



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# Vaccinia Virus Produces Late mRNAs by Discontinuous Synthesis

Christine Bertholet, Erwin Van Meir,  
Béatrice ten Heggeler-Bordier, and Riccardo Wittek  
Institut de Biologie Animale  
Université de Lausanne  
Bâtiment de Biologie  
CH-1015 Lausanne, Switzerland

## Summary

We describe the unusual structure of a vaccinia virus late mRNA. In these molecules, the protein-coding sequences of a major late structural polypeptide are preceded by long leader RNAs, which in some cases are thousands of nucleotides long. These sequences map to different regions of the viral genome and in one instance are separated from the late gene by more than 100 kb of DNA. Moreover, the leader sequences map either upstream or downstream of the late gene, are transcribed from either DNA strand, and are fused to the late gene coding sequence via a poly(A) stretch. This demonstrates that vaccinia virus produces late mRNAs by tagging the protein-coding sequences onto the 3' end of other RNAs.

## Introduction

The 5' flanking sequences of cellular genes transcribed by RNA polymerase II contain conserved sequence elements that are responsible for correct initiation of transcription and regulation of gene expression (Breathnach and Chambon, 1981). In the case of DNA viruses that utilize the host-cell transcription machinery, the 5' flanking regions of the genes closely resemble those of the host cell.

Vaccinia virus, a member of the poxvirus family, belongs to the small group of DNA viruses that replicate in the cytoplasm of the host cell. Gene expression in vaccinia virus is temporally well regulated and occurs in two distinct phases. Early genes are transcribed before the complete uncoating of the viral DNA, and some encode enzymes that are subsequently used for DNA replication. Late genes encode most structural polypeptides and are expressed after DNA replication. Vaccinia virus thus offers a unique opportunity to study eukaryotic transcription and mechanisms of gene regulation that presumably have evolved independently from those of the host cell.

All enzymes required for the production of mRNAs are carried in the virion (reviewed in Moss, 1985), and, although the mRNAs are made in the cytoplasm, they show such typical eukaryotic features as methylated caps (Weir and Moss, 1975) and poly(A) tails (Kates and Beeson, 1970). The mRNAs produced early in infection have an average size of about 1500 nucleotides (Oda and Joklik, 1967; Cooper et al., 1981; Mahr and Roberts, 1984). Furthermore, these molecules are not spliced (Wittek et al.,

1980; Cooper et al., 1981), and processing at the 5' end has also been ruled out (Venkatesan and Moss, 1981).

Late transcription is characterized by several unusual features. First, a large fraction of late RNA can self-hybridize to form double-stranded structures, indicating extensive symmetrical transcription late in infection (Colby and Duesberg, 1969; Duesberg and Colby, 1969; Colby et al., 1971; Boone et al., 1979). Second, late mRNAs are on average about twice as long as early mRNAs and very heterogeneous in size (Oda and Joklik, 1967; Cooper et al., 1981; Mahr and Roberts, 1984). These properties have been explained by a failure of the virus to terminate transcription specifically late in infection (Moss, 1985).

Several studies have demonstrated that the sequences involved in gene regulation in vaccinia virus reside in the 5' flanking region of early and late genes (Weir and Moss, 1984; Bertholet et al., 1985; Cochran et al., 1985). Recent analyses (Bertholet et al., 1986; Hänggi et al., 1986) have revealed unexpected features of the putative promoter region of a strongly expressed late gene that encodes a major structural polypeptide of molecular weight 11,000 (11K gene). Surprisingly, very short stretches, of about 20 nucleotides, preceding the translation initiation codon are sufficient to regulate late gene expression. All mutations around the putative mRNA start site, however, have abolished transcription. In several late genes the putative transcription initiation site has been located within, or very close to, the highly conserved sequence TAAATG, which also includes the translation initiation codon (Weir and Moss, 1984; Bertholet et al., 1985; Hirt et al., 1986; Weinrich and Hruby, 1986). Thus late mRNAs appear to have extremely short nontranslated leader sequences. In all these studies, however, the 5' ends of the mRNAs were mapped by the nuclease S1 procedure. In this communication we show that the 5' end of the 11K mRNA, as defined by this technique, does not represent the true 5' end. Instead, the protein-coding sequences of the late gene are preceded by long, polyadenylated RNAs.

