Advances in VIRUS RESEARCH

Edited by

MAX A. LAUFFER FREDERIK B. BANG

KARL MARAMOROSCH KENNETH M. SMITH

VOLUME 26

Advances in
VIRUS RESEARCH

Edited by

MAX A. LAUFFER

Andrew Mellon Professor of
Biophysics
University of Pittsburgh
Pittsburgh, Pennsylvania

FREDERIK B. BANG

Department of Pathobiology
The Johns Hopkins University
Baltimore, Maryland

KARL MARAMOROSCH

Waksman Institute
of Microbiology
Rutgers University
New Brunswick, New Jersey

KENNETH M. SMITH

Cambridge, England

VOLUME 26

1981

ACADEMIC PRESS

A Subsidiary of Harcourt Brace Jovanovich, Publishers

New York London Toronto Sydney San Francisco

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 53–11559

ISBN 0–12–039826–5

PRINTED IN THE UNITED STATES OF AMERICA

81 82 83 84 9 8 7 6 5 4 3 2 1

CONTRIBUTORS TO VOLUME 26

Numbers in parentheses indicate the pages on which the authors' contributions begin.

SELINA CHEN-KIANG, *The Rockefeller University, New York, New York 10021 (1)*

G. CROIZIER, *Station de Recherches de Pathologie Comparée Saint-Christol, Gard, France (117)*

L. HIRTH, *Institut de Biologie Moléculaire et Cellulaire du CNRS, Strasbourg, France (145)*

ARNOLD J. LEVINE, *Department of Microbiology, School of Basic Health Sciences, State University of New York at Stony Brook, Stony Brook, New York 11794 (65)*

J. F. LONGWORTH, *Entomology Division, Department of Scientific and Industrial Research, Auckland, New Zealand (117)*

WARREN MALTZMAN,* *Department of Microbiology, School of Basic Health Sciences, State University of New York at Stony Brook, Stony Brook, New York 11794 (65)*

JOSEPH R. NEVINS, *The Rockefeller University, New York, New York 10021 (1)*

N. PLUS, *Station de Recherches de Pathologie Comparée Saint-Christol, Gard, France (117)*

C. REINGANUM, *Plant Research Institute, Burnley, Victoria 3121, Australia (117)*

K. E. RICHARDS, *Institut de Biologie Moléculaire et Cellulaire du CNRS, Strasbourg, France (145)*

P. D. SCOTTI, *Entomology Division, Department of Scientific and Industrial Research, Auckland, New Zealand (117)*

ILAN SELA, *Virus Laboratory, The Hebrew University, Faculty of Agriculture, Rehovot, Israel (201)*

G. STREISSLE, *Institute of Immunology and Oncology, Bayer AG, D-5600 Wuppertal 1, Federal Republic of Germany (37)*

*Present address: Department of Biological Sciences, Douglass College, Rutgers University, New Brunswick, New Jersey 08903.

CONTENTS

Tobacco Mosaic Virus: Model for Structure and Function of a Simple Virus

L. HIRTH AND K. E. RICHARDS

Plant–Virus Interactions Related to Resistance and Localization of Viral Infections

ILAN SELA

PROCESSING OF ADENOVIRUS NUCLEAR RNA TO mRNA

Joseph R. Nevins and Selina Chen-Kiang

The Rockefeller University
New York, New York

I. Introduction

The mechanism as well as the regulation of the biogenesis of adenovirus mRNA has been the subject of intensive research over the past several years principally because it provides an excellent model for the events which occur in the uninfected mammalian cell. The usefulness of adenovirus as a model system is illustrated by the following points:

1. Adenovirus is a double-stranded DNA virus which replicates in the nucleus of infected cells. The virus carries no known enzymes into the cell and thus is completely dependent upon the cellular machinery for at least the initial transcription of its genome. All currently available evidence suggests that the transcription of the genome is indeed performed by cellular enzymes (Ledinko, 1971; Wallace and Kates,

ISBN 0-12-039826-5

1972; Chardonnet *et al.*, 1972; Price and Penman, 1972; Weinmann *et al.*, 1974, 1976). Furthermore, adenovirus mRNA found in the cytoplasm both early and late in infection possesses all of the modifications known to occur in cellular RNA. These include a 5′ methylated cap structure (Sommer *et al.*, 1976; Moss and Koczot, 1976; Wold *et al.*, 1976; Hashimoto and Green, 1976), internal m^6A residues (Sommer *et al.*, 1976; Moss and Koczot, 1976) which are contained within the general sequence m^6ApC or Apm^6ApC (J. R. Nevins, S. Sommer, and J. E. Darnell, unpublished data), and a 3′ terminal poly(A) segment (Philipson *et al.*, 1971). As all of these modifications are common to both adenovirus and cellular mRNAs it would appear likely that cellular enzymes are responsible for performing these modifications.

