate labeled with ^{32}P by decapping and kinasing.

In addition, the 3' ends of the IVa2 and major late leaders are indicated.


Hinf1
l GATTCCGGGGAGTAGGCATCCGCGCCGCAGGCCCGCAGACGGTCTCGCATTCCACGAGCCAGGTGAGCTCTGGCCGTTCGGGGTCAAAAACCAGGTTTCC
r CTAAGGCCCCTCATCCGTAGGCGCGGCGTCCGGGCGTCTGCCAGAGCGTAAGGTGCTCGGTCCACTCGAGACCGGCAAGCCCCAGTTTTTGGTCCAAAGG
10 20 30 40 50 60 70 80 90 100
CCCATGCTTTTTGATGCGTTTCTTACCTCTGGTTTCCATGAGCCGGTGTCCACGCTCGGTGACGAAAAGGCTGTCCGTGTCCCCGTATACAGACTTGAGA
GGGTACGAAAAACTACGCAAAGAATGGAGACCAAAGGTACTCGGCCACAGGTGCGAGCCACTGCTTTTCCGACAGGCACAGGGGCATATGTCTGAACTCT
110 End of Leader 140 150 160 170 180 190 200
Xho I
Ava II
GGCCTGTCCTCGAGCGGTGTTCCGCGGTCCTCCTCGTATAGAAACTCGGACCACTCTGAGACGAAGGCTCGCGTCCAGGCCAGCACGAAGGAGGCTAAGT
CCGGACAGGAGCTCGCCACAAGGCGCCAGGAGGAGCATATCTTTGAGCCTGGTGAGACTCTGCTTCCGAGCGCAGGTCCGGTCGTGCTTCCTCCGATTCA
-CCGGACAGGAGCUCGCCACAAGGCGCCAGGAGGAGCAUAUCUUUGAGCCUGGUGAGACUpppG7m
270 280 -31 300
pUGAGACUp* :Spot 18B-5
pGACUp* :Spot 18A-2
pCUp* :Spot 18B-1
U Terminus
3′←5′
IVa2 mRNA
-CCGGACAGGAGCUCGCCACAAGGCGCCAGGAGGAGCAUAUCUUUGAGCCUGGUGAGAGApppG7m
pUGAGAp* :Spot 18B-4
pGAGAp* :Spot 18A-3
pGAp* :Spot 18A-1
A Terminus
3′←5′
GGGAGGGGTAGCGGTCGTTGTCCACTAGGGGGTCCACTCGCTCCAGGGTGTGAAGACACATGTCGCCCTCTTCGGCATCAAGGAAGGTGATTGGTTTATA
CCCTCCCCATCGCCAGCAACAGGTGATCCCCCAGGTGAGCGAGGTCCCACACTTCTGTGTACAGCGGGAGAAGCCGTAGTTCCTTCCACTAACCAAATAT
310 320 330 340 350 360 370 380 390 400
5′→3′
MAJOR LATE mRNA
Spot 18B-2: *pACp
Spot 18B-3: *pACUp
Spot 18A-4: *pACUCUCUUCCGp
-31
7mGpppACUCUCUUCCGCAUCGCUGUCUGCGAGGGCC-
GGTGTAGGCCACGTGACCGGGTGTTCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCACTCTCTTCCGCATCGCTGTCTGCGAGGGCC
CCACATCCGGTGCACTGGCCCACAAGGACTTCCCCCCGATATTTTCCCCCACCCCCGCGCAAGCAGGAGTGAGAGAAGGCGTAGCGACAGACGCTCCCGG
410 420 430 440 450 460 470 480 490 500


End of leader
AGCTGTTGGGGTGAGTACTCCCTCTCAAAAGCGGGCATGACTTCTGCGCTAAGATTGTCAGTTTCCAAAAACGAGGAGGATTTGATATTCACCTGGCCCG
TCGACAACCCCACTCATGAGGGAGAGTTTTCGCCCGTACTGAAGACGCGATTCTAACAGTCAAAGGTTTTTGCTCCTCCTAAACTATAAGTGGACCGGGC
510 520 530 540 550 560 570 580 590 600
Sac II
CGGTGATGCCTTTGAGGGTGGCCGCGTCCATCTGGTCAGAAAAGACAATCTTTTTGTTGTCAAGCTTGGTGGCAAACGACCCGTAGAGGGCGTTGGACAG
GCCACTACGGAAACTCCCACCGGCGCAGGTAGACCAGTCTTTTCTGTTAGAAAAACAACAGTTCGAACCACCGTTTGCTGGGCATCTCCCGCAACCTGTC
610 620 630 640 650 660 670 680 690 700
CAACTTGGCGATGGAGCGCAGGGTTTGGTTTTTGTCGCGATCGGCGCGCTCCTTGGCCGCGATGTTTAGCTGCACGTATTCGCGCGCAACGCACCGCCAT
GTTGAACCGCTACCTCGCGTCCCAAACCAAAAAACAGCGCTAGCCGCGCGAGGAACCGGCGCTACAAATCGACGTGCATAAGCGCGCGTTGCGTGGCGGTA
710 720 730 740 750 760 770 780 790 800
TCGGGAAAGACGGTGGTGCGCTCGTCGGGCACCAGGTGCACGCGCCAACCGCGGTTGTGCAGGGTGACAAGGTCAACGCTGGTGGCTACCTCTCCGCGTA
AGCCCTTTCTGCCACCACGCGAGCAGCCCGTGGTCCACGTGCGCGGTTGGCGCCAACACGTCCCACTGTTCCAGTTGCGACCACCGATGGAGAGGCGCAT
810 820 830 840 850 860 870 880 890 900
GGCGCTCGTTGGTCCAGCAGAGGCGGCCGCCCTTGCGCGAACAGAATGGCGGTAGTGGGTCTAGCTGCGTCTCGTCCGGGGGGTCTGCGTCCACGGTAAA
CCGCGAGCAACCAGGTCGTCTCCGCCGGCGGGAACGCGCTTGTCTTACCGCCATCACCCAGATCGACGCAGAGCAGGCCCCCCAGACGCAGGTGCCATTT
910 920 930 940 950 960 970 980 990 1000
Sma I
GACCCGGG
CTGGGCCC

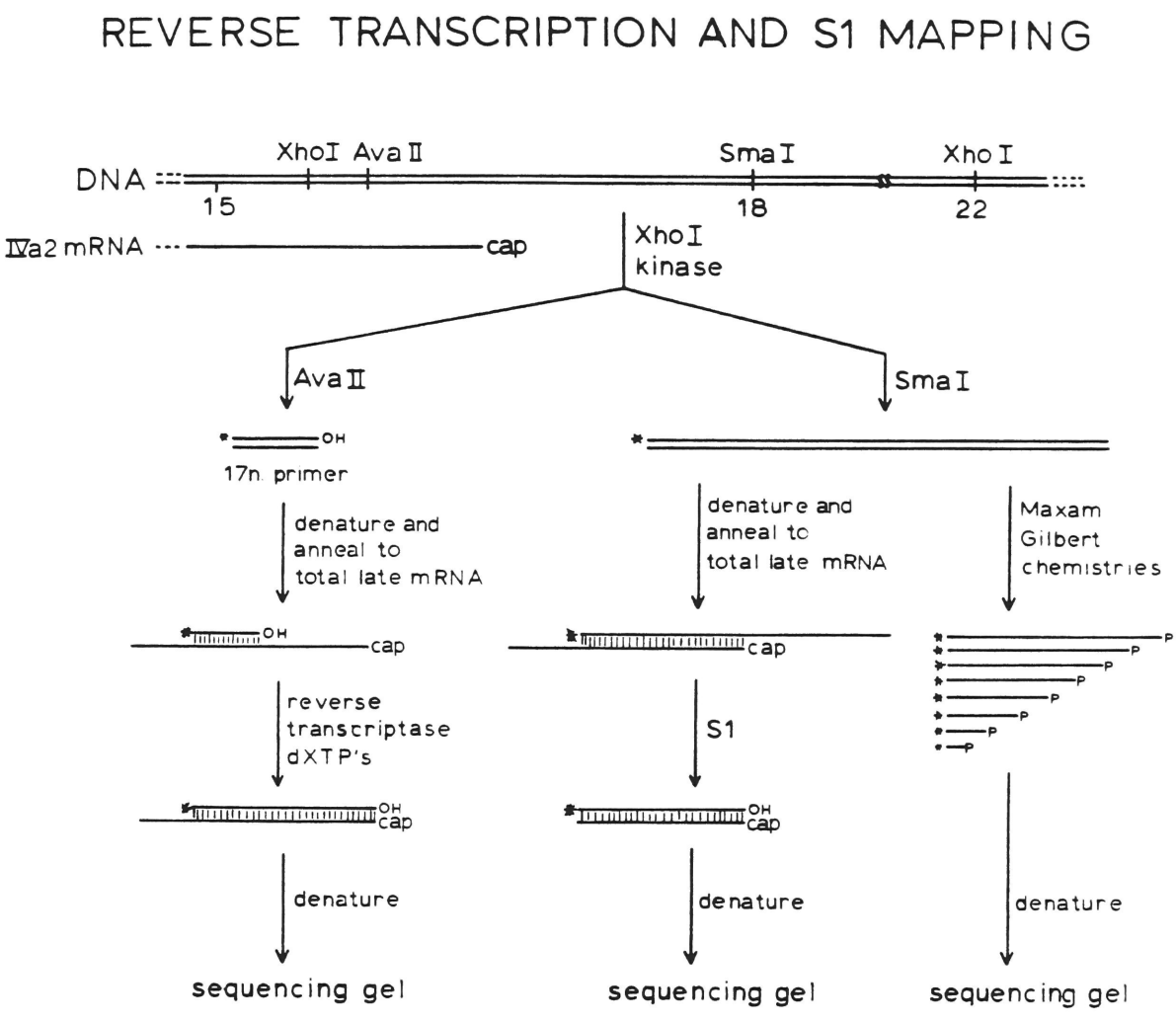


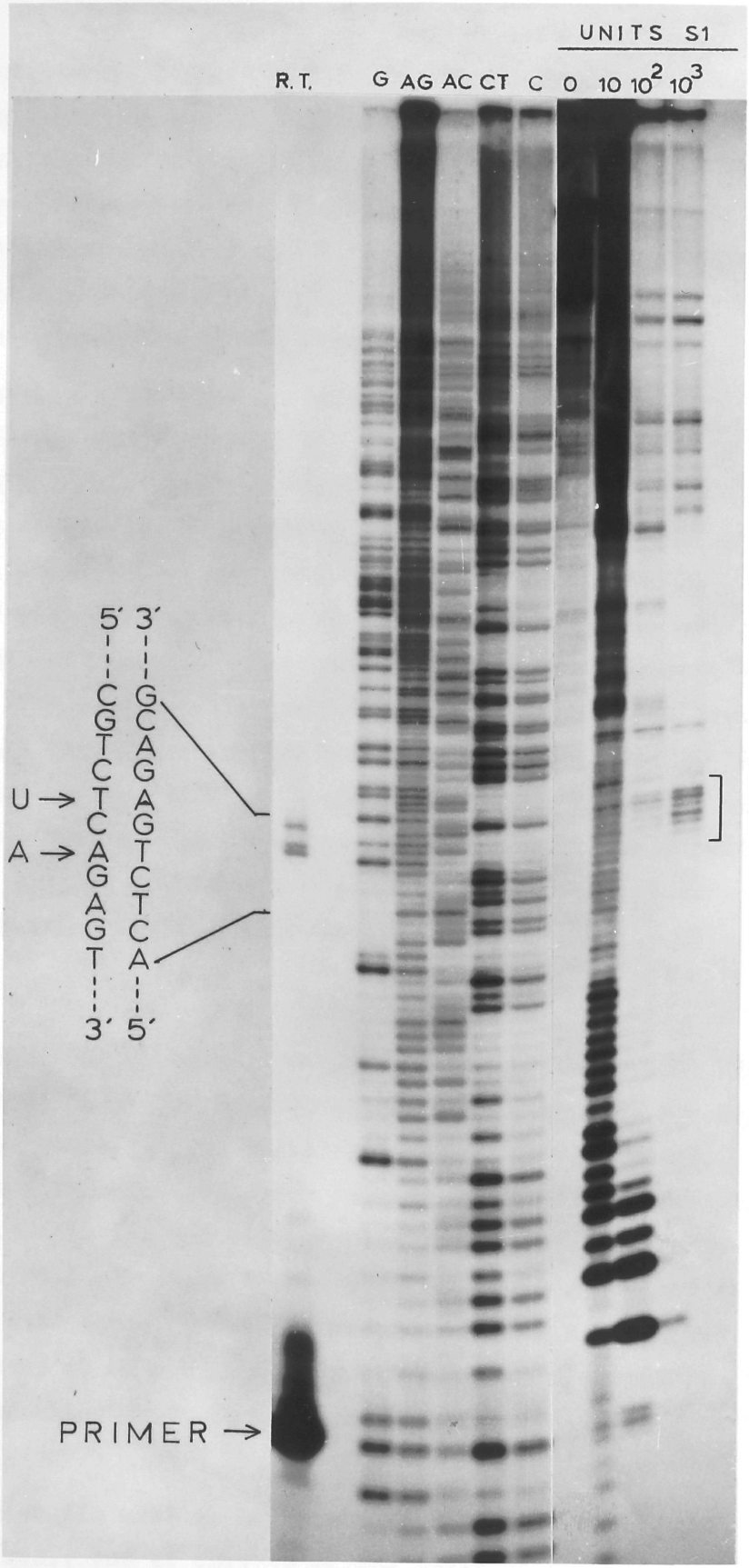
Figure 21. Reverse Transcription and S1 Mapping Protocol

The mapping of the 5' terminus of IVa2 mRNA by reverse transcription and S1 is shown schematically. The sequencing gel pattern from this experiment is shown in Figure 22.

Figure 22. Reverse Transcription and S1 Mapping of the IVa2 mRNA 5' Terminus

The reaction products from the combined DNA sequencing and S1 and reverse transcription mapping experiment, shown schematically in Figure 21 and discussed in detail in the text, were electrophoresed on a 13% acrylamide-8M urea thin sequencing gel. The 5 lanes marked G, AG, AC, CT and C contain the products from the Maxam-Gilbert chemistries (Maxam and Gilbert, 1980) cleaving at G, A+G, A>C, C+T and C, respectively. The leftmost lane contains the reverse transcription reaction products. Note that most of the primer was not extended. The DNA sequence in the region of the full length reverse transcripts is shown and the positions of the IVa2 mRNA 5' termini mapped by direct RNA sequencing are indicated.

The righthand 4 lanes show the DNA resistant to 0, 10, 100 and 1000 units of S1. The bracket indicates a cluster of S1 resistant bands which accumulate with increasing S1 digestion and approximately comigrate with the reverse transcription products. Note that reverse transcript and S1 bands identifying a given residue will be longer by 1 nucleoside than the corresponding chemical degradation product, requiring a small mobility correction (Moss and Birnstiel, 1979; Sollner-Webb and Reeder, 1979).


UNITS S1
R.T.
G AG AC CT C
0 10 10^2 10^3
5′ 3′
CGTCTCAGAGT
GCAGAGTCTCA
U →
A →
3′ 5′
PRIMER →

that reverse transcript bands identifying a given residue will be longer by 1 nucleoside than the corresponding chemical degradation product, requiring a small mobility correction. Three major extension products are seen, corresponding to the U and A termini, and also to a G terminus one nucleotide downstream from the A terminus. No in vivo terminus was found corresponding to the G position, and the extra band may result from premature termination of the reverse transcript.

For S1 mapping, the same 5' labeled DNA fragment used for Maxam-Gilbert sequencing was hybridized to late mRNA, digested with nuclease S1 to remove all single stranded tails, and electrophoresed alongside the sequencing reactions. The pattern obtained using increasing S1 concentrations is shown in the righthand four lanes of Figure 22. Notice that a cluster of bands approximately comigrating with the reverse transcription bands are the major resistant products after the most vigorous S1 digestion. The same mobility correction which applies to reverse transcript bands also applies to S1 bands (Moss and Birnstiel, 1979; Sollner-Webb and Reeder, 1979). In addition the cap structure may protect a few additional nucleotides from S1 digestion. Thus direct RNA sequencing and reverse transcription and S1 mapping all identify the same narrow region as the IVa2 cap site.

Early Region II

Early region IIa encodes a 72K DNA binding protein (DBP) (Van der Vliet et al., 1975; Lewis et al., 1976) which is necessary for DNA replication (Van der Vliet and Sussenbach, 1975; Horwitz et al., 1979). This protein is also responsible for repression of EIV transcription late in infection (Nevins and Winkler, 1980). Expression of the DBP begins late in the early phase from a promoter at 75 m.u., and as infection proceeds, mRNA encoding the DBP continues to be synthesized, but from new promoters at 72 m.u. and 86 m.u. (Chow et al., 1980). This early to late promoter shift is blocked by cycloheximide but not by Ara C (Chow et al., 1980).

Early region IIb mRNA is expressed in the early phase, also from a promoter at 75 m.u., but at only 1% of the level of the DBP mRNA

(Stillman et al., 1981). Its expression is independent of DNA replication. The IIb mRNA encodes the protein which is covalently linked to the 5' termini of adenovirus DNA (Stillman et al., 1981) and which is proposed to prime Ad2 DNA synthesis (Rekosh et al., 1977).

The First Leader and 5' Termini of the DBP mRNA Synthesized at Early Times After Infection

The spliced structure of the early DBP mRNA is shown in Figure 23 (species a; Chow et al., 1979). Note that the first leader, estimated to be 50 nucleotides long (Berk and Sharp, 1978), is the only exonic RNA encoded within the Eco RI F fragment. Therefore this leader should be the exclusive product in the ribonuclease resistant hybrid of early mRNA to Eco RI F. Early mRNA, labeled with ^{32}P *in vivo* and cleaved to approximately 350 to 500 nucleotide long fragments with RNAse T1, was hybridized to Eco RI F. Unpaired RNA was trimmed from the hybrid with RNAse T1, and the nuclease resistant fraction of the hybrid was eluted and fingerprinted after complete RNAse T1 digestion (shown in Figure 24A). This fingerprint has the complexity of a short specific RNA sequence consistent with the estimated size of the leader RNA. In Figure 24B, a hybrid obtained in a similar manner, but without nuclease pre-treatment on the filter, was RNAse T1 fingerprinted. This fingerprint shows increased complexity relative to Figure 24A. The RNAse A fingerprint of the hybrid (trimmed with RNAse A) is given in Figure 24C. Spots from fingerprints of Figures 24A and 24C were subjected to reciprocal redigestion with RNAse A or RNAse T1, and also with RNAse T2. The majority of these spots could be completely hydrolyzed by RNAse T2 (indicating that they are derived from internal positions of the mRNA). The results of secondary analysis of the spots are given in Table II. Spots 4 and 14 of the T1 fingerprint in Figure 24A, spots 4, 14, and 15 of the T1 fingerprint in Figure 24B, and spots 11 and 12 of the RNAse A fingerprint in Figure 24C yielded components which were resistant to RNAse T2 hydrolysis to mononucleotides (indicative of an inverted nucleotide cap structure), and are considered separately below.

Although the data in Table II are not sufficient to determine the

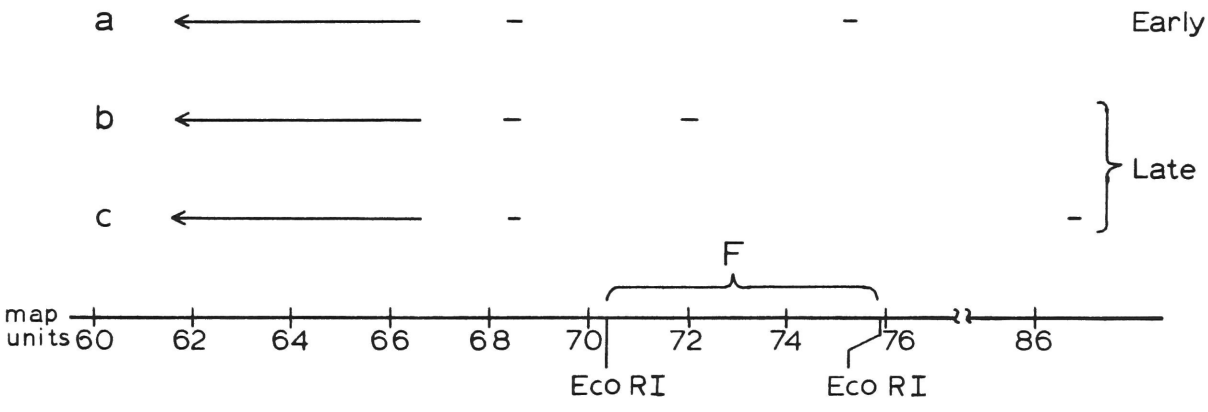


Figure 23. mRNAs of Early Region IIa

The structures of the DNA binding protein mRNA at early and late times are shown (Chow et al., 1979; 1980). The first leaders of the early DNA binding protein mRNA (encoded at 75 m.u.) and 95% of the late DNA binding protein mRNA (encoded at 72 m.u.) are both encoded within the Eco RI F fragment.

Figure 24. RNAse T1 and RNAse A Fingerprints of EIIa Leader RNA

Polyadenylated cytoplasmic RNA was isolated from cells labeled with ^{32}P in the presence of cycloheximide 0-5 hours post-infection, cleaved with RNAse T1 to approximately 350-500 nucleotides, and hybridized to Eco RI F DNA.

(A) shows the RNAse T1 fingerprint of the RNAse T1 trimmed hybrid.

(B) shows the RNAse T1 fingerprint of the untrimmed hybrid. Numbered spots correspond with the major spots in panel (A).

(C) shows the RNAse A fingerprint of the RNAse A trimmed hybrid.

Results of the secondary analyses are presented in Table II and correlated with the DNA sequence in Figure 25. The structures of capped oligonucleotides from these fingerprints are shown in Figure 26.

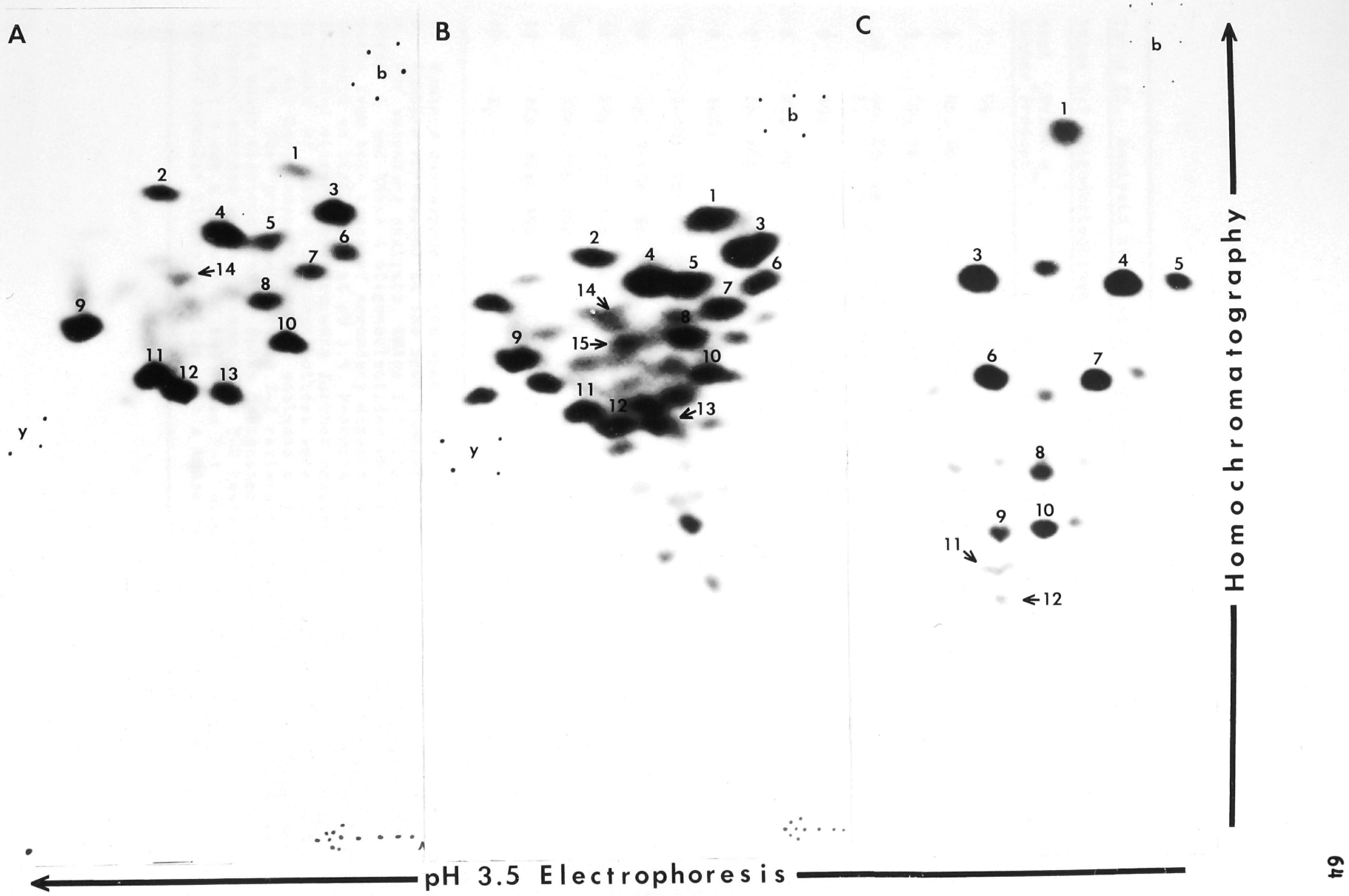

A
B
C
b
y
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
pH 3.5 Electrophoresis
Homochromatography

Table II. Analysis of RNAse T-1 and RNAse A Oligonucleotides of Region IIa

RNAse T-1 Oligonucleotides			RNAse A Oligonucleotides	
Spot Number[a]	RNAse A Product[c]	Molarity	Spot Number[b]	RNAse T-1 Product[c]
1	Gp	2-3	1	Up
2	Up, Gp	1-2	2	Cp
3	Cp, Gp	~3	3	Gp, Up
4[d]	Up, Cp, Gp X_1	2 <1	4	Gp, Cp
5	AGp	1-2	5	ACp
6	2Cp, Gp	1	6	Gp, AUp
7	Cp, AGp	1	7	AGp, Cp
8	AAGp	1	8	Gp, AGp, Cp
9	3-4Up, Cp, Gp	1	9	2Gp, AGp, Cp
10	Up, 3-4Cp, Gp	1	10	Gp, AAGp, Cp
11	2Up, 2Cp, AUp, Gp	1	11	X_3, AGp, Cp
12	3Up, 3Cp, AGp	1	12	X_4, X_5, AGp, Cp
13	ACp, AUp, AGp	1		
14	X_2	<1		

a. Numbers correspond to the spot numbers of Figure 24A.
b. Numbers correspond to the spot numbers of Figure 24C.
c. For secondary analysis, RNAse T-1 oligonucleotides were redigested with RNAse A and RNAse A oligonucleotides were redigested with RNAse T-1. Products from both types of secondary digests were fractionated by electrophoresis on DEAE paper at pH 3.5. Products from the RNAse A redigestions of RNAse T-1 oligonucleotides were further analysed by RNAse T-2 digestion. Aliquots of RNAse A oligonucleotides were digested with RNAse T-2 directly. All base compositions were analysed by electrophoresis on 540 paper at pH 3.5. The presence of RNAse T-2 resistant components was determined by DEAE paper electrophoresis. Spots designated X are RNAse A or RNAse T-1 oligonucleotides which yielded RNAse T-2 resistant components.
d. Spot 4 was a mixture of the RNAse T-1 oligonucleotide (U,C)Gp and a second submolar oligonucleotide with a RNAse T-2 resistant component.

structure of the hybridized RNA, these analyses permitted the localization of its template within the nucleotide sequence of the Eco RI F fragment (Galibert, Herisse and Courtois, 1979). The Eco RI F fragment was scanned for a region whose sequence would account for the observed RNAse T1 and RNAse A oligonucleotides obtained from the mRNA hybrid. Only one contiguous DNA sequence was found which could serve as template for these products. This region was located on the leftward DNA strand of Eco RI F at 75.05 m.u. (nucleotides 1393-1460). Correlation of the RNA oligonucleotides with the DNA sequence is given in Figure 25. For each RNAse T1 product listed in Table II, a corresponding predicted RNAse T1 oligonucleotide may be found within the DNA sequence between residues 1464 and 1390 of Eco RI F. Furthermore, for all RNAse T1 oligonucleotides predicted by this interval, spots of equivalent composition could be found in the RNAse T1 fingerprint. A similar correlation is shown for the RNAse A oligonucleotides of Figure 24C, although this fingerprint contains a slightly higher level of background products. One RNAse A oligonucleotide, spot 9 of Figure 24C, had no equivalent in this region of the DNA. This spot, and also spots with RNAse T2 resistant components, are considered below.

Localization of Splicing Site In the mature cytoplasmic messenger RNA, the first leader is linked to a second leader whose sequence is transcribed at coordinate 68.8 (Kitchingman et al., 1977; Berk and Sharp, 1978; Chow et al., 1979). The nucleotide sequence of the mRNA will diverge from that of the DNA at the 3' end of the first leader at the splice point. Its location in the template was estimated by comparing the DNA sequence with the products in fingerprints of the leader. Because RNAse T1 spot 7 of Figure 24A, CAG, and RNAse A spot 8 of Figure 24B, AGGC, are present, the leader must extend beyond residue 1402 at least as far as residue 1399. The first characteristic oligonucleotides predicted by the DNA which are not obtained from the leader are the RNAse A oligonucleotides GAGGGAGC (residues 1390 to 1381) and GGGU (residues 1394 to 1391). Because these products are absent the leader cannot extend beyond nucleotide 1392. From the estimated molarity (given in Table II) of RNAse T1 spot 4, CUG, it is highly likely

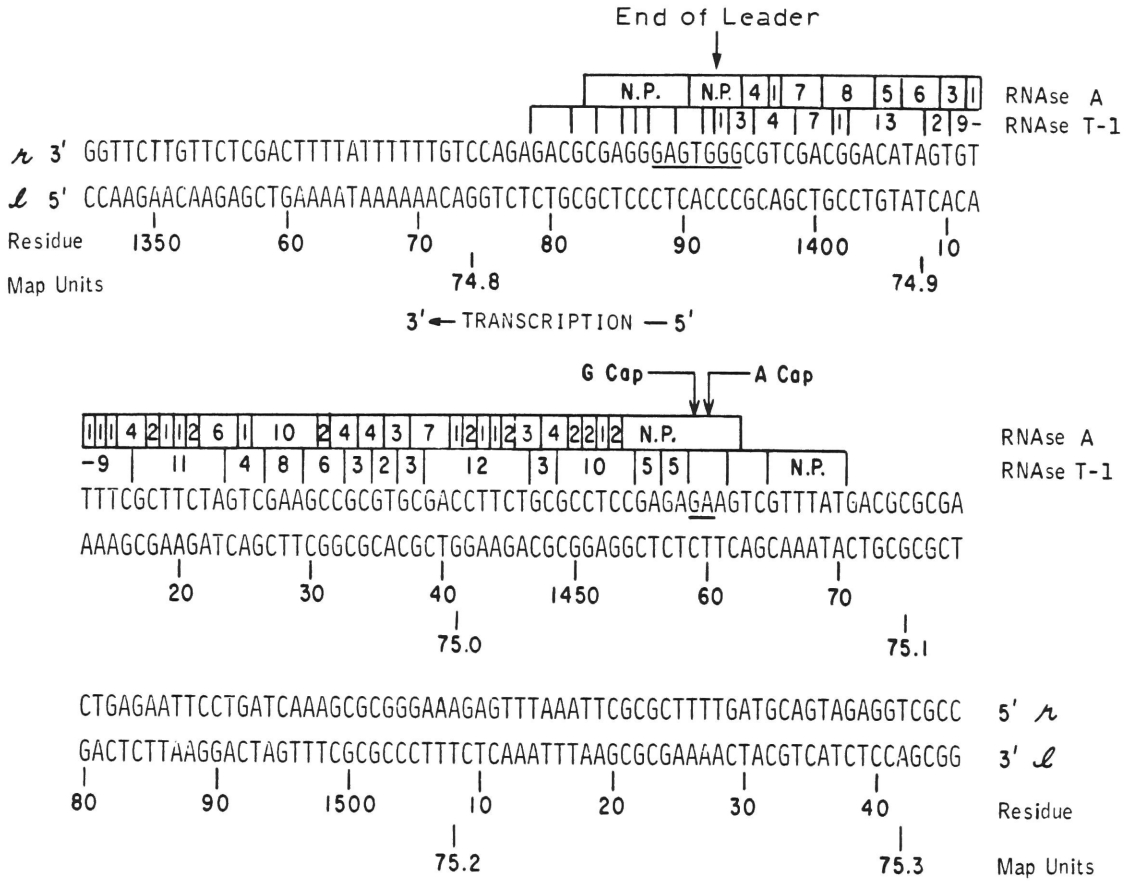


Figure 25. Alignment of the RNAse T1 and RNAse A Oligonucleotides from the EIIa Leader with the DNA Sequence of the Hae III C Subfragment of Eco RI F

The DNA sequence of the Hae III C subfragment of the Eco RI F fragment is taken from Galibert et al. (1979). Residue numbers are relative to the left end of the Eco RI F fragment which was taken to be 0. Map coordinates are relative to the ends of the Eco RI F fragment which were taken to be 70.7 and 75.9. Transcription of the DBP mRNA is leftward. The two capped 5' termini encoded at residues 1459 and 1460 (underlined) are discussed in Results. The positions of the oligonucleotides of Figures 24A, B, and C and Table II are indicated above the sequence. N.P. designates oligonucleotides not found in the fingerprints of Figure 24. The splice donor consensus sequence GGGTGAG and the 3' end of the leader are also indicated.

that residue 1396 is also included within the first leader. Together these results place the splice point between residues 1392 and 1396. The sequence GGGTGAG, similar to the consensus splice donor sequence, is found at this site, so the end of the first leader is probably at nucleotide 1393. Note that one RNAse A product, spot 9 of Figure 24C, (AG,2G)C, which appears reproducibly in high molar yield, cannot be accommodated by the DNA structure. This would be expected for an oligonucleotide that spanned the splice point and derived its 3' residues from the second leader.

Localization of the 5' Terminus The RNAse T1 oligonucleotide CCUCCG, spot 5 of Figure 24A (residues 1454 to 1449), is present in the leader. However, neither the RNAse T1 hexanucleotide UAUUUG (residues 1470 to 1465) nor the RNAse A nonanucleotide GAAGAGAGC (residues 1462 to 1454) are found. These results localize the 5' end between residues 1462 and 1455, a region which contains a long stretch of purine nucleotides. As shown in Table II, two oligonucleotides released by RNAse A, spots 11 and 12 of Figure 24C, with structure given in Figure 26, were the only RNAse A fingerprint spots which contained components resistant to RNAse T2 cleavage. These spots migrated roughly as 6-7' mers and therefore were rich in purines. To confirm the presence of a 5' cap structure in these oligonucleotides, spots 11 and 12 were digested with a mixture of P1 nuclease and alkaline phosphatase and electrophoresed on 3MM paper. Figure 27 shows that spot 11 yielded a resistant product which co-migrated with $^{7m}GpppG_m$ while spot 12 released a product with the mobility of $^{7m}GpppA_m$. In addition to these cap cores, faster moving products are seen. These products were tentatively identified as cap cores with ring opened inverted ^{7m}G's.

These results demonstrate two distinct capped RNAse A oligonucleotides present in the nuclease resistant mRNA hybrid to Eco RI F. To further analyze these termini, early mRNA was 5' labeled in vitro and hybridized to Eco RI F. Digestion of this hybrid selected RNA with RNAse A yielded two products, spots 1 and 2 shown in Figure 28A, which had mobilities comparable to the capped RNAse A spots previously found in vivo in Figure 24C. With nuclease P1, spots 1 and 2 yielded pG_m and

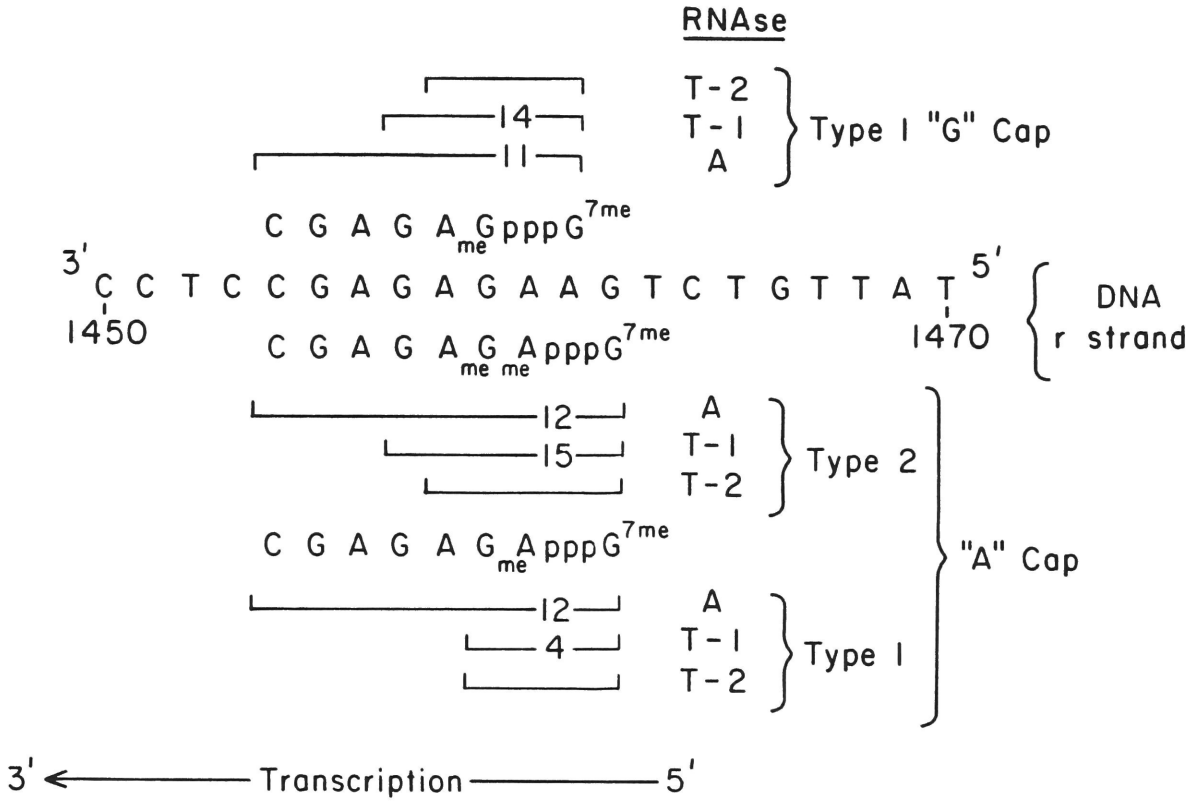


Figure 26. Capped Oligonucleotides Released from the DBP mRNA by RNAse A, RNAse T1 and RNAse T2

A segment of the DNA r strand from Figure 25 coding for the DBP message 5' terminus is shown. Aligned with this sequence are the structures of the capped RNAse T1 oligonucleotides from Figures 24A and B and capped RNAse A oligonucleotides from Figure 24C. Oligonucleotide numbers identify the corresponding spots in these fingerprints. The predicted RNAse T2 redigestion products of the RNAse T1 and RNAse A oligonucleotides are shown.