## Results

### Primer Extension of the 11K RNA

To define the 5' end of the 11K late mRNA by an alternative procedure, we performed primer extension experiments. As a control, we included the thymidine kinase (TK) early mRNA, which showed identical map positions for the 5' end by both S1 analysis and primer extension (Bajszar et al., 1983; Weir and Moss, 1983). Appropriate 5' end-labeled DNA fragments (Figure 1A) were hybridized to early or late RNA from infected cells and then either extended with reverse transcriptase or treated with nuclease S1 (Figure 1B). For the TK mRNA, primer extension and S1 nuclease analysis yielded DNA fragments of identical length and of the expected size. A protected fragment of 300 nucleotides was obtained by the nuclease S1 procedure with the 11K mRNA. This places the putative 5' end



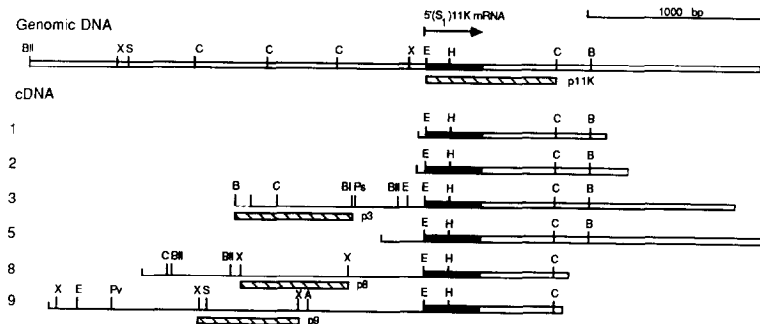


Figure 2. Restriction Maps of Genomic DNA and cDNA Clones

EcoRI (E), HindIII (H), ClaI (C), and BamHI (B) restriction sites were mapped on genomic DNA and cDNA clones. Sequences upstream of the 11K gene sequences in genomic DNA and cDNA clones were also analyzed with the following enzymes: Avall (A), BglI (BI), BglII (BII), PstI (Ps), PvuII (Pv), Sall (S), and XbaI (X). Genomic DNA sequences are represented by boxes; the black parts show the 11K coding sequence. The direction of transcription of the 11K mRNA is indicated by the arrow. Plasmid DNA at the 5' end of cDNA 3 is represented by a dashed line. Hatched boxes (p3, p8, and p9) show the positions of the hybridization probes used in subsequent experiments (Figures 3 and 4).

possibility that they represent cellular RNAs that are induced upon virus infection. RNA was isolated from noninfected or infected cells and then hybridized to <sup>32</sup>P-labeled probes isolated from the 5' end region, upstream of the 11K coding sequences of cDNA clones 3, 8, and 9 (see Figure 2). As a control, a fragment containing the coding region of the 11K gene and 3' flanking sequences was also hybridized to various nitrocellulose-bound RNAs.

Weak hybridization to RNA from uninfected cells was only observed when the filters were washed at low stringency (Figure 3). In contrast, the probes from cDNA clones 3 and 8 hybridized strongly to both early and late RNA from infected cells. Whereas these probes hybridized somewhat more strongly to early than to late RNA, the opposite is true for the probe derived from cDNA 9. Finally, the probe specific for the 11K gene hybridized strongly to late RNA, as expected, but also to early RNA. This may be due to the presence of long early RNAs of unknown significance; these RNAs have been found to be transcribed from various parts of the vaccinia virus genome (Wittek et al., 1980; Cooper et al., 1981), and they may traverse early or late genes on the same or opposite DNA strand.

#### Hybridization of cDNA-Derived Probes to Viral DNA

The hybridization experiment suggested that the sequences upstream of the 11K coding sequences are transcribed from the vaccinia virus genome. To confirm this, and to identify the regions from which they are transcribed, the cDNA-derived probes used in the previous experiment were hybridized to cloned DNA restriction fragments representing the entire vaccinia virus genome (Figure 4). Each probe hybridized to only one fragment; in the case of probes p3 and p8, the fragments map more than 100 kb apart on the genome. Moreover, with respect to the position of the 11K gene, which is located at the junction of the HindIII F and HindIII E fragments (Wittek et al., 1984) and is transcribed from left to right, probe p3 hybridized to sequences located upstream (see below), whereas probes p8 and p9 clearly hybridized to sequences downstream of it.