2. The adenovirus genome, a double-stranded linear DNA molecule of 35,000 base pairs (for the type C adenoviruses) (Green, 1970), is well defined. The cleavage sites for Ad-2 DNA for more than 12 restriction endonucleases have now been established. In addition, considerable portions of the adenovirus genome have been sequenced.

Purified adenovirus DNA can be prepared in large quantities so that defined segments of the genome are available in sufficient amounts for most experiments. Furthermore, with the advent of recombinant DNA techniques, it is possible to obtain large quantities of virtually any region of the viral genome in pure form.

3. The fraction of total RNA synthesis in the infected cell which is viral-specific is quite high. For instance, transcription of a given early transcription unit, as measured in a 5′ pulse-label with [^{3}H]uridine, amounts to approximately 0.1–0.2% of the total nuclear RNA synthesis. Transcription of the late region amounts to more than 20% of the total RNA synthesis. This, therefore, allows very short labeling periods to be employed in the measurement of RNA synthesis such that even the earliest events of mRNA manufacture can be studied. When one is attempting to analyze the processes of mRNA biogenesis this fact becomes *very* important. For instance, due to the high level of adenovirus RNA synthesis in infected cells, viral transcription can be measured by labeling for as short as 30 seconds with [^{3}H]uridine. By so doing, it is possible to exclude totally the effects of posttranscriptional processes in the measurement of transcription rates. Furthermore, labeling for such short periods allows the identification of the *initial* products of transcription.

In this article we will not attempt to cover all of the aspects of the replication of adenovirus (for reviews on the molecular biology of adenovirus see Flint and Broker, 1980, Wold *et al.*, 1978, and Ziff, 1980). Rather, we will present a more limited picture of the processes

involved in the production of a group of specific mRNAs. Since the most comprehensive information available is that dealing with the production of the late adenovirus mRNAs, it is that subject on which we will focus our attention, making reference when appropriate to other systems.

II. Definition of Adenovirus Transcription Units

A. Rationale and Techniques for Defining Transcription Units

A transcription unit may be defined as that region of a genome between a start signal and a stop signal for RNA polymerase (Darnell, 1978). The RNA which results from the transcription of a particular transcription unit is the primary transcript and thus is the substrate for subsequent processing reactions. It is therefore most important for an understanding of mRNA biogenesis to define accurately transcription units for mRNA production. To this end, a considerable body of evidence has accumulated concerning the adenovirus transcription units. Basically, two techniques have been employed for defining transcription units: size analysis of nascent-labeled RNA and UV transcription mapping.

1. Nascent Chain Analysis

Since the transcription of a particular transcription unit proceeds from a single initiation site,[1] then at any given instant, the RNA molecules in the process of being formed (nascent) represent an array of molecules each with common 5′ termini but of varying lengths extending from the initiation site (see Fig. 1). Since all of the molecules have a common 5′ terminus then the length of the molecules will depend on how far their growing 3′ end is from the initiation site. If a labeled precursor ([^{3}H]uridine) is added for a period of time which is short compared to the total synthesis time of the primary transcript, only the ends of the growing chains (3′ termini) will be labeled (see Fig. 1). When an array of nascent-labeled molecules is then separated according to size and hybridized to a set of DNA fragments derived from the transcription unit, then both the origin of transcription as well as the limit of transcription can be determined. The shortest

[1] For the purposes of this article, we have used the term "transcription initiation site" to indicate the genomic site where transcription actually begins. The term "promoter" is used to designate a region of the genome containing sequences required for polymerase recognition and binding.