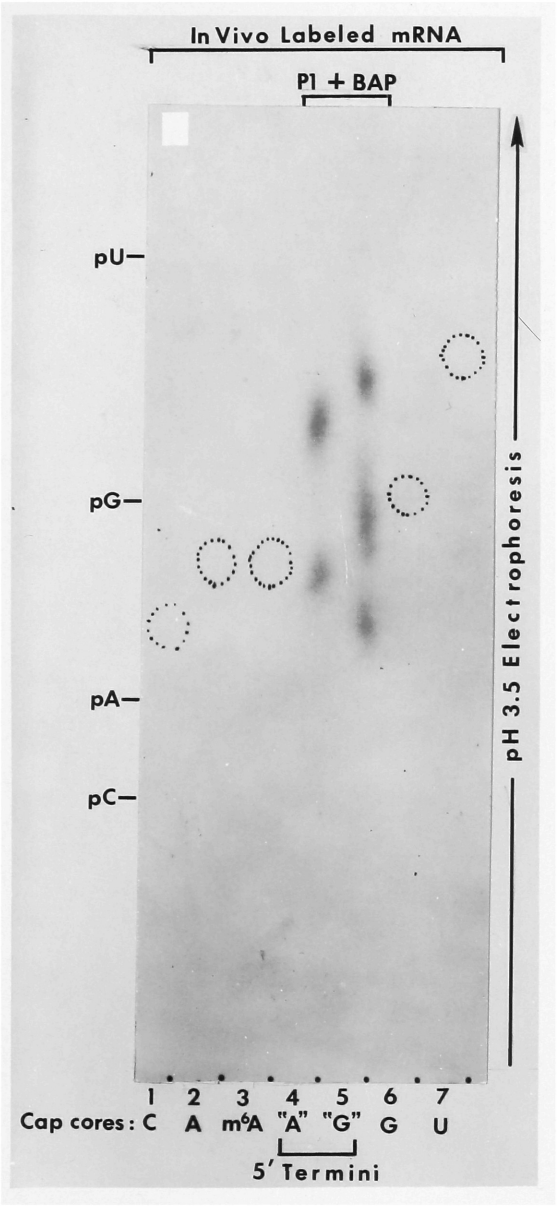


Figure 27. Secondary Analysis of *in Vivo* Labeled RNAse A Cap Oligonucleotides from the 5' Termini of the DBP mRNA

Spots 11 and 12 from Figure 24C (RNAse A fingerprint of *in vivo* ^{32}P labeled RNA hybridized to Eco RI F) were digested with nuclease P1 and bacterial alkaline phosphatase and electrophoresed on 3MM paper at pH 3.5 in lanes 5 and 4, respectively . Cold cap core standards ($^{7m}GpppX_m$ where X is C, A, m^6A, G, and U) were run in lanes 1, 2, 3, 6, and 7 respectively. The highest mobility spots in lanes 4 and 5 presumably have the ring opened form of ^{7m}G and are discussed in Results. The spot with the lowest mobility in lane 5 has not been identified.

Figure 28. In Vitro Kinasing and Sequencing of the 5' Terminal RNAse A Oligonucleotides of the DBP mRNA

(A) Polyadenylated cytoplasmic RNA isolated at 5 hours post-infection from $4X10^6$ HeLa cells infected with Ad2 and incubated in the presence of cycloheximide was chemically decapped and labeled at the 5' end by the polynucleotide kinase transfer of $^{32}PO_4$ from γ-^{32}P-ATP. The kinased RNA was then hybridized to filter bound Eco RI F DNA, eluted, and RNAse A fingerprinted.

In panels (B) and (C), spots 2 and 1, respectively, were partially digested with RNAse T2 and fingerprinted. The sequences deduced for spots 2 and 1 are -AGAGCp and -GAGCp, respectively.

Homochromatography
pH 3.5 Electrophoresis
A
B
C
b
y
1
2
A
G
A
G
C
G
A
G
C

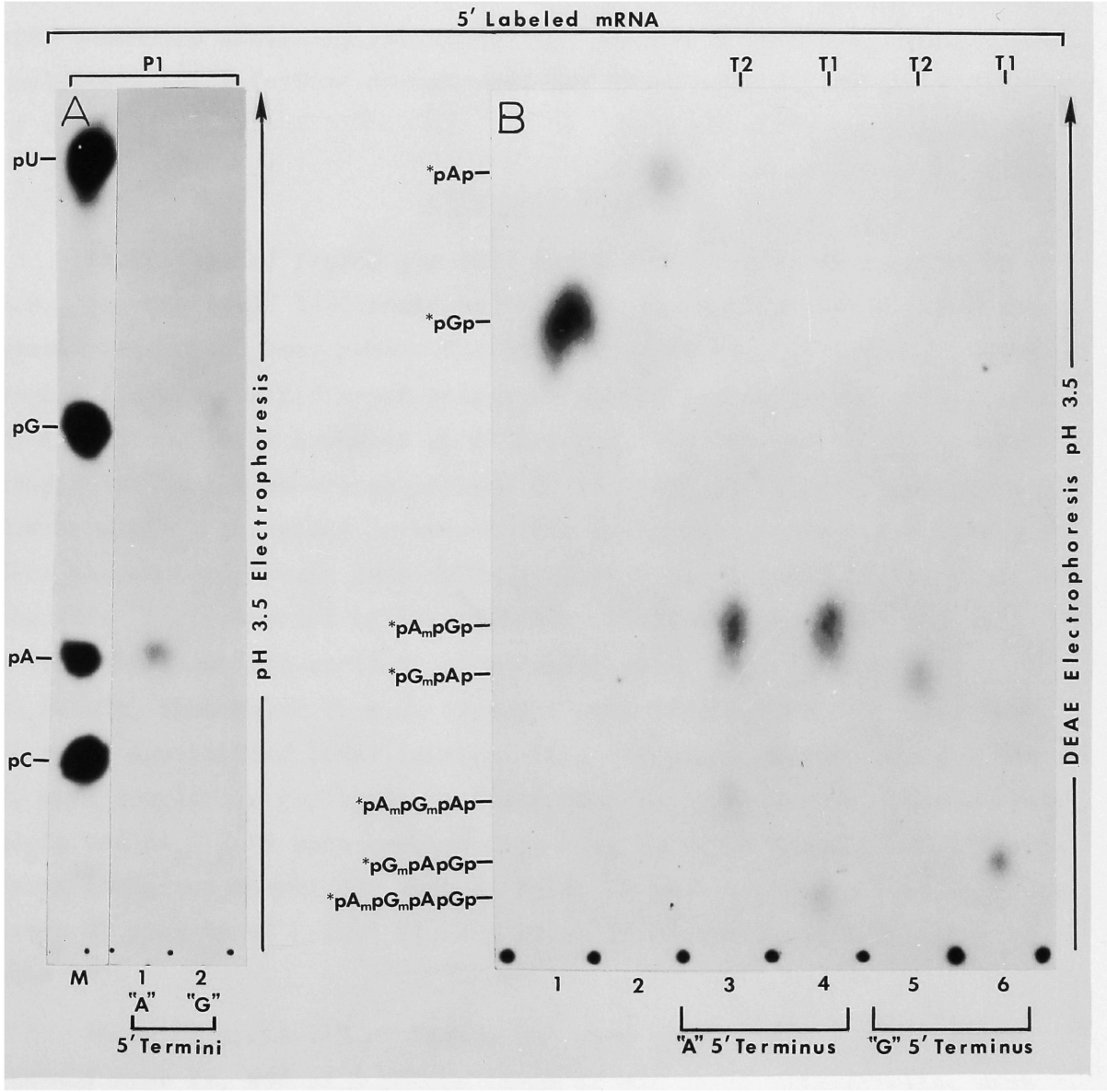
pA_m respectively, identifying their 5' termini as shown in Figure 29A lanes 2 and 1, respectively. DEAE paper electrophoresis of RNAse T2 and RNAse T1 digests shown in Figure 29B lanes 3-6, confirmed pG_mAG- at the 5' terminus of spot 1, and $pA_mG_{(m)}AG$- at the 5' end of spot 2 when mobility was interpreted according to the M-value rules of Sanger et al. (1965). Nuclease resistant oligonucleotides expected from the 2'-O-methyl pattern of both type I and type II cap structures were found for spot 2, but only type I products for spot 1. Sequences at the 3' end of spots 1 and 2 were determined by the method of Galibert, Sedat and Ziff (1974) through two dimensional fractionation of partial RNAse T2 digests as shown in Figures 28B and C. For both spots the sequence -GAGCp was found, suggesting that these capped oligonucleotides overlapped in structure.

In contrast to the in vivo labeled capped RNAse A oligonucleotides, in vivo labeled capped RNAse T1 products, spots 4 and 14 of Figure 24A, with structures shown in Figure 26, were obtained from nuclease trimmed hybrids in low yield. These same products appeared, however, in the fingerprint of the untrimmed hybrid, Figure 24B, in increased yield. The latter fingerprint gave a third capped oligonucleotide, spot 15, also shown in Figure 26. The RNAse T2 resistant moieties were isolated from these spots and treated with P1 nuclease. Spots 4 and 15 yielded $^{7m}GpppA_m$ cap cores while spot 14 gave $^{7m}GpppG_m$ (data not shown). In all three P1 nuclease digestions, only pG and pA mononucleotides were released.

Alignment of Capped Oligonucleotides with the DNA Sequence The sequences AGAG- and -AGAGC found, respectively, at the 5' and 3' ends of the capped RNAse A product with 5' A can be accommodated within the DNA sequence at only one position, between residues 1460 and 1454. From the corresponding sequence data for the capped oligonucleotide with 5' G, I conclude that the second capped sequence is also encoded at this site, but is one residue shorter at the 5' end. These assignments, summarized in Figure 26, are consistent with the relative mobilities of the various capped oligonucleotides in the fingerprints of Figure 24 and Figure 28A, when interpreted according to the rules of Galibert et al. (1974) and

Figure 29. Secondary Analysis of <u>in Vitro</u> Labeled 5' Terminal Oligonucleotides from the DBP mRNA

Spots 1 and 2 from Figure 28A (RNAse A fingerprint of <u>in vitro</u> decapped and kinased RNA hybridizing to Eco RI F) were redigested with nuclease P1 and electrophoresed on 540 paper at pH 3.5 (Panel A) to determine the 5' terminal nucleotide and with RNAse T2 and RNAse T1 (Panel B) and electrophoresed on DEAE paper at pH 3.5 to determine the next 2-3 nucleotides. A nuclease P1 digest of ^{32}P labeled RNA was run in the lane labeled M as 5'XMP standards. Lanes 1 and 2 have pGp and pAp standards respectively. The two redigestion products in lanes 3 and 4 result from the absence or presence of a 2'-O-methyl group on the 5' penultimate nucleotide of each RNA molecule depending upon whether it had a type I or a type II cap structure, respectively. Two 5' termini can be deduced from this analysis: $pA_mG_{(m)}AG-$ from spot 2 and pG_mAG- from spot 1. Interpretation of these results is aided by Figure 26.


5′ Labeled mRNA
P1
A
pU
pG
pA
pC
pH 3.5 Electrophoresis
M
1
2
"A"
"G"
5′ Termini
T2
T1
T2
T1
B
*pAp
*pGp
*pAmpGp
*pGmpAp
*pAmpGmpAp
*pGmpApGp
*pAmpGmpApGp
DEAE Electrophoresis pH 3.5
1
2
3
4
5
6
"A" 5′ Terminus
"G" 5′ Terminus

are also in agreement with the results of nuclease redigestions given in Figures 28B and C, and Figure 29, and in Table II.

Thus the DBP mRNA has a heterogeneous 5' terminus with a major A terminus and a G terminus 1/4 as abundant which is encoded one nucleotide downstream. Overexposure of the fingerprint in Figure 28A, along with secondary analyses, indicates that another A terminus, encoded one nucleotide still further downstream, may be present at 1/4 the abundance of the G terminus.

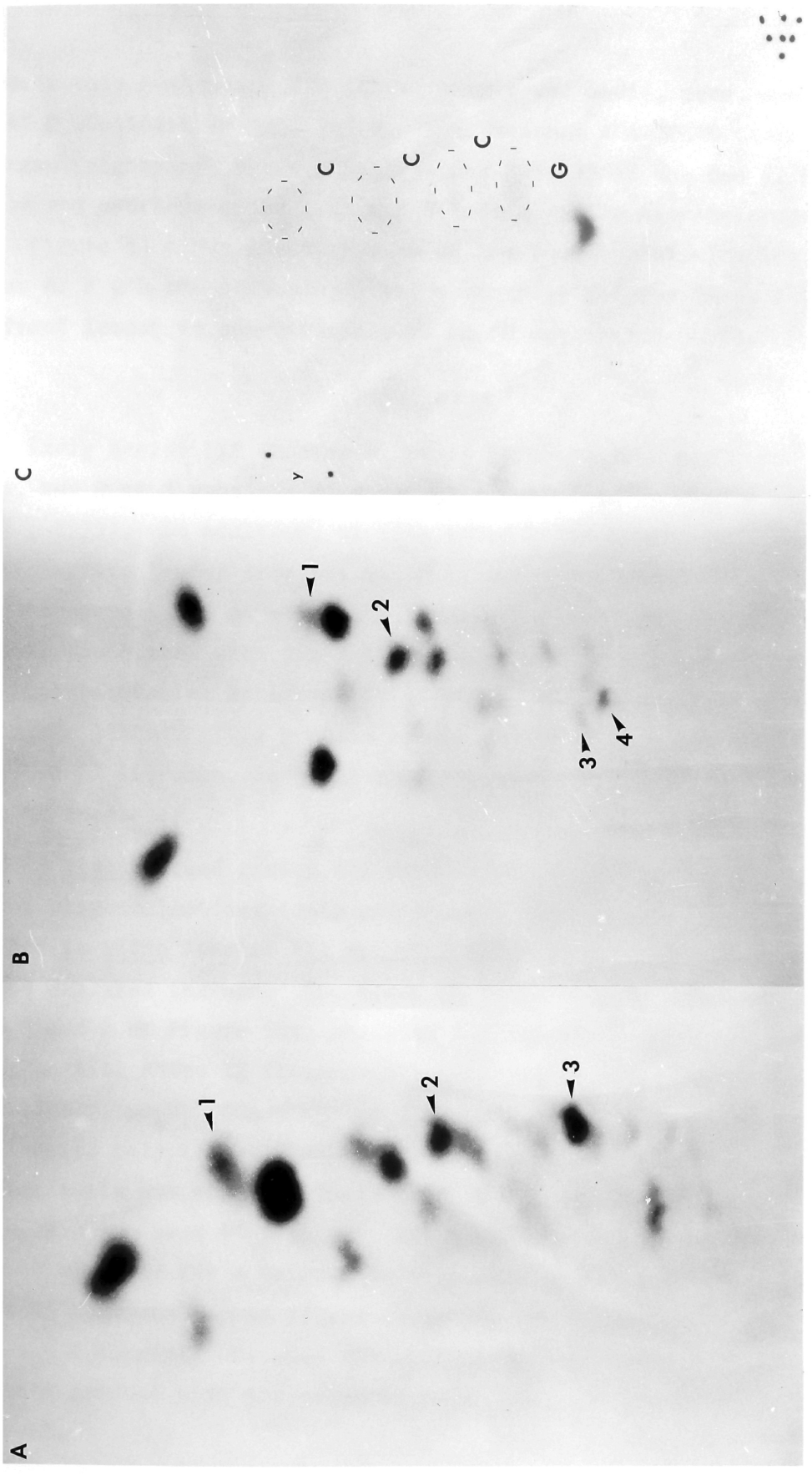
EIIa Late mRNA

At late times region IIa mRNA has a first leader which maps at 72 m.u., approximately 1000 residues from the early first leader, but the second leader and body remain the same (Chow et al., 1979,1980). Thus region IIa may use different promoters during the early and late stages of infection. mRNA isolated at 24 hrs p.i. was labeled in vitro, and hybridized to the separated strands of Eco RI F to resolve the region IIa termini (transcribed leftward) from background spots which result from abundant rightward late mRNAs. Spots 4 and 3 found in low yield in the RNAse A fingerprint of the l strand hybrid in Figure 30B are the previously described early 5' oligonucleotides, $pA_mG_{(m)}AGAGCp$ and $pG_mAGAGCp$, encoded at 75 m.u. (compare with Figure 28A). Two new short (2 and 3 nucleotides long) late specific oligonucleotides, spots 1 and 2, were completely resistant to RNAse T2. Their sequences, respectively pA_mCp and $pA_mG_{(m)}Cp$ were deduced from data in Table I and from DEAE analyses (data not shown; for spot 2, RNAse T2 gave a pattern like combined lanes 47 plus 49 of Figure 11, and RNAse T1 analysis was like lanes 48 plus 50).

RNAse T1 spots 1-3 of Figure 30A were pA_mGp, $pA_mC_{(m)}AGp$ (predominately type I), and $pA_mG_mCCCCGp$ (type II cap only) respectively, with structures deduced from data in Table I, from Figure 11 lanes 45-50, and from Figure 30C. Spots 1 and 3 are types I and II products from a single terminus with sequence $^{7m}GpppA_mG_{(m)}CCCCG-$. Spot 2 is from a second late terminus, $^{7m}GpppA_mC_{(m)}AG-$. The donor splice site of the late region IIa first leader maps at coordinate 72, (Chow et al., 1979)

Figure 30. The 5' Termini of the Late Form of Region IIa mRNAs

Late mRNA 5' labeled *in vitro* by polynucleotide kinase and fragmented with alkali was hybridized to the separated *l* strand of Eco RI F on a nitrocellulose filter. The RNAse T1 and A fingerprints of the eluted hybrid are shown in (A) and (B), respectively. Numbered spots are mRNA 5' terminal oligonucleotides. Spots 3 and 4 in (B) correspond to spots 1 and 2, respectively, in Figure 28A and thus are a contribution from the early form of the region IIa message also present in the preparation. Spot 3 of (A), partially digested with RNAse T2, was fingerprinted in (C). The dashed circles locate spots visible in the original autoradiogram. The sequence at the 3' end of spot 3 deduced from (C) is -CCCGp.


A
1
2
3
B
1
2
3
4
C
C
C
C
G
y

approximately nucleotide 270 (Chow, Broker and Lewis, pers. comm.) of Eco RI F (Galibert et al., 1979). The sequence ACAGCCCCG occurs just upstream (rightward) between nucleotides 315-323 of the Eco RI F r strand and overlaps spots 1, 2 and 3 into a single microheterogeneous 5' end. Figure 31 shows the alignment of the two termini with the sequence of Eco RI F and the probable splice donor site for the first leader. The first leader is approximately 68 to 70 nucleotides long.

EIII mRNA

Early Region III encodes a family of mRNAs with a common 5' terminus, but 2 or 3 possible 3' ends, as illustrated in Figure 32. (Chow et al., 1979). In addition, at late times these mRNAs are produced with the tripartite leader from the major late transcription unit. This region encodes a 19K glycoprotein (Persson, Signas and Philipson, 1979) which is associated with the cell membrane and as a complex with the cell transplantation antigens (Kvist et al., 1978; Persson, Jansson and Philipson, 1980a). This protein is encoded by mRNA species a, b and c of Figure 32 (Persson, Jornvall and Zabielski, 1980b) and is translated as a 16K precursor.

In vivo labeled region III mRNA selected by Eco RI D gave at least two cap oligonucleotides (data not shown), but the fingerprint was complex and in vitro labeled RNA selected by 75.9-78.15 DNA (Eco RI D/Bgl II) was analyzed instead. The major in vitro labeled RNAse T1 products, spots 1 and 2 of Figure 33A, analyzed in Table I, Figure 11 lanes 11-14, and by partial RNAse T2 fingerprinting in Figure 33B and 33C were, respectively, $pA_mU_{(m)}UCAGp$ and $pG_mU_{(m)}AUUCAGp$. These oligonucleotides were labeled only if the mRNA was decapped prior to phosphatasing and kinasing (data not shown). These form an overlapping oligonucleotide pair with a two base 5' stagger. The arrow in Figure 33A indicates the position expected for a third potential oligonucleotide, $pU_mA_{(m)}UUCAG$, from a 5' terminus capped at the uridylate residue intervening between the G and A termini. Because the spot at this position gave no RNAse T2 resistant product with the sequence $pU_mA_{(m)}U-$, the potential U terminus is absent.

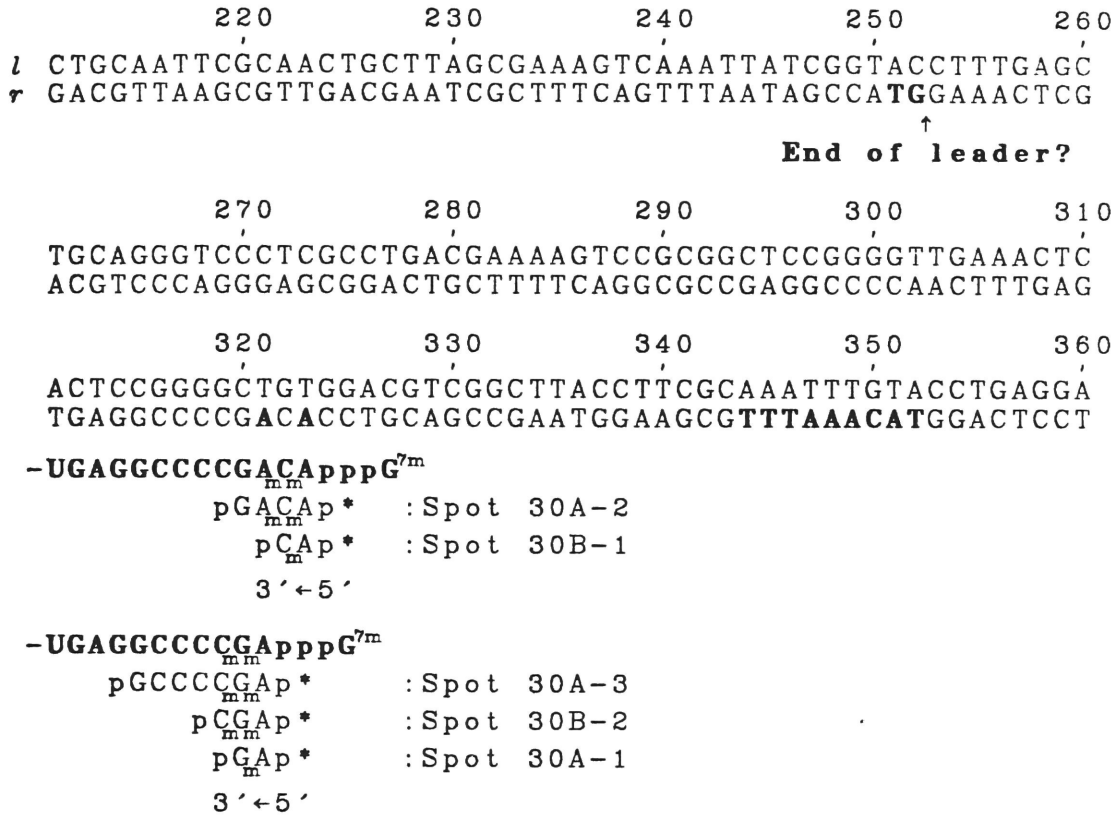


Figure 31. DNA Sequences Encoding the Late DNA Binding Protein mRNA 5' Terminus and First Leader

The two late DNA binding protein mRNA 5' termini have been aligned with the sequence of Eco RI F at nucleotides 321 and 323 (sequence and numbering of Galibert et al. (1979)). In addition, all of the 5' terminal oligonucleotides from Figure 30A and B have been aligned with the RNA. The TATA box at nucleotides 344-352 and a possible splice donor site encoded at nucleotide 253 are also indicated. Transcription is off the l strand and is right to left and bottom to top to correspond with the map in Figure 1.

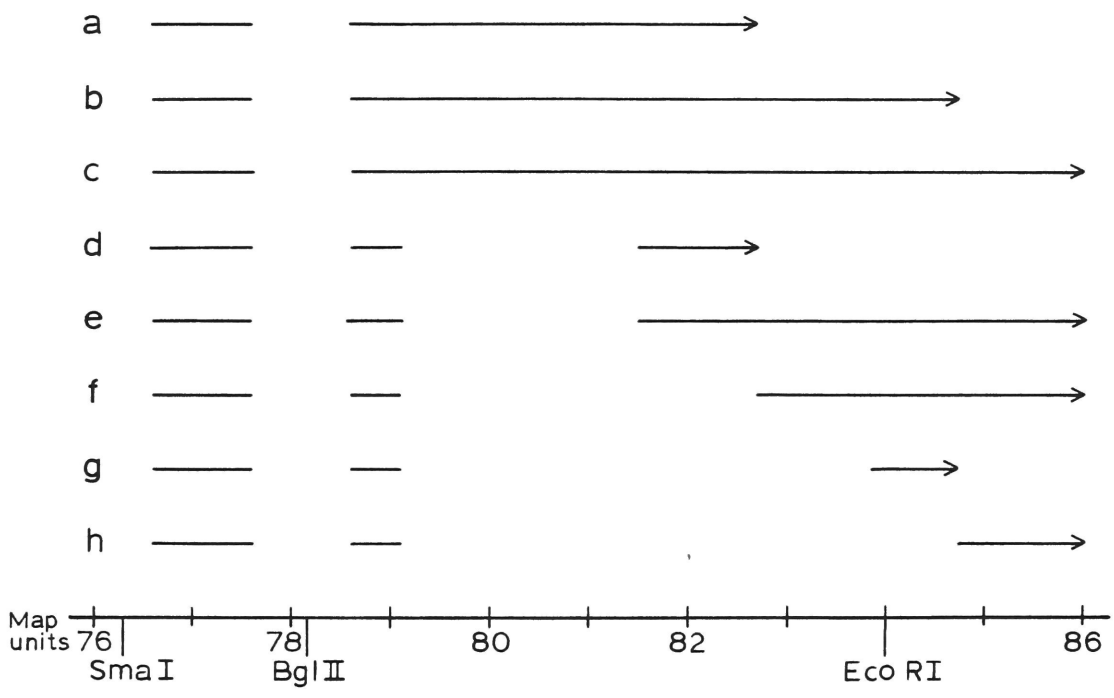
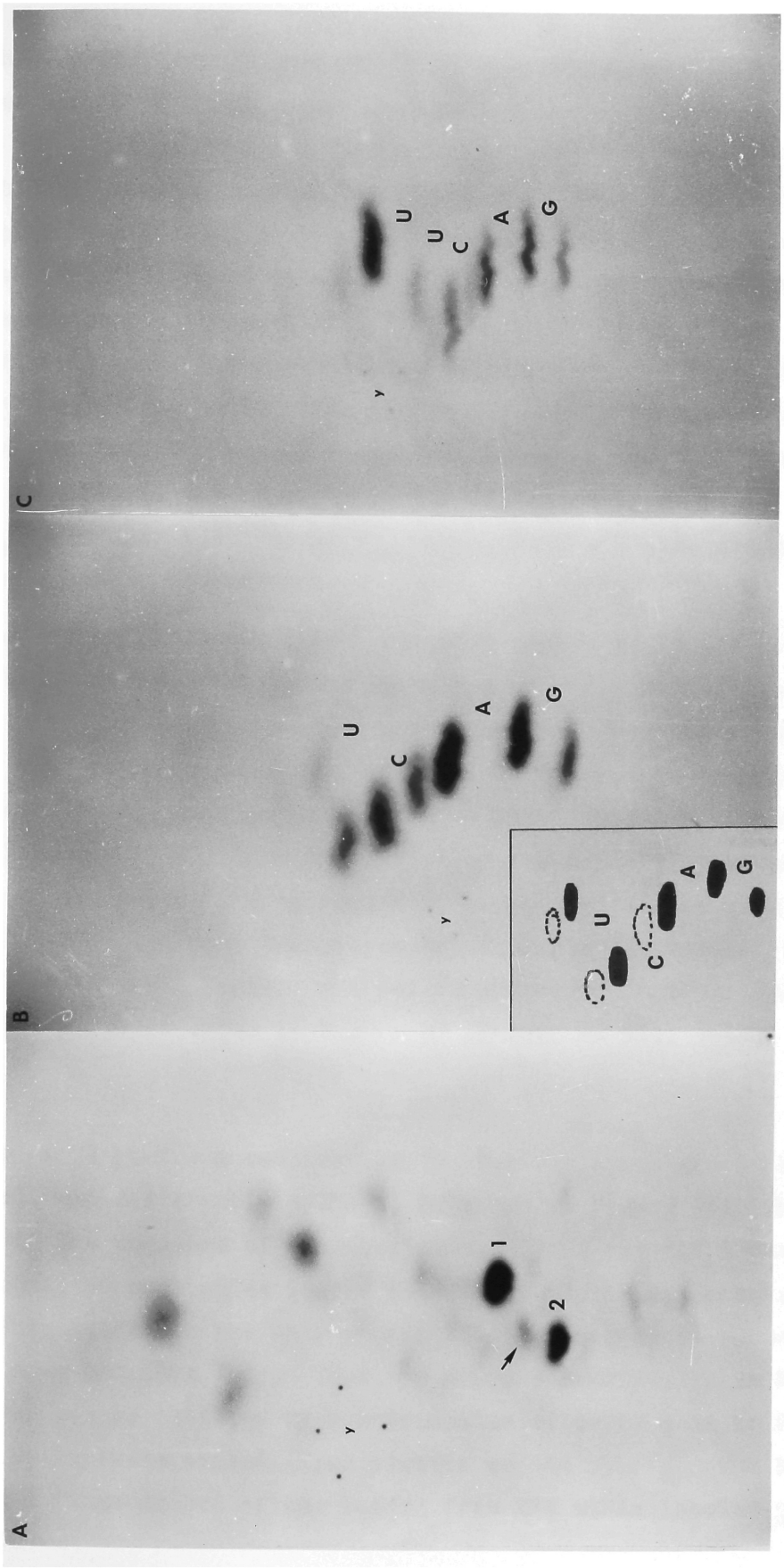


Figure 32. mRNAs of Early Region III

EIII mRNAs were mapped by electron microscopy (Chow et al., 1979). Note that the first leader is encoded within the Eco RI D/Bgl II fragment (75.9-78.15 m.u.).

Figure 33. The 5' Termini of Region III mRNA

Early mRNA was decapped, kinased, alkali fragmented and hybridized to 75.9-78.15 DNA (Eco RI D/Bgl II). The RNAse T1 fingerprint of the eluted hybrid is shown in (A). Spots 1 and 2 (both 5' terminal oligonucleotides), when partially digested with RNAse T2, gave the fingerprints shown in (B) and (C), respectively. The inset in (B) is a diagram distinguishing the main trail of spots (solid) from which the sequence is deduced, from paired minor products shown as dashed spots. The latter are oligonucleotides with the same sequence as the solid spots but with altered charge, and are discussed in the legend to Figure 12. The sequences at the 3' end of spots 1 and 2 deduced from (B) and (C) are -UCAGp and -UUCAGp, respectively.


A
1
2
y
B
U
C
A
G
y
U
C
A
G
C
C
U
U
C
A
G
y

The Ad2 DNA from 75 m.u. to 78.55 m.u. was restriction mapped and sequenced by the strategy shown in Figure 34 and discussed in Appendix B. The DNA sequence is shown in Figure 35. The sequence GTATTCAG is found at only one position on either strand of this region. This is from nucleotides 551-558 on the l (non-sense) strand or approximately 76.6 m.u., the coordinate to which EIII mRNA 5' termini were mapped by electron microscopy (Chow et al., 1979). The sequence ATTCAG can be accommodated 2 nucleotides downstream from nucleotides 553-558 and also on the r strand from nucleotides 842 to 837. This latter position is unlikely to encode the major A terminus, since no known leftward mRNA 5' termini map at this position. Thus the EIII mRNA 5' termini, $^{7m}GpppG_{m}U_{(m)}AUUCAG-$ and $^{7m}GpppA_{m}U_{(m)}UCAG-$, form a 2 base staggered microheterogeneous 5' terminus.

The end of the first leader has been mapped to 77.6 m.u. (Chow et al., 1979). The sequence GGUGAG encoded at nucleotides 922 to 927 is similar to the consensus splice donor sequence. Thus the first leader probably ends at nucleotide 922. The only AUG found in the first leader occurs near the splice site at nucleotide 841. This AUG, however, is not utilized for translation initiation of the EIII/19K glycoprotein (Persson et al., 1980b), even though it is present in the mRNA. It should also be noted that the polyadenylation signal, AAUAAA, for the late L4 3' coterminal family of mRNAs is encoded at nucleotides 1147 to 1152 of the sequence shown in Figure 35 (Fraser et al., 1981).

EIV mRNA

Region IV yields a complex family of mRNAs which are 5' and 3' coterminal, but differently spliced, as shown in Figure 36. When synthesized in the presence of cycloheximide, 95% of the EIV mRNAs have an approximately 50 nucleotide leader encoded at 99.1 m.u. (Chow et al., 1979). This leader is the only exonic RNA from region IV encoded within 98.3-100 m.u. DNA (Sma I K). Thus, as shown schematically in Figure 36, this leader can be isolated by hybridization of early mRNA to Sma I K followed by nuclease trimming and elution of the hybrid. The RNAse T1 and RNAse A fingerprints of the leader from EIV mRNAs labeled in vivo

Figure 34. Strategy for DNA Sequencing of the Region Encoding the EIIa and EIII 5' Termini

Nucleotide numbers are those of the sequence presented in Figure 35. The positions of the EIIa and EIII leaders are indicated. Mapping of the Sma I sites at 78.5 and 76.5 is discussed in Appendix B. Rightward arrows indicate sequence data obtained from the l strand and leftward arrows from the r strand.