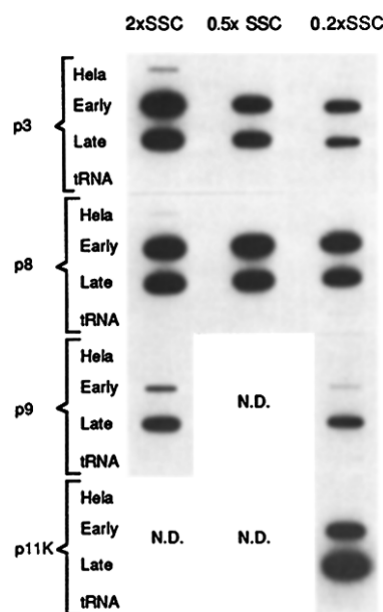


Figure 3. Hybridization of cDNA Probes to RNA from Infected and Noninfected Cells

<sup>32</sup>P-labeled DNA probes (p3, p8, p9, p11K) derived from the 5' region of the corresponding cDNA clones or from genomic DNA (see Figure 2) were hybridized to RNA isolated from uninfected (HeLa) or infected (Early, Late) cells, or to tRNA immobilized on a nitrocellulose membrane. After hybridization, parallel strips were washed at 38°C in 50% formamide, 0.1% SDS, and the indicated SSC concentration (see Experimental Procedures). N.D., not determined.

#### Size Heterogeneity of the 11K RNA Population

The previous experiments demonstrated that various RNAs transcribed from different regions of the viral genome serve as leader RNAs. The 11K mRNA population should thus exhibit considerable length heterogeneity. This was tested by Northern blot analysis (Figure 5). When a DNA fragment from the 11K coding region was used as a hybridization probe (p11K), two long RNA species were detected with early RNA. These presumably represent RNAs transcribed from the DNA strand oppo-

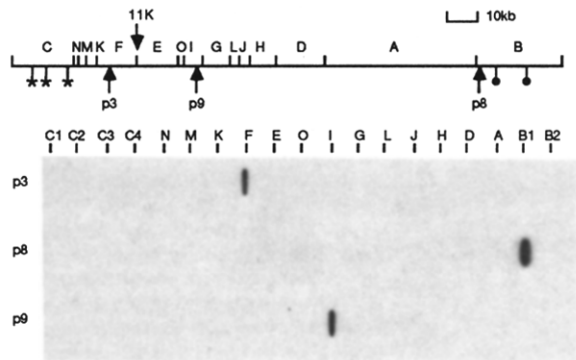
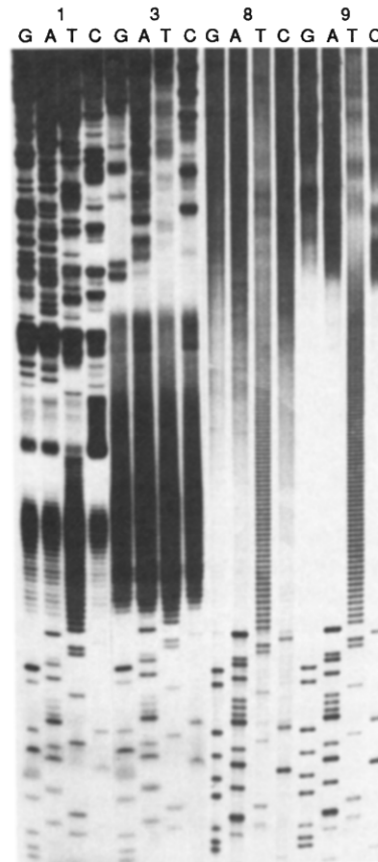


Figure 4. Hybridization of <sup>32</sup>P-labeled cDNA Probes to Vaccinia Virus DNA

Cloned DNA restriction fragments representing the entire vaccinia genome were immobilized on a nitrocellulose membrane, and parallel strips were hybridized to the indicated probes (p3, p8, p9; see Figure 2). A HindIII map of the vaccinia virus genome is also shown, and the fragments to which the particular probes hybridized are indicated. The map position of the 11K gene is also shown. C1-C4 represent four consecutive fragments produced by EcoRI (asterisks) digestion of HindIII fragment C. B1 and B2 are HindIII-Sall (closed circles) and Sall-Sall fragments that together make up almost the entire unique sequences of the HindIII B fragment. (The sequences of the long inverted terminal repeat of the vaccinia virus DNA are in fragments C1 and C2, at the left-hand end).