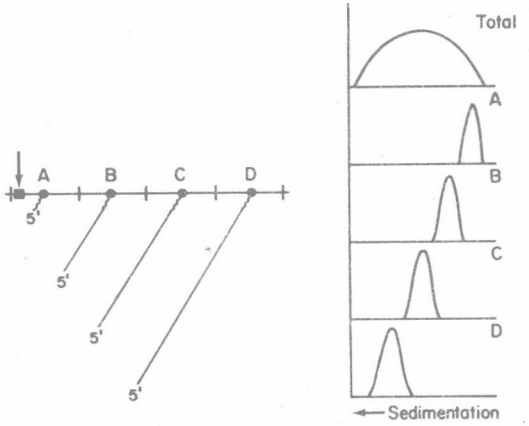

FIG. 1. Schematic representation of a nascent chain analysis. In the left figure, a DNA genome is depicted divided into restriction fragments A, B, C, and D. The initiation site for a single transcription unit is located as shown in fragment A (■) and transcription proceeds to the right terminating in fragment D. A short pulse of [^{3}H]uridine labels only the growing 3′ termini of the nascent chains (wavy line); the remainder of the chain is not labeled (straight line). After purification, these molecules are then separated by means of a sucrose gradient (right figure) and the fractions from the gradient are hybridized to each of the DNA fragments. Radioactivity in ribonuclease resistant hybrids is then scored.

nascent chain will hybridize to a DNA fragment containing the initiation site. The farther a fragment is located from the initiation site, the longer will be the RNA chain which it detects, due to the fact that only the termini of the growing chain are labeled.

2. *UV Transcription Mapping*

When cells are exposed to ultraviolet light, thymine–thymine dimers are formed in the cellular DNA. When an RNA polymerase molecule encounters such a dimer during transcription it terminates and falls off the DNA; the reinitiation of transcription, however, is not affected (Michalke and Bremer, 1969; Sauerbier *et al.*, 1970). Therefore, for a given transcription unit in a DNA molecule, the probability of successful transcription of a particular sequence in the transcription unit is a function of the distance of that sequence from the initiation site. This probability decreases exponentially as a function of distance form the initiation site (for a review see Sauerbier, 1976). Transcription of sequences within a particular transcription unit near the initiation site will be more resistant to UV than the transcription of sequences near the termination site. A given transcription unit can thus

be defined by the survival of transcriptional activity after UV irradiation.

B. Early Adenovirus Transcription Units

The expression of the adenovirus genome during lytic infection can be separated into two phases defined by the replication of the viral DNA. Early genes are expressed prior to DNA replication and late genes are only expressed once DNA replication commences. Hybridization studies demonstrated that the early mRNAs were encoded by four widely separated regions of the viral DNA (Sharp *et al.*, 1974; Pettersson *et al.*, 1976; Buttner *et al.*, 1976). Two of these regions were from the rightward-reading strand (regions 1 and 3) and the other two regions were from the leftward-reading strand (regions 2 and 4). However, the mRNA maps determined in the early studies left open the question as to the nature of the transcription units responsible for the production of the early mRNAs: Did each early region represent a single transcription unit, were there more than one transcription unit for each region, or were there two large transcriptional units, one for each strand, reading in opposite directions?

Employing the technique of nascent chain analysis Evans *et al.* (1977) were able to demonstrate the existence of at least four distinct transcriptional units involved in the production of early mRNA. In this work, isolated nuclei from early adenovirus-infected cells were pulse-labeled *in vitro* so as to obtain sufficient radioactivity in the nascent labeled RNA chain. Their analysis indicated the presence of rightward-reading transcription initiation sites in the DNA fragments of coordinates 0–4 and coordinates 76–83; leftward-reading initiation sites were detected in the 59–76 and 90–100 fragments. Berk and Sharp (1977a) utilized the UV transcription mapping technique to demonstrate the same four early transcriptional units. Their approach was to irradiate purified virions with increasing doses of UV and then to measure levels of accumulated RNA after infection of cells with these virions. In addition, their results also indicated that the target size for the early region 2 mRNA was approximately three times larger than expected based on the size of the mature mRNA. This, therefore, presented the possibility that this mRNA was derived from a larger precursor, a possibility which was suggested earlier by the detection of nuclear poly(A) + RNA which was larger than the mRNA (Craig and Raskas, 1976).