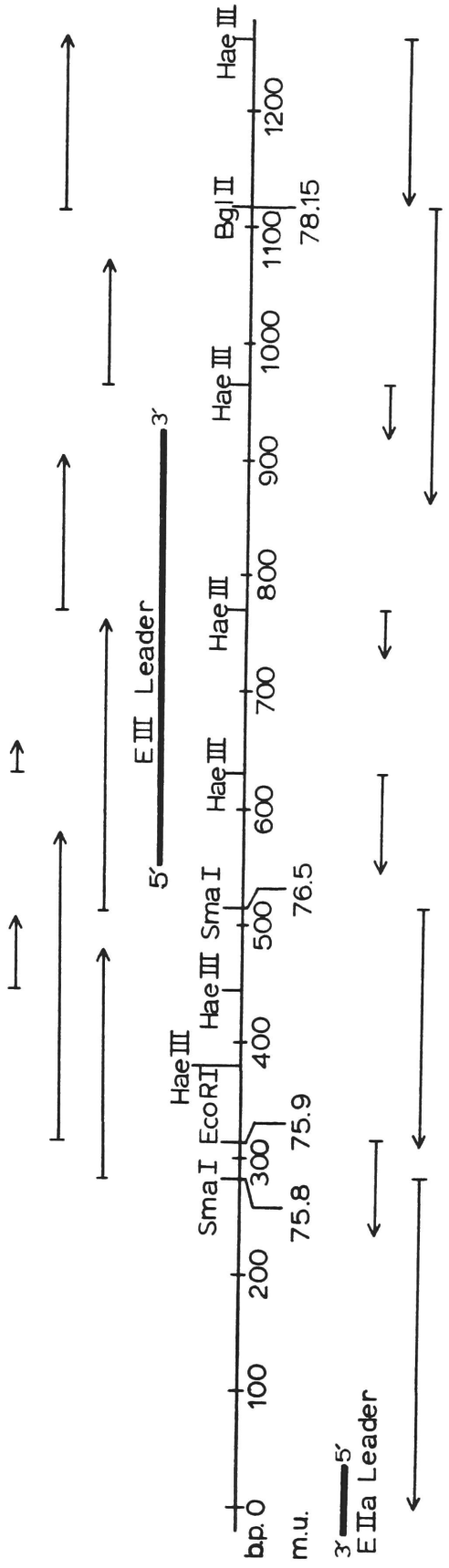
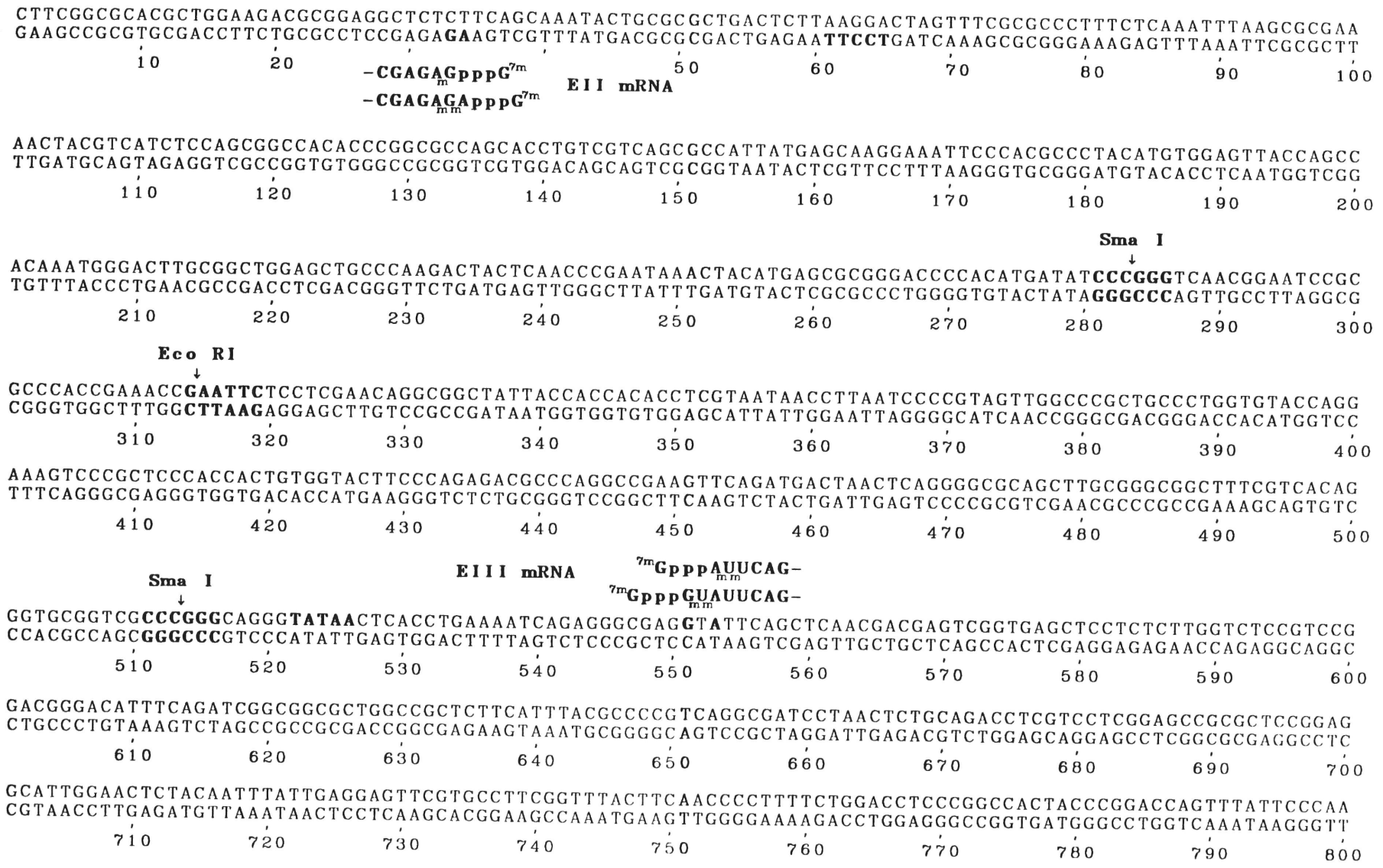
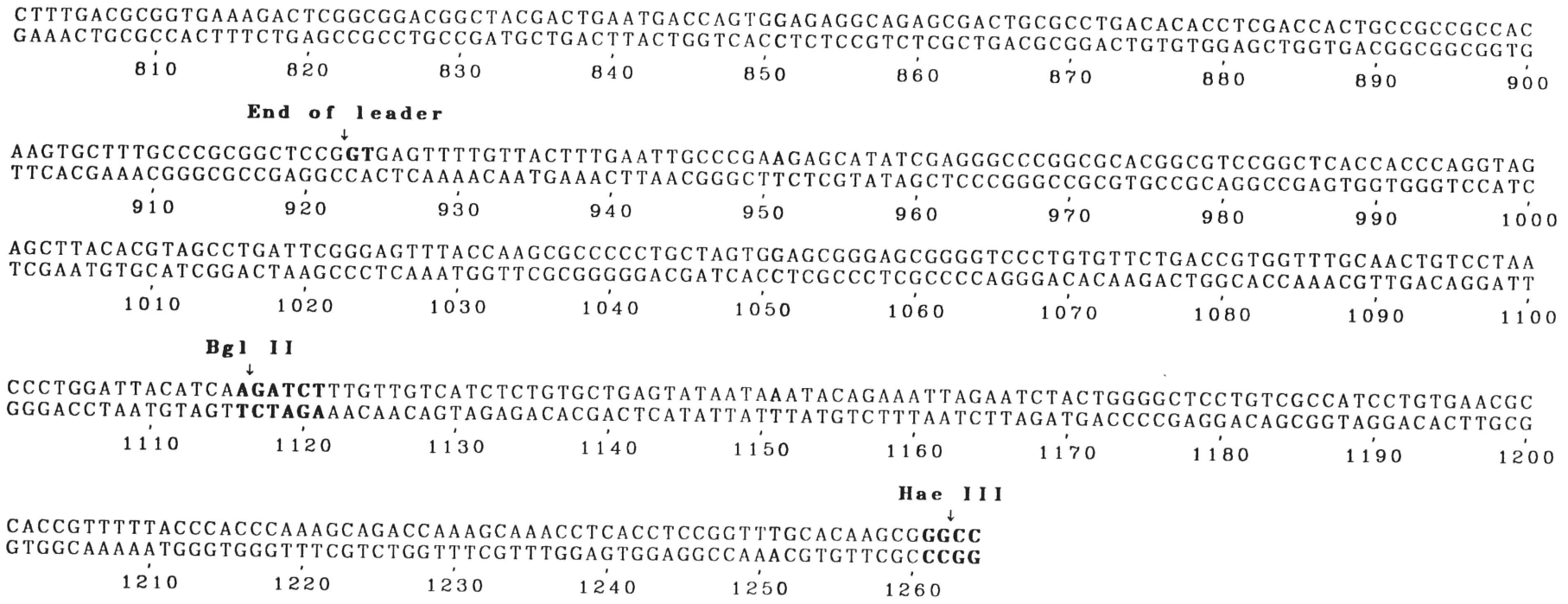

bp. 0
100
200
300
400
500
600
700
800
900
1000
1100
1200
m.u.
75.8
75.9
76.5
78.15
Sma I
EcoRI
Hae III
Hae III
Sma I
Hae III
Hae III
Hae III
Bgl II
Hae III
E III Leader
5'
3'
3'
5'
E IIa Leader

Figure 35. DNA Sequence Encoding the EIIa and EIII mRNA 5' Termini

The EIIa and EIII mRNAs have been aligned with the DNA sequence. The 3' end of the EIII leader is probably at nucleotide 922. The polyadenylation signal (AAUAAA) for the L4 late 3' coterminal family of mRNAs is encoded from nucleotides 1147 to 1152 (Fraser et al., 1981).


l CTTCGGCGCACGCTGGAAGACGCGGAGGCTCTCTTCAGCAAATACTGCGCGCTGACTCTTAAGGACTAGTTTCGCGCCCTTTCTCAAATTTAAGCGCGAA
r GAAGCCGCGTGCGACCTTCTGCGCCTCCGAGAGAAGTCGTTTATGACGCGCGACTGAGAATTCCTGATCAAAGCGCGGGAAAGAGTTTAAATTCGCGCTT
10 20
-CGAGAGpppG7m
-CGAGAGAApppG7m
EII mRNA
50 60 70 80 90 100
AACTACGTCATCTCCAGCGGCCACACCCGGCGCCAGCACCTGTCGTCAGCGCCATTATGAGCAAGGAAATTCCCACGCCCTACATGTGGAGTTACCAGCC
TTGATGCAGTAGAGGTCGCCGGTGTGGGCCGCGGTCGTGGACAGCAGTCGCGGTAATACTCGTTCCTTTAAGGGTGCGGGATGTACACCTCAATGGTCGG
110 120 130 140 150 160 170 180 190 200
Sma I
ACAAATGGGACTTGCGGCTGGAGCTGCCCAAGACTACTCAACCCGAATAAACTACATGAGCGCGGGACCCCACATGATATCCCGGGTCAACGGAATCCGC
TGTTTACCCTGAACGCCGACCTCGACGGGTTCTGATGAGTTGGGCTTATTTGATGTACTCGCGCCCTGGGGTGTACTATAGGGCCCAGTTGCCTTAGGCG
210 220 230 240 250 260 270 280 290 300
Eco RI
GCCCACCGAAACCGAATTCTCCTCGAACAGGCGGCTATTACCACCACACCTCGTAATAACCTTAATCCCCGTAGTTGGCCCGCTGCCCTGGTGTACCAGG
CGGGTGGCTTTGGCTTAAGAGGAGCTTGTCCGCCGATAATGGTGGTGTGGAGCATTATTGGAATTAGGGGCATCAACCGGGCGACGGGACCACATGGTCC
310 320 330 340 350 360 370 380 390 400
AAAGTCCCGCTCCCACCACTGTGGTACTTCCCAGAGACGCCCAGGCCGAAGTTCAGATGACTAACTCAGGGGCGCAGCTTGCGGGCGGCTTTCGTCACAG
TTTCAGGGCGAGGGTGGTGACACCATGAAGGGTCTCTGCGGGTCCGGCTTCAAGTCTACTGATTGAGTCCCCGCGTCGAACGCCCGCCGAAAGCAGTGTC
410 420 430 440 450 460 470 480 490 500
Sma I
EIII mRNA
7mGpppAUUCAG-
7mGpppGUAUUCAG-
GGTGCGGTCGCCCCGGGCAGGGTATAACTCACCTGAAAATCAGAGGGCGAGGTATTCAGCTCAACGACGAGTCGGTGAGCTCCTCTCTTGGTCTCCGTCCG
CCACGCCAGCGGGGCCCGTCCCATATTGAGTGGACTTTTAGTCTCCCGCTCCATAAGTCGAGTTGCTGCTCAGCCACTCGAGGAGAGAACCAGAGGCAGGC
510 520 530 540 550 560 570 580 590 600
GACGGGACATTTCAGATCGGCGGCGCTGGCCGCTCTTCATTTACGCCCCGTCAGGCGATCCTAACTCTGCAGACCTCGTCCTCGGAGCCGCGCTCCGGAG
CTGCCCTGTAAAGTCTAGCCGCCGCGACCGGCGAGAAGTAAATGCGGGGCAGTCCGCTAGGATTGAGACGTCTGGAGCAGGAGCCTCGGCGCGAGGCCTC
610 620 630 640 650 660 670 680 690 700
GCATTGGAACTCTACAATTTATTGAGGAGTTCGTGCCTTCGGTTTACTTCAACCCCTTTTCTGGACCTCCCGGCCACTACCCGGACCAGTTTATTCCCAA
CGTAACCTTGAGATGTTAAATAACTCCTCAAGCACGGAAGCCAAATGAAGTTGGGGAAAAGACCTGGAGGGCCGGTGATGGGCCTGGTCAAATAAGGGTT
710 720 730 740 750 760 770 780 790 800


CTTTGACGCGGTGAAAGACTCGGCGGACGGCTACGACTGAATGACCAGTGGAGAGGCAGAGCGACTGCGCCTGACACACCTCGACCACTGCCGCCGCCAC
GAAACTGCGCCACTTTCTGAGCCGCCTGCCGATGCTGACTTACTGGTCACCTCTCCGTCTCGCTGACGCGGACTGTGTGGAGCTGGTGACGGCGGCGGTG
810 820 830 840 850 860 870 880 890 900
End of leader
AAGTGCTTTGCCCGCGGCTCCGGTGAGTTTTGTTACTTTGAATTGCCCGAAGAGCATATCGAGGGCCCGGCGCACGGCGTCCGGCTCACCACCCAGGTAG
TTCACGAAACGGGCGCCGAGGCCACTCAAAACAATGAAACTTAACGGGCTTCTCGTATAGCTCCCGGGCCGCGTGCCGCAGGCCGAGTGGTGGGTCCATC
910 920 930 940 950 960 970 980 990 1000
AGCTTACACGTAGCCTGATTCGGGAGTTTACCAAGCGCCCCCTGCTAGTGGAGCGGGAGCGGGGTCCCTGTGTTCTGACCGTGGTTTGCAACTGTCCTAA
TCGAATGTGCATCGGACTAAGCCCTCAAATGGTTCGCGGGGGACGATCACCTCGCCCTCGCCCCAGGGACACAAGACTGGCACCAAACGTTGACAGGATT
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
Bgl II
CCCTGGATTACATCAAGATCTTTGTTGTCATCTCTGTGCTGAGTATAATAAATACAGAAATTAGAATCTACTGGGGCTCCTGTCGCCATCCTGTGAACGC
GGGACCTAATGTAGTTCTAGAAACAACAGTAGAGACACGACTCATATTATTTATGTCTTTAATCTTAGATGACCCCGAGGACAGCGGTAGGACACTTGCG
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
Hae III
CACCGTTTTTACCCACCCAAAGCAGACCAAAGCAAACCTCACCTCCGGTTTGCACAAGCGGGCC
GTGGCAAAAATGGGTGGGTTTCGTCTGGTTTCGTTTGGAGTGGAGGCCAAACGTGTTCGCCCGG
1210 1220 1230 1240 1250 1260

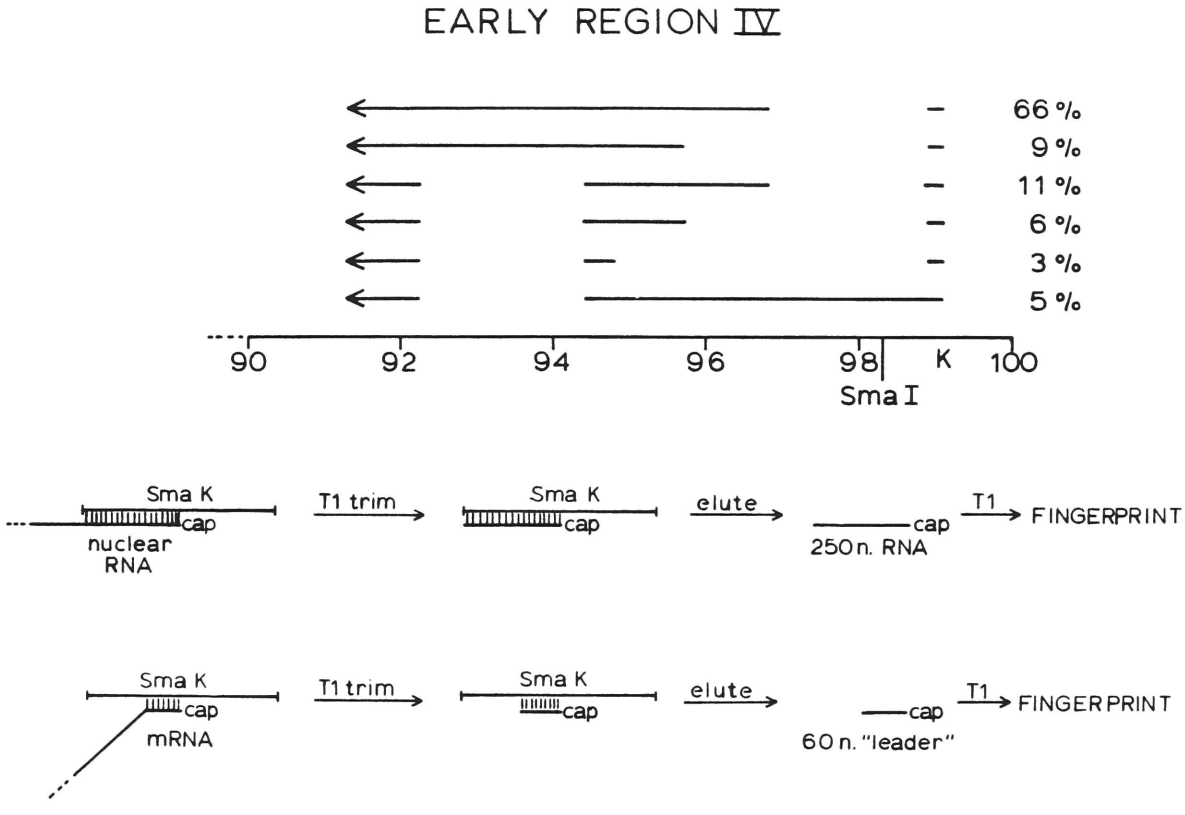


Figure 36. mRNAs of Early Region IV

The various mRNAs of EIV mapped by electron microscopy (Chow et al., 1979) are shown, along with the relative amounts of each when synthesized in the presence of cycloheximide. The bottom two lines show the strategy used to obtain the fingerprints of Figure 37 A and B.

Figure 37. Fingerprint Analysis of RNA from Early Region IV

Nuclear and polyadenylated cytoplasmic RNA was prepared from Ad2 infected cells labeled with ^{32}P for the 5 hour period following infection in the presence of cycloheximide added at 30 minutes post infection. In (A), total nuclear RNA was hybridized to 98.3-100 DNA (Sma I K), RNAse T1 trimmed, eluted, and RNAse T1 fingerprinted. In (B) and (C), polyadenylated cytoplasmic RNA was hybridized to Sma I K, RNAse T1 or RNAse A trimmed, eluted, and RNAse T1 or A fingerprinted, respectively. The secondary analyses of numbered spots in (A)-(C) are presented in Table III. RNAse T2 analyses demonstrating that RNAse T1 spot 8 and spots a, b and c of (B) are cap oligonucleotides are given in panel (G). RNAse A spot 9 was also a cap oligonucleotide by this test.

(D) and (E) compare the region IV 5' terminal oligonucleotides labeled in vitro from two adenovirus strains. Ad2, analyzed in (D), gives five U terminated minor spots plus a major A terminated spot. Ad5, analyzed in (E), yields an additional U terminated product. These differences are correlated with the sequences of the Ad2 and Ad5 promoter regions in the text.

In (F), spot 1 of (D) was partially digested with RNAse T2 and fingerprinted. The deduced sequence at the 3' end of spot 1 is -ACUGp.

In (G), minor capped oligonucleotides a, b and c and the major capped oligonucleotide, spot 8, all from (B) were digested with RNAse T2 and electrophoresed on DEAE paper at pH 3.5. The RNAse T2 resistant portions of the oligonucleotides containing the cap structure are identified. Spot 8 from the nucleus in (A) also gave RNAse T2 resistant material (not shown).

with ^{32}P are shown in Figure 37B and C, respectively. Unspliced nuclear RNA, however, is colinear with the DNA. Therefore when nuclear RNA is similarly selected, an approximately 250 nucleotide long RNA fragment is obtained, as shown schematically in Figure 36. The greater complexity of the nuclear RNA RNAse T1 fingerprint shown in Figure 37A is consistent with this. Secondary analyses presented in Table III showed that all RNAse T1 spots from the mRNA leader were also present in the nuclear RNA fingerprint. Spots 8 of Figures 37A and 37B (analyzed by RNAse T2 in Figure 37G) and spot 9 of Figure 37C are the only major cap oligonucleotides. Both cap spots 8 gave AC, U, and G with RNAse A (not shown). The RNAse T2 type I cap from spot 8 released pC with nuclease P1 digestion and pC and pA with combined nuclease P1 and pyrophosphatase digestion. The type II cap released pC and 2 moles of pA with combined P1 plus pyrophosphatase digestion. The structure of spot 8 was deduced to be $^{7m}GpppA_mC_{(m)}ACUGp$. Minor spots a, b and c in Figure 37B gave RNAse T2 resistant products, shown in Figure 37G, indicating they too were caps. In addition, minor UU and UUU caps were found in darker exposures of the RNAse A fingerprint in Figure 37C.

The r strand of Ad5 between 98.3-100 (Steenbergh and Sussenbach, 1979) contains the hexanucleotide ACACTG at nucleotide 331, which corresponds to the region IV cap. Because Ad2 spots 9 and 10 of Figure 37B could not be aligned with Ad5, I suspected strain differences and determined the Ad2 Sma I K sequence. Restriction mapping and sequencing of Sma I K is described in detail in Appendix C. All mRNA oligonucleotides of Table III may be aligned between residues +1 and +62 of the DNA sequence shown in Figure 38. In agreement, an RNAse T1 15' mer predicted between -8 to +6 which spans the cap site is absent. The presence of RNAse T1 spot 9 and the absence of RNAse A product AAGGC (nucleotides 65-69) and RNAse T1 product UAAG (nucleotides 64-67) suggest that the consensus splice donor sequence GGUAAG encoded between nucleotides 62-67 contains the splice site, and that the leader ends at nucleotide 62, adjacent to GU.

All spots in Figure 37A which were exclusive to the nucleus and longer than 5 nucleotides could be aligned in Figure 38, downstream from

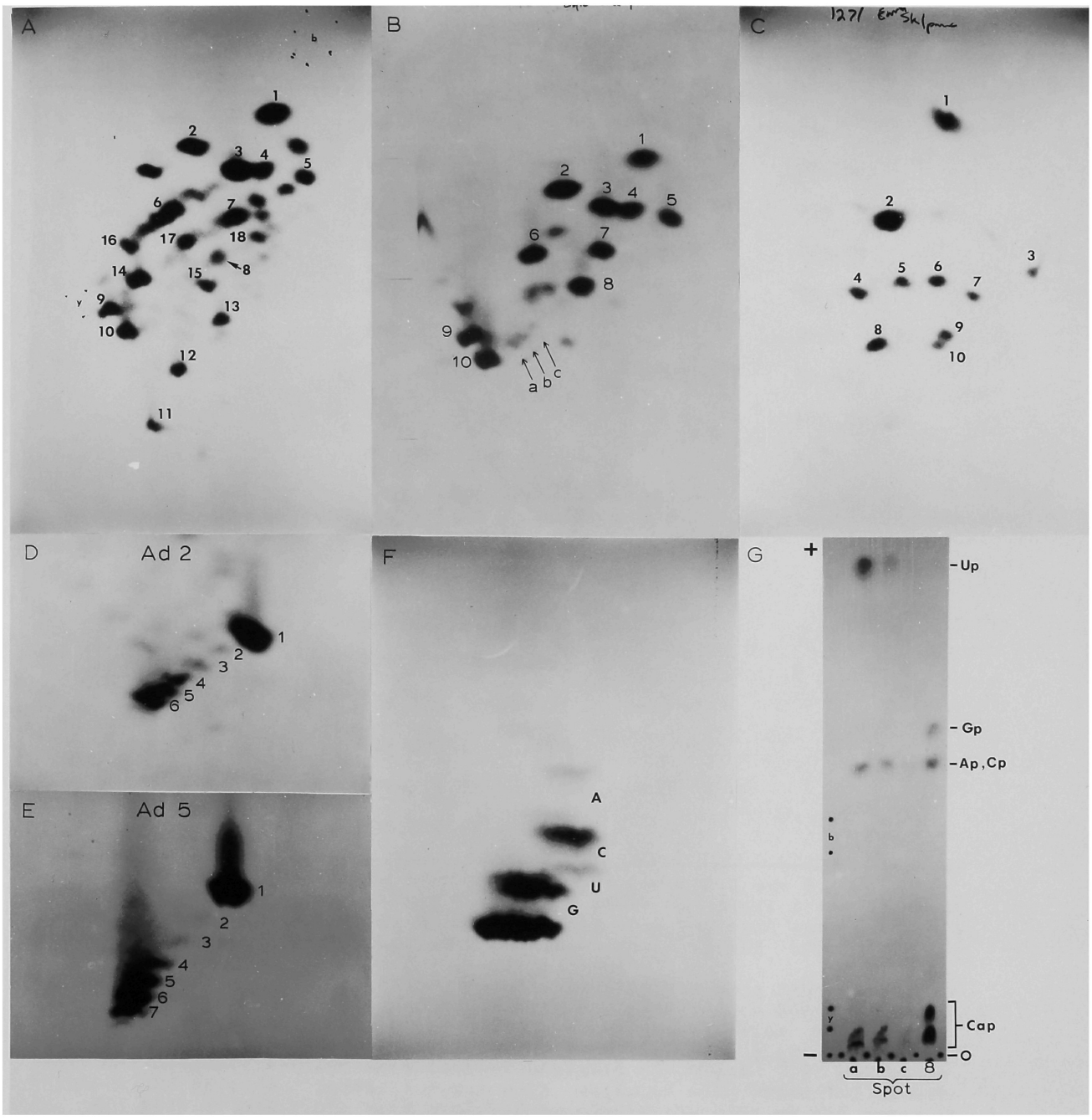

A
B
C
D
Ad 2
E
Ad 5
F
A
C
U
G
G
+
−
Up
Gp
Ap, Cp
Cap
O
Spot
a
b
c
8

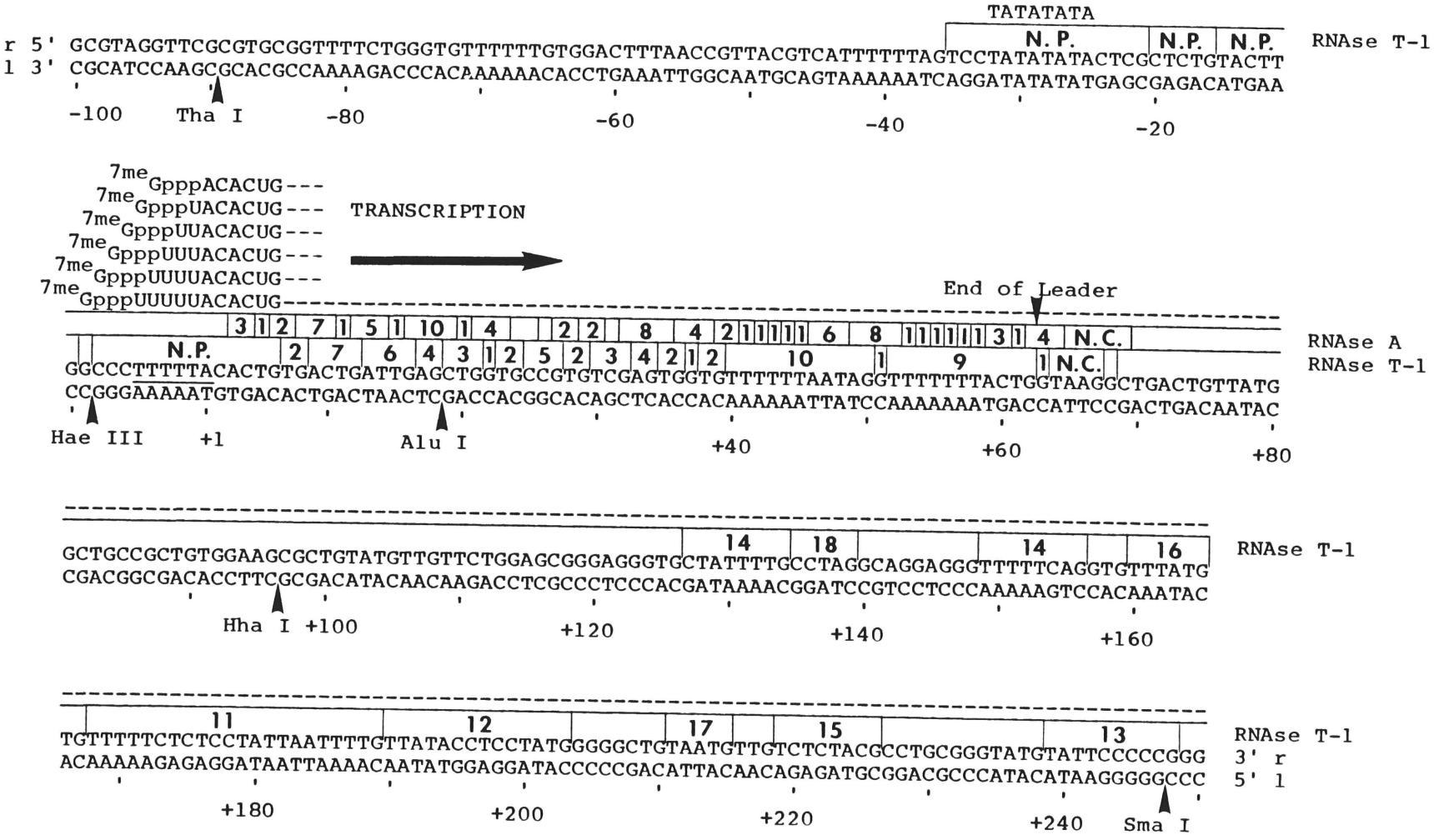
Table III. Analysis of RNAse T-1 and RNAse A Oligonucleotides of Region IV

RNAse T-1 Oligonucleotides			RNAse A Oligonucleotides	
Spot Number[a]	Nuclear or Cytoplasmic[b]	RNAse A Products[c]	Spot Number[d]	RNAse T-1 Products[c]
1	N,C	Gp	1	Up
2	N,C	Up,Gp	2	Gp,Up
3	N,C	Up,Cp,Gp	3	ACp
4	N,C	AGp	4	2Gp,Up
5	N,C	2Cp,Gp	5	Gp,AUp
6	N,C	AUp,Up,Gp	6	AAUp
7	N,C	ACp,Up,Gp	7	Gp,ACp
8	N,C	X_1,ACp,Up,Gp	8	AGp,Gp,Up
9	N,C	ACp, 9Up,Gp	9	X_2[e]
10	N,C	AAUp,6-8Up,AGp	10	AGp,Gp,Cp
11	N	AAUp,AUp,Up>Cp,Gp		
12	N	ACp,2AUp,Cp,Up,Gp		
13	N	AUp,Cp>Up,Gp		
14[f]	N	AUp,AGp,Cp,Up,Gp		
15	N	ACp,2Cp,3Up,Gp		
16	N	AUp,Up,Gp		
17	N	AAUp,Up,Gp		
18	N	AGp,2Cp,Up		

a. Numbers correspond to the spot numbers of Figure 37 A and B.
b. N indicates that the spot was found in a RNAse T-1 fingerprint of total nuclear RNA hybridized to Sma K. C indicates that the spot was found in a RNAse T-1 fingerprint of cytoplasmic poly(A)-containing RNA hybridized to Sma K.
c. For secondary analysis, RNAse T-1 oligonucleotides were redigested with RNAse A and RNAse A oligonucleotides were redigested with RNAse T-1. Products from both types of secondary digests were fractionated by electrophoresis on DEAE paper at pH 3.5. Products from the RNAse A redigestions of RNAse T-1 oligonucleotides were further analysed by RNAse T-2 digestion. Aliquots of RNAse A oligonucleotides were digested with RNAse T-2 directly. All base compositions were analysed by electrophoresis on 540 paper at pH 3.5. The presence of RNAse T-2 resistant components was determined by DEAE paper electrophoresis. Products designated X are RNAse A or RNAse T-1 oligonucleotides which yielded RNAse T-2 resistant components.
d. Numbers correspond to the spot numbers of Figure 37C.
e. X_2 when digested with T-2 releases Cp plus T-2 resistant cap material.
f. Spot 14 is a doublet of the two isomers CUAUUUUGp and UUUUUCAGp.

Figure 38. Alignment of the Heterogeneous Cap Structures and Internal Oligonucleotides of Ad2 Region IV mRNA and Nuclear RNA with the Ad2 DNA Sequence

Oligonucleotides from the fingerprints of Region IV nuclear RNA and mRNA, shown in Figures 37A and 37B-C respectively, were aligned with the DNA sequence of Ad2 between coordinates 99.3 and 98.3 m.u. (presented in Appendix C). Products from both RNAse A and RNAse T1 digests are shown and are identified by spot number. Alignment was on the basis of spot secondary analyses given in Table III. DNA residues are numbered relative to the major A cap site. RNAse T1 oligonucleotides predicted by the DNA to be either traversing or preceding the cap site but not found in either nuclear or messenger RNA are marked N.P. RNAse A and RNAse T1 oligonucleotides absent from mRNA and used to define a donor splice site in mRNA are marked N.C. The five minor U terminated caps and the major A terminated cap found in Figure 37A, B, and D are shown. An upstream TATATATA box is marked. Note that the orientation of the DNA is the reverse of Figure 1 so that transcription is from left to right. The Sma I site at nucleotide 247 is the boundary of Sma I K of Figure 1.


TATATATA
r 5' GCGTAGGTTCGCGTGCGGTTTTCTGGGTGTTTTTTGTGGACTTTAACCGTTACGTCATTTTTTAGTCCTATATATACTCGCTCTGTACTT
1 3' CGCATCCAAGCGCACGCCAAAAGACCCACAAAAAACACCTGAAATTGGCAATGCAGTAAAAAATCAGGATATATATGAGCGAGACATGAA
N.P.
N.P.
N.P.
RNAse T-1
-100
Tha I
-80
-60
-40
-20
7meGpppACACUG---
7meGpppUACACUG---
7meGpppUUACACUG---
7meGpppUUUACACUG---
7meGpppUUUUACACUG---
7meGpppUUUUUACACUG----------
TRANSCRIPTION
End of Leader
3 1 2 7 1 5 1 10 1 4 2 2 8 4 2 1 1 1 1 1 6 8 1 1 1 1 1 1 3 1 4 N.C.
RNAse A
N.P. 2 7 6 4 3 1 2 5 2 3 4 2 1 2 10 1 9 1 N.C.
RNAse T-1
GGCCCTTTTTACACTGTGACTGATTGAGCTGGTGCCGTGTCGAGTGGTGTTTTTTAATAGGTTTTTTTACTGGTAAGGCTGACTGTTATG
CCGGGAAAAATGTGACACTGACTAACTCGACCACGGCACAGCTCACCACAAAAAATTATCCAAAAAAATGACCATTCCGACTGACAATAC
Hae III
+1
Alu I
+40
+60
+80
14 18 14 16
RNAse T-1
GCTGCCGCTGTGGAAGCGCTGTATGTTGTTCTGGAGCGGGAGGGTGCTATTTTGCCTAGGCAGGAGGGTTTTTCAGGTGTTTATG
CGACGGCGACACCTTCGCGACATACAACAAGACCTCGCCCTCCCACGATAAAACGGATCCGTCCTCCCAAAAAGTCCACAAATAC
Hha I
+100
+120
+140
+160
11 12 17 15 13
RNAse T-1
TGTTTTTCTCTCCTATTAATTTTGTTATACCTCCTATGGGGGCTGTAATGTTGTCTCTACGCCTGCGGGTATGTATTCCCCCGGG
3' r
ACAAAAAGAGAGGATAATTAAAACAATATGGAGGATACCCCCGACATTACAACAGAGATGCGGACGCCCATACATAAGGGGGCCC
5' 1
+180
+200
+220
+240
Sma I

the leader, placing them within the first region IV intron.

When the EIV cap structures were analyzed by in vitro labeling, the RNAse T1 fingerprint in Figure 37D gave one major spot and a family of five minor related spots. These spots were labeled only if the mRNA was decapped prior to phosphatasing and kinasing (data not shown) and are therefore all derived from capped 5' termini. Nuclease P1 released a pA 5' end from the major spot 1 and pU from spots 2-6 (Table I). From additional secondary analyses given in Figure 11 lanes 15-16 and in Figure 37F, the major spot 1 was $pA_mC_{(m)}ACUGp$, identified with the major in vivo cap, spot 8 of Figure 37B. The 5' sequence of spot 2 was $pU_mA_{(m)}C-$, spot 3 was $pU_mU_{(m)}AC-$, and spots 4-6 were $pU_mU_{(m)}U-$, all deduced from Table I and Figure 11 lanes 17-26. It can be concluded that spots 2-6 are from a staggered series of capped termini with 5' uridylates encoded by the run of T residues (nucleotides -1 to -5) which precedes the major A terminus. This explains the origin of the three U termini previously mapped to region IV by Hashimoto and Green (1980).

Ad5 region IV also has a series of T residues which precedes its major A cap site (Steenbergh and Sussenbach, 1979). However, relative to Ad2, one T has been inserted creating a run of six T's in Ad5 versus five T's in Ad2. This displaces the cap site one residue further from the TATA box:

TATATATACTCGCTCTGTACTTGGCCCTTTTTACACTG :Ad2
TATATATACTCGCTCTGCACTTGGCCCTTTTTTACACTG :Ad5

Despite this displacement, in Figure 37E the A terminus remains the major Ad5 terminus. Also, all six T positions (one more than Ad2) are represented in the minor product trail. With both Ad2 and Ad5, the abundance of a U terminus increases as its distance from the A cap site increases. Spots with the mobility of the U capped products are also seen with in vivo label (Figure 37B). The lower three (spots a-c) migrating like the 3 UUU... termini, spots 4, 5 and 6 of Figure 37D, were shown in Figure 37G to contain RNAse T2 resistant material. The very low mobility of these RNAse T2 cap cores is consistent with UU and UUU cap cores.