GenomicDNA: \_\_\_\_\_  
 ... GTTTTTTCTATGCTATA**AAATGAATTCTCATT**...  
 cDNA:  
 ... AAAAAAAAAAAAAAAAAA**AAATGAATTCTCATT**...

Figure 6. Nucleotide Sequence around the Junction of the 11K Coding Sequence and 5' Flanking Region of cDNA Clones 1, 3, 8, and 9

A synthetic oligonucleotide complementary to nucleotides 44 to 65 downstream of the G of the 11K ATG of the noncoding strand was used as a primer for sequencing. Plasmid DNA was used to sequence clones 1 and 3; sequences for clones 8 and 9 were obtained from fragments cloned into single-stranded phage DNA (see Figure 2 for clone designations). The genomic sequence (mRNA-like strand) around the 11K ATG is compared with the corresponding region of the cDNA (bottom). Boldface, underlined letters indicate common sequences, and the bar represents the DNA region to which the 5' end of the 11K mRNA was mapped by nuclease S1 analysis.

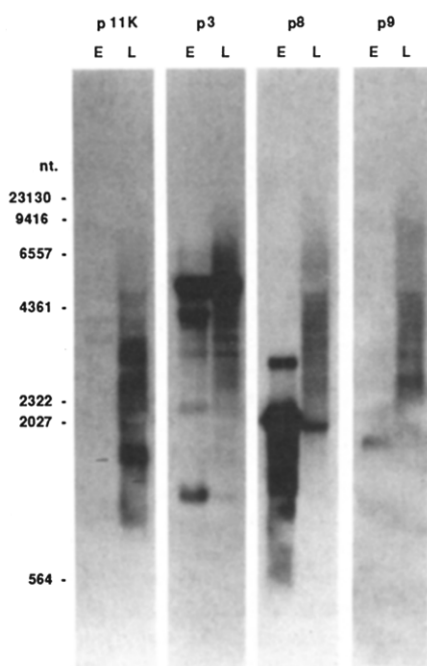


Figure 5. Hybridization of Size-Fractionated RNA to Radioactively Labeled DNA Fragments

Early (E) or late (L) RNA was denatured, electrophoresed in a 1% agarose gel, and transferred to a nitrocellulose membrane. Parallel strips were hybridized to the indicated DNA fragments (p11K, p3, p8, and p9; see also Figure 2). Glyoxylated HindIII fragments of  $\lambda$  DNA were used as length standards. Their sizes in nucleotides (nt.) are indicated at left.

site to the 11K gene previously detected by nuclease S1 analysis (Wittek et al., 1984). With late RNA, a smear was observed over a wide size range. This size heterogeneity is certainly in part due to the well-established length heterogeneity at the 3' end of late RNAs (reviewed by Moss, 1985). On the basis of the previous experiments, however, it is clear that the 5' end also contributes to the overall size heterogeneity. Total early and late RNAs were also hybridized to probes p3, p8, and p9 derived from the leader RNAs. Particularly with probes p8 and p9, smears were seen with late RNA, demonstrating that late RNAs encoded in these parts of the genome also vary in size. Contrary to what one might have expected, probe p11K did not