Two different approaches were employed to achieve a more detailed analysis of the number and locations of the early transcription initia-

tion sites. These techniques do not define the boundaries of the transcription units but only attempt to locate the initiation sites. Sehgal *et al.* (1979) used the drug 5,6 dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) (for a review see Tamm and Sehgal, 1978) to induce premature termination of RNA transcripts. By so doing, they were able to cause the accumulation of short RNA chains, representing sequences adjacent to the initiation site, which could then serve as a probe to determine the location of initiation sites. These short chains produced in early adenovirus-infected cells were hybridized to a variety of viral DNA restriction fragments. The results of these experiments confirmed the initiation sites identified by nascent-chain analysis and UV-transcription mapping. However, this method also allowed the definition of an additional early initiation site, not previously detected, mapping between genome coordinates 3 and 11 in the rightward direction as well as an initiation site at map position 16 reading in the leftward direction. Wilson *et al.* (1979) employed high-dose UV irradiation to obtain similar results. Rather than measuring the sensitivity of transcription of various regions of the DNA genome to UV inactivation they employed very high doses of UV to once again produce only short RNAs. These short RNAs were then mapped as to their origin in the genome. Their results confirmed the initiation sites detected previously as well as one additional initiation site at approximately map position 8. This presumably would be the site for initiation of synthesis of the mRNA coding for protein IX (Pettersson and Mathews, 1977; Spector *et al.*, 1979), which appears to be a protein produced beginning at an intermediate time after infection (Persson *et al.*, 1978; Spector *et al.*, 1979). What becomes clear from these studies is that the four blocks of early transcription, which had been previously identified by hybridization studies, represent discrete transcriptional units, each possessing an independent site for transcription initiation, with the exception of early region 1 which actually consists of two transcription units at early times and three transcription units at intermediate and late times. A map depicting the locations of the early adenovirus transcription units is shown in Fig. 2. The two region 1 transcription units have now been designated 1A (1.5–4.4 map units) and 1B (4.4–11 map units). The precise locations of the early transcription initiation sites have been determined by first obtaining the 5′ sequence of the various early mRNAs and then locating these sequences in the appropriate regions of the DNA (see Table I) (Baker and Ziff, 1980; Alestrom *et al.*, 1980) making the assumption that the start site for transcription is the sequence encoding the mRNA 5′ terminus (see Section III,B). Also included in the map is the transcriptional unit mapping between

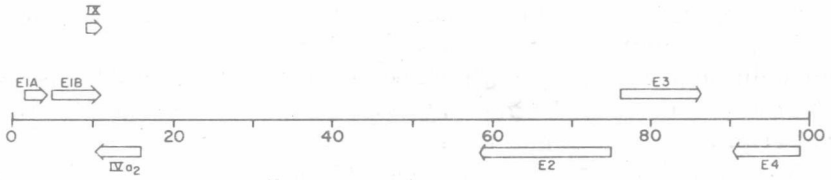

FIG. 2. Map of the adenovirus genome depicting the locations of the early adenovirus transcription units as well as the transcription units active at intermediate times postinfection (those producing the mRNAs for protein IX and protein IVa_2). One map unit is equal to 350 base pairs.

genome coordinates 16 and 11 reading in the leftward direction. This produces the mRNA for protein IVa_2 (Chow and Broker, 1978) which, like the transcription unit for protein IX, is believed to be active beginning at intermediate times after infection.

Finally, two other regions of the adenovirus genome have recently been shown to be transcribed during early infection. The site of the mutations in a single complementation group of temperature-sensitive mutants of adenovirus type 5 (*H5ts36, H5ts69, H5ts149*) has been localized between genome coordinates 18.5 and 22.0 (Galos *et al.*, 1979). Since these are early mutants (i.e., no DNA replication or late gene expression at the nonpermissive temperature) these results would suggest that an additional transcription unit is functional during early infection. Furthermore, cytoplasmic poly(A) + viral mRNA was detected which hybridized to the leftward-reading strand of a DNA

TABLE I

LOCATIONS OF THE EARLY ADENOVIRUS TRANSCRIPTION INITIATION SITES

Transcription unit	Direction of transcription	Initiation site[a]
1A	Rightward	1.40[b]
1B	Rightward	4.86[b]
2	Leftward	75.05[b]
3	Rightward	76.60[b]
4	Leftward	99.10[b]
IX[c]	Rightward	9.60[d]

[a] The precise locations of the initiation sites were obtained by comparing the sequence of the 5′ termini of the early mRNA to the DNA sequence in a region expected to contain the initiation site.

[b] Baker and Ziff (1980).

[c] Becomes active at intermediate times (about 6 hours) postinfection

[d] Alestrom *et al.* (1980).

fragment of coordinates 19.8–23.5. The nature of the transcription unit for this mRNA (such as the location of the site for transcription initiation and termination) remains to be determined.