In Vitro Transcription of Ad5 EIV

I wanted to analyze pulse labeled nascent RNA from region IV to determine whether the U caps were present in the same abundance in nascent RNA as mRNA and also whether there was any transcription upstream of the cap site. I first tried to pulse label RNA *in vitro* in isolated nuclei, but found it possible only to incorporate sufficient label to see the major spots in a long exposure. I was therefore fortunate to enter into a collaboration with Andy Fire of MIT to analyze EIV transcription in a soluble whole-cell extract (Manley et al., 1980). In this collaboration, Andy performed all of the *in vitro* transcription reactions, and I analyzed the structures of the transcripts.

To obtain a defined transcript for analysis, cloned DNA containing the region IV promoter was cleaved with a restriction enzyme so that transcription from the promoter would produce a run-off product of known length. This run-off product could then be isolated by acrylamide gel electrophoresis. The EIV DNA used was an Ad5 Eco RI B (84-100 m.u.) clone cleaved with Sma I. Transcripts starting at the EIV cap site and transcribed from the *l* strand will be 250 nucleotides long. A control, which provided an internal marker, was the Ad2 major late promoter DNA cut with Sac II to yield a 130 nucleotide run-off product (Figure 20). Figure 39 shows an acrylamide gel of the *in vitro* transcripts from the major late promoter DNA (lane M) and from EIV labeled with α-^{32}P-UTP, α-^{32}P-ATP, and α-^{32}P-GTP (lanes U, A and G, respectively). Each of the 250 nucleotide run-off products was electroeluted, EtOH precipitated, and RNAse T1 fingerprinted. The fingerprints of the α-^{32}P-UTP, α-^{32}P-ATP, and α-^{32}P-GTP labeled transcripts are shown in Figures 40 B, C and D, respectively. Figure 40A shows for comparison a fingerprint of a mixture of *in vivo* ^{32}P labeled Ad2 nuclear and messenger RNA hybridized to Ad2 Sma I K. Notice that the fingerprint of the UTP labeled transcripts closely resembles that of the Ad2 *in vivo* labeled RNA while the fingerprints of the transcripts labeled with ATP or GTP have a higher background of other spots. This is not surprising, since the gel patterns (Figure 39) show that the UTP labeled run-off product is the most intensely labeled relative to the background in the lane. The numbered

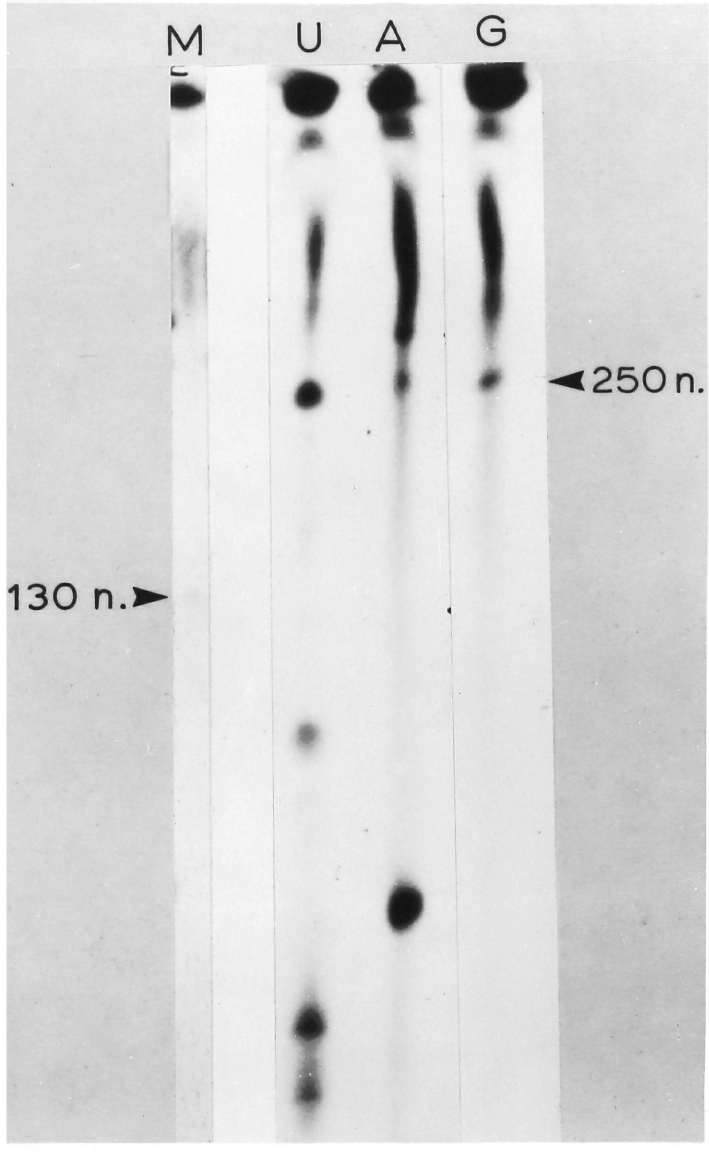


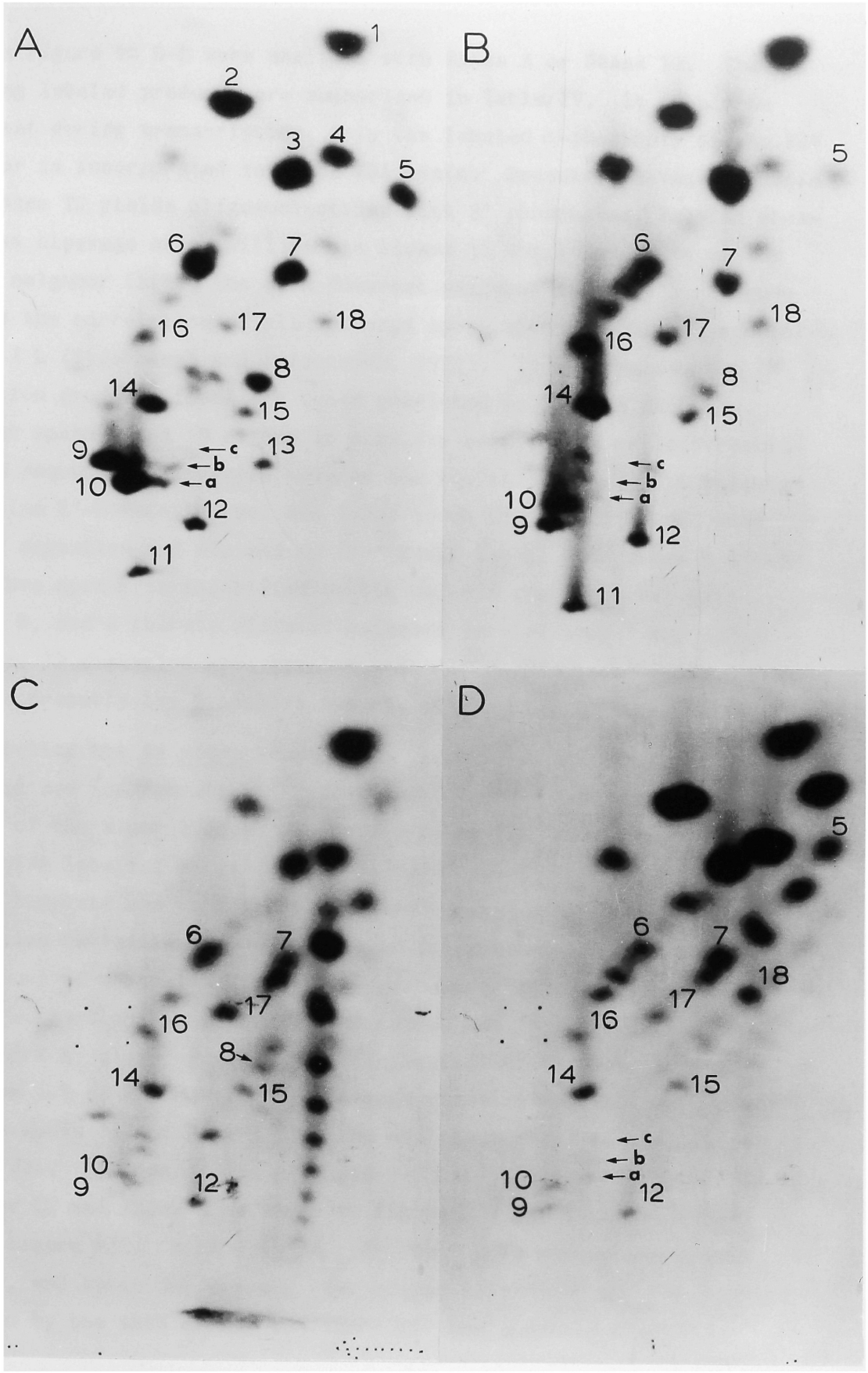
Figure 39. Gel Electrophoresis Separation of EIV *in Vitro* Transcripts

Cloned Ad5 Eco RI B (84-100 m.u.) DNA cut with Sma I was used as template in the cell-free transcription system of Manley et al. (1980). The *in vitro* transcription products were separated by electrophoresis on a 7 M urea-8% acrylamide gel. Run-off products are expected to be 250 nucleotides long. Lanes marked U, A and G contain products from reactions labeled with α-^{32}P-UTP, ATP and GTP, respectively. The 250 nucleotide run-off products are indicated. The lane marked M contains *in vitro* transcripts from major late promoter DNA cut with Sac II to give a 130 nucleotide run-off product (see Figure 20).

Figure 40. Fingerprinting of EIV *in Vitro* Run-off Products

(A) Early Ad2 mRNA labeled with ^{32}P *in vivo* was hybridized to Ad2 Sma I K, eluted and RNAse T1 fingerprinted. The mRNA either had a large nuclear contamination or a high proportion of the bottom mRNA species shown in Figure 36. Spot numbers correspond to those in Figures 37 A and B and 38 and Table III. This fingerprint is shown because its spot mobilities most closely resemble those of the fingerprints in panels B-D.

(B),(C),(D) The 250 nucleotide runoff products from the gel shown in Figure 39 were electroeluted, EtOH precipitated, digested with RNAse T1 and fingerprinted. The run-off products labeled with α-^{32}P-UTP, α-^{32}P-ATP and α-^{32}P-GTP are shown in panels (B), (C) and (D), respectively. The results of analysis of numbered (and lettered) oligonucleotides with RNAse A or RNAse T2 are summarized in Table IV. Oligonucleotides are aligned with the Ad5 sequence in Figure 41. Spot 8 is the major A cap and spots a, b and c are three of the minor U caps.


A
1
2
3
4
5
6
7
16
17
18
8
14
15
9
c
13
b
10
a
12
11
B
5
6
7
18
16
17
8
14
15
c
b
10
a
9
12
11
C
6
7
17
16
8
14
15
10
9
12
D
5
6
7
18
17
16
14
15
c
b
10
a
12
9

spots in Figure 40 B-D were analyzed with RNAse A or RNAse T2. The resulting labeled products are summarized in Table IV. It should be noted that during transcription, only the labeled α-phosphate of the NTP precursor is incorporated into the RNA chain. Because cleavage by RNAse A and RNAse T2 yields oligonucleotides with 3' phosphates, labeled phosphates at cleavage sites will remain linked to the 3' position of the nearest neighbor (hence the term "nearest neighbor analysis"). Figure 41 shows the correlation of all numbered spots with the DNA sequence of Ad5 Sma I L (Steenbergh and Sussenbach, 1979). In most cases, the redigestion products found are those predicted by the DNA sequence. Note that spots 9 and 10 differ in mobility between Ad2 and Ad5 resulting from sequence differences between the two strains. Spot 8 released types 0 (no 2'-O-methylation) and I cap cores (α-^{32}P-ATP label) with RNAse T2 digestion and ACp and Gp (α-^{32}P-UTP label) with RNAse A digestion. Thus spot 8 is identified as the major A cap oligonucleotide. Spots a, b, and c (barely visible) released labeled type I cap cores with RNAse T2 digestion when labeled with α-^{32}P-UTP or α-^{32}P-GTP. These spots are probably the 3 longest UUU cap oligonucleotides.

Labeling the <u>in vitro</u> transcripts with α-^{32}P-nucleoside triphosphates did not incorporate sufficient radioactivity for adequate analysis of the minor U caps. Therefore I decided to try the decapping and kinasing labeling procedure on unlabeled <u>in vitro</u> transcripts. This time the template was an Ad5 Eco RI B clone recut with Bgl II. A parallel reaction containing an α-^{32}P-labeled triphosphate indicated that the total amount of unlabeled transcripts was approximately 10^{-3} picomoles. The entire reaction mixture, unfractionated, was 5' labeled, hybridized to Ad2 Sma I K, eluted and RNAse T1 fingerprinted, as shown in Figure 42. Spots 1-7 of Figure 42 have the same relative mobilities and intensities as spots 1-7 of Figure 37E, the analogous RNAse T1 fingerprint of Ad5 EIV mRNA. The identities of spots 1-7 were confirmed by analysis with RNAse T2 and RNAse A as shown in Figure 43A and nuclease P1 as shown in Figure 43B. Spot 1 is $pA_{(m)}C$-, spot 2 is contaminated with spots 3-7, and spots 3-7 are $pU_{(m)}U$-. Other very minor $pU_{(m)}U$- termini (indicated by the thin arrows in figure 42) and possibly a minor

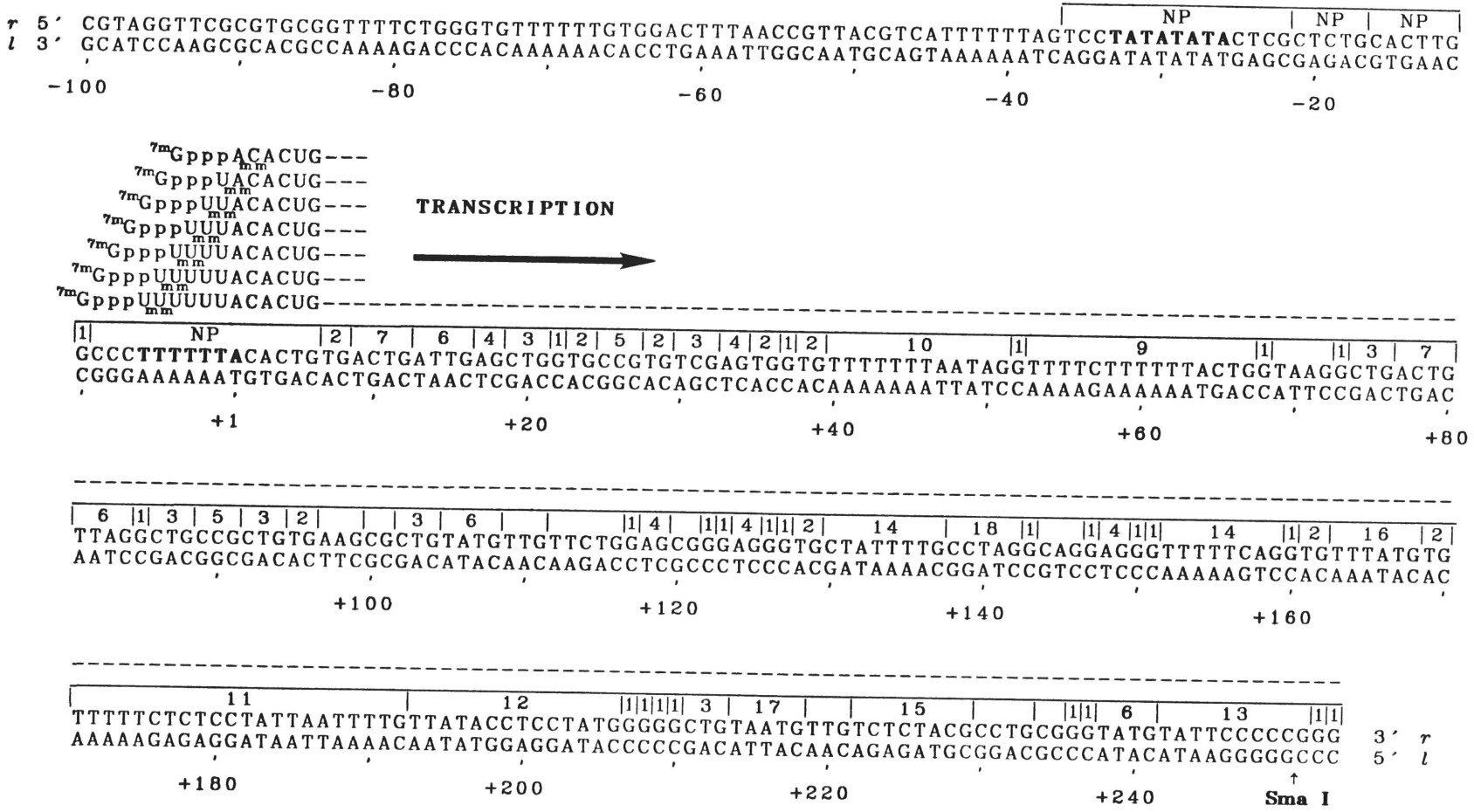
Table IV. Analysis of RNAse T1 Oligonucleotides from Ad5 EIV RNA Transcribed In Vitro

	RNAse A Products[b]		
Spot Number[a]	α-^{32}P-UTP Label[c]	α-^{32}P-ATP Label[c]	α-^{32}P-GTP Label[c]
1	NA	NA	NA
2	NA	NA	NA
3	NA	NA	NA
4	NA	NA	NA
5	Gp	NA	Cp
6	AUp,Gp	Up,Gp	AUp,Up
7	ACp,Gp	Gp	[AAGp,AGp,Gp,(AUp,Cp,Up)][d]
8	ACp,Gp	[Type 0 and I cap cores][e]	NP
9	ACp,Cp,Up	Up	Up,Gp
10	AAUp,Up,(ACp,Cp)[f]	AAUp,Up	AGp
11	AAUp,AUp,Cp,Up,(Gp)	NA	NA
12	AUp,Cp,(Up)	AUp,Up	AUp,Gp
13	NP	NP	NP
14	AUp,Up,Cp	AAUp[d],Cp,Up	Up,AGp
15	Cp	Up	ACp
16	AUp,Gp,Up	Up	AUp,(Up)[d]
17	AAUp,Gp	AAUp,Up	AAUp
18	Cp	NA	AGp
a	[Cp,Up,cap I core][e]	NP	[cap I core,Up][e]
b	[Cp,Up,cap I core][e]	NP	[cap I core,Up][e]
c	[Cp,Up,cap I core][e]	NP	[cap I core,Up][e]

a. Numbers correspond to the spot numbers of Figure 40.
b. For secondary analysis, RNAse T-1 oligonucleotides were redigested with RNAse A. Products from the secondary digests were fractionated by electrophoresis on DEAE paper at pH 3.5.
c. These are the products resulting from RNAse A redigestion of the RNAse T1 oligonucleotides from the fingerprints shown in Figure 40 B, C and D, of RNA labeled in vitro with α-^{32}P-UTP, α-^{32}P-ATP or α-^{32}P-GTP, respectively. NA indicates that the corresponding spot was not analysed; NP indicates that the spot was not present in the fingerprint. Products in parentheses were barely detectable.
d. This product(s) is not predicted by the Ad5 sequence in Figure 41.
e. These products are the result of redigestion with RNAse T2.
f. Spot 10 is contaminated with spot 9.

Figure 41. Alignment of Oligonucleotides from the Ad5 EIV *in Vitro* Transcripts with the Ad5 DNA Sequence

The Ad5 DNA sequence presented here is from Steenbergh and Sussenbach (1979). RNAse T1 oligonucleotides from the fingerprints in Figure 40 B-D have been aligned with the DNA. Alignment is on the basis of the secondary analyses presented in Table IV. The major A cap and six U caps identified in Figures 42-43 are indicated by sequence rather than by number. Oligonucleotides traversing or preceding the caps but not found in the run-off products are marked NP. Spot 13, found in nuclear RNA, is absent from the *in vitro* run-off products since the template is truncated within the spot. Transcription is from left to right since the orientation of the DNA is the opposite of that of the maps in Figures 1 and 36 but is the same as that of Figure 38.


r 5′ CGTAGGTTCGCGTGCGGTTTTCTGGGTGTTTTTTGTGGACTTTAACCGTTACGTCATTTTTTAGTCCTATATATACTCGCTCTGCACTTG
l 3′ GCATCCAAGCGCACGCCAAAAGACCCACAAAAAACACCTGAAATTGGCAATGCAGTAAAAAATCAGGATATATATGAGCGAGACGTGAAC
NP
NP
NP
-100
-80
-60
-40
-20
7mGpppACACUG---
7mGpppUACACUG---
7mGpppUUACACUG---
7mGpppUUUACACUG---
7mGpppUUUUACACUG---
7mGpppUUUUUACACUG---
7mGpppUUUUUUACACUG----------
TRANSCRIPTION
GCCCTTTTTTACACTGTGACTGATTGAGCTGGTGCCGTGTCGAGTGGTGTTTTTTTAATAGGTTTTCTTTTTTACTGGTAAGGCTGACTG
CGGGAAAAAATGTGACACTGACTAACTCGACCACGGCACAGCTCACCACAAAAAAATTATCCAAAAGAAAAAATGACCATTCCGACTGAC
+1
+20
+40
+60
+80
TTAGGCTGCCGCTGTGAAGCGCTGTATGTTGTTCTGGAGCGGGAGGGTGCTATTTTGCCTAGGCAGGAGGGTTTTTCAGGTGTTTATGTG
AATCCGACGGCGACACTTCGCGACATACAACAAGACCTCGCCCTCCCACGATAAAACGGATCCGTCCTCCCAAAAAGTCCACAAATACAC
+100
+120
+140
+160
TTTTTCTCTCCTATTAATTTTGTTATACCTCCTATGGGGGCTGTAATGTTGTCTCTACGCCTGCGGGTATGTATTCCCCCGGG 3′ r
AAAAAGAGAGGATAATTAAAACAATATGGAGGATACCCCCGACATTACAACAGAGATGCGGACGCCCATACATAAGGGGGCCC 5′ l
+180
+200
+220
+240
Sma I

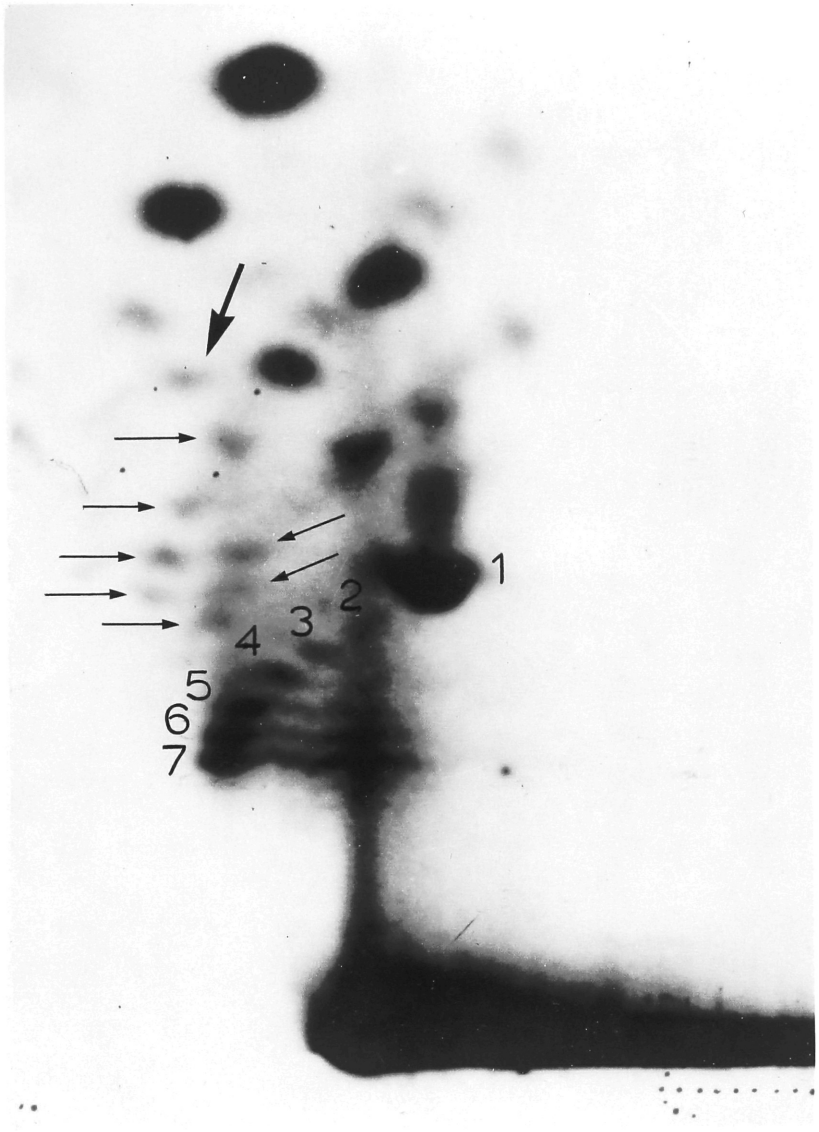


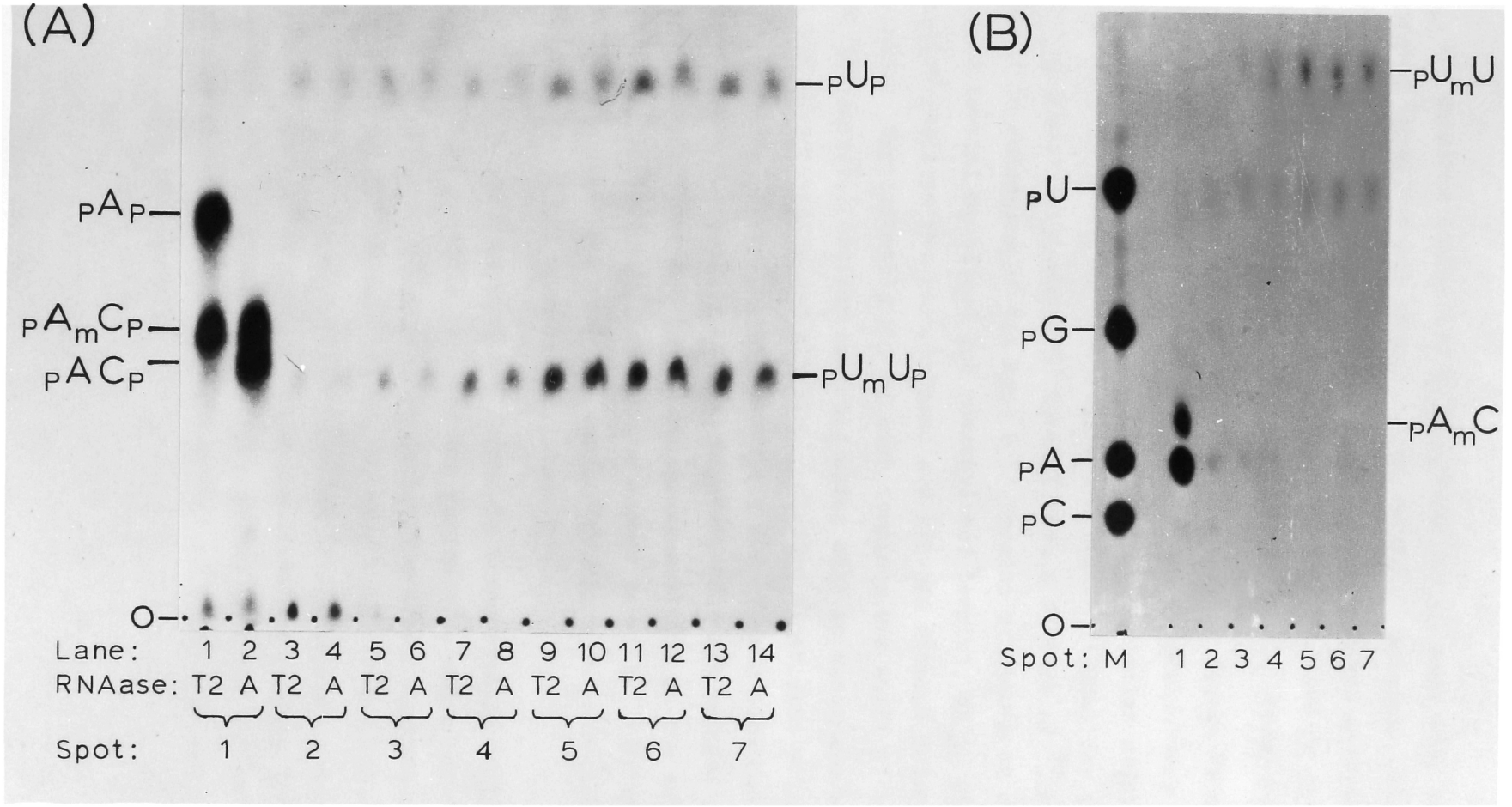
Figure 42. The 5' Termini of the Ad5 EIV *in Vitro* Transcripts

Total unlabeled and unfractionated *in vitro* transcription products synthesized using cloned Ad5 Eco RI B DNA cut with Bgl II as template were chemically decapped, phosphatased, kinased, hybridized to Ad2 Sma I K, eluted and T1 fingerprinted. Spot 1 and spots 2-7 (analyzed in Figure 43) are the major A terminus and the minor U termini, respectively, found for Ad5 mRNA (Figure 37E). The seven spots indicated by thin arrows are other pU_mU- termini (analysis not presented). The spot indicated by the thick arrow is probably pG_mUG (analysis not shown). The radioactivity remaining at the origin of the second dimension is 5' labeled DNA.

Figure 43. RNAse T2 and A and Nuclease P1 Analysis of *in Vitro* Transcript 5' Termini

(A) Spots 1-7 from Figure 42 were eluted, digested with RNAse T2 or A and electrophoresed at pH 3.5 on DEAE paper. Compare the pattern with Figure 11 lanes 15-26. Only some of the molecules have the type I methylation and no molecules are type II. Radioactivity remaining at the origin of lanes 1-4 is 5' labeled DNA.

(B) Spots 1-7 from Figure 42 were digested with nuclease P1 and electrophoresed at pH 3.5 on 540 paper, shown in lanes 1-7, respectively. Lane M has ^{32}P labeled 5' mononucleotide markers.


(A)
pUp
pAp
pAmCp
pACp
pUmUp
O
Lane: 1 2 3 4 5 6 7 8 9 10 11 12 13 14
RNAase: T2 A T2 A T2 A T2 A T2 A T2 A T2 A
Spot: 1 2 3 4 5 6 7
(B)
pUmU
pU
pG
pA
pC
pAmC
O
Spot: M 1 2 3 4 5 6 7

$pG_{(m)}UG-$ terminus (indicated by the thick arrow) were also identified in Figure 42 (analyses not shown). None of the 5' terminal oligonucleotides were seen when a DNA minus control reaction was analyzed (data not shown), indicating that the 5' termini are synthesized de novo by a template dependent process and were not present in the whole-cell extract.

None of the 5' terminal oligonucleotides in Figure 42 had a type II methylation pattern and at most 50% of the molecules were type I, as shown by the RNAse T2 analyses in Figure 43. A similar result was obtained by Yamamoto, de Crombrugghe and Pastan (1980b) for the RSV LTR in vitro transcripts, whose 5' termini were a mixture of $^{7m}GpppG_m-$ and GpppG-. To determine if the type 0 5' termini of Figure 42 represented uncapped termini or capped but unmethylated termini, an in vitro transcription reaction was phosphatased and kinased without prior chemical decapping. Approximately 20% of each terminus was still 5' labeled (data not shown). Similar controls using mRNA as substrate indicated that the phosphatasing and kinasing procedure alone does not label capped termini. Therefore approximately 20% of each terminus is uncapped. The possibility exists, however, that all termini are capped during transcription, but that a decapping activity in the extract is responsible for decapping some of the capped termini. Furuichi et al. (1976) have pointed out that the capping reaction is potentially reversible until the G has been 7-methylated.

If the U and A termini are formed by transcription initiation, then it might be possible to alter the ratios of U and A termini by changing the UTP and ATP concentrations. Therefore, instead of using the usual 50 μM NTP concentrations, we used 1 μM UTP and 500 μM ATP. The K_m for UTP in the elongation reaction is approximately 1 μM (A. Fire, pers. comm.), so transcription should not be greatly inhibited. As expected, the 5' labeling procedure showed that the U termini were essentially absent and the absolute amount of A terminus was probably increased (data not shown).

DISCUSSION

The sequences of the 5' termini of all Ad2 mRNAs have been determined, in some cases through the first leader. In addition, the sequences of the DNA encoding the mRNA 5' termini and first leaders have been determined. Comparison of these sequences gives some insight into the sequences important for transcription initiation and splicing.

Microheterogeneity and Transcription Initiation

Ad2 mRNA 5' Termini Figure 44 summarizes the 5' termini found for all major Ad2 mRNAs. [By convention, the DNA strand which is shown is the non-coding strand and has the same sense as the mRNA. The residues which correspond to mRNA 5' termini are referred to as the cap sites.] Seven out of nine transcription units encode mRNAs with multiple 5' termini. The major late mRNAs may also have additional termini, but at a very low level (<1%). It has been shown for the EIa, EIb, PIX, major late, EIII and EIV transcription units that the same 5' termini are present in mRNA and RNA synthesized in cell-free in vitro transcription systems (Lee and Roeder, 1981; A. Fire, C. Baker, E. Ziff and P. Sharp, manuscript submitted).

The evidence that eukaryotic mRNA 5' termini are produced by an initiation event is given in the Introduction. The best evidence of this for adenovirus mRNAs is the work of Hagenbuchle and Schibler (1981) in which they showed that the major late cap could be labeled in the in vitro transcription system with β-^{32}P-ATP. I have investigated this question for the microheterogeneous EIV mRNA 5' termini. There is no evidence for transcription upstream of the cap site in either nuclei labeled in vivo or RNA transcribed in vitro, as shown by the absence of the oligonucleotides marked NP in Figures 38 and 41 in the fingerprints shown in Figures 37A and 40B, respectively. Therefore if EIV mRNA is cleaved from a precursor, cleavage must be very rapid and must occur during transcription of the RNA molecule. Indirect evidence that the U termini of Ad5 EIV are formed by transcription initiation comes from the observation that changing the proportions of ATP and UTP in the in vitro

"TATA" BOX	CAPS	GENE (COORDINATE)
GTG**TATTTATA**CCCGGTGAGTTCCTCAAGAGGCC**A**CTCTTGAGTG		Ad2 EIa (1.4)
GGG**TATATAAT**GCGCCGTGGGCTAATCTTGGTT**A**C**A**TCTGACCTC		Ad2 EIb (4.7)
GAA**TATATAA**GGTGGGGGTCTCATGTAGTTTT**GT**A**T**CTGTTTTGC		Ad2 Protein IX (9.8)
TCCTTCGTGCTGGCCTGGACGCGAGCCTTCGTC**T**C**A**GAGTGGTCC		Ad2 IVa2 (15.9)
GGC**TATAAAA**GGGGGTGGGGGCGCGTTCGTCCTC**A**CTCTCTTCCG		Ad2 Major Late (16.4)
AGG**TACAAATTT**GCGAAGGTAAGCCGACGTCC**A**C**A**GCCCCGGAGT		Ad2 EIIa – late (72)
TAGTCCTTAAGAGTCAGCGCGCAGTATTTGCTGA**AG**AGAGCCTCC		Ad2 EIIa (75)
GGG**TATAA**CTCACCTGAAAATCAGAGGGCGAG**GT**A**T**TCAGCTCAA		Ad2 EIII (76.6)
TCC**TATATATA**CTCGCTCTGTACTTGGCCC**TTTTTA**CACTGTGAC		Ad2 EIV (99.1)
TCC**TATATATA**CTCGCTCTGCACTTGGCCC**TTTTTTA**CACTGTGA		Ad5 EIV (99.1)

Figure 44. DNA Sequences Preceding Adenovirus mRNA Cap Sites

The DNA sequences shown are the non-coding strand and therefore have the same sense as the mRNA. Adenovirus genes are designated by the map positions of their 5' termini as well as by the region or protein product. The TATA box homology as well as the residues corresponding to mRNA 5' termini are shown in bold type. The TCCTT homology is underlined. The DNA sequences for the Ad2 EIa, EIb, IVa2, EIII and EIV cap regions and the 5' termini of all but the Ad2 major late mRNAs are from this thesis. The Ad2 major late cap region and 5' terminus was previously reported by Ziff and Evans (1978). The DNA sequences for the Ad2 PIX, Ad2 EIIa and EIIa-late and Ad5 EIV cap regions are from Alestrom et al. (1980), Galibert et al. (1979) and Steenbergh and Sussenbach (1979), respectively.

transcription reaction produce changes in the relative amounts of the U and A termini. Transcription in the presence of equal concentrations of ATP and UTP (50 μM each) yielded a ratio of approximately 4:3 for the amount of the A terminus to the total amount of the 6 U termini. Transcription in the presence of 500 μM ATP and 1 μM UTP, however, yielded no detectable U termini and a possible increase in the absolute amount of the A terminus. The K_m for UTP in the elongation reaction is approximately 1 μM (A. Fire, pers. comm.). The K_m for NTPs at the initiation site of the E. coli RNA polymerase is approximately 10 fold higher than the K_m at the elongation site (Krakow, Rhodes and Jovin, 1976). If this is also true for eukaryotic RNA polymerase II, then the results of this experiment suggest that the U termini are formed by transcription initiation. Conversely, if we assume that the U termini are produced by transcription initiation, then we can deduce that the K_m for initiation is higher than that for elongation.

The Initiation Region The heterogeneous 5' termini of individual mRNAs shown in Figure 44 are encoded within a narrow region (called the cap site region or initiation region) of 7 nucleotides or less. The presence of multiple termini encoded within a narrow region has been called microheterogeneity. Multiple termini encoded over a large region, observed for SV40 and polyoma late mRNAs (Figure 45 and references therein), is called macroheterogeneity. The location of the initiation region with respect to an upstream sequence element, the TATA box, is discussed below.