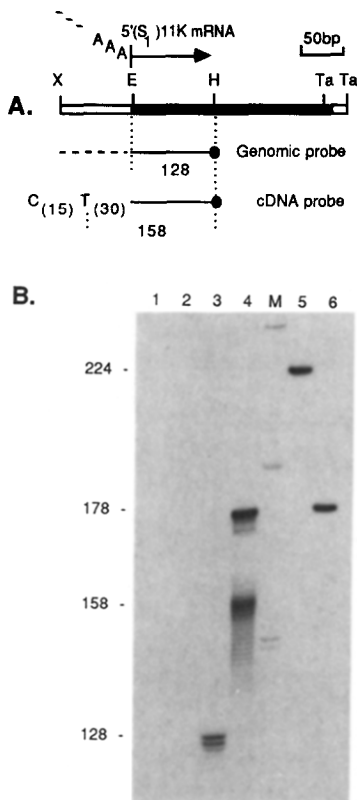


Figure 7. Nuclease S1 Protection Experiment of a cDNA-Derived DNA Fragment

The diagram (A) shows the map location of the 11K coding sequences (black box) and the direction of transcription of the mRNA as well as a part of the 5' poly(A) sequence. The 30 T and 15 C residues immediately upstream of the sequence AAATG (complementary strand), which comprises the 11K ATG, are also indicated in the cDNA probe. (B) The end-labeled (solid circles in [A]) hybridization probes were analyzed on a sequencing gel either directly (lane 5, genomic probe; lane 6, cDNA-derived probe) or after they were hybridized to early RNA and treated with S1 nuclease (lanes 1 and 2). The probes were also hybridized to late RNA, and the S1-resistant material obtained from the genomic probe (lane 3) and cDNA probe (lane 4) was analyzed. The sizes in nucleotides of DNA fragments are shown at left. M, pUC9 HpaII fragments used as size markers. Symbols for restriction sites: X, XbaI; E, EcoRI; H, HindIII; Ta, TaqI.

detect smears that comigrated with those detected with the leader-RNA-derived probes. This is not surprising since, as shown by cDNA cloning, the length heterogeneity contributed by the 5' end of the 11K RNA population results from fusion of different leader RNAs to the 11K coding region. In addition, a given leader RNA as represented by probe p3, p8, or p9 may be fused to various other RNAs, of which the 11K RNA population presumably only represents a small subset. Furthermore, some RNAs detected by the leader RNA probes may themselves be acceptor RNAs to which other leader RNAs are fused.

#### Sequence at the Junction of Leader RNA and 11K Coding Region

To characterize the junction at the 11K and foreign sequences, the relevant portions of the cDNAs were se-

quenced. The cDNA clones 3, 8, and 9 were chosen for this analysis, as was clone 1, which contained only a short stretch of additional DNA upstream of the EcoRI site (Figure 6). An identical sequence, corresponding to the 11K gene (Bertholet et al., 1985), was obtained for all four cDNA clones up to the position of the ATG codon. Beyond this point, clone one had a long poly(A) stretch (on the mRNA-like strand) of about 40 residues, which led to the typical problems in the sequencing reactions described for long homopolymer stretches (Smith, 1980) and which presumably also caused the problems in the sequencing of the RNA (Figure 1C). This region was followed by a poly(G) tract, as expected from the cDNA cloning protocol used. The cDNA clones 8 and 9 also contained long poly(A) sequences of at least 80 residues. This is less obvious for cDNA 3, in which only the first three bases can clearly be identified as A residues.

In the three long cDNAs, the homopolymer stretches appeared to be followed by other sequences, although the actual sequences could not be read in this part of the gel. Appropriate fragments from the 5' end of the cDNA clones were therefore subcloned, and in each case stretches of 300 to 400 nucleotides were sequenced in the opposite direction (not shown). This confirmed that the leader RNAs contain additional sequences upstream of the poly(A) tract.

#### Nuclease S1 Protection of a cDNA-Derived Probe

The primer extension and cDNA cloning experiments described above depended on the use of reverse transcriptase. To rule out the possibility that the unusual structure at the 5' end of the 11K mRNA as revealed by these experiments is due to an artifact generated by reverse transcriptase, we searched for means independent of this enzyme to confirm the observed structures. A nuclease S1 protection experiment was performed using as hybridization probe a DNA fragment derived from cDNA clone 1 (Figure 7A). This fragment was 5' end-labeled at the HindIII site within the coding sequences of the 11K gene, and it contained a stretch of 30 T residues (complementary strand) immediately upstream of the AAATG genomic sequence (see also Figure 6) comprising the 11K translation initiation codon. This sequence is preceded by 15 C residues resulting from the cDNA cloning procedure, and the fragment was isolated after cleavage at the PstI site just upstream of this sequence. A second probe consisted of a XbaI-HindIII fragment derived from genomic DNA and was also labeled at the HindIII site. The genomic (Figure 7B, lane 5) and cDNA-derived probes (lane 6) were analyzed either directly or after hybridization to early RNA and S1 treatment (lanes 1 and 2). As expected, no protected fragments were observed with early RNA and either probe. Note that in contrast to the probe used for hybridization to RNA (Figure 3), the S1 probe was labeled on one strand only and therefore did not detect the early RNA transcribed from the DNA strand opposite to the 11K gene (Wittek et al., 1984). Total late RNA from infected cells protected fragments of 127 and 128 nucleotides of the genomic probe, consistent with the previous experiments, which demonstrated that the sequence complementarity