In addition, it now appears that a portion of the major late transcription unit is transcribed during early infection prior to DNA replication. Transcription originating at the late initiation site at map position 16 can be detected at a rate which is nearly equal to that of any of the other early transcription units (Shaw and Ziff, 1980; Nevins and Wilson, submitted). However, unlike the expression of this transcription unit during late infection (detailed in Section IV,B), transcription during early infection terminates near map position 60–70 (Nevins and Wilson, submitted). From kinetic labeling measurements, it appears that the major mRNA product of this transcription is an L1 mRNA (see Fig. 5) (Shaw and Ziff, 1980; Nevins and Wilson, submitted) which was previously detected by electron microscopy (Chow *et al.*, 1979). The RNA appears to differ, however, from those made late in infection in that it often contains an additional leader sequence, termed the "i" leader, derived from map position 22.0–23.1 (Chow *et al.*, 1979).

C. Late Adenovirus Transcription Unit

The initial studies concerning the definition of adenovirus transcription units dealt with the production of late viral RNA. It was first demonstrated that mRNA produced late in infection was derived from sequences in the rightward-reading strand of the DNA between map positions 30 and 91 (Sharp *et al.*, 1974; Pettersson *et al.*, 1976). The initial indication that this region was part of a single transcriptional unit came from a nascent chain analysis of late adenovirus nuclear RNA (Bachenheimer and Darnell, 1975). In this work, the largest nascent chains were RNA molecules with termini near the right end of the genome and were of a size consistent with molecules initiating somewhere in the left-hand portion of the genome. A more refined analysis was obtained using isolated nuclei prepared from late-infected cells to produce nascent-labeled chains (Weber *et al.*, 1977). Pulse-labeled RNA was sedimented in a sucrose gradient and hybridized to a variety of restriction fragments. The results obtained in this work produced the picture that the majority of late transcription derived from a single large transcription unit initiating at approximatley map position 10–20 and extending rightward to near the end of the genome. That is, the shortest nascent chains were located in a frag-

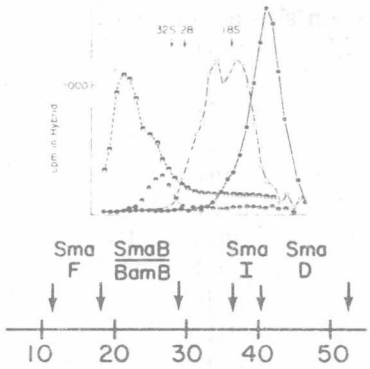

FIG. 3. Nascent chain analysis of *in vitro* labeled RNA produced from the late adenovirus promoter. RNA labeled in isolated nuclei and then separated in a sucrose gradient was hybridized to the DNA fragments *Sma*F (●), *Sma*B/*Bam*B (○), *Sma*I (◓), and *Sma*D (▭). From Evans *et al.* (1977).

ment of coordinates 11–18 and the longest nascent chains were in the 90–100 fragment. UV-transcription mapping was also employed to define the late transcriptional unit (Goldberg *et al.*, 1977). In this case infected cells were irradiated with various doses of UV and then labeled briefly with [^{3}H]uridine and the RNA was then hybridized to fragments throughout the genome. The results once again indicated that the entire region between map units 20 and 100 was transcribed from a single promoter and thus represented a single transcriptional unit. A further refinement of the nascent chain analysis was employed to map more accurately the late adenovirus transcription initiation site (Evans *et al.*, 1977). Nascent chains were labeled both *in vitro* in isolated nuclei and *in vivo* and analyzed by sucrose gradient centrifugation. This time, however, the analysis was performed only on the very small nascent chains so as to achieve resolution of the molecules nearest the site of transcription initiation (see Fig. 3). These experiments placed the start site for late transcription at 16 ± 0.5 map units.

Since only a small fraction of the viral nuclear RNA produced late in infection exits to the cytoplasm (Philipson *et al.*, 1974; Nevins and Darnell, 1978b) it still remained a possibility that the cytoplasmic viral mRNAs in fact did not derive from this large transcriptional unit but rather from smaller ones possessing independent promoters. This possibility was tested in an experiment which employed UV-transcription mapping to determine the origin of the late cytoplasmic viral mRNAs (Goldberg *et al.*, 1978). If the cytoplasmic mRNAs de-

rived from the initiation site at map position 16, then their UV sensitivity should be equal to the corresponding sequences in the nucleus. If, however, the mRNAs derived from independent transcription initiation sites then their sensitivity to UV should be less than the nuclear sequence, especially for the fiber mRNA which is at the greatest distance from the start site for late transcription. In each case it was observed that the synthesis of the mRNA sequence found in the cytoplasm had a UV target size as great as the synthesis of the corresponding sequence in the nucleus (see Fig. 4). Thus, the functional mRNAs coded from sequences between 30 and 100 in the DNA genome produced late in adenovirus infection all derive from transcription initiated at map position 16. This experiment, therefore, established the fact that these late mRNAs must derive by processing of a larger transcript involving endonucleolytic chain scissions.