Purine Preference When the distribution of capped residues is compared from site to site, a clear preference for purine termini is evident. Except for regions IV and IVa2, all termini are purines, A or G. Furthermore, with cap site regions which contain both purines and pyrimidines, purines are used with much higher frequency. For example, in Figure 33, separation of the early region III termini by fingerprinting gave two major capped purine termini with 5' G and A, respectively. However, the T residue which lies between the purines (Figure 44) does not yield a corresponding U cap as shown in Figure 33. This is also true of early region Ib, the PIX mRNA, and the late form of region IIa

mRNA, shown in Figures 12E, 12F, and 30. In each of these examples, the observed purine termini are separated by a pyrimidine residue which is not represented as a capped terminus.

Other viral and cellular genes, shown in Figure 45, also exhibit a strong preference for purines at mRNA 5' termini. Chicken ovalbumin mRNA has a microheterogeneous 5' terminus similar to that of Ad2 PIX and EIII mRNA (Malek et al., 1981). Kamen et al. (1980) found 2 major purine termini for polyoma early mRNAs. Flavell et al. (1979) found at least seven 5' termini, all purines, for polyoma late mRNAs. Three of these have been mapped to a 4 nucleotide region shown in Figure 46 (Flavell et al., 1980). Note that in all three examples potential pyrimidine termini were either utilized at very low levels, or not at all. If RNA initiation takes place at the cap site, within a short denatured DNA segment (as is the case for E. coli RNA polymerase; see Siebenlist et al., 1980) the observed 5' heterogeneity could reflect a scanning of the denatured cap region by the polymerase during a search for the thermodynamically most favorable start site.

If all of the termini identified in this study are considered, 14 are A caps, 3 are G's, and 6 are low molarity U's. Only very low levels of C caps were found. These data therefore suggest that the order of nucleotide preference for forming the capped residue in Ad2 is: A>G>>U>C. This preference parallels the frequency of 5' terminal residues in transcripts initiated by E. coli RNA polymerase where it reflects the binding affinities for the initiating nucleotide at the initiation site of the polymerase molecule (see Rosenberg and Court, 1979).

In Ad2, the width of the cap region is 1-3 nucleotides, except for region IV for which it is up to 7 nucleotides wide (largely as a consequence of a broad region which yields U caps). The narrow width of the proposed scanning suggests that the enzymes in the putative initiation complex (i.e. the polymerase) are precisely positioned (discussed further below). Although in this discussion I suggest that the polymerase itself recognizes the DNA in the vicinity of the cap site, I note

"TATA BOX"	CAP	GENE
GGG**CATAAAA**GGCAGAGCAGGGCAGCTGCTGCTT**A**CACTTGCTTT		Rabbit β-Globin
GAG**CATATAA**GGTGAGGTAGGATCAGTTGCTCCTC**A**CATTTGCTT		Mouse β-Globin Major
AAG**CATAAA**CCCTGGCGCGCTCGCGGCCCGGC**A**CTCTTCTGGTCC		Human α2-Globin
CAG**TATAAAAA**GGTTCAACTTTTTCAAATCAGC**A**TCAGTTCGGTT		Silk Fibroin
GGC**TATATATT**CCCCAGGGCTCAGCCAGTGTCT**G**T**A**C**A**TACAGCT		Chicken Ovalbumin
CTC**TATAAAA**GGGGAAGAAAGAGGCTCCGCAGCC**A**TCACAGACCC		Chicken Conalbumin
TTG**TATTTAA**GTGCCTAGCTCGATACAATAAAC**G**CCATTTGACCA		RSV LTR
AGG**TATAAATA**GCCACCAAAACGCTGCTGGGC**A**TCC**A**TTCAAGTC		*S. purpuratus* H2A
GTG**TAACAATA**CTCGGTGCAATCCGGTTGAGGC**A**TC**A**TTCGCTTA		*S. purpuratus* H4
TTT**TATTTAT**GCAGAGGCCGAGGCCGCCTCG**G**CCTCT**G**AGCT**A**TT		SV40 Early
TGA**TATAATTA**GCCCCAACCGCCTCTTCCC**G**CCTC**A**TTTCAGCCT		Polyoma Early
GAAGGTACCTAACCAAGTTCCTCTTTCAGAGGTT**A**TTTCAGGCCA		SV40 Late
GAGCCTGGGGACTTTCCACACCCTAACTGACAC**A**C**A**TTCCACAGC		SV40 Late
CGTCAGTTAGTCCACTTCCTGCTTAACTGACTT**GA**C**A**TTTTCTAT		Polyoma Late

Figure 45. DNA Sequences Preceding Other Viral and Cellular mRNA Cap Sites

The DNA sequences shown are the non-coding strands and therefore have the same sense as the mRNA. The TATA box homology and residues corresponding to mRNA 5' termini are shown in bold type. The sources for the sequences and cap sites are: rabbit β-globin sequence (Van Ooyen et al., 1979) and cap site (Lockard and RajBhandary, 1976; Baralle, 1977b); mouse β-globin major sequence (Konkel, Tilghman and Leder, 1978) and cap site (Baralle and Brownlee, 1978; also see Appendix D); human α2-globin sequence (Liebhaber, Goossens and Kan, 1980) and cap site (Baralle, 1977a; Chang et al., 1977); silk fibroin (Tsujimoto and Suzuki, 1979a and b); chicken ovalbumin sequence (Gannon et al., 1979; Benoist et al., 1980) and cap site (McReynolds et al., 1978; Malek et al., 1981); chicken conalbumin (Cochet et al., 1979); rous sarcoma virus long terminal repeat (Yamamoto et al., 1980a and b); S. purpuratus H2A and H4 (Sures, Levy and Kedes, 1980); SV40 sequence (Fiers et al., 1978; Reddy et al., 1978); SV40 early cap sites (Reddy et al., 1979; Thompson et al, 1979; Haegeman and Fiers, 1980); SV40 late cap sites (Ghosh et al., 1978a and b; Haegeman and Fiers, 1978; Canaani et al., 1979); polyoma sequence (Soeda et al., 1980); polyoma early cap sites (Kamen et al., 1980); and polyoma late cap sites (Flavell et al., 1979; 1980).

that in Ad2 capping is temporally (and potentially mechanistically) very tightly coupled to RNA initiation (Babich et al., 1980). A component of the capping machinery could align the polymerase in the initiation complex instead.

Homologies at and Upstream from Ad2 Cap Sites

As discussed above, cap sites correspond to transcription initiation sites. In addition, sequences surrounding the cap site may correspond to classical promoters. Promoters have been defined genetically (for prokaryotes) as the loci of mutations which enhance or repress transcription (Jacob et al., 1964). Prokaryotic promoter mutations fall within an approximately 35 nucleotide long DNA segment, upstream from the RNA initiation site, which makes close contacts with RNA polymerase during transcription initiation (for reviews see Rosenberg and Court, 1979; Siebenlist et al., 1980).

Figure 44 summarizes the sequences found near Ad2 cap sites. Figure 45 gives the corresponding sequences for several other viral and cellular genes. The only sequence homology common to most cap regions is the AT rich TATA box first noted by Goldberg and Hogness for the histone genes of D. melanogaster (Goldberg, 1979). This TATA box lies -27 to -33 nucleotides upstream from the cap site and has the consensus sequence $TATA\begin{matrix}A\\T\end{matrix}A\begin{matrix}T\\A\end{matrix}$ (Corden et al., 1980). The cap regions of Ad2 IVa2 and EIIa and the polyoma and SV40 late mRNAs have no recognizable TATA boxes, however.

If the sequences are lined up by the first T of the TATA box, the initiation regions all lie the same distance downstream from the TATA box. This situation is analogous to that found for prokaryotic promoters, where the initiation sites lie within a narrow region downstream from the AT rich Pribnow box, the binding site for the core polymerase (Rosenberg and Court, 1979). The Pribnow box and initiation sites are spaced about 1 turn of the DNA helix apart, allowing the polymerase to interact with only one side of the DNA helix (Siebenlist et al., 1980). In eukaryotes, the TATA box and initiation region lie approximately 3 turns of the helix apart, also suggesting that the polymerase molecule

may interact with only one side of the helix.

To investigate the hypothesis that the TATA box determines the location of the initiation region, I used the two closely related adenovirus strains, Ad2 and Ad5. Both strains have a large TATA box (TATATATA) upstream of the EIV cap site. In the cap region of EIV, the only sequence differences between Ad2 and Ad5 are the change of a T to a C halfway between the TATA box and the cap site, and the insertion of an extra T at the cap site (compare the Ad2 sequence in Figure 38 with the Ad5 sequence in Figure 41). If the TATA box fixes the site of initiation, two consequences of these sequence differences may be imagined. Displacing the A cap site from the TATA box might decrease the frequency of A termini in Ad5 because the A cap site would be shifted away from the initiation site favored for steric reasons by the RNA polymerase. Alternatively, the polymerase could scan the broader region of potential U caps, but still use the A terminus for the majority of starts because purines are preferred for initiation. With the latter model, the Ad5 EIV mRNA should have six U-termini instead of the five found for Ad2, plus a major A cap. When Ad5 region EIV mRNA was assayed by the 5' labeling technique, the A terminus was still preferred, but six U termini were found in contrast to Ad2 which had 5 U termini (compare Figures 37D and E). Note that the U termini most distant from the A cap are reproducibly the most intense. This was also found with the in vivo EIV capped products shown in Figure 37B. As with EIb, PIX and EIII discussed above, the close proximity of a possible purine start may greatly decrease the frequency of use of a potential pyrimidine start. Rosenberg and Court (1979) have discussed the question of steric constraint versus purine preference for E. coli RNA polymerase.

The Role of the TATA Box In Vivo The role of the TATA box in transcription initiation has been further elucidated in other labs by studies of the expression of isolated genes and their mutants both in vivo and in vitro. As shown by deletion mutants, the TATA box is not essential for in vivo expression of the SV40 early, polyoma early and histone H2A genes when assayed for mRNA or protein production (Benoist and Chambon, 1980; Bendig, Thomas and Folk, 1980; Grosschedl and

Birnstiel, 1980a). Therefore the TATA box is not a promoter in vivo by the classical definition (Jacob et al., 1964).

The TATA box appears to act as a selector sequence which specifies the initiation region. If small deletions are made between the TATA box and initiation sites of the SV40 early genes, the 5' termini of mRNA made in vivo are encoded 27-34 base pairs downstream from the first T of the TATA box, rather than at the same absolute position in the DNA (Gluzman, Sambrook and Frisque, 1980; Ghosh et al., 1981; Benoist and Chambon, 1981). These deletions can extend up to 2 nucleotides downstream from the TATA box without destroying the selector function. In addition the cap site can be deleted entirely. The new 5' ends will be encoded the same distance downstream from the TATA box within sequences which were previously far downstream. Therefore it appears that the site of initiation is specified not by the sequences at the initiation site, but by the TATA box upstream.

The selector function of the TATA box is further demonstrated by analyzing the 5' termini of mRNA made in vivo from TATA deletion mutants of histone H2A (Grosschedl and Birnstiel, 1980a) and SV40 early genes (Benoist and Chambon, 1981). In both cases, mRNA 5' termini are encoded at multiple new sites (for SV40, these may correspond to very minor wild type start sites) over a large region. The presence of a TATA box in the wild type gene apparently suppresses initiation at these sites and restricts initiation to a narrow region downstream. The genes for SV40 and polyoma late mRNAs, shown in Figure 45, lack TATA boxes. In both cases the mRNAs have multiple 5' termini encoded over a large (approximately 400 nucleotide) region of the DNA (see references given in Figure 45). However, the genes for Ad2 EIIa and IVa2, both lacking TATA boxes (Figure 44), produce mRNAs whose 5' termini are microheterogeneous. Perhaps other sequence elements, such as the TCCTT sequence found in both upstream regions at about the same position as the TATA box, can also function as a selector element. It is worth mentioning that the sequence TCAGAGTGG at the IVa2 cap site and the upstream sequence TCCTT are completely conserved between adenoviruses 2, 3, 5, 7 and 12 (representing adenovirus groups A, B and C) whereas the rest of the

sequence in the region has diverged somewhat (J. Engler and T. Broker, pers. comm.). Also most of the macroheterogeneous 5' termini of the yeast iso-1-cytochrome c gene lie downstream from either "TATA"-like sequences or TCCTT sequences (Faye at al, 1981).

If the sequence TCCTT can function as a selector element, this could explain the bimodal distribution of U and A termini in region IV (see Figures 37D, 37E and 45). The TATA box and upstream sequences of EIV (TCCTATATATA) could be viewed as the fusion of TCCTT and TATATATA with the deletion of a single T. The TCCTT element would then specify the upstream U termini and the TATA box the major A terminus.

Function of the TATA Box In Vitro Both the whole-cell extracts of Manley et al. (1980) and the S100 extracts of Weil et al. (1979) initiate transcription accurately on DNA containing the major late promoter (Manley et al., 1980; Weil et al., 1979), the mouse β-globin gene (Luse and Roeder, 1980), the human β-, ε- and α-globin genes (Proudfoot et al., 1980) and conalbumin and ovalbumin genes (Wasylyk et al., 1980b). The first experiments studying deletion mutants in in vitro transcription systems suggested that the TATA box was both necessary and sufficient for specific transcription initiation in vitro. By assaying for the synthesis of specific run-off products from deletion mutants, Hu and Manley (1981) have shown that the TATA box plus approximately 15-17 nucleotides on either side is sufficient for near maximal expression of the Ad2 major late promoter. In similar experiments, Corden et al. (1980) have shown that sequences from -12 to -32 nucleotides from the cap site for Ad2 major late and -10 to -44 nucleotides for conalbumin are sufficient for expression in vitro. In both cases, however, these sequences are the maximum necessary for in vitro transcription since all possible deletion mutants were not examined. The work of Mathis and Chambon (1981) has shown that in the SV40 early region, downstream deletions extending to within 2 nucleotides of the TATA box do not destroy the selector function of the TATA box. However, deletions which remove the sequences immediately downstream from the TATA box decrease the efficiency of transcription in vitro. A point mutation of the conalbumin TATA box to TAGA decreases the level of in vitro transcription to 5%

of the wild type control reaction, but does not alter the RNA 5' termini as shown by the production of a wild type run-off product (Wasylyk et al., 1980a).

Deletion of the early SV40 TATA box eliminates transcription when assayed by run-off products. However, if transcription is assayed by S1 mapping, there is no net decrease in transcription over the wild type control and the RNA has the same multiple 5' termini found for mRNA produced from the TATA deletion mutant in vivo (Mathis and Chambon, 1981). It has not been reported whether the TAGA mutant of conalbumin also produces multiple 5' termini encoded over a large region.

The late SV40 transcription unit, which lacks a TATA box and encodes macroheterogeneous 5' termini, produces RNAs in vitro with the same 5' termini as late mRNA (Handa et al., 1981). It has been reported that the Ad2 EIIa and IVa2 transcription units, which also lack TATA boxes, are nonfunctional in vitro by the run-off assay (Hu and Manley, 1981; Lee and Roeder, 1981). A. Fire, however, observes an EIIa run-off product being produced at approximately 1/10 the level found for the other Ad2 early transcription units. We have demonstrated, using 5' labeling experiments similar to those described above for Ad5 EIV, that the EIIa transcription unit produces RNA in vitro with the same A and G 5' termini (data not shown) found for EIIa mRNA (spots 2 and 1 of Figure 28A). An additional major CU terminus and other possible minor termini not found in mRNA from EIIa were also found for EIIa RNA synthesized in vitro. Thus the selector element which is functional in vivo for EIIa is also functional in vitro, although it may not be as dominant over other weak selector elements as the TATA box is.

It can be concluded that the TATA box is not an essential component of a promoter either in vivo or in vitro. However, it functions both in vivo and in vitro as a selector element specifying the initiation site(s). Therefore the TATA box can be considered to be one functional element of a multi-component promoter. Other sequences or structural elements must also be important for promoter function and are discussed below.

Sequences Upstream from the TATA Box

There are no obvious sequence homologies upstream from the TATA box which are common to several of the adenovirus transcription units. Limited regions of homology can be found between pairs of sequences (e.g. Ia and EIV or IVa2 and EIIa), but their functional significance is difficult to ascertain without genetic studies. A computer aided search for homologies is presently underway in collaboration with Bruce Erickson (Rockefeller University). A sequence homology (with consensus sequence $GG^{C}_{T}CAATCT$) has been found at approximately -70 nucleotides with respect to the cap sites of many genes (Benoist et al., 1980; Efstratiadis et al., 1980). Similar sequences can be found at approximately -70 nucleotides of Ad2 EIa (GGTCAAAGT) and Ad2 major late (GGTTTATAG) genes but it is difficult to identify this homology in the sequences upstream from the other Ad2 transcription units.

Although the TATA box itself does not appear to be a promoter *in vivo*, there is increasing evidence that sequences upstream may at least be part of the promoter. In addition, these sequences may be responsible for gene specific regulation (e.g. see Jones and Kafatos, 1980).

Grosschedl and Birnstiel (1980a, 1980b) have studied the effects of deletions of sequences upstream from the TATA box of the histone H2A gene of *P. miliaris* on the expression of the gene in *Xenopus laevis* oocytes. Deletion of the highly conserved H2A specific sequence block at approximately -70 to -100 nucleotides enhanced mRNA synthesis by a factor of 2. Deletion of the spacer DNA (which is AT rich and contains palindromic sequences) upstream of -100 nucleotides reduced mRNA synthesis to 1/15 to 1/20 of the wild type control. Inversion of this DNA lead to a 4-5 fold increase in transcription. This sequence element has been called a "modulator" by Grosschedl and Birnstiel. The modulator element apparently requires the presence of the selector (TATA box) element and initiator element (cap site region) for transcription, since the inverted spacer mutant does not synthesize nonsense strand RNA.

A similar situation exists for SV40, which has a 72 base pair tandem repeat located at -116 to -188 and -189 to -261 nucleotides with

respect to the early cap site. Benoist and Chambon (1981) and Gruss, Dhar and Khoury (1981) have shown that at least one complete 72 base pair repeat is necessary for *in vivo* expression of the SV40 early genes.

The 72 base pair repeats of SV40 are not required for expression of the SV40 genes in the *in vitro* transcription systems (Mathis and Chambon, 1981). Likewise, sequences upstream from the TATA box are not required for *in vitro* expression of the Ad2 major late promoter (Hu and Manley, 1981; Corden et al., 1980) or of conalbumin (Corden et al., 1980).

The apparent discrepancy between sequences required *in vivo* and those required *in vitro* for transcription may be a reflection of the state of the DNA in the two environments. In cells the DNA is complexed with proteins in a well defined structure, chromatin. The DNA in the *in vitro* transcription systems does not appear to be in a chromatin structure (Mathis and Chambon, 1981). As mentioned in the Introduction, transcriptionally active genes appear to be in a different chromatin structure than inactive genes. Perhaps more relevant to this discussion, however, are the observations that DNA upstream from and at the 5' ends of genes is in a different conformation than the DNA within a gene. Samal et al. (1981) have shown that the nucleosomes are phased on the intergenic spacer DNA but not on the transcribed DNA of the *D. melanogaster* histone gene repeat. In addition, the 5' ends of all histone genes are in an exposed conformation, as shown by hypersensitivity to DNAse I and micrococcal nuclease. The nuclease hypersensitivity of the 5' ends of the *Drosophila* heat shock genes (Wu, 1980; Keene et al., 1981) and SV40 early and late genes (Scott and Wigmore, 1978; Waldeck et al., 1978; Varshavsky, Sundin and Bohn, 1979) has also been demonstrated. In addition, the region between the early and late SV40 mRNA 5' ends appears to be free of nucleosomes when examined by electronmicroscopy (Saragosti, Moyne and Yaniv, 1980). It is possible that this region is the only site accessible to RNA polymerase. This hypothesis is supported by the experiments of Jakobovits et al. (1980). Transcription of naked SV40 DNA by *E. coli* RNA polymerase produces RNA which is predominantly complementary to the early strand and which has initiated

randomly. However, transcription with E. coli RNA polymerase (in the presence of α-amanitin to inhibit endogenous RNA polymerase) of the SV40 minichromosome, isolated from cells late after infection, produces RNA which is predominantly complementary to the late strand and which initiates at the late promoter.

Thus DNA sequences far upstream from the 5' ends of a gene could control expression of a gene by phasing the nucleosomes or in some other way determining chromatin structure. Alternatively it may be the DNA in the immediate vicinity of a gene which determines the chromatin structure and activity of a gene. Either way, chromatin structure is likely to be an important feature of promoters.

Regulation of Protein IX and Late Region IIa mRNA

It is of interest that region IIa is expressed at early times through an mRNA with 5' ends at 75 m.u. and at late times through an mRNA with 5' ends at 72 m.u. (Chow et al., 1979, 1980). By fingerprinting region IIa mRNA at early and late times, the representation of the 75 m.u. and 72 m.u. caps at these two stages of infection can be assessed. In the fingerprint of 5' labeled mRNA isolated at early times (Figure 28A), the coordinate 75 5' terminal oligonucleotides are readily detected, while the coordinate 72 5' terminal oligonucleotides are absent. Note that the coordinate 72 5' termini are absent even though the template for these termini lies within an active transcription unit. In contrast, at late times, as shown in Figure 30B, the reverse is true. The coordinate 72 5' terminal oligonucleotides are abundant and the coordinate 75 5' terminal oligonucleotides are present only at very low levels.

An analogous result was obtained with PIX mRNA. The TATA box and cap sites for PIX mRNA lie wholely within the EIb transcription unit. The PIX caps were not detected in mRNA 5 hours post-infection with a cycloheximide block (Figure 12B). Under these conditions, the EIb transcription unit is active, and EIb caps are readily detected in the cytoplasm (Figure 12A). In contrast, at late times the PIX caps were abundant (Figure 12C). Thus, as with region IIa-late, the PIX promoter lies

within a second active transcription unit and does not yield (stable) mRNA during the early stage of infection, but is active during the late stage.

A similar example has been found in mice by Young, Hagenbuchle and Schibler (1981). The mouse salivary α-amylase gene encodes an mRNA with a 50 nucleotide leader encoded 7.5 kb upstream from the coding region. In the liver, the α-amylase mRNA coding region is transcribed from the same gene as the salivary mRNA, but the 168 nucleotide leader is encoded only 4.5 kb upstream from the the coding region, within DNA which acts as intron in the salivary gland.

Wilson et al. (1979) have shown that the PIX promoter is inactive early after infection, but is functional at late times. However, the shifts in 5' termini of the DBP mRNAs and α-amylase mRNAs have not yet been demonstrated to be at the level of transcription initiation. In the case of the α-amylase gene, the α-amylase mRNA is 2% of the cytoplasmic polyadenylated RNA in the salivary gland and only 0.02% in the liver (Young et al., 1981). This 100 fold difference in gene expression could be obtained by expression of the α-amylase gene through two different tissue specific promoters which differ highly in activity. The PIX promoter, and perhaps the late DBP promoter and liver α-amylase promoter, is an example of an inactive promoter within an active transcription unit. Thus the chromatin structure (or other DNA organization) required for a transcriptionally active template may not be sufficient to permit RNA initiation.

Homologies at Ad2 Splice Donor Sites

The putative splice donor sequences from Figures 16, 20, 25, 31 and 38 are shown in Figure 46, along with the consensus sequence from Rogers and Wall (1980) and Lerner et al. (1980). Ad2 splice donor sequences are similar to those from other viral and cellular genes. Therefore the mechanism of splicing in Ad2 is probably similar to that in other systems.

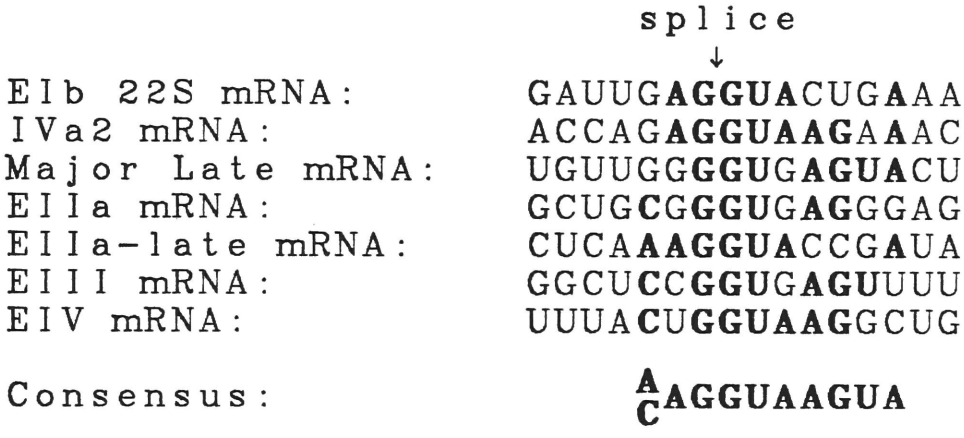


Figure 46. Sequences at Splice Donor Sites

The RNA sequences at the first splice donor site of the precursors to the EIb 22S, IVa2, major late, EIIa, EIIa-late, EIII and EIV mRNAs are taken from the DNA sequences in Figures 16, 20, 25, 31 and 38, respectively. The consensus splice donor sequence of Rogers and Wall (1980) and Lerner et al. (1980) is shown on the bottom line. Residues which match the consensus nucleotide at that position are shown in bold.

Conclusions

Most of the promoter structures of Ad2 share in common with cellular promoters the presence of the TATA box and the preference for purine termini. My data suggest that the precise initiation site is determined 1) by steric factors of the polymerase which measure approximately 30 nucleotides from the TATA box to define an initiation region, and 2) a scanning of this initiation region for preferred purine starts. However, two cap regions lack evident TATA boxes, and may rely on other sequence elements or factors for their specification. A complete understanding of the mechanism of RNA initiation, and its control, will require analysis of both the enzymology and the chromatin structures utilized in RNA initiation.

EXPERIMENTAL PROCEDURES

Preparation of Ad2 Lysate

Virus from a single Ad2 plaque (from M.-T. Hsu) was used to infect approximately 10^7 HeLa cells grown in monolayer culture. After approximately 1 1/2 weeks, the HeLa cells became rounded and detached from the petri dish. The cells were harvested, pelleted at 1k RPM and resuspended in 5 ml culture medium without serum. The suspension was freeze-thawed 6 times and centrifuged at 2k RPM for 5 minutes to remove cell debris. The supernatant was kept frozen as the primary lysate.

One ml of the primary lysate was passaged twice on HeLa monolayer cultures to produce a higher titer secondary lysate. This lysate was quantitated by preparation of purified virus (discussed below) and by fluorescent focus assay (Philipson, 1961; Thiel and Smith, 1967; with the aid of H. Ginsberg). The ratio of particles to pfu's was approximately 20 for the lysate. This lysate was then used to infect approximately 10^{10} HeLa cells in spinner culture at an moi of 12 pfu/cell to prepare a lysate stock for subsequent virus preparations and purified virus for Ad2 mRNA preparations. Virus was harvested at 48 hours p.i. A typical yield of virus from cells infected with lysate at low moi was approximately 4×10^4 particles/cell.

Preparation of Purified Virus

HeLa cells in spinner culture at a density of approximately 8×10^8 cells/l were infected at an moi less than 50 pfu/cell by addition of lysate. After incubation for 4 hours, cells were diluted to 4×10^8 cells/l with medium plus 5% FCS. At 48 hrs p.i. virus was isolated by the procedure of Doerfler (1969). Infected cells were centrifuged, washed with cold PBS and resuspended in 0.01 M Tris-HCl pH 8.1 at 2×10^7 cells/ml. This cell suspension was freeze-thawed twice, sonicated, and centrifuged to remove cell debris. The supernatant was extracted twice with trichlorotrifluoroethane and layered onto preformed 1.2-1.45 gms/ml CsCl gradients. The gradients were spun in a SW27 rotor for 2 hrs at RT and 20K RPM. The virus band was collected by bottom puncture,

diluted with an equal volume of 0.01 M Tris pH 8.1, and again layered on preformed CsCl gradients. This time the gradients were spun in a SW27 rotor for 17 hrs at 22K RPM and 4°C. The collected virus was quantitated by dilution of an aliquot into 0.1 X SSC, 0.5% SDS and the absorbance at 260 nm obtained ($1A_{260}=10^{12}$ particles/ml). Purified virus in CsCl was stored at 4°C for up to one month and used for infection at high multiplicity for early mRNA preparations.

Preparation of Purified Ad2 DNA

DNA was extracted from virions by the procedure of Pettersson and Sambrook (1973). CsCl banded virus was diluted into 5 volumes of a solution of 0.05 M Tris pH 7.5, 0.001 M EDTA, 0.5% SDS and 1 mg/ml pronase which had been self-digested for 30' at 37°C to remove DNAse and incubated at 37°C for an additional 30'. The DNA was extracted twice with 1/4 volume of $CHCl_3$-1% isoamylalcohol plus 1 volume 0.1 X SSC saturated phenol and once with $CHCl_3$ and then EtOH precipitated. DNA was recovered by centrifugation, dissolved in H_2O and stored at -20°C.

Preparation of Ad2 DNA Restriction Fragments

Purified Ad2 DNA was digested with various restriction enzymes under the appropriate conditions for each enzyme and electrophoresed on 1.4% agarose slab gels containing 1 X E buffer. DNA bands were localized with ethidium bromide and UV light. DNA was extracted from the agarose either by dissolution of the gel in 5 M $NaClO_4$ at 65°C and isolation on hydroxyapatite (Lewis et al., 1975) or by electroelution in 1 X E buffer. DNA fragments were stored in H_2O at -20°C.

Binding of DNA to Nitrocellulose Filters

A DNA fragment in 200 λ of H_2O was denatured by addition of 20 λ 1 M NaOH and incubation at RT for 15'. Then 8 ml ice-cold 2 M NaCl was added and the DNA was allowed to flow slowly through a nitrocellulose filter (Millipore; 0.45 μm filters were used for fragments larger than 600 b.p. long and 0.22 μm filters for those less than 600 b.p.). Filters were washed with 6 X SSC and baked at 80°C for 4 hrs under

vacuum. Closed circular plasmid DNA was boiled 2' after addition of NaOH to nick the molecules to allow strand separation after denaturation. Baked filters were stored under vacuum prior to use. Filters used for preparative hybridization were rebaked after elution of the RNA and reused.

DNA Strand Separation

Approximately 1 γ of a purified DNA restriction fragment in H_2O (<25 λ) was made 0.1 N NaOH with addition of 1 N NaOH and incubated at RT for 15'. An equal volume of cold H_2O and 2/13 volume of bromophenol blue in 100% glycerol were added. The DNA was immediately loaded onto a 0.7 cm X 15 cm precooled (4°C) tube gel composed of 1.4% agarose in 1/2 X E buffer without ethidium bromide. Electrophoresis was at 40 V and 4°C. The gel was stained with ethidium bromide to localize the bands, which were cut out and electroeluted into 1 X E buffer. The electroeluted DNA was made 5 X SSC, incubated at 65°C for 2 hrs to reanneal any contamination from the other strand, and loaded directly onto nitrocellulose filters.

^{32}P in Vivo Labeled Early mRNA

Approximately $4X10^8$ exponentially growing HeLa cells were centrifuged, taken up in serum-free medium at a concentration of 10^7 cells/ml, infected with 2-10,000 particles/cell of purified virus, and incubated 30' at 37°C. The cells were then washed 3X with phosphate free medium, resuspended in 150 ml phosphate-free medium plus 5% dialyzed fetal calf serum plus 100 mCi $^{32}PO_4$, and incubated at 37°C. Cycloheximide was added after 30' to a final concentration of 25 γ/ml. After 5 hours incubation, cells were washed in PBS, resuspended in 7 ml RSB, allowed to swell 10', and dounced 10 times. Nuclei were spun out at 2k for 5' and washed with 3 ml RSB. The combines supernatants were spun at 10k for 20' to remove mitochondria. This supernatant was made up to 0.15M NaCl, 0.01 M EDTA, 0.01M Tris HCl pH 7.4, and 0.7% SDS, extracted twice with phenol saturated with the same plus an equal volume of $CHCl_3$, and then with $CHCl_3$ plus 1% isoamylalcohol. The cytoplasmic RNA was

precipitated with EtOH. To prepare poly(A)-containing mRNA, the precipitated cytoplasmic RNA was dissolved in 10 ml ETS, heated at 65°C for 5', diluted with 10 ml 0.4 NETS, and run onto a 3 ml poly(U)-sepharose column. This column was washed with 0.2 NETS, ETS, 10% formamide-ETS and step eluted with 90% formamide-ETS. The RNA was then EtOH precipitated in the presence of 100 γ yeast carrier RNA. Prior to hybridization, the mRNA was partially broken with RNAse T1 to approximately 350-500 nucleotides. The mRNA was dissolved in 10 ml ET, heated to 37°C and incubated for 5' with one unit of RNAse T1. SDS was added to 0.2% to stop the reaction. The RNA was extracted with phenol and twice with $CHCl_3$ with 1% isoamylalcohol, made up to 0.2 NETS with SDS and NaCl, and run over a poly(U)-sepharose column to remove the 3' end portions of the mRNA (approximately 25 %). The flow through contained the non-3' portion and was EtOH precipitated and used for hybridization and subsequent analysis of mRNA 5' termini.

RNA was extracted from nuclei as follows. Nuclei were washed in 4 ml RSB containing magic (3.3% DOC and 6.6% Tween-80) and centrifuged. The nuclear pellet was then resuspended in 4 ml HSB with 10^{-4} M $CaCl_2$ and 400 γ DNAse I and incubated at 37°C for two minutes with constant mixing by pipetting. The solution was then made to 10 ml of 0.1 M NaAc pH 5, 0.02 M EDTA, 0.2% SDS and extracted with 10 ml phenol at 65°C for 10 min., 1:1 phenol-$CHCl_3$, and $CHCl_3$ and then EtOH precipitated. The precipitated RNA was then dissolved in 1 ml DNAse buffer (0.1 M NaCl, 0.05 M Tris pH 7.4, 5 mM $MgCl_2$) plus 100 γ DNAse I and digested at 37°C for 1 hr. The reaction was stopped by addition of EDTA and phenol and Et_2O extracted. After addition of 1 volume EtOH the RNA was run onto a CF11 cellulose column, washed with 50% STE-50% EtOH (STE=0.1 M NaCl, 0.05 M Tris pH 7.4, 1 mM EDTA), eluted with STE-0.2% SDS, and EtOH precipitated. The RNA was collected by centrifugation and then used for hybridization.

^{32}P in Vivo Labeled Late Ad2 mRNA

HeLa cells were infected with $2X10^3$ particles per cell of purified virus and incubated for 14 hours at 37°C and $8X10^8$ cells per liter.

Cells were then washed with phosphate free medium, resuspended in phosphate free medium plus 5% FCS plus 100 mCi $^{32}PO_4$ and incubated for 3 hours at 37°C (no cycloheximide). Cytoplasmic polyadenylated RNA was then extracted and prepared as for early polyadenylated RNA.

5' Kinased Early and Late Ad2 mRNA

Unlabeled early mRNA was prepared similarly to ^{32}P labeled mRNA, except that the infected cells ($4X10^8$ cells/l) were incubated in medium containing phosphate and with 25 γ/ml cycloheximide added at 30' p.i. Cells were harvested at 5 hrs p.i. Unlabeled late mRNA was isolated from cells infected with lysates, incubated without cycloheximide and harvested at 24 hrs p.i.

Early or late unlabeled polyadenylated cytoplasmic RNA was 5' labeled by kinasing using a modification of the method of Lockard and RajBhandary (1976). RNA from $4X10^8$ infected cells was dissolved in 95 λ H_2O to which was added 5 λ 0.4 M $NaIO_4$ (freshly dissolved in H_2O) and incubated at room temperature for 15 minutes in the dark. SDS was added and the RNA was EtOH precipitated. The RNA pellet was then dissolved in 25 λ H_2O, 2.5 λ 2% SDS and 50 λ 0.1 M lysine pH 8.6, 0.2% SDS and incubated at 45°C for 2 hrs. The RNA was then ethanol precipitated several times to remove SDS, dissolved in 50 λ 0.05 M Tris HCl pH 9.0, 1mM $MgCl_2$, 2.7 u/ml calf intestinal alkaline phosphatase (RNAse free; gift of Gregory Gasic) and incubated at 37°C for 30'. The enzyme was then inactivated with EDTA, the solution was extracted with phenol and Et_2O, and the RNA was EtOH precipitated. To kinase, 250-500 μCi γ-ATP (3000 Ci/mmol) dissolved in 45 λ H_2O plus 5 λ 10X kinase buffer (500 mM Tris HCl pH 9, 100 mM $MgCl_2$, 50 mM DTT, 1mM spermidine, 1mM EDTA) was added to the RNA pellet. Approximately 10 units T4 polynucleotide kinase (P-L Biochemicals or NEN) was added and the reaction incubated at 37°C for 30'. The sample was EtOH precipitated several times to remove unincorporated γ-ATP. Prior to hybridization, the RNA was fragmented with alkali by dissolving in 100 λ ETS, adding 25 λ 1N NaOH and incubating on ice for 30'. The solution was neutralized with 50 λ 1M unbuffered HEPES and the RNA EtOH precipitated.

Preparation and Analysis of EIV in Vitro RNA

RNA was transcribed *in vitro* using whole-cell extracts (Manley et al., 1980) from uninfected cells. Reaction conditions were similar to those given in Handa et al. (1981). The EIV DNA template was a pBR clone of the Ad5 Eco RI B (84-100 m.u.) fragment. The DNA was cut with Sma I for the production of α-^{32}P-NTP labeled run-off products. The NTP concentrations for each reaction were: U label, 50 μM ATP, CTP and GTP and 100 μM UTP; A label, 75 μM ATP, 50 μM CTP and GTP and 5 μM UTP; and G label, 50 μM ATP, CTP and UTP and 29 μM GTP. Each reaction was phenol extracted and electrophoresed on a 7 M urea-8% acrylamide gel. The run-off bands were electroeluted, ethanol precipitated and RNAse T1 fingerprinted as described below.