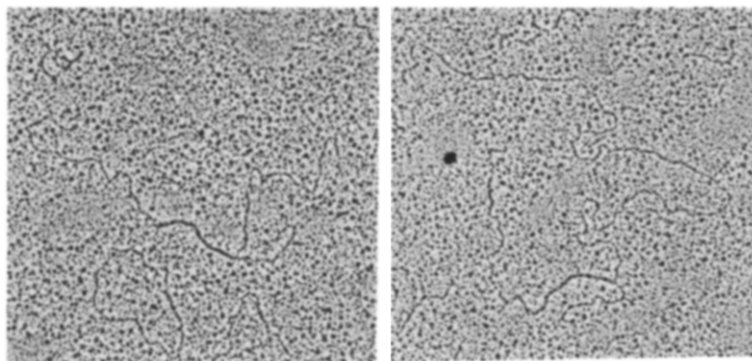
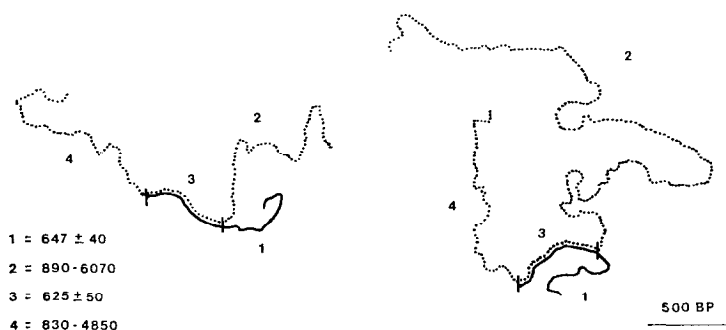


Figure 8. Electron Micrographs of DNA-RNA Hybrids

Hybrid-selected 11K mRNA was annealed to a *Clal* DNA fragment of about 1200 bp containing the 11K gene as well as 5' and 3' flanking sequences. Two typical molecules are shown. Average sizes and standard deviations for regions 1 and 3 are indicated, as are the size ranges for regions 2 and 4. A total of 75 molecules presenting these structures were photographed, and the values were taken from 25 molecules in which there was no ambiguity with respect to the ends of each region. Solid lines, DNA; dashed lines, RNA.



between the genomic DNA and 11K mRNA ends just upstream of the 11K ATG codon. Significantly, protected fragments of about 158 nucleotides were observed with the cDNA-derived probe, as well as some reannealing of the probe (Figure 7B, lane 4). These bands can best be explained by the presence of a poly(A) sequence upstream of the 11K ATG codon. The ladder pattern in the lower part of the gel probably results from nuclease S1 cleavage at internal positions within the poly(A) sequence. However, very little, if any, material was observed at the position of the protected fragments obtained with the genomic probe, suggesting that the majority of the 11K mRNA molecules contain such a poly(A) stretch upstream of the translation initiation codon.