III. Structural Aspects of the Late Transcription Unit

A. Structure of the Late mRNAs

All of the late adenovirus mRNAs have been shown to possess a 3′ poly(A) segment, a 5′ methylated cap structure, and internally methylated adenosine residues. Hybridization experiments performed by several laboratories demonstrated that the late viral mRNAs were heterogeneous in size and derived from DNA sequences between 30 and 91 map units. The structure and genomic origin of specific mRNAs has been determined in two ways: (1) size fractionation of labeled mRNAs followed by hybridization to specific DNA probes and (2) electron microscopy of DNA–RNA hybrids. The first procedure indicated the existence of approximatley 13 discrete late mRNAs deriving from genome coordinates 30–91 (McGrogan and Raskas, 1977, 1978; Nevins and Darnell, 1978a; Tal *et al.*, 1974). In addition these experiments led to the finding that various DNA fragments from genome coordinates 30–91 hybridized multiple species of RNA which exceeded the coding capacity of the DNA fragment (McGrogan and Raskas, 1978; Nevins and Darnell, 1978a). By hybridizing gel-fractionated mRNA to a series of adjacent DNA fragments it was concluded that the late mRNAs fall into five families of mRNAs, the RNAs within a family sharing sequences. Furthermore, the hybridization experiments suggested that the RNAs within a family were 3′ coterminal and RNA fingerprint analysis confirmed this conclusion for one of the mRNA families (Ziff

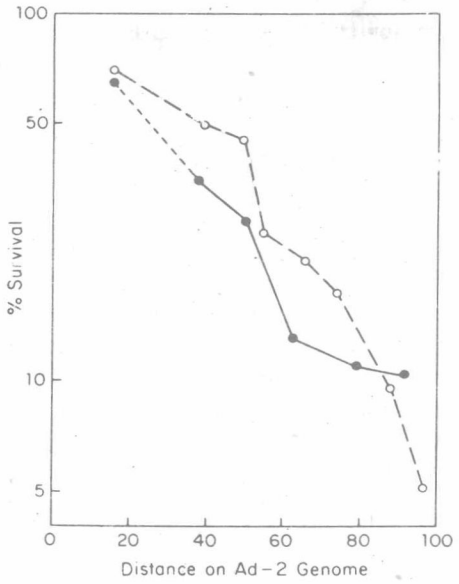

FIG. 4. UV inactivation of labeling of late adenovirus nuclear and cytoplasmic RNA. Nuclear and cytoplasmic RNAs prepared from control cells and cells irradiated with UV were hybridized to viral DNA fragments of the indicated genome coordinates. Nuclear RNA (○); cytoplasmic RNA (●). From Goldberg *et al.* (1978).

and Fraser, 1978). Thus within a given family of mRNAs, the individual species share sequences at their 3′ termini adjacent to the poly(A) segment. However, the 3′ terminal sequences for each of the five families of late mRNAs are distinct (Fraser and Ziff, 1978). In addition, two of the five 3′ terminal sequences, and possibly all five, contain the hexanucleotide sequence AAUAAA (Fraser and Ziff, 1978) which is common to most poly(A) + mRNAs (Proudfoot and Brownlee, 1976).

The visualization of DNA–RNA hybrids in the electron microscope also served to establish the genomic origin of the late mRNAs (Chow *et al.*, 1977a) and, in addition, led to a most significant discovery concerning the structure of the late adenovirus mRNAs. It was determined that the late viral mRNAs were constructued from sequences not contiguous in the DNA (Berget *et al.*, 1977; Chow *et al.*, 1977b). Although the main bodies of the late mRNAs are encoded by sequences between 30 and 91 map units, each mRNA also possesses short sequences deriving from map positions 16, 19, and 26. Thus each late mRNA is a composite of sequences deriving from four separate regions of the genome. This finding, then, gave an explanation to the earlier