For analysis of *in vitro* transcripts by the 5' labeling procedure, the Ad5 Eco RI B clone cut with Bgl II was used as template. No α-^{32}P-NTP was added and each NTP was 50 μM. The RNA was not fractionated prior to analysis. The *in vitro* transcripts were 5' labeled by the same procedure used for mRNA, except that the RNA was not alkali treated prior to hybridization. Ad2 Sma I K DNA was used for hybridization.

Hybridization

All RNA was hybridized to nitrocellulose filter bound DNA, prepared as described above. All filters contained 100 γ equivalents of specific viral DNA or 200 γ equivalents of a cloned viral DNA fragment.

RNA was dissolved in 2 X TESS and hybridized at 65°C for 24-48 hrs. For any RNA preparation, all filters were incubated together in one vial. Filters were washed with 2 X SSC twice for 30' at 64°C. Hybridized RNA was released from the filters by three successive 1 minute elutions with 1 ml H_2O each at 100°C. RNA was precipitated with EtOH in the presence of 10 γ yeast carrier RNA. In some cases hybrids were trimmed on the filter prior to elution by incubation with 5 u/ml RNAse T1 or 2.5 γ/ml RNAse A at 37°C for 30 minutes, followed by nuclease inactivation with 0.15 M sodium iodoacetate in 2 X SSC at 37°C for 1 hour.

RNA Sequencing

Eluted hybrid RNA was digested with 1 γ (5 units) RNAse T1 or 1 γ RNAse A in ET at 37°C for 30' and fractionated by two-dimensional fingerprinting according to the method of Brownlee (1972). Oligonucleotides were eluted from the homoplate with 30% TEC (30% TEA in H_2O bubbled with CO_2 until a single phase results). TEC was removed by repeated lyophilizations. Redigestions were done at 37°C for 1 hr with 10 λ of 0.5 mg/ml RNAse A plus 25 u/ml RNAse T2 in RNAse T2 buffer (15 mM NH_4Ac pH 4.5, 2 mM EDTA); 10 λ 0.5 mg/ml RNAse A in ET; 10 λ 1 mg/ml RNAse T1 in ET; 10 λ of 0.25 mg/ml nuclease P1 in 10 mM NaAc pH 6, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$; 10 λ of 1.5 units/ml nucleotide pyrophosphatase in 30 mM Tris pH 7.4, 2 mM $MgCl_2$; and 25 λ of 20 units/ml BAP in 15 mM Tris pH 8.1, 1 mM $MgCl_2$. Base compositions on products from RNAse A redigestion were obtained by digestion with 10 λ of 5 u/ml RNAse T2 in RNAse T2 buffer. Partial RNAse T2 sequencing was done by digestion with 5 λ of 50 u/ml RNAse T2 in ET at 25±5°C for 45' followed by 2-D fingerprinting. Base compositions, and nuclease P1 and pyrophosphatase digestions were electrophoresed at pH 3.5 on 540 paper (Whatman); all other digests were electrophoresed on DE81 paper (Whatman) at pH 3.5 (Brownlee, 1972).

Unlabeled P1 cap core markers ($^{7m}GpppX_m$) were obtained from P-L Biochemicals. RNAse T1 and RNAse T2 were from Sankyo (supplied by Calbiochem), Nuclease P1 from Yamasha Shoyu Co., nucleotide pyrophosphatase (Type II) from Sigma, BAPF from Worthington and RNAse A from Calbiochem.

Reverse Transcription and S1 Mapping

Ad2 DNA (100 γ) was digested with restriction endonuclease Xho I, phosphatased, and kinased. The Xho I F fragment (15.5-22) was isolated and half was recut with Sma I and half with Ava II. Sma I produced a 15.5-18 fragment which was used as the S1 probe, and also for Maxim-Gilbert chemistries (see below) for sequence markers. Ava II produced a 17 nucleotide reverse transcription primer labeled at 15.5. Reverse transcription was by the procedure of Lamb and Lai (1980). Half of the primer was co-precipitated with 10^8 cell equivalents of cold late Ad2

mRNA and 50 γ carrier RNA. This DNA-RNA mixture was taken up in 100 λ 80% formamide, 0.4 M NaCl, 0.01 M PIPES-HCl pH 6.4, 2 mM EDTA and heated denatured at 75°C for 10 minutes, then hybridized at 61°C for 16 hours. Two volumes of H_2O were added and the material EtOH precipitated. Formamide was removed with two more EtOH precipitations. The hybrid was then dissolved in 400 λ reverse transcriptase reaction mix (50 mM Tris HCl pH 8.3, 50 mM KCl, 10 mM DTT, 10 mM $MgCl_2$, 40 μg/ml Actinomycin D, and 600 μM of each of the four cold deoxynucleotides), 50 units AMV reverse transcriptase (Dr. J. Beard, Life Sciences Inc., St. Petersberg, Fla.) was added, and incubated for 1 hr at 37°C. The reaction was phenol extracted, EtOH precipitated, hydrolysed with 0.03 M NaOH at 68°C for 10 minutes, and desalted on a Sephadex G-25 column. The DNA was then EtOH precipitated and taken up in 80% formamide-buffer-dye mixture before loading on a sequencing gel.

S1 mapping was by the methods of Berk and Sharp (1977a) as modified by Weaver and Weissmann (1979) and Manley et al. (1979a). Approximately 10 γ-equivalents of 5' labeled probe DNA was coprecipitated with late mRNA from $4X10^7$ cells and 15 γ carrier RNA, taken up in 30 λ 80% formamide, 0.4 M NaCl, 0.01 M PIPES-HCl pH 6.4, 2 mM EDTA, denatured at 75°C for 10 minutes and hybridized at 55°C for 12 hours. Aliquots consisting of 10 λ of hybridization reaction mix were then diluted into 100 λ cold S1 buffer (0.28 M NaCl, 30 mM NaAc pH 4.5, 1 mM $ZnSO_4$, 5% glycerol, 20 γ/ml denatured sonicated calf thymus DNA) containing 10, 100, or 1000 units S1 nuclease (Boehringer-Mannheim Biochemicals) and incubated at 30°C for 40 minutes. Digestion was stopped by addition of 1 λ 0.5 M EDTA and 10 γ carrier RNA, followed by EtOH precipitation. The sequencing chemistries, reverse transcripts and S1 resistant DNA fragments were run on a 0.5 mm thick 13% acrylamide-8M urea sequencing gel.

DNA Sequencing

All sequencing was done using DNA isolated directly from purified virus and not with cloned DNA fragments. All DNA was 5' labeled. Best kinasing results were obtained using pure viral DNA which had not been

through an agarose gel, or DNA fragments out of agarose gels which had been purified twice on HAP. BAPF (Worthington) was dialyzed against 0.05 M Tris pH 8.5, 0.01 M $MgCl_2$ to remove all traces of $(NH_4)_2SO_4$, since $(NH_4)_2SO_4$ is ethanol precipitable and NH_4^+ ions greatly inhibit PNK. It is also advisable to avoid PO_4^- contamination, since it is also EtOH precipitable and inhibits both BAP and PNK. Kinasing of protruding or flush 5' termini of DNA utilized 100-200 γ-equivalents of phosphatased DNA, 250 μCi γ-ATP (>2000 Ci/mmol), and 10 units PNK in the same kinase buffer used above for RNA. Incubation was for 60' at 37°C. The use of pH 9 buffer favors the forward reaction over the reverse reaction (van de Sande, Kleppe and Khorana, 1973). Kinasing of some flush and all recessed 5' ends was according to the procedure of Maxam and Gilbert (1980) and used 500 μCi of γ-ATP. All kinased fragments were recut with a second restriction endonuclease to generate DNA fragments labeled on only one end. The actual strategy used to sequence each region is shown in Figures 13, 17, 34 and C-2. DNA sequencing was by the method of Maxam and Gilbert (1977) modified for thin gels (Sanger and Coulson, 1978; Maxam and Gilbert, 1980).

LIST OF BUFFERS

DNAse I Buffer	0.1 M NaCl 0.05 M Tris pH 7.4 5 mM $MgCl_2$
1 X E Buffer	0.04 M Tris pH 7.9 0.005 M NaAC 0.001 M EDTA
ET	0.01 M EDTA 0.01 M Tris pH 7.4
ETS	0.01 M EDTA 0.01 M Tris pH 7.4 0.2 % SDS
HSB	0.5 M NaCl 0.01 M Tris pH 7.4 0.05 M $MgCl_2$
10 X Kinase Buffer	500 mM Tris-HCl pH 9 100 mM $MgCl_2$ 50 mM DTT 1 mM spermidine 1 mM EDTA
Magic	1 volume 10% deoxycholate 2 volumes 10% Tween-80
0.2 NETS	ETS plus 0.2 M NaCl
0.4 NETS	ETS plus 0.4 M NaCl
PBS	0.133 M NaCl 0.0025 M KCl 0.14 M Na phosphate (dibasic) 0.001 M K phosphate (monobasic)

RSB	0.01 M Tris-HCl pH 7.4
	0.01 M NaCl
	0.0015 M $MgCl_2$
1 X SSC	0.15 M NaCl
	0.015 M Na citrate pH 7.0
STE	0.1 M NaCl
	0.05 M Tris pH 7.4
	1 mM EDTA
2 X TESS	0.01 M TES pH 7.4
	0.3 M NaCl
	0.01 M EDTA
	0.2% SDS

APPENDIX A

Nuclease P1 Partials

Nuclease P1 is an enzyme which hydrolyzes 3'-5' phosphodiester bonds to yield 5' mononucleotides (Fujimoto, Kuninaka and Yoshino, 1969). Since cleavage does not proceed via a 2',3'-cyclic intermediate, it should not be inhibited by 2'-O-methyl groups. Therefore it is an ideal enzyme for determination of the 5' mononucleotide of 5' terminal oligonucleotides. Figure A-1 shows the P1 digests of several of the 5' terminal oligonucleotides. Note that P1 digestion usually yields more than one product and that only one of these products comigrates with a 5' mononucleotide. It was observed that the mobility pattern of products produced by termini with the same primary sequence was always the same. I therefore wondered if the products could have a monomer, dimer, trimer relationship to each other. Figure A-2 lane 1 shows the products produced by P1 digestion of the EIa 5' terminal oligonucleotide, $pA_mC_{(m)}U$-. Products a, b and c were eluted and redigested with the same P1 concentration (shown in Figure A-2 lanes 2-4, respectively). It can be seen that product a remained totally unchanged, while b was partially converted to a, and c to b and a. Thus a,b and c are probably pA_m, pA_mC_m and pA_mC_mU, respectively.

The presence of 2'-O-methyl groups is apparently inhibitory to phosphodiester bond cleavage by nuclease P1. This is further illustrated in Figure A-2, lanes 5-8. The EIa type I and II RNAse T2 products (Figure 11 lane 1) were digested with nuclease P1, as shown in Figure A-2 lanes 5 and 6, respectively. The oligonucleotide pA_mCp yielded pA_m only while pA_mC_mUp yielded pA_m, pA_mC, and pA_mC_mU. Similarly the major late type I oligonucleotide, pA_mCp (Figure 18B spot 2), yields mostly pA_m with some pA_mC with P1, while the type II oligonucleotide pA_mC_mUp (Figure 18B spot 3) yields mostly pA_mC_mU and less $pA_mC_{(m)}$ and pA_m (Figure A-2 lanes 7-8). These data are consistent with the observations of Yamada and Ishikura (1975) that methylation of the 2'-OH, especially on a pyrimidine, greatly reduced the susceptibility of the phosphodiester bond to P1 cleavage. They found that a 500 fold increase in

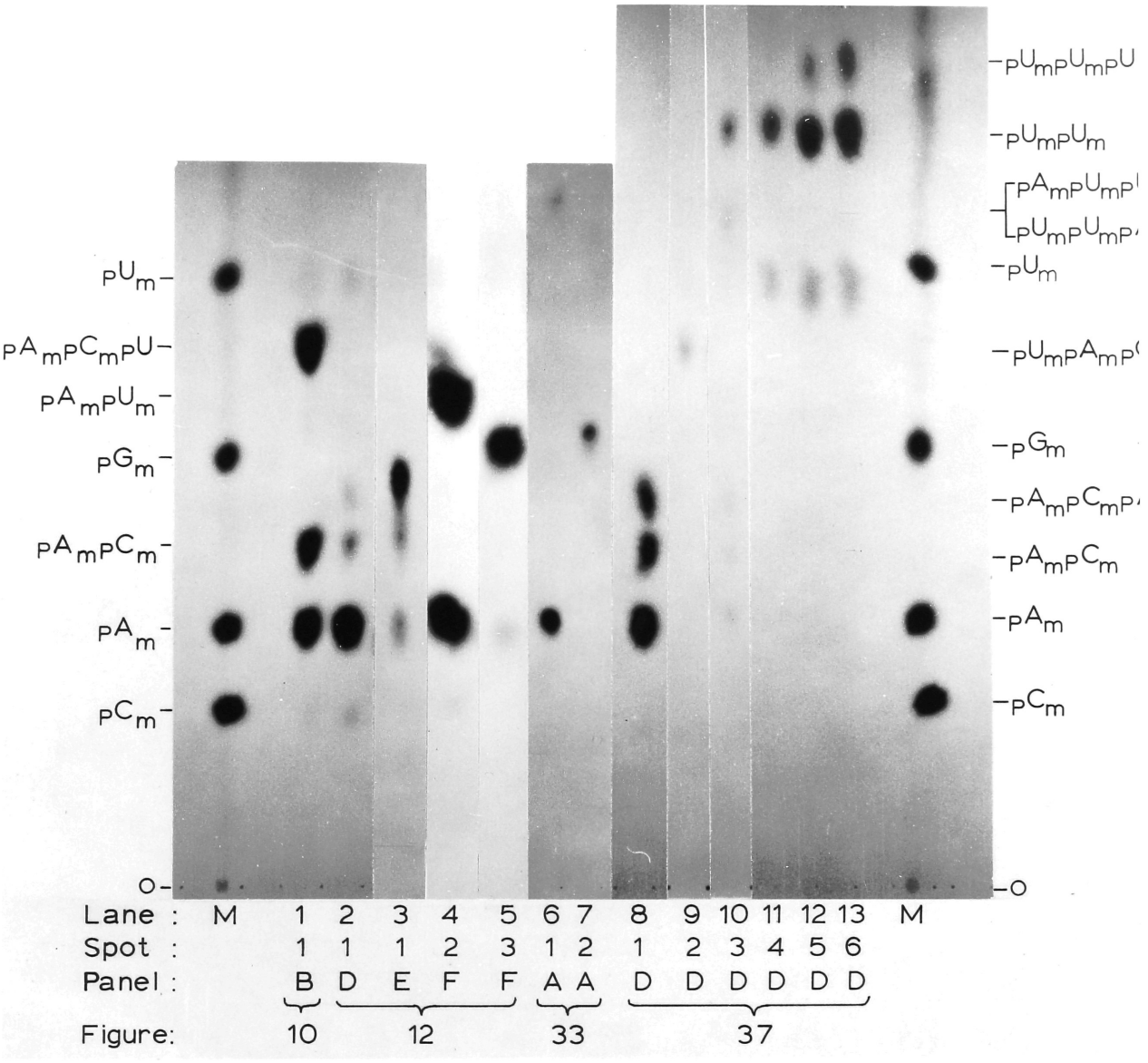


Figure A-1. P1 Analysis of <u>in Vitro</u> Labeled 5' Terminal Oligonucleotides

The 5' terminal oligonucleotides from Figures 10B (EIa), 12 D, E and F (EIb, PIX), 33A (EIII) and 37D (EIV) were digested with nuclease P1 and electrophoresed on 540 paper at pH 3.5 alongside 5' mononucleotide markers (lanes M). The origin of each spot is identified below each lane by figure, panel and spot number. The identities of all digestion products are indicated at the sides of the figure. These analyses have been summarized in Table I of Results.

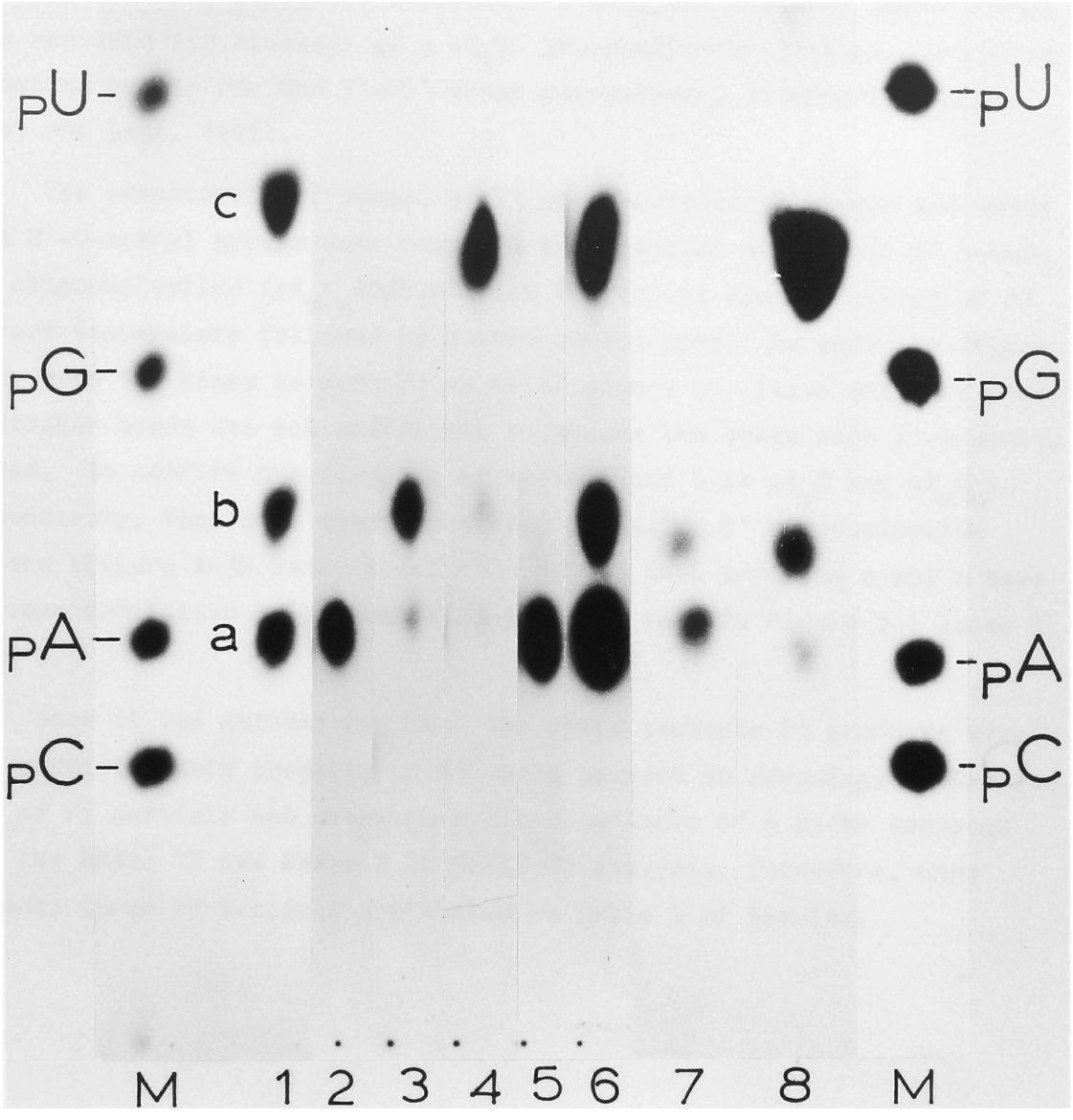


Figure A-2. Analysis of Nuclease P1 Partials

All analyses presented here have been electrophoresed on 540 paper at pH 3.5 alongside 5' mononucleotide markers (lanes M). Lane 1 is the nuclease P1 digestion of the EIa 5' terminal oligonucleotide shown in Figure A-1 lane 1. In lanes 2-4, products a, b and c from lane 1 were redigested with nuclease P1 under the same conditions. Note that b yields a, and c yields b and a. In lanes 5-6, the types I and II RNAse T2 oligonucleotides from Figure 11 lane 1 were digested with nuclease P1. In lanes 7-8, spots 2 and 3 (pA_mCp and pA_mC_mUp, respectively) from the fingerprint in Figure 18 B were digested with nuclease P1.

P1 concentration, increased incubation times, and elevated temperatures were required for cleavage of a $-C_mU-$ phosphodiester linkage. A similar situation exists for the 3'->5' venom exonuclease, phosphodiesterase I (Gray and Lane, 1967).

The sensitivity of normal 3'-5' phosphodiester linkages and those with 2'-O-methyl groups were compared by digestion of the EIb 5' terminal oligonucleotide ($pA_mC_mAUCUGp$) with increasing concentrations of P1 at room temperature followed by homochromatography. As shown in Figure A-3A, even 100 times as much P1 as is necessary to cleave ordinary phosphodiester bonds was not sufficient to cleave the bonds with 2'-O-methyl groups. To confirm the identity of spots a and b as pA_mC and pA_mC_mA, respectively, they were electrophoresed alongside 5' mononucleotide markers (Figure A-3B lanes a and b). Notice that products a and b have the same mobilities as the extra P1 products seen in Figure A-1 lanes 2, 3 and 8.

Once it was established that the extra nuclease P1 products were just partials, this property of P1 could be used to advantage. The pattern of P1 partials was often more characteristic of a given sequence than the RNAse T2 and RNAse A or RNAse T1 analysis. Therefore, when present, these P1 partials are listed in Table I of Results.

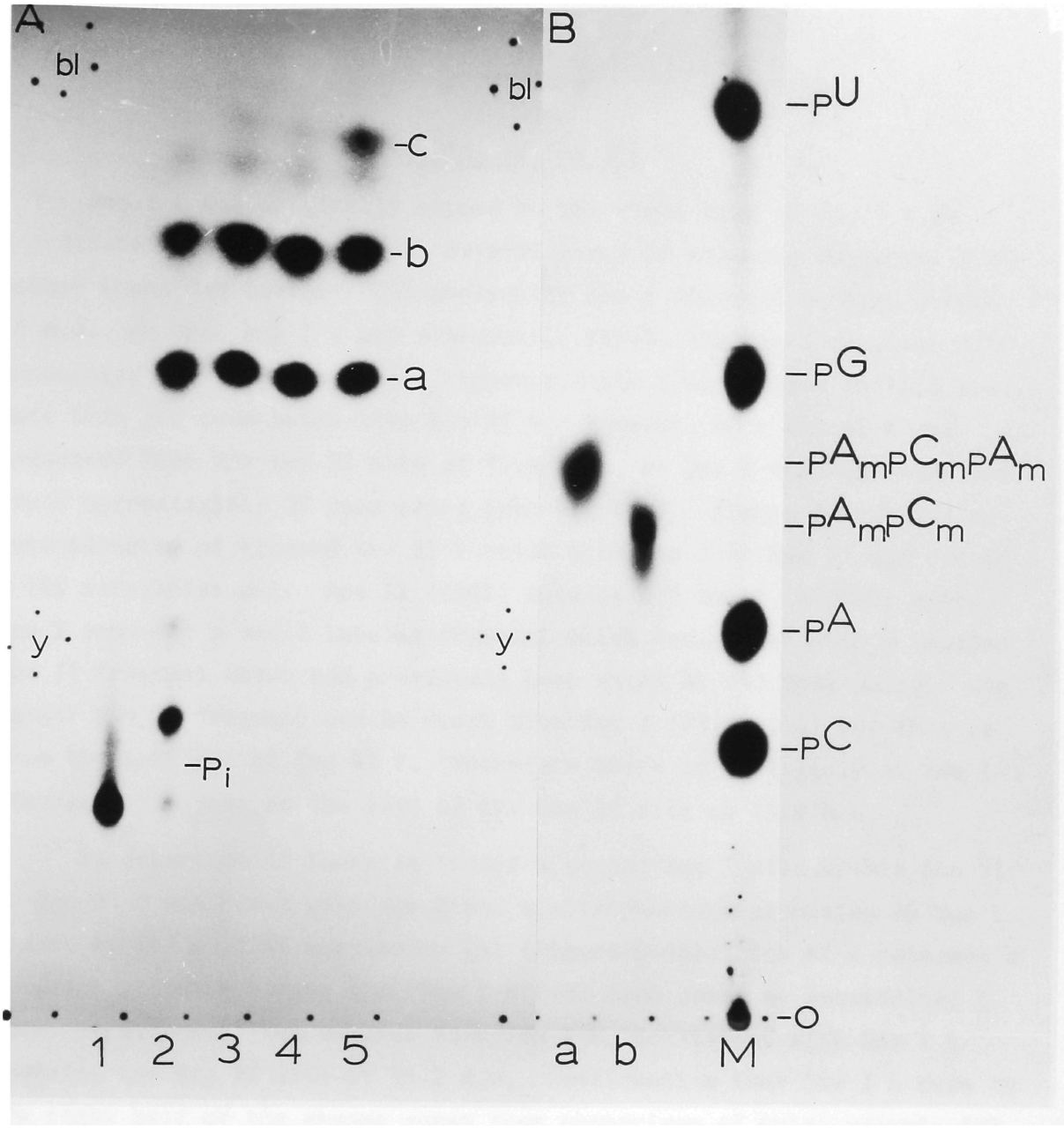


Figure A-3. Susceptibility of Phosphodiester Bonds With and Without 2'-O-Methylation to Cleavage with Nuclease P1

(A) In lanes 1-5, 5% of spot 1 ($pA_mC_{(m)}$AUCUGp) from Figure 12D was digested with 0, 0.0125, 0.125, 0.25 and 1.25 Y nuclease P1 at pH 6.0 and 26°C for 30 minutes. Each sample was spotted on a DEAE plate and developed by homochromatography. The first major spot in lane 2 is $pA_mC_{(m)}$AUCUG produced by the 3' phosphatase activity of nuclease P1.

(B) Products a and b from lane 5 of (A) were electrophoresed on 540 paper at pH 3.5 alongside 5' mononucleotide markers (lane M). They have mobilities similar to the faster two products in lanes 2, 3 and 8 of Figure A-1.

APPENDIX B

Mapping of Sma I L

Sma I L was originally mapped to the right side of Sma I F at coordinate 18 by Marc Zabeau. Several lines of evidence suggested that either there was another ~200 nucleotide Sma I fragment mapping around 76 m.u., or that Sma I L was mismapped. First, the Sma I cleavage site separating the two very large fragments A and C was mapped at 76.8 m.u., more than 300 base pairs into Eco RI D. However, when Eco RI F was sequenced from the Eco RI site at 75.9 m.u., an Sma I cleavage site was found approximately 35 base pairs into Eco RI F. Figure B-1A shows an autoradiogram of kinased Eco RI F recut with Sma I or Hpa II and run on a 10% acrylamide gel. Hpa II (CCGG) cuts at all Sma I (CCCGGG) sites. Sma I produces a small labeled fragment which comigrates with a labeled Hpa II fragment which had previously been sized at <40 base pairs. The larger Hpa II fragment can be recut with Kpn I (71.4 m.u.) and thus is from the left end of Eco RI F. Therefore there is definitely an Sma I cleavage site just to the left of the Eco RI site at 75.9 m.u.

To determine if there is indeed a second Sma I site within Eco RI D, Eco RI D was recut with Sma I and electrophoresed alongside an Sma I digest of Ad2 on a 6% acrylamide gel (Figure B-1B). Eco RI D releases a fragment slightly larger than Sma I M (180 base pairs by sequencing; E. Ziff, pers. comm.) but smaller than Sma I L, consistent with Sma I L spanning the Eco RI site at 75.9 m.u. Confirmation that Sma I L maps in the right half of the genome comes from comparison of total genomic Ad2 DNA and Eco I A (0-59 m.u.) cut with Sma I (Figure B-1C). Both digests produce Sma I M, but only total genomic Ad2 DNA releases Sma I L.

Sma I L was sequenced completely and shown to match the sequences flanking the Eco RI site at 75.9 and maps from 75.8-76.4 m.u.

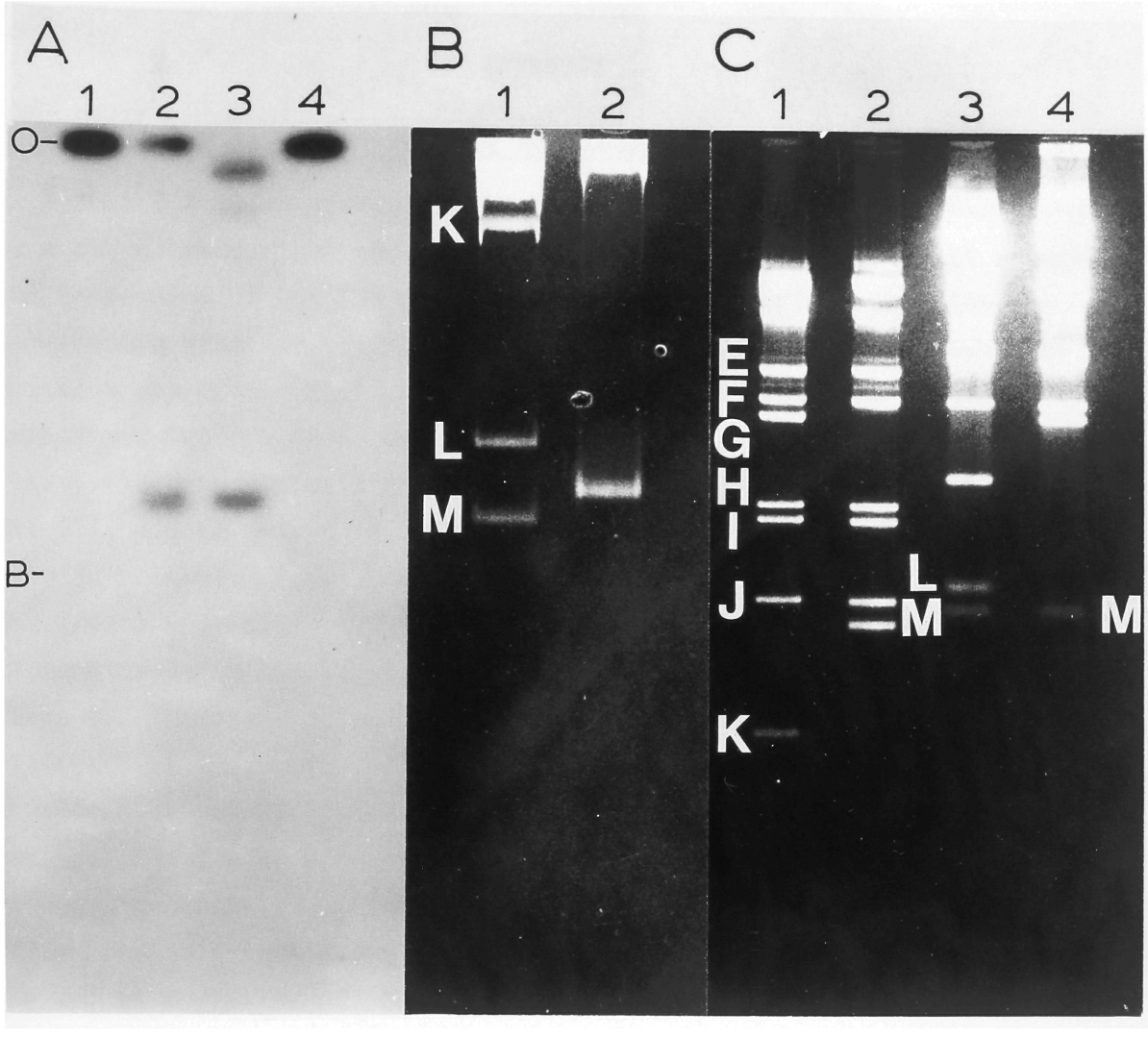


Figure B-1. Mapping of Sma I L

(A) Eco RI F fragment was kinased and electrophoresed on a 10% acrylamide gel without treatment (lanes 1 and 4), or cleaved with Sma I (lane 2) or Hpa II (lane 3). B indicates the position of the bromophenol blue dye marker.

(B) Total Ad2 DNA (lane 1) and Eco RI D DNA (lane 2) were digested with Sma I and electrophoresed on a 6% acrylamide gel. The Sma I K, L and M fragments are indicated.

(C) Total Ad2 DNA (lanes 1 and 3) and Eco RI A DNA (lanes 2 and 4) were digested with Sma I and electrophoresed on a 1.4% agarose gel. Lanes 1 and 2 were electrophoresed longer than lanes 3 and 4.

DNA fragments were visualized in (A) by autoradiography and in (B) and (C) by ethidium bromide and UV light.

APPENDIX C

Restriction Mapping and Sequencing of Sma I K

Sma I K is an approximately 600 base pair DNA fragment which comes from the extreme right end of the genome. On one strand (l) it encodes the 5' termini of the EIV mRNAs and on the other (r) it contains the termination site for the major late transcription unit. A rough restriction map of the right end of the genome was available (Roberts, Arrand and Keller, 1974) and showed the following order of restriction sites, starting at the right terminus and going leftward: Hha I, Hph I, Hae III, Hpa II and Hpa I. To map these restriction sites more accurately, unlabeled Sma I K was digested with Hpa II, Hae III, Hph I or Hha I or a combination of these enzymes and run on a 6% acrylamide gel (Figure C-1). A Hpa II digest of Eco RI F was run alongside for markers (lane 9), the sizes of which were calculated from the Eco RI F sequence (Galibert et al., 1979). The DNA fragments were visualized with ethidium bromide and UV light and are shown in Figure C-1. The double digest with Hha I plus Hph I was also done, but is not shown. The size of each DNA fragment was calculated from a semilog plot of base pairs versus distance (not shown) and is presented in Table C-I. Note that the sum of the sizes of the fragments produced by each digest is between 580 and 595 base pairs, giving an estimate of the size of the Sma I K fragment. When Sma I K was kinased and recut with Hae III, the 320 base pair fragment had only half the radioactivity of the 260 base pair fragment. Since 50-80% of the Ad2 genome 5' termini are resistant to kinasing (Carusi, 1977), this tentatively identifies the 320 base pair Hae III fragment as the terminal fragment. Further recutting of the 320 base pair kinased Hae III fragment with Hha I and Hph I gave labeled 75 and 100 base pair fragments, respectively (data not shown), indicating that there are Hha I and Hph I sites 75 and 100 base pairs, respectively, from the right terminus (consistent with the map of Roberts et al., 1974). The only map consistent with these data is shown in Figure C-2A.

Several other restriction sites were mapped using partial

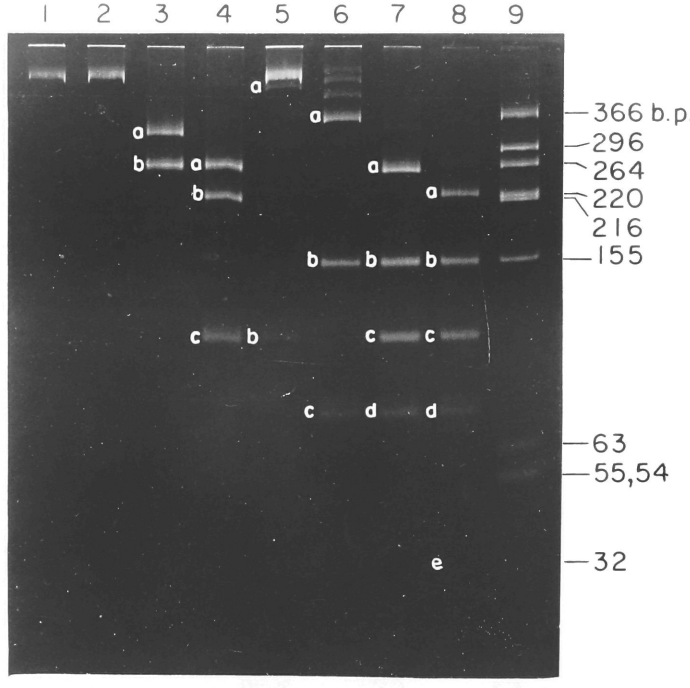


Figure C-1. Restriction Endonuclease Mapping of Unlabeled Sma I K

Pure Sma I K fragment was digested with Hpa II (lane 2), Hae III (lane 3), Hae III plus Hph I (lane 4), Hph I (lane 5), Hha I (lane 6), Hae III plus Hha (lane 7), and Hae III plus Hph I plus Hha I (lane 8). Untreated Sma I K was run in lane 1. Eco RI F digested with Hpa II was run in lane 9 for size markers (sizes were calculated from the sequence of Galibert et al. (1979)). Fragments in each lane were named in order of decreasing size and have been listed in Table C-I with their calculated sizes.

Table C-I. Sizes of Sma I K Restriction Subfragments

	Fragments[a]					
Enzymes	a	b	c	d	e	Total
Hae III	320	260				580
Hph I	480	105				595
Hha I	350	155	75			580
Hae III + Hph I	260	230	105			595
Hae III + Hha I	260	155	105	75		595
Hha I + Hph I	320	155	75	30		580
Hae III + Hph I + Hha I	230	155	105	75	30	595

a. These fragments are those produced by digestion of Sma I K with the corresponding enzymes, as shown in Figure C-1. Fragments have been named alphabetically in order of decreasing size. Fragment sizes are given in nucleotides.