#### Electron Microscopy of DNA-RNA Hybrids

To confirm that the 11K mRNA molecules also contain sequences upstream of the poly(A) tract, we performed electron microscopy of DNA-RNA hybrids. A *Clal* DNA fragment consisting of about 700 bp of DNA downstream of the 11K ATG codon and about 600 bp upstream of the translation initiation codon was isolated. This DNA was hybridized to 11K RNA purified by hybrid selection to DNA of the 11K coding region, and the resulting molecules were analyzed by electron microscopy. Many molecules with a "Y" structure characteristic of the two hybrids shown in Figure 8 were examined. Four different regions can be distinguished, two of which were constant in size and two of which varied in size. We interpret the different regions as follows: Region 1, which is constant in size, represents the single-stranded DNA of the 11K 5' flanking sequence. Region 2 varies in size from 890 to 6070 nucleotides and

represents the leader RNA. The double-stranded region 3 results from hybridization of the 11K coding sequence and 3' flanking sequence to the corresponding region of the RNA. Finally, region 4 is the 3' portion of the RNA. The fact that this region again varies considerably in length is not unexpected from the 3' length heterogeneity of late vaccinia virus transcripts (see Moss, 1985). Thus the results obtained by electron microscopy are consistent with the unusual structure of the 11K mRNA proposed on the basis of the previous experiments.

#### Map Position and Direction of Transcription of Leader RNAs

The following experiment was designed to map precisely the genes encoding the leader RNAs on the vaccinia virus DNA and to determine their direction of transcription. The restriction sites present in cDNA clones 3, 8, and 9 were first mapped within the large *HindIII* fragments to which the corresponding cDNA-derived probes had hybridized (see also Figure 4). Figure 9A shows the map positions of these sequences within the *HindIII* fragments F, B, and I. These are shown in the correct left-to-right orientation with respect to the conventional orientation of the genome (Figure 9B). Thus, the sequences in cDNA clone 3 map to the left-hand half of the *HindIII* F fragment, whereas the 11K gene is located at the extreme right-hand end (Wittek et al., 1984). The sequences in cDNA clone 8 map at the left-hand end of the 30 kb *HindIII* B fragment, and those present in clone 9 map toward the right of the *HindIII* I fragment. Furthermore, from the 5'-to-3' polarity of the leader RNAs established by restriction analysis of the cDNA clones (see also Figure 2), and as indicated by the poly(A) se-

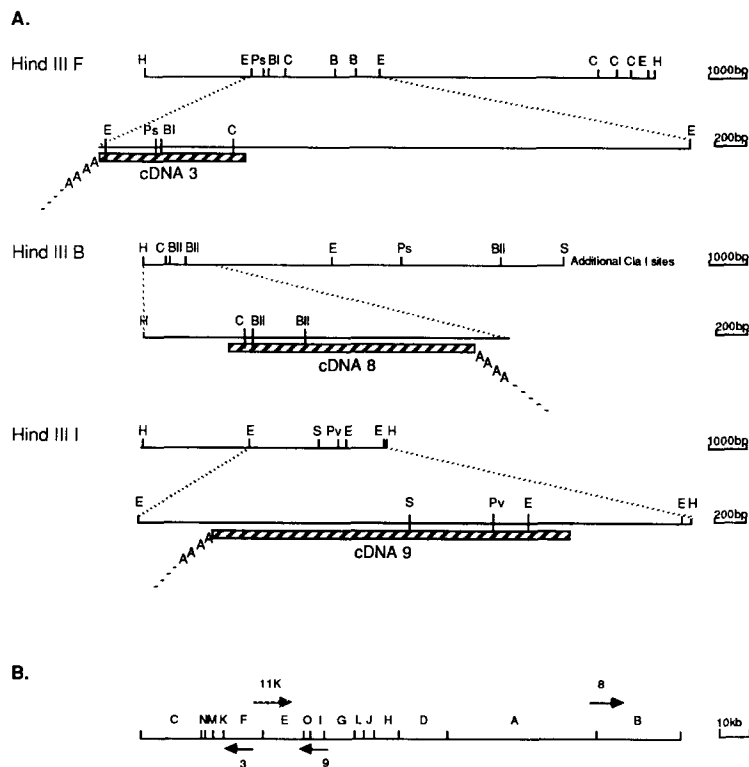


Figure 9. Map Positions and Directions of Transcription of the Leader RNAs Found in Three cDNA Clones

(A) shows the map locations (hatched boxes) of the DNA sequences upstream of the 11K ATG contained in cDNA clones 3, 8, and 9 within the HindIII fragments identified by the hybridization experiment (see also Figure 4). Only the left-hand portion up to the position of the first Sall site (S) is shown for the HindIII B fragment. The 3' end of each leader RNAs is indicated in the cDNAs by a poly(A) sequence. Symbols for restriction sites are as in Figure 2. (B) shows the map positions and direction of transcription of RNAs represented in cDNA clones 3, 8, and 9 (solid arrows). A HindIII map of the vaccinia virus genome is shown, and the map position and direction of transcription of the 11K mRNA are indicated (stippled arrow).

quence (Figure 9A), it is obvious that the leader RNAs in cDNA clones 3 and 9 are transcribed from the leftward-reading strand of the vaccinia virus genome (Figure 9B). On the other hand, the 11K mRNA is transcribed in the opposite direction (Witek et al., 1984), as is the leader RNA of cDNA clone 8.