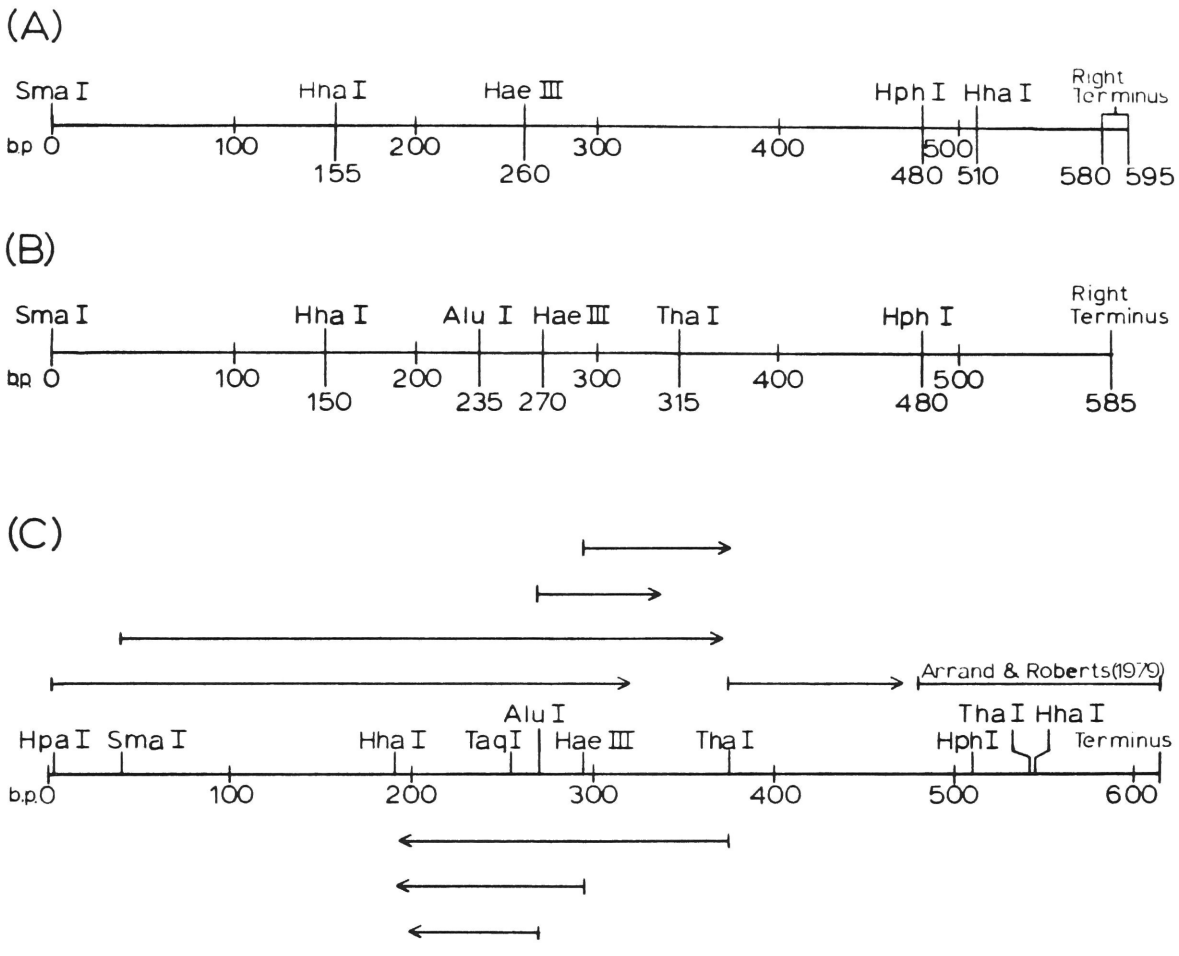


Figure C-2. Sma I K Restriction Endonuclease Cleavage Maps and Sequencing Strategy.

(A) shows the restriction endonuclease cleavage map derived from the experiment shown in Figure C-1 and Table C-I.

(B) shows the restriction endonuclease cleavage map derived from the experiment shown in Figure C-3.

(C) shows the strategy used to sequence the Hpa I G fragment. Rightward arrows represent sequence data for the l strand and leftward arrows for the r strand. The terminal repeat was not sequenced.

restriction digests of end labeled Sma I K by the method of Smith and Birnstiel (1976). An Sma I digest of Ad2 virion DNA was kinased, and the Sma I K fragment was recut with Hph I to give a fragment labeled on only one end. The terminal 105 base pair Hph I fragment was found to be totally unlabeled under these conditions, however, so the recutting was unnecessary. The labeled Hph I fragment was partially digested with various restriction enzymes, run on a 5% acrylamide gel with SV40/Hind III/Kpn I size markers, and autoradiographed (Figure C-3). The size of each fragment was once again determined by a semilog plot. Since each fragment is labeled at the Sma I site, the order of restriction enzyme cut sites and their exact positions were determined from the fragment sizes, and are shown in the map in Figure C-2B. The Hph I site was calculated to be 480 nucleotides from the Sma I site. Adding this to the 105 base pairs of the small Hph I fragment gives a 585 base pair estimate for the size of Sma I K. There appear to be no cut sites for Mbo I, Eco RII, Hinf I, Hinc II or Pst I. Cut sites to the right of the Hph I site will not be detected, of course, by this analysis.

The sequence of the Hpa I G fragment, which contains the Sma I K fragment, was determined using 5' labeling and Maxam-Gilbert chemistries (Maxam and Gilbert, 1977). The strategy used in sequencing is shown in Figure C-2C. The long and short 8% acrylamide thin gels showing the sequence of the r strand in the region of the EIV TATA box, cap site, and first leader are shown in Figure C-4. The sequence of Hpa I G is shown in Figure C-5. The sequence of nucleotides 482-615 is from Arrand and Roberts (1979). The ten nucleotides from 472-481 were not identified because of exonuclease present in the Hph I used to recut the Tha I fragment. The sequence of Ad2 Sma I K has subsequently been reported by Shinagawa, Padmanabhan and Padmanabhan (1980).

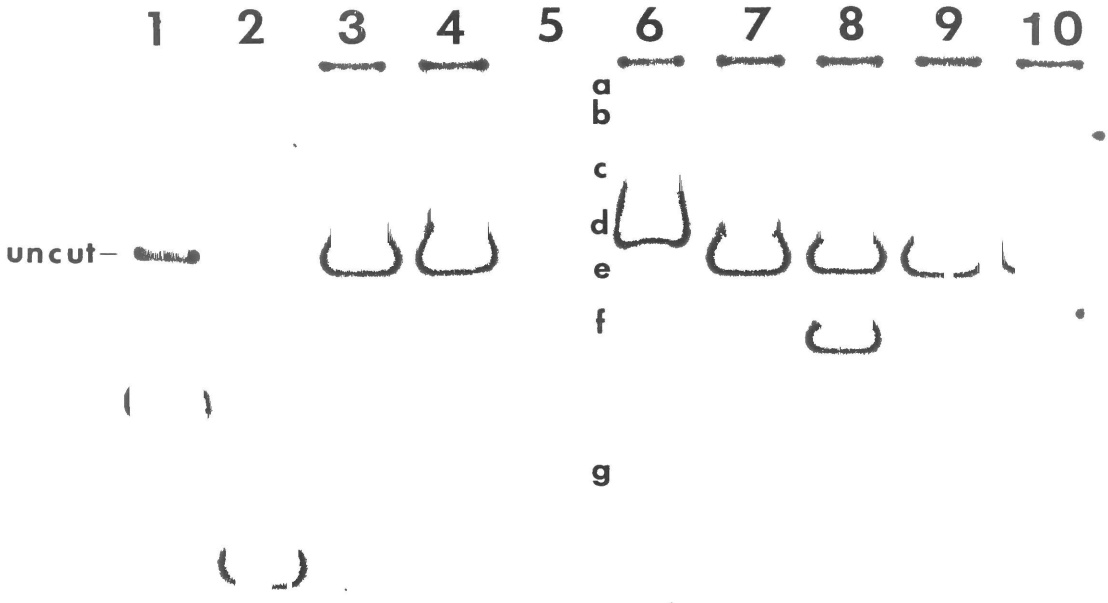


Figure C-3. Restriction Endonuclease Mapping by Partial Digestion of 5' Labeled DNA

Sma I K fragment was 5' labeled and recut with Hph I to produce a DNA fragment labeled only at the Sma I cleavage site (98.3 m.u.). This fragment was then digested with Hae III (lane 1), Hha I (lane 2), Alu I (lane 3), Mbo I (lane 4), Eco RII (lane 6), Hinf I (lane 7), Tha I (lane 8), Hinc II (lane 9) and Pst I (lane 10) under conditions such that digestion would give partials (Smith and Birnstiel, 1976). Kinased SV40 DNA digested with Hind III and Kpn I was run in lane 5 for size markers. The sizes of fragments a-g in lane 5 are 1768, 1169, 752, 526, 447, 349 and 215, respectively (Reddy et al., 1978). The restriction map derived from this experiment is shown in Figure C-2B.

Figure C-4. DNA Sequencing Gels

Sma I K was digested with Tha I, kinased, and recut with Hha I. The fragment encoding the EIV 5' terminus and first leader was subjected to the Maxam-Gilbert chemistries (Maxam and Gilbert, 1977, 1980) and electrophoresed on 8% acrylamide-8M urea thin gels. Lanes marked G, AG, AC, CT and C have the cleavage specificities G, A+G, A>C, C+T and C, respectively. The sequence of the EIV TATA box and the site encoding the EIV 5' termini are indicated in (A), the shorter electrophoretic run. The sequence encoding the splice donor site is indicated in (B), the longer electrophoretic run.

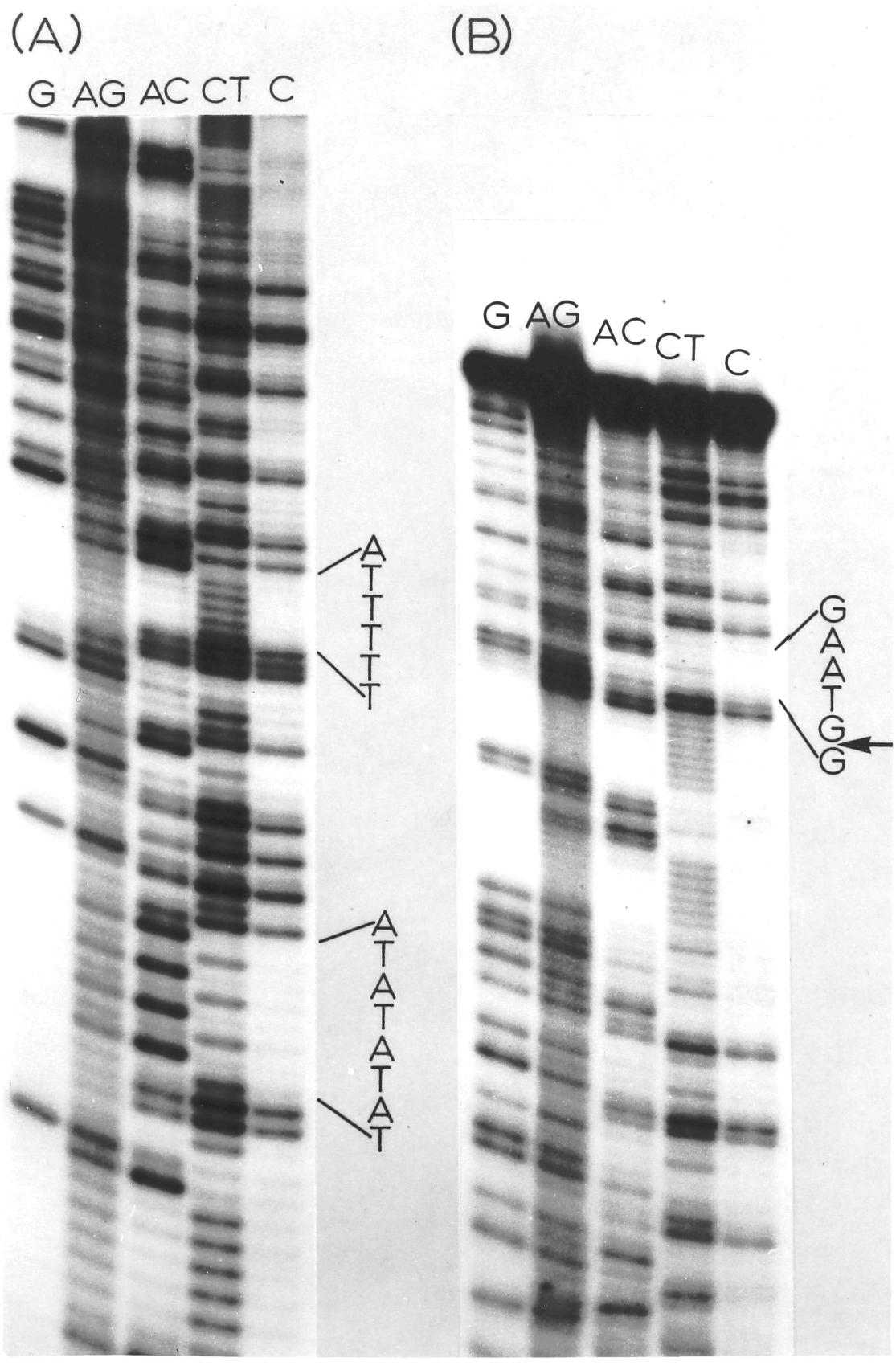
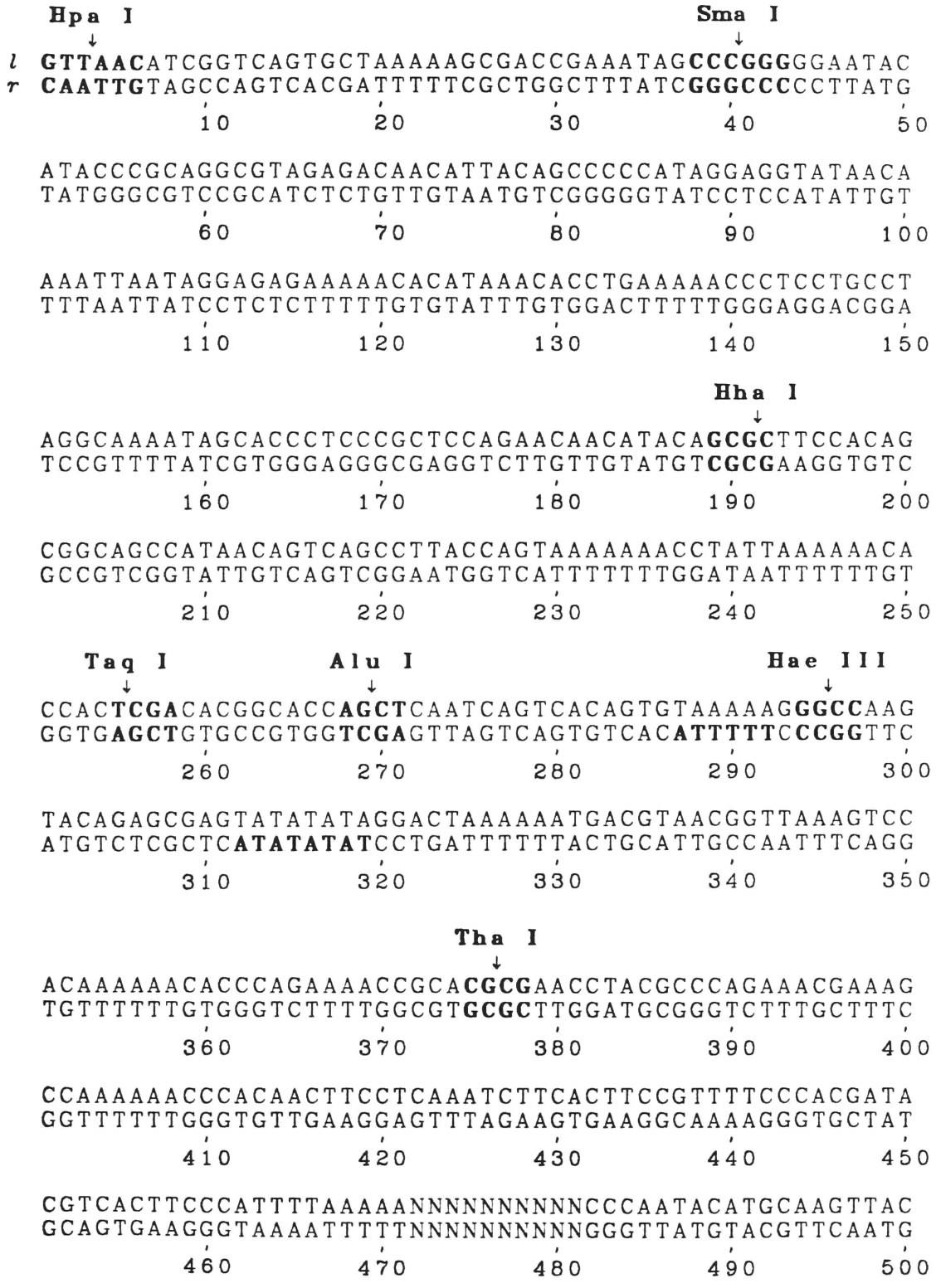
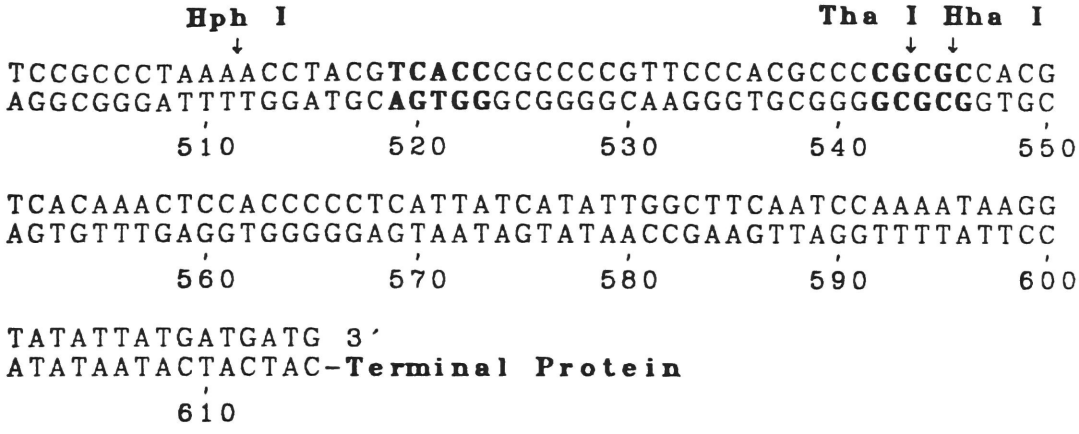

(A)
G AG AC CT C
ATTTTT
ATATAT
(B)
G AG AC CT C
GAATGG

Figure C-5. DNA Sequence of the Ad2 Hpa I G Fragment

The DNA sequence from 98.2 to 99.6 m.u. was determined by the strategy shown in Figure C-2C. The sequence of the terminal repeat (nucleotides 482-615) is taken from Arrand and Roberts (1979). Nucleotides 472-481 were not determined due to exonuclease contamination of the Hph I used in recutting. The EIV 5' termini are encoded at nucleotides 287-292 on the l strand.

The DNA sequence of the Sma I K fragment has also been determined by Shinagawa et al. (1980).


Hpa I
Sma I
l GTTAACATCGGTCAGTGCTAAAAAGCGACCGAAATAGCCCGGGGGAATAC
r CAATTGTAGCCAGTCACGATTTTTCGCTGGCTTTATCGGGCCCCCTTATG
10 20 30 40 50
ATACCCGCAGGCGTAGAGACAACATTACAGCCCCCATAGGAGGTATAACA
TATGGGCGTCCGCATCTCTGTTGTAATGTCGGGGGTATCCTCCATATTGT
60 70 80 90 100
AAATTAATAGGAGAGAAAAACACATAAACACCTGAAAAACCCTCCTGCCT
TTTAATTATCCTCTCTTTTTGTGTATTTGTGGACTTTTTGGGAGGACGGA
110 120 130 140 150
Hha I
AGGCAAAATAGCACCCTCCCGCTCCAGAACAACATACAGCGCTTCCACAG
TCCGTTTTATCGTGGGAGGGCGAGGTCTTGTTGTATGTCGCGAAGGTGTC
160 170 180 190 200
CGGCAGCCATAACAGTCAGCCTTACCAGTAAAAAAACCTATTAAAAAACA
GCCGTCGGTATTGTCAGTCGGAATGGTCATTTTTTTGGATAATTTTTTGT
210 220 230 240 250
Taq I
Alu I
Hae III
CCACTCGACACGGCACCAGCTCAATCAGTCACAGTGTAAAAAGGGCCAAG
GGTGAGCTGTGCCGTGGTCGAGTTAGTCAGTGTCACATTTTTCCCGGTTC
260 270 280 290 300
TACAGAGCGAGTATATATAGGACTAAAAAATGACGTAACGGTTAAAGTCC
ATGTCTCGCTCATATATATCCTGATTTTTTACTGCATTGCCAATTTCAGG
310 320 330 340 350
Tha I
ACAAAAAACACCCAGAAAACCGCACGCGAACCTACGCCCAGAAACGAAAG
TGTTTTTTGTGGGTCTTTTGGCGTGCGCTTGGATGCGGGTCTTTGCTTTC
360 370 380 390 400
CCAAAAAACCCACAACTTCCTCAAATCTTCACTTCCGTTTTCCCACGATA
GGTTTTTTGGGTGTTGAAGGAGTTTAGAAGTGAAGGCAAAAGGGTGCTAT
410 420 430 440 450
CGTCACTTCCCATTTTAAAAANNNNNNNNNNCCCAATACATGCAAGTTAC
GCAGTGAAGGGTAAAATTTTTNNNNNNNNNNGGGTTATGTACGTTCAATG
460 470 480 490 500


Hph I
Tha I
Hha I
TCCGCCCTAAAACCTACGTCACCCGCCCCGTTCCCACGCCCCGCGCCACG
AGGCGGGATTTTGGATGCAGTGGGCGGGGCAAGGGTGCGGGGCGCGGTGC
510
520
530
540
550
TCACAAACTCCACCCCCTCATTATCATATTGGCTTCAATCCAAAATAAGG
AGTGTTTGAGGTGGGGGAGTAATAGTATAACCGAAGTTAGGTTTTATTCC
560
570
580
590
600
TATATTATGATGATG 3′
ATATAATACTACTAC-Terminal Protein
610

APPENDIX D

Analysis of the Mouse β-Globin mRNA 5' Terminus

Since 5' terminal microheterogeneity seems to be the rule in adenovirus-2, I wanted to examine a cellular mRNA for 5' microheterogeneity. I picked mouse β-globin major for several reasons. The mouse β-major gene has been cloned and sequenced (Konkel, Tilghman and Leder, 1978). Thus specific DNA probes could be obtained for isolation of the 5' terminus of the β-globin mRNA. The 5' terminus of the mouse β-globin major mRNA has been sequenced by reverse transcription (Baralle and Brownlee, 1978). The sequence at the cap site is similar to that of Ad2 EIb, so a second minor A terminus might be expected. The β-globin mRNA could be obtained in fairly large amounts from cells grown in culture since β-globin mRNA is 0.1% of total mRNA in Friend erythroleukemia cells 3 days after induction with DMSO (Friend, et al., 1971; E. Hofer, pers. comm.).

Total polyadenylated cytoplasmic RNA from 10^9 Friend cells 3 days after induction was chemically decapped, phosphatased and kinased. The 5' labeled β-globin major mRNA was isolated by hybridization to a cloned Alu I fragment encoding the first ~165 nucleotides of the mRNA. The T1 fingerprint of this hybrid RNA is shown in Figure D-1A. The major spot (indicated by the thick arrow) was analyzed with nuclease P1 (panel B) and RNAse T2 and RNAse A (panel C) and shown to be $pA_mC_{(m)}AU-$. Although RNAse T2 partial fingerprinting was not done, this spot is most likely the predicted $pA_mC_{(m)}AUUUGp$ β-major 5' terminal oligonucleotide. A minor spot (thin arrow), possibly $pA_mC_{(m)}AC-$ (data not shown), has 2'-O-methylations and thus also comes from an mRNA 5' terminus. It cannot be encoded within this cap region and may represent the 5' terminus of a cross-hybridizing globin mRNA. No spot is seen migrating with the mobility expected for $pA_mU_{(m)}UUGp$ (just below the yellow dye). Therefore the β-major globin mRNA 5' terminus does not appear to be microheterogeneous.

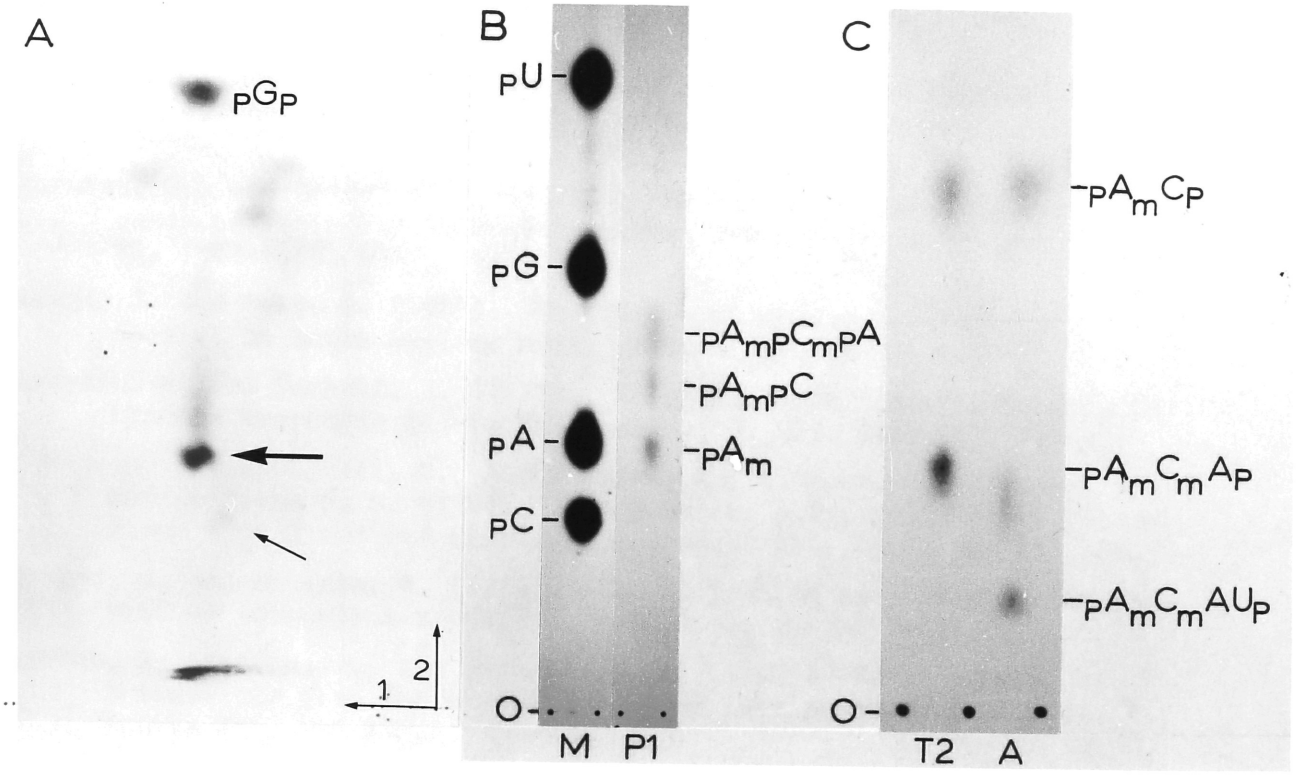


Figure D-1. Analysis of the Mouse β-Globin mRNA 5' Terminus

(A) Total cytoplasmic polyadenylated RNA from 10^9 Friend cells three days post-induction with DMSO was 5' labeled in vitro, hybridized to a cloned Alu I fragment of mouse DNA encoding the first 165 nucleotides of the β-globin mRNA, eluted and T1 fingerprinted. The thick arrow indicates the major 5' terminal oligonucleotide ($pA_mC_{(m)}AUUUGp$) analyzed further in panels (B) and (C). The thin arrow indicates a very minor 5' terminal oligonucleotide, perhaps $pA_mC_{(m)}AC$-.

In (B) and (C), the major 5' terminal oligonucleotide of (A) was digested with nuclease P1 and RNAse T2 and A, respectively. Lane M in (B) indicates marker 5' mononucleotides. Enzymes are indicated below lanes and products beside the panels. (B) is electrophoresis on 540 paper and (C) on DEAE paper, both at pH 3.5.

REFERENCES

Abelson, J. (1979). RNA processing and the intervening sequence problem. Ann. Rev. Biochem. 48, 1035-1069.

Abraham, G., Rhodes, D., and Banerjee, A. (1975). Novel initiation of RNA synthesis in vitro by vesicular stomatitis virus. Nature 255, 37-40.

Abraham, G., and Banerjee, A. (1976). Sequential transcription of the genes of vesicular stomatitis virus. Proc. Natl. Acad. Sci. USA 73, 1504-1508.

Adams, J. and Cory, S. (1975). Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. Nature 255, 28-33.

Adesnik, M. and Darnell, J. (1972). Biogenesis and characterization of histone messenger RNA in HeLa cells. J. Mol. Biol. 67, 397-406.

Alestrom, P., Akusjarvi, G., Perricaudet, M., Mathews, M., Klessig, D., and Pettersson, U. (1980). The gene for polypeptide IX of adenovirus type 2 and its unspliced messenger RNA. Cell 19, 671-681.

Arrand, J. and Roberts, R. (1979). The nucleotide sequences at the termini of adenovirus-2 DNA. J. Mol. Biol. 128, 577-594.

Babich, A., Nevins, J., and Darnell, J. E. (1980). Early capping of transcripts from the adenovirus major late transcription unit. Nature 287, 246-248.

Baker, C.C. and Ziff, E.B. (1980). Biogenesis, structures, and sites of encoding of the 5' termini of adenovirus-2 mRNAs. Cold Spring Harbor Symp. Quant. Biol. 44, 415-428.

Baralle, F. (1977a). Complete nucleotide sequence of the 5' noncoding region of human α- and β-globin mRNA. Cell 12, 1085-1095.

Baralle, F. (1977b). Complete nucleotide sequence of the 5' noncoding region of rabbit β-globin mRNA. Cell 10, 549-558.

Baralle, F. and Brownlee, G. (1978). AUG is the only recognizable signal sequence in the 5' non-coding regions of eukaryotic mRNA. Nature 274, 84-87.

Bastos, R. and Aviv, H. (1977). Globin RNA precursor molecules: Biosynthesis and processing in erythroid cells. Cell 11, 641-650.

Bendig, M., Thomas, T. and Folk, W. (1980). Regulatory mutants of polyoma virus defective in DNA replication and the synthesis of early proteins. Cell 20, 401-409.

Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980). The ovalbumin gene--sequence of putative control regions. Nucleic Acids Res. 8, 127-142.

Benoist, C. and Chambon, P. (1980). Deletions covering the putative promoter region of early mRNAs of simian virus 40 do not abolish T-antigen expression. Proc. Natl. Acad. Sci. USA 77, 3865-3869.

Benoist, C. and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature 290, 304-310.

Berget, S., Moore, C., and Sharp, P. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. USA 74, 3171-3175.

Berk, A., Lee, F., Harrison, T., Williams, J. and Sharp, P. (1979). Pre-early adenovirus 5 gene product regulates synthesis of early viral messenger RNAs. Cell 17, 935-944.

Berk, A. and Sharp, P. (1977a). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12, 721-732.

Berk, A. and Sharp, P. (1977b). Ultraviolet mapping of the adenovirus-2 early promoters. Cell 12, 45-55.

Berk, A. and Sharp, P. (1978). Structure of the adenovirus 2 early mRNAs. Cell 14, 695-711.

Blanton, R. and Carter, T. (1979). Autoregulation of adenovirus type 5 early gene expression. III. Transcription studies in isolated nuclei. J. Virol. 29, 458-465.

Brawerman, G. (1976). Characteristics and significance of the polyadenylate sequence in mammalian messenger RNA. Prog. Nuc. Ac. Res. and Mol. Biol. 17, 117-148.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eukaryotic protein-coding nuclear split genes. Ann. Rev. Biochem., in press.

Brownlee, G. (1972). Determination of sequences in RNA. In Laboratory Techniques in Biochemistry and Molecular Biology, 3, Part 1, T.S. Work and E. Work, eds. (Amsterdam: North-Holland/ American Elsevier).

Canaani, D., Kahana, C., Mukamel, A. and Groner, Y. (1979). Sequence heterogeneity at the 5' termini of late simian virus 40 19S and 16S mRNAs. Proc. Natl. Acad. Sci. USA 76, 3078-3082.

Carusi, E. (1977). Evidence for blocked 5' termini in human adenovirus DNA. Virology 76, 380-394.

Chambon, P. (1975). Eukaryotic nuclear RNA polymerases. Ann. Rev. Biochem. 44, 613-638.

Chambon, P. (1978). Summary: The molecular biology of the eukaryotic genome is coming of age. Cold Spring Harbor Symp. Quant. Biol. 42, 1209-1234.

Chang, J., Temple, G., Poon, R., Neumann, K. and Kan, Y.W. (1977). The nucleotide sequences of the untranslated 5' regions of human α- and β-globin mRNAs. Proc. Natl. Acad. Sci. USA 74, 5145-5149.

Chardonnet, Y., Gazzolo, L. and Pogo, B. (1972). Effect of α-amanitin on adenovirus 5 multiplication. Virology 48, 300-304.

Chow, L., Broker, T., and Lewis, J. (1979). Complex splicing patterns of RNA from the early regions of adenovirus-2. J. Mol. Biol. 134, 265-303.

Chow, L., Gelinas, R., Broker, T., and Roberts, R. (1977a). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell 12, 1-8.

Chow, L., Lewis, J., and Broker, T. (1980). RNA transcription and splicing at early and intermediate times after adenovirus-2 infection. Cold Spring Harbor Symp. Quant. Biol. 44, 401-414.

Chow, L., Roberts, J., Lewis, J., and Broker, T. (1977b). A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. Cell 11, 819-836.

Chow, L. and Broker, T. (1978). The spliced structures of adenovirus-2 fiber message and other late mRNAs. Cell 15, 497-510.

Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. and Chambon, P. (1979). Organization and sequence studies of the 17-piece chicken conalbumin gene. Nature 282, 567-574.

Contreras, R. and Fiers, W. (1981). Initiation of transcription by RNA polymerase II in permeable, SV40-infected or noninfected, CV1 cells; evidence for multiple promoters of SV40 late transcription. Nucleic Acids Res. 9, 215-236.

Corden, J., Engelking, H. and Pearson, G. (1976). Chromatin-like organization of the adenovirus chromosome. Proc. Natl. Acad. Sci. USA 73, 401-404.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980). Promoter sequences of eukaryotic protein-coding genes. Science 209, 1406-1414.

Cory, S. and Adams, J. (1975). The modified 5'-terminal sequences in messenger RNA of mouse myeloma cells. J. Mol. Biol. 99, 519-547.

Curtis, P., Mantei, N. and Weissmann, C. (1978). Characterization and kinetics of synthesis of 15S β-globin RNA, a putative precursor of β-globin mRNA. Cold Spring Harbor Symp. Quant. Biol. 42, 971-984.

Darnell, J. (1979). Transcription units for mRNA production in eucaryotic cells and their DNA viruses. Prog. in Nucleic Acid Res. and Molecular Biol. 22, 327-353.

Desrosiers, R., Friderici, K. and Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff Hepatoma cells. Proc. Natl. Acad. Sci. USA 71, 3971-3975.

Doerfler, W. (1969). Nonproductive infection of baby hamster kidney cells (BHK21) with adenovirus type 12. Virology 38, 587-606.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. Cell 20, 313-319.

Efstratiadis, A., Posakony, J., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R., DeRiel, J., Forget, B., Weissman, S., Slightom, J., Blechl, A., Smithies, O., Baralle, F., Shoulders, C. and Proudfoot, N. (1980). The structure and evolution of the human β-globin gene family. Cell 21, 653-668.

Eggerding, F. and Raskas, H. (1978). Effect of protein synthesis inhibitors on viral mRNAs synthesized early in adenovirus type 2 infection. J. Virol. 25, 453-458.

Ensinger, M., Martin, S., Paoletti, E., and Moss, B. (1975). Modification of the 5'-terminus of mRNA by soluble guanylyl and methyl transferases from vaccinia virus. Proc. Natl. Acad. Sci. USA 72, 2525-2529.

Evans, R., Fraser, N., Ziff, E., Weber, J., Wilson, M., and Darnell, J. (1977). The initiation sites for RNA transcription in Ad2 DNA. Cell 12, 733-739.

Faye, G., Leung, D., Tatchell, K., Hall, B. and Smith, M. (1981). Deletion mapping of sequences essential for in vivo transcription of the iso-1-cytochrome c gene. Proc. Natl. Acad. Sci. USA 78, 2258-2262.

Felsenfeld, G. (1978). Chromatin. Nature 271, 115-122.

Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G., and Ysebaert, M. (1978). Complete nucleotide sequence of SV40 DNA. Nature 273, 113-120.

Flavell, A., Cowie, A., Arrand, J. and Kamen, R. (1980). Localization of three major capped 5' ends of polyoma virus late mRNAs within a single tetranucleotide sequence in the viral genome. J. Virol. 33, 902-908.

Flavell, A., Cowie, A., Legon, S. and Kamen, R. (1979). Multiple 5' terminal cap structures in late polyoma virus RNA. Cell 16, 357-371.

Flint, S. J. (1980). Structure and genomic organization of adenoviruses. In DNA Tumor Viruses, Part 2 (J. Tooze, ed.), pp.383-441, CSH Laboratory, New York.

Flint, S.J. and Broker, T. (1980). Lytic infection by adenoviruses. In DNA Tumor Viruses, part 2 (J. Tooze, ed.), pp. 443-546, CSH Laboratory, New York.

Flint, S.J. and Weintraub, H. (1977). An altered subunit configuration associated with the actively transcribed DNA of integrated adenovirus genes. Cell 12, 783-794.

Ford, J. and Hsu, M.-T. (1978). Transcription pattern of in vivo-labeled late simian virus 40 RNA: Equimolar transcription beyond the mRNA 3' terminus. J. Virol. 28, 795-801.

Fraser, N., Baker, C., Moore, M. and Ziff, E. (1981). The poly A sites of adenovirus type 2 early transcription units. Manuscript in preparation.

Fraser, N., Nevins, J., Ziff, E. and Darnell, J. (1979). The major late adenovirus type-2 transcription unit: Termination is downstream from the last poly(A) site. J. Mol. Biol. 129, 643-656.

Fraser, N., Sehgal, P. and Darnell, J. (1978). DRB-induced premature termination of late adenovirus transcription. Nature 272, 590-593.

Fraser, N. and Ziff, E. (1978). RNA structures near poly(A) of adenovirus-2 late messenger RNAs. J. Mol. Biol. 124, 27-51.

Freeman, A., Black, P., Vanderpool, E., Henry, P., Austin, J. and Huebner, R. (1967). Transformation of primary rat embryo cells of adenovirus type 2. Proc. Natl. Acad. Sci. USA 58, 1205-1212.

Friend, C., Scher, W., Holland, J. and Sato, T. (1971). Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: Stimulation of erythroid differentiation by dimethyl sulfoxide. Proc. Natl. Acad. Sci. USA 68, 378-382.

Fujimoto, M., Kuninaka, A. and Yoshino, H. (1969). Specificity of a nuclease from Penicillium citrinum. Agric. Biol. Chem. 33, 1517-1518.

Furuichi, Y. (1978). "Pretranscriptional capping" in the biosynthesis of cytoplasmic polyhedrosis virus mRNA. Proc. Natl. Acad. Sci. USA. 75, 1086-1090.

Furuichi, Y., Muthukrishnan, S., Tomasz, J., and Shatkin, A. (1976). Mechanism of formation of reovirus mRNA 5'-terminal blocked and methylated sequence, $^{7m}GpppG_{m}pC$. J. Biol. Chem. 251, 5043-5053.

Furuichi, Y., and Shatkin, A. (1976). Differential synthesis of blocked and unblocked 5'-termini in reovirus mRNA: Effect of pyrophosphate and pyrophosphatase. Proc. Natl. Acad. Sci. USA 73, 3448-3452.

Furuichi, Y., and Shatkin, A. (1977). 5'-Termini of reovirus mRNA: Ability of viral cores to form caps post-transcriptionally. Virology 77, 566-578.

Galibert, F., Herisse, J., and Courtois, F. (1979). Nucleotide sequence of the Eco RI F fragment of adenovirus 2 genome. Gene 6, 1-22.

Galibert, F., Sedat, J., and Ziff, E. (1974). Direct determination of DNA nucleotide sequences: Structure of a fragment of bacteriophage øX174 DNA. J. Mol. Biol. 87, 377-407.

Galos, R., Williams, J., Binger, M. and Flint, S. (1979). Location of additional early gene sequences in the adenovirus chromosome. Cell 17, 945-956.

Gannon, F., O'Hare, K., Perrin, F., LePennec, J., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979). Organization and sequences at the 5' end of a cloned

complete ovalbumin gene. Nature 278, 428-434.

Garel, A. and Axel, R. (1976). Selective digestion of transcriptionally active avalbumin genes from oviduct nuclei. Proc. Natl. Acad. Sci. USA 73, 3966-3970.

Gelinas, R. and Roberts, R. (1977). One predominant 5'-undecanucleotide in adenovirus 2 late messenger RNAs. Cell 11, 533-544.

Ghosh, P., Lebowitz, P., Frisque, R. and Gluzman, Y. (1981). Identification of a promoter component involved in positioning the 5' termini of simian virus 40 early mRNAs. Proc. Natl. Acad. Sci. USA 78, 100-104.

Ghosh, P., Reddy, V., Swinscoe, J., Choudary, P., Lebowitz, P., and Weissman, S. (1978a). The 5'-terminal leader sequence of late 16S mRNA from cells infected with Simian Virus 40. J. Biol. Chem. 253, 3643-3647.

Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P., and Weissman, S.M. (1978b). The heterogeneity and 5'-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. 126, 813-846.

Gidoni, D., Kahana, C., Canaani, D. and Groner, Y. (1981). Specific *in vitro* initiation of transcription of simian virus 40 early and late genes occurs at the various cap nucleotides including cytidine. Proc. Natl. Acad. Sci. USA 78, 2174-2178.

Gluzman, Y., Sambrook, J., and Frisque, R. (1980). Expression of early genes of origin-defective mutants of Simian Virus 40. Proc. Natl. Acad. Sci. USA 77, 3898-3902.

Goldberg, M. (1979). Ph.D. Thesis, Stanford University.

Goldberg, S., Weber, J., and Darnell, J. (1977). The definition of a large viral transcription unit late in Ad-2 infection of HeLa cells: mapping by effects of ultraviolet irradiation. Cell 10, 617-621.

Goldberger, R. (1979). Strategies of genetic regulation in prokaryotes. In *Biological Regulation and Development*, Vol. 1 (R. Goldberger, ed.) pp. 1-18, Plenum Press, New York.

Gray, M. and Lane, B. (1967). Studies of the sequence distribution of 2'-O-methylribose in yeast soluble ribonucleates. Biochim. Biophys. Acta 134, 243-257.

Green, M. (1970). Oncogenic viruses. Ann. Rev. Biochem. 39, 701-756.

Groner, Y., Gilboa, E. and Aviv, H. (1978). Methylation and capping of RNA polymerase II primary transcripts by HeLa nuclear homogenates. Biochemistry 17, 977-982.

Grosschedl, R. and Birnstiel, M. (1980a). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. USA 77, 1432-1436.

Grosschedl, R. and Birnstiel, M. (1980b). Spacer DNA sequences upstream of the TATAAATA sequence are essential for promotion of H2A histone gene transcription in vivo. Proc. Natl. Acad. Sci. USA 77, 7102-7106.

Gruss, P., Dhar, R. and Khoury, G. (1981). Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. USA 78, 943-947.

Guha, A. (1975). An exoribonuclease in bovine brain. Brain Research 83, 65-79.

Haegeman, G. and Fiers, W. (1978). Localization of the 5' terminus of late SV40 mRNA. Nucleic Acids Research 5, 2359-2371.

Haegeman, G. and Fiers, W. (1980). Characterization of the 5'-terminal cap structures of early simian virus 40 mRNA. J. Virol. 35, 955-961.

Hagenbuchle, O., Bovey, R. and Young, R. (1980). Tissue-specific expression of mouse α-amylase genes: Nucleotide sequence of isoenzyme mRNAs from pancreas and salivary gland. Cell 21, 179-187.

Hagenbuchle, O. and Schibler, U. (1981). Mouse β-globin and adenovirus-2 major late transcripts are initiated at the cap site in vitro. Proc. Natl. Acad. Sci. USA 78, 2283-2286.

Hamer, D. and Leder, P. (1979). Splicing and the formation of stable RNA. Cell 18, 1299-1302.

Handa, H., Kaufman, R., Manley, J., Gefter, M. and Sharp, P. (1981). Transcription of simian virus 40 DNA in a HeLa whole cell extract. J. Biol. Chem. 256, 478-482.

Hashimoto, S. and Green, M. (1976). Multiple methylated cap sequences in adenovirus type 2 early mRNA. J. Virol. 20, 425-435.

Hashimoto, S. and Green, M. (1980). Adenovirus 2 early messenger RNA-genome mapping of 5' terminal RNAse T1 oligonucleotides and heterogeneity of 5' termini. J. Biol. Chem. 255, 6780-6788.

Hendler, S., Furer, E., and Srinivasan, P. (1970). Synthesis and chemical properties of monomers and polymers containing 7-methylguanine and an investigation of their substrate or template properties for bacterial deoxyribonucleic acid or ribonucleic acid polymerase. Biochemistry 9, 4141-4153.

Herisse, J., Courtois, G. and Galibert, F. (1980). Nucleotide sequence of the Eco RI D fragment of adenovirus 2 genome. Nucleic Acids Res. 8, 2173-2192.

Hilleman, M. and Werner, J. (1954). Recovery of new agent from patients with acute respiratory illness. Proc. Soc. Exp. Biol. and Med. 85, 183-188.

Hofer, E. and Darnell, J. (1981). The primary transcription unit of the mouse β-major globin gene. Cell 23, 585-593.

Horwitz, M., Kaplan, L., Abboud, M., Maritato, J., Chow, L., and Broker, T. (1979). Adenovirus DNA replication in soluble extracts of infected cell nuclei. Cold Spring Harbor Symp. Quant. Biol. 43, 769-780.

Hu, S.-L. and Manley, J. (1981). DNA sequence required for initiation of transcription in vitro from the major late promoter of adenovirus 2. Proc. Natl. Acad. Sci. USA 78, 820-824.

Jacob, F., Ullman, A., and Monod, J. (1964). Le promoteur, element genetique necessaire a l'expression d'un operon. C. R. Acad. Sci. 258, 3125-3128.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318-356.

Jakobovits, E., Saragosti, S., Yaniv, M. and Aloni, Y. (1980). Escherichia coli RNA polymerase in vitro mimics simian virus 40 in vivo transcription when the template is viral nucleoprotein. Proc. Natl. Acad. Sci. USA 77, 6556-6560.

Jawetz, E., Kimurs, S., Nicholas, A., Thygeson, P. and Hanna, L. (1958). A new type of APC virus from epidemic kerato-conjunctivus. Science 122, 1190-1191.

Jones, C. and Kafatos, F. (1980). Structure, organization and evolution of developmentally regulated chorion genes in a silk moth. Cell 22, 855-867.

Jones, N. and Shenk, T. (1979). An adenovirus type 5 early gene function regulates expression of other early viral genes. Proc. Natl. Acad. Sci. USA 76, 3665-3669.

Kahana, C., Gidoni, D., Canaani, D. and Groner, Y. (1981). Simian virus 40 early mRNAs in lytically infected and transformed cells contain six 5'-terminal caps. J. Virol. 37, 7-16.

Kamen, R., Favaloro, J., Jat, P., Treisman, R. and Folk, W. (1980). The 5' ends of polyoma virus early region transcripts. CSH DNA Tumor Virus Abstracts, p. 51.

Kedes, L. (1979). Histone genes and histone messengers. Ann. Rev. Biochem. 48, 837-870.

Keene, M., Corces, V., Lowenhaupt, K. and Elgin, S. (1981). DNAse I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription. Proc. Natl. Acad. Sci. USA 78, 143-146.

Keith, J., Ensinger, M. and Moss, B. (1978). HeLa cell RNA (2'-o-methyladenosine-N^6-)-methyltransferase specific for the capped 5'-end of messenger RNA. J. Biol. Chem. 253, 5033-5039.

Kitchingman, G., Lai, S., and Westphal, H. (1977). Loop structures in hybrids of early RNA and the separated strands of adenovirus DNA. Proc. Natl. Acad. Sci. USA 74, 4392-4395.

Kitchingman, G. and Westphal, H. (1980). The structure of adenovirus 2 early nuclear and cytoplasmic RNAs. J. Mol. Biol. 137, 23-48.

Klessig, D. (1977). Two adenovirus mRNAs have a common 5' terminal leader sequence encoded at least 10 kb upstream from their main coding regions. Cell 12, 9-21.

Konkel, D., Tilghman, S. and Leder, P. (1978). The sequence of the chromosomal mouse β-globin major gene: Homologies in capping, splicing and poly(A) sites. Cell 15, 1125-1132.

Krakow, J., Rhodes, G. and Joviin, T. (1976). RNA polymerase: Catalytic mechanisms and inhibitors. In RNA Polymerase (Losick, R. and Chamberlin, M., eds.) pp. 127-157, Cold Spring Harbor Laboratory, New York.

Kvist, S., Ostberg, L., Persson, H., Philipson, L. and Peterson, P. (1978). Molecular association between transplantation antigens and cell surface antigen in adenovirus-transformed cell lines. Proc. Natl. Acad. Sci. USA 75, 5674-5678.

Lamb, R.A. and Lai, C. (1980). Sequence of interrupted and uninterrupted mRNAs and cloned full-length DNA coding for the two overlapping non-structural proteins of influenza virus. Cell 21, 475-485.

Ledinko, N. (1971). Inhibition by α-amanitin of adenovirus 12 replication in human embryo kidney cells and of adenovirus transformation of hamster cells. Nature New Biol. 233, 247-248.

Lee, D. and Roeder, R. (1981). Transcription of adenovirus 2 genes in a cell-free system: Apparent heterogeneity of initiation of some promoters. Mol. Cell. Biol., in press.

Legon, S., Flavell, A., Cowie, A. and Kamen, R. (1979). Amplification in the leader sequence of late polyoma virus mRNAs. Cell 16, 373-388.

LeMeur, M., Glanville, N., Mandel, J., Gerlinger, P., Palmiter, R. and Chambon, P. (1981). The ovalbumin gene family: Hormonal control of X and Y gene transcription and mRNA accumulation. Cell 23, 561-571.

Lerner, M., Boyle, J., Mount, S., Wolin, S. and Steitz, J. (1980). Are snRNPs involved in splicing? Nature 283, 220-224.

Lewis, J., Atkins, J., Anderson, C., Baum, P. and Gesteland, R. (1975). Mapping of late adenovirus genes by cell-free translation of RNA selected by hybridization to specific DNA fragments. Proc. Natl. Acad. Sci. USA 72, 1344-1348.

Lewis, J., Atkins, J., Baum, P., Solem, R., Gesteland, R., and Anderson, C. (1976). Location and identification of the genes for adenovirus type 2 early polypeptides. Cell 7, 141-151.

Lewis, J. and Mathews, M. (1980). Control of adenovirus early gene expression: a class of immediate early products. Cell 21, 303-313.

Liebhaber, S., Goossens, M. and Kan, Y.W. (1980). Cloning and complete nucleotide sequence of human 5'-α-globin gene. Proc. Natl. Acad. Sci. USA 77, 7054-7058.

Lindberg, U., Persson, T. and Philipson, L. (1972). Isolation and characterization of adenovirus messenger ribonucleic acid in productive infection. J. Virol. 10, 909-919.

Lockard, R. and RajBhandary, U. (1976). Nucleotide sequences at the 5' termini of rabbit alpha and beta globin mRNA. Cell 9, 747-760.

Luse, D, and Roeder, R. (1980). Accurate transcription initiation on a purified mouse β-globin DNA fragment in a cell-free system. Cell 20, 691-699.

Maat, J., Van Beveren, C. P., and Van Ormondt, H. (1980). The nucleotide sequence of adenovirus type 5 early region E1: the region between map positions 8.0 (HindIII site) and 11.8 (SmaI site). Gene 10, 27-38.

Maat, J. and Van Ormondt, A. (1979). The nucleotide sequences of the transforming HindIII-G fragment of adenovirus type 5 DNA. The region between positions 4.5 (HpaI site) and 8.0 (HindIII site). Gene 6, 75-90.

MacAllister, R., Nicholson, M., Lewis, A., Macpherson, I. and Huebner, R. (1969). Transformation of rat embryo cells by adenovirus type 1. J. Gen. Virol. 4, 29-36.

Malek, L.T., Eschenfeldt, W.H., Munns, T.W., and Rhoads, R.E. (1981). Heterogeneity of the 5' terminus of hen ovalbumin messenger ribonucleic acid. Nucleic Acids Res. 9, 1657-1673.

Manley, J., Fire, A., Cano, A., Sharp, P., and Gefter, M. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA 77, 3855-3859.

Manley, J., Sharp, P., and Gefter, M. (1979a). RNA synthesis in isolated nuclei: Identification and comparison of Adenovirus 2 encoded transcripts synthesized *in vitro* and *in vivo*. J. Mol. Biol. 135, 171-197.

Manley, J., Sharp, P., and Gefter, M. (1979b). RNA synthesis in isolated nuclei: *In vitro* initiation of adenovirus 2 major late mRNA precursor. Proc. Natl. Acad. Sci. USA 76, 160-164.

Mathis, D. and Chambon, P. (1981). The SV40 early region TATA box is required for accurate *in vitro* initiation of transcription. Nature 290, 310-315.

Maxam, A. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74, 560-564.

Maxam, A. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Methods in Enzymology 65, 499-560.

McBride, W. and Weiner, A. (1964). In vitro transformation of hamster kidney cells by human adenovirus type 12. Proc. Soc. Exp. Biol. Med. 115, 870-874.

McKnight, S. (1980). The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. 8, 5949-5964.

McKnight, G.S. and Palmiter, R. (1979). Transcriptional regulation of the ovalbumin and conalbumin genes by steroid hormones in chick oviduct. J. Biol. Chem. 254, 9050-9058.

McReynolds, L., O'Malley, B., Nisbet, A., Fothergill, J., Givol, D., Fields, S., Robertson, M., and Brownlee, G. (1978). Sequence of chicken ovalbumin mRNA. Nature 273, 723-728.

Mizumoto, K., and Lipmann, F. (1979). Transmethylation and transguanylylation in 5'-RNA capping system isolated from rat liver nuclei. Proc. Natl. Acad. Sci. USA 76, 4961-4965.

Moss, B., Gershowitz, A., Wei, C., and Boone, R. (1976). Formation of the guanylylated and methylated 5'-terminus of vaccinia virus mRNA. Virology 72, 341-351.

Moss, T. and Birnstiel, M. (1979). The putative promoter of a Xenopus laevis ribosomal gene is duplicated. Nucleic Acids Res. 6, 3733-3743.

Nagata, S., Mantei, N. and Weissmann, C. (1980). The structure of one of the eight or more distinct chromosomal genes for human interferon-α. Nature 287, 401-408.

Neu, H. and Heppel, L. (1964). Nucleotide sequence analysis of polyribonucleotides by means of periodate oxidation followed by cleavage with an amine. J. Biol. Chem. 239, 2927-2934.

Nevins, J. (1979). Processing of late adenovirus nuclear RNA to mRNA: Kinetics of formation of intermediates and demonstration that all events are nuclear. J. Mol. Biol. 130, 493-506.

Nevins, J., Blanchard, J.-M. and Darnell, J.E. (1980). Transcription units of adenovirus type 2: Termination of transcription beyond the poly(A) addition site in Early Regions 2 and 4. J. Mol. Biol. 144, 377-386.

Nevins, J. and Chen-Kiang, S. (1981). Processing of adenovirus nuclear RNA to mRNA. Advances in Virus Research 26, 1-35.

Nevins, J. and Darnell, J. (1978a). Groups of adenovirus type 2 mRNA's derived from a large primary transcript: Probable nuclear origin and possible common 3' ends. J. Virol. 25, 811-823.

Nevins, J. and Darnell, J. (1978b). Steps in the processing of Ad2 mRNA: Poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell 15, 1477-1493.

Nevins, J.R., Ginsberg, H.S., Blanchard, J.-M., Wilson, M.C., and Darnell, J.E. (1979). Regulation of the primary expression of the

early adenovirus transcription units. J. Virol. 32, 727-733.

Nevins, J.R. and Winkler, J.J. (1980). Regulation of early adenovirus transcription: A protein product of early region 2 specifically represses region 4 transcription. Proc. Natl. Acad. Sci. USA 77, 1893-1897.

Nicolson, M. and McAllister, R. (1972). Infectivity of human adenovirus-1 DNA. Virology 48, 14-21.

Parsons, J. and Green, M. (1971). Biochemical studies on adenovirus multiplication: XVIII. Resolution of Early virus-specific RNA species in Ad2 infected and transformed cells. Virology 45, 154-162.

Perricaudet, M., Akusjarvi, G., Virtanen, A. and Pettersson, U. (1979) Structure of two spliced mRNAs from the transforming region of human subgroup C adenoviruses. Nature 281, 694-696.

Perricaudet, M., Le Moullec, J. and Pettersson, U. (1980). Predicted structure of two adenovirus tumor antigens. Proc. Natl. Acad. Sci. USA 77, 3778-3782.

Perry, R., Kelley, D., Friderici, K. and Rottman, F. (1975). Methylated constituents of heterogeneous nuclear RNA: Presence in blocked 5' terminal structures. Cell 6, 13-19.

Perry, R. and Kelley, D. (1976). Kinetics of formation of 5' terminal caps in mRNA. Cell 8, 433-442.

Persson, H., Jansson, M. and Philipson, L. (1980a). Synthesis and genomic site for an adenovirus type 2 early glycoprotein. J. Mol. Biol. 136, 375-394.

Persson, H., Jornvall, H. and Zabielski, J. (1980b). Multiple mRNA species for the precursor to an adenovirus-encoded glycoprotein: Identification and structure of the signal sequence. Proc. Natl. Acad. Sci. USA 77, 6349-6353.

Persson, H., Mathisen, B., Philipson, L., and Pettersson, U. (1979). A maturation protein in adenovirus morphogenesis. Virology 93, 198-208.

Persson, H., Pettersson, U., and Mathews, M. (1978). Synthesis of a structural adenovirus polypeptide in the absence of viral DNA replication. Virology 90, 67-79.

Persson, H., Signas, C. and Philipson, L. (1979). Purification and characterization of an early glycoprotein from adenovirus type 2-infected cells. J. Virol. 29, 938-948.

Pettersson, U. and Mathews, M. (1977). The gene and messenger RNA for adenovirus polypeptide IX. Cell 12, 741-750.

Pettersson, U. and Sambrook, J. (1973). Amount of viral DNA in the genome of cells transformed by adenovirus type 2. J. Mol. Biol. 73, 125-130.

Philipson, L. (1961). Adenovirus assay by the fluorescent cell-counting procedure. Virology 15, 263-268.

Philipson, L., Pettersson, U., Lindberg, U., Tibbetts, C., Vennstrom, B. and Persson, T. (1975). RNA synthesis and processing in adenovirus-infected cells. Cold Spring Harbor Symp. Quant. Biol. 39, 447-456.

Price, R. and Penman, S. (1972). Transcription of the adenovirus genome by an α-amanitine-sensitive ribonucleic acid polymerase in HeLa cells. J. Virol. 9, 621-626.

Proudfoot, N., Shander, M., Manley, J., Gefter, M. and Maniatis, T. (1980). Structure and in vitro transcription of human globin genes. Science 209, 1329-1336.

Proudfoot, N. and Brownlee, G.G. (1976). 3' Non-coding region sequences in eukaryotic messenger RNA. Nature 263, 211-214.

Reddy, V., Ghosh, P., Lebowitz, P., Piatak, M., Weissman, S. (1979). Simian virus 40 early mRNAs: I. Genomic localization of 3' and 5' termini and two major splices in mRNA from transformed and lytically infected cells. J. Virol. 30, 279-296.

Reddy, V., Thimmappaya, B., Dhar, R., Subramanian, K., Zain, B., Pan, J., Ghosh, P., Celma, M., and Weissman, S. (1978). The genome of simian virus 40. Science 200, 494-502.

Rekosh, D., Russel, W., Bellet, A., and Robinson, A. (1977). Identification of a protein linked to the ends of adenovirus DNA. Cell 11, 283-295.

Roberts, R., Arrand, J. and Keller, W. (1974). The length of the terminal repetition in adenovirus-2 DNA. Proc. Natl. Acad. Sci. USA 71, 3829-3833.

Roeder, R. (1976). Eukaryotic nuclear RNA polymerases. In RNA Polymerase (Losick, R. and Chamberlin, M., eds.), pp. 285-329, Cold Spring Harbor Laboratory, New York.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. and Wall, R. (1980). Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chains. Cell 20, 303-312.

Rogers, J. and Wall, R. (1980). A mechanism for RNA splicing. Proc. Natl. Acad. Sci. USA 77, 1877-1879.

Roop, D., Kristo, P., Stumph, W., Tsai, M. and O'Malley, B. (1981). Structure and expression of a chicken gene coding for U1 RNA. Cell 23, 671-680.

Roop, D., Tsai, M. and O'Malley, B. (1980). Definition of the 5' and 3' ends of transcripts of the ovalbumin gene. Cell 19, 63-68.

Rosenberg, M. (1974). Isolation and sequence determination of the 3'-terminal regions of isotopically labeled RNA molecules. Nucleic Acids Research 1, 653-671.

Rosenberg, M., Wiebers, J., and Gilham, P. (1972). Studies on the interactions of nucleotides, polynucleotides, and nucleic acids with dihydroxyboryl-substituted celluloses. Biochemistry 11, 3623-3628.

Rosenberg, M. and Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. 13, 319-353.

Rowe, W., Huebner, R., Gilmore, L., Parrott, R. and Ward, T. (1953). Isolation of a cytoplasmic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Proc. Soc. Exp. Biol. and Med. 84, 570-573.

Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S. and Darnell, J. (1980a). The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. Cell 19, 69-78.

Salditt-Georgieff, M., Harpold, M., Sawicki, S., Nevins, J. and Darnell, J. (1980b). Addition of poly(A) to nuclear RNA occurs soon after RNA synthesis. J. Cell. Biology 86, 844-848.

Salditt-Georgieff, M., Jelinek, W., Darnell, J., Furuichi, Y., Morgan, M. and Shatkin, A. (1976). Cell 7, 227-237.

Samal, B., Worcel, A., Louis, C. and Schedl, P. (1981). Chromatin structure of the histone genes of D. melanogaster. Cell 23, 401-409.

Sanger, F., Brownlee, G.G., and Barrell, B.G. (1965). A two-dimensional fractionation procedure for radioactive nucleotides. J. Mol. Biol. 13, 373-398.

Sanger, F. and Coulson, A. (1978). The use of thin acrylamide gels for DNA sequencing. FEBS Letters 87, 107-110.

Saragosti, S., Moyne, G. and Yaniv, M. (1980). Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. Cell 20, 65-73.

Schibler, U. and Perry, R. (1976). Characterization of the 5' termini of hnRNA in mouse L cells: Implications for processing and cap formation. Cell 9, 121-130.

Schott, H., Rudloff, E., Schmidt, P., Roychoudhury, R., and Kossel, H. (1973). A dihydroxyboryl-substituted methacrylic polymer for the column chromatographic separation of mononucleotides, oligonucleotides, and transfer ribonucleic acids. Biochem. 12, 932-938.

Scott, W. and Wigmore, D. (1978). Sites in simian virus 40 chromatin which are preferentially cleaved by endonucleases. Cell 15, 1511-1518.

Sehgal, P., Fraser, N., and Darnell, J. (1979). Early Ad-2 transcription units: Only promoter-proximal RNA continues to be made in the presence of DRB. Virology 94, 185-191.

Shatkin, A. (1976). Capping of eukaryotic mRNAs. Cell 9, 645-653.

Shaw, A. and Ziff, E.B. (1980). Transcripts from the adenovirus-2 major late promoter yield a single family of 3' coterminal mRNAs during early infection and five families at late times. Cell 22, 905-916.

Shenk, T. (1981). Transcriptional control regions: Nucleotide sequence requirements for initiation by RNA polymerase II and III. Current Topics Microbiology and Immunology, in press.

Shinagawa, M., Padmanabhan, R.V., and Padmanabhan, R. (1980). The nucleotide sequence of the right-hand terminal SmaI-K fragment of adenovirus type 2 DNA. Gene 9, 99-114.

Siebenlist, U., Simpson, R., and Gilbert, W. (1980). E. coli RNA polymerase interacts homologously with two different promoters. Cell 20, 269-281.

Smith, H. and Birnstiel, M. (1976). A simple method for DNA restriction site mapping. Nucleic Acids Res. 3, 2387-2398.

Soeda, E., Arrand, J., Smolar, N., Walsh, J. and Griffin, B. (1980). Coding potential and regulatory signals of the polyoma virus genome. Nature 283, 445-453.

Sollner-Webb, B. and Reeder, R. (1979). The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in X. laevis. Cell 18, 485-499.

Spector, D., McGrogan, M., and Raskas, H. (1978). Regulation of the appearance of cytoplasmic RNAs from region 1 of the adenovirus 2 genome. J. Mol. Biol. 126, 395-414.

Stalder, J., Larsen, A., Engel, J., Dolan, M., Groudine, M. and Weintraub, H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAse I. Cell 20, 451-460.

Steenbergh, P. H. and Sussenbach, J.S. (1979). The nucleotide sequence of the right-hand terminus of adenovirus type 5 DNA: implications for the mechanisms of DNA replication. Gene 6, 307-318.

Stillman, B., Lewis, J., Chow, L., Mathews, M., and Smart, J. (1981). Identification of the gene and mRNA for the adenovirus terminal protein precursor. Cell 23, 497-508.

Sures, I., Levy, S. and Kedes, L. (1980). Leader sequences of *Strongylocentrotus purpuratus* histone mRNAs start at a unique heptanucleotide common to all five histone genes. Proc. Natl. Acad. Sci. USA 77, 1265-1269.

Testa, D., Chanda, P., and Banerjee, A. (1980). Unique mode of transcription *in vitro* by vesicular stromatitis virus. Cell 21, 267-275.

Thiel, J. and Smith, K. (1967). Fluorescent focus assay of viruses on cell monolayers in plastic petri plates. Proc. Soc. Exper. Biol. Med. 125, 892-895.

Thomas, G. and Mathews, M. (1980). DNA replication and the early to late transition in adenovirus infection. Cell 22, 523-533.

Thompson, J., Radonovich, M., and Salzman, N. (1979). Characterization of the 5'-terminal structure of simian virus 40 early mRNAs. J. Virol. 31, 437-446.

Tooze, J. (1980). Restriction endonuclease cleavage maps of group-C adenoviruses 1, 2, and 5. In DNA Tumor Viruses, Part 2 (J. Tooze, ed.) pp. 931-935, Cold Spring Harbor Laboratory, New York.

Trentin, J., Yabe, Y. and Taylor, G. (1962). The quest for human cancer viruses. Science 137, 835-841.

Tsuda, M., Ohshima, Y. and Suzuki, Y. (1979). Assumed initiation site of fibroin gene transcription. Proc. Natl. Acad. Sci. USA 76, 4872-4876.

Tsujimoto, Y. and Suzuki, Y. (1979a). Structural analysis of the fibroin gene at the 5' end and its surrounding regions. Cell 16, 425-436.

Tsujimoto, Y. and Suzuki, Y. (1979b). The DNA sequence of Bombyx mori fibroin gene including the 5' flanking, mRNA coding, entire intervening and fibroin protein coding regions. Cell 18, 591-600.

Van Ooyen, A., Van den Berg, J., Mantei, N. and Weissmann, C. (1979). Comparison of total sequence of a cloned rabbit β-globin gene and its flanking regions with a homologous mouse sequence. Science 206, 337-344.

Van Ormondt, H., Maat, J., DeWaard, A., and Van der Eb, A. J. (1978). The nucleotide sequence of the transforming HpaI-E fragment of adenovirus type 5 DNA. Gene 4, 309-328.

Van Ormondt, H., Maat, J. and Van Beveren, C. (1980). The nucleotide sequence of the transforming early region E1 of adenovirus type 5 DNA. Gene 11, 299-309.

Van de Sande, J., Kleppe, K. and Khorana, H. (1973). Reversal of bacteriophage T_4 Induced polynucleotide kinase action. Biochemistry 12, 5050-5055.

Van der Eb, A., Mulder, C., Graham, F. and Houweling, A. (1977). Transformation with specific fragments of adenovirus DNAs I. Isolation of specific fragments with transforming activity of adenovirus 2 and 5 DNA. Gene 2, 115-132.

Van der Eb, A., van Ormondt, H., Schriers, P., Lupker, J., Jochemsen, H., van den Elsen, P., DeLeys, R., Maat, J., van Beveren, C., Dijkema, R. and de Waard, A. (1980). Structure and function of the transforming genes of human adenoviruses and SV40. Cold Spring Harbor Symp. Quant. Biol. 44, 383-399.

Van der Vliet, P., Levine, A., Ensinger, M., and Ginsberg, H. (1975). Thermolabile DNA binding proteins from cells infected by a temperature-sensitive mutant of adenovirus defective in viral DNA synthesis. J. Virol. 15,, 348-354.

Van der Vliet, P. and Sussenbach, J. (1975). An adenovirus type 5 gene function required for initiation of viral DNA replication. Virology 67, 415-426.

Varshavsky, A., Sundin, O., and Bohn, M. (1979). A stretch of "late" SV40 viral DNA about 400 b.p. long which includes the origin of replication is specifically exposed in SV40 minichromosomes. Cell 16, 453-466.

Venkatesan, S. and Moss, B. (1980). Donor and acceptor specificities of HeLa cell mRNA guanylyltransferase. J. Biol. Chem. 255, 2835-2842.

Wagner, M., Sharp, J. and Summers, W. (1981). Nucleotide sequence of the thymidine kinase gene of herpes simplex type 1. Proc. Natl. Acad. Sci. USA 78, 1441-1445.

Waldeck, W., Fohring, B., Chowdhury, K., Gruss, P. and Sauer, G. (1978). Origin of replication in papovavirus chromatin is recognized by endogenous endonuclease. Proc. Natl. Acad. Sci. USA 75, 5964-5968.

Wallace, R. and Kates, J. (1972). State of adenovirus 2 deoxyribonucleic acid in the nucleus and its mode of transcription, studies with isolated viral deoxyribonucleic acid-protein complexes and isolated nuclei. J. Virol. 9, 627-635.

Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teoule, R., and Chambon, P. (1980a). Specific *in vitro* transcription of conalbumin gene is drastically decreased by single-point mutation in TATA box homology sequence. Proc. Natl. Acad. Sci. USA 77, 7024-7028.

Wasylyk, B., Kedinger, C., Corden, J., Brison, O. and Chambon, P. (1980b) Specific *in vitro* initiation of transcription on conalbumin and ovalbumin genes and comparison with adenovirus-2 early and late genes. Nature 285, 367-373.

Weaver, R. and Weissman, C. (1979). Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. 7, 1175-1193.

Weber, J., Blanchard, J.-M., Ginsberg, H. and Darnell, J. (1980). Order of polyadenylic acid addition and splicing events in early adenovirus mRNA formation. J. Virol. 33, 286-291.

Weber, J., Jelinek, W., and Darnell, J. (1977). The definition of a large viral transcription unit late in Ad2 infection of HeLa cells: mapping of nascent RNA molecules labeled in isolated nuclei. Cell 10, 611-616.

Wei, C. and Moss, B. (1977). 5'-Terminal capping of RNA by guanylyltransferase from HeLa cell nuclei. Proc. Natl. Acad. Sci. USA 74, 3758-3761.

Weil, P.A., Luse, D., Segall, J., and Roeder, R. (1979). Selective and accurate initiation of transcription at the Ad2 major late

promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell 18, 469-484.

Weinmann, R., Jaehning, J., Raskas, H. and Roeder, R. (1976). Viral RNA synthesis and levels of DNA-dependent RNA polymerases during replication of adenovirus 2. J. Virol. 17, 114-126.

Weinmann, R., Raskas, H. and Roeder, R. (1974). Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. Proc. Natl. Acad. Sci. USA 71, 3426-3430.

Weintraub, H. and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science 193, 848-856.

Weisbrod, S., Groudine, M. and Weintraub, H. (1980). Interaction of HMG 14 and 17 with actively transcribed genes. Cell 19, 289-301.

Whitfeld, P. and Markham, R. (1953). Natural configuration of the purine nucleotides in ribonucleic acids. Nature 171, 1151-1152.

Williams, J., Tsang, A. and Mahbubani, H. (1980). A change in the rate of transcription of a eukaryotic gene in response to cyclic AMP. Proc. Natl. Acad. Sci. USA 77, 7171-7175.

Wilson, M., Fraser, N., and Darnell, J. (1979). Mapping of RNA initiation sites by high doses of UV irradiation. Evidence for three independent promoters within the left 11% of the Ad-2 genome. Virology 94, 175-184.

Wilson, M., Nevins, J., Blanchard, J.-M., Ginsberg, H., and Darnell, J. E. (1980). Metabolism of mRNA from the transforming region of adenovirus 2. Cold Spring Harbor Symp. Quant. Biol. 44, 447-455.

Winicov, I., and Perry, R. (1976). Synthesis, methylation, and capping of nuclear RNA by a subcellular system. Biochemistry 15, 5039-5046.

Wu, C. (1980). The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNAse I. Nature 286, 854-860.

Yamada, Y. and Ishikura, H. (1975). Identification of a modified nucleoside in *Escherichia coli* $tRNA^{ser}$ as 2'-O-methylcytidine. Biochim. Biophys. Acta 402, 285-287.

Yamamoto, T., Jay, G. and Pastan, I. (1980a). Unusual features in the nucleotide sequence of a cDNA clone derived from the common region of avian sarcoma virus mRNA. Proc. Natl. Acad. Sci. USA 77, 176-180.

Yamamoto, T., de Crombrugghe, B. and Pastan, I. (1980b). Identification of a functional promoter in the long terminal repeat of rous sarcoma virus. Cell 22, 787-797.

Young, R., Hagenbuchle, O. and Schibler, U. (1981). A single mouse α-amylase gene specifies two different tissue-specific mRNAs. Cell 23, 451-458.

Zamenhof, S. and Eichhorn, H. (1967). Study of microbial evolution through loss of biosynthetic functions: Establishment of "defective" mutants. Nature 216, 456-458.

Ziff, E. and Evans, R. (1978). Coincidence of the promoter and capped 5' terminus of RNA from the adenovirus-2 major late transcription unit. Cell 15, 1463-1475.

Ziff, E.B. (1980). Transcription and RNA processing by the DNA tumour viruses. Nature 287, 491-499.

End