**Discussion**

We have shown that late in infection vaccinia virus produces chimeric mRNAs consisting of the coding sequences for a major structural polypeptide downstream of poly(A) sequences. We have also found cDNA clones that contain only the poly(A) stretch, and thus lack the sequences preceding it. It is possible that these molecules result from premature termination of reverse transcription during cDNA cloning. It is known, at least for the Klenow fragment of DNA polymerase, that long homopolymer stretches are not easily copied (Smith, 1980). This interpretation is also consistent with the results obtained by electron microscopy, where we have found no evidence for RNA molecules that had only very short sequences upstream of the 11K coding sequences. On the contrary, most molecules observed possessed even longer leader RNAs than those found by cDNA cloning. Both procedures also confirmed the 3' length heterogeneity of late RNA. We have sequenced the 3' ends of five cDNA clones and have found stretches of between 30 and 100 A residues (not shown). These ends, therefore, most likely represent the true 3' ends and did not result from priming at internal positions in the RNA molecules. On the other

hand, it is clear from the data presented here that the 5' end also contributes significantly to the overall length heterogeneity of late RNA.

An important question is whether this unusual mechanism of RNA synthesis is unique to the particular late gene examined. A preliminary analysis of cDNA clones for another late gene that we have recently mapped and sequenced (Hirt et al., 1986) has shown that they also contain a poly(A) stretch and additional sequences upstream of the translation initiation codon that are not transcribed from DNA in the immediate vicinity of the gene.

Perhaps the most intriguing aspect of the bizarre structure of the 11K mRNA concerns its translation. In most eukaryotic mRNAs the AUG closest to the 5' end is used for translation initiation (Kozak, 1978), although exceptions to this rule, particularly for mRNAs of viral origin (reviewed in Kozak, 1986), have been described. If the 11K polypeptide is indeed translated from the mRNAs described in this paper, the translation initiation codon would be located in several cases thousands of nucleotides downstream of the 5' end. Furthermore, the leader sequences certainly contain several AUGs. One might therefore expect a poor translation efficiency of the 11K sequences. However, the opposite appears to be true. From the abundance of the polypeptide, the 11K gene is judged to be one of the most strongly expressed vaccinia virus genes, since its product contributes about 10% to the total protein mass present in purified virions (Sarov and Joklik, 1972; Moss, 1974). Furthermore, in vitro translation of hybrid-selected RNA, which was used to map the 11K gene (Witek et al., 1984), yields large amounts of the polypeptide, indicating either



that the mRNA is present at high concentration or that it is particularly well translated. We cannot exclude the possibility that the 11K polypeptide is translated from mRNAs with other structures than those reported here. However, the results of this study strongly suggest that the observed structures are representative of the 11K mRNA population. In fact, a preliminary experiment (not shown), in which we were able to translate the 11K polypeptide from RNA that was first selected by hybridization to the cDNA of one of the leader RNAs and then on the 11K coding sequences, suggests that such chimeric RNAs may indeed be functional 11K mRNAs.

Another interesting question is whether the leader RNA sequences are also translated. At present this question cannot be answered, since in the previous mapping of the 11K gene (Wittek et al., 1984) we used immunoprecipitation of the *in vitro* translation products of hybrid-selected RNA to identify the 11K polypeptide. Any additional polypeptides translated from 11K mRNA would therefore have been lost. It should be borne in mind, however, that these unusual mRNAs are translated in infected cells, and results obtained in a reticulocyte lysate might not reflect the *in vivo* situation.

The polyadenylated sequences preceding the late gene coding region most likely represent early or late mRNAs. This interpretation is in agreement with the results of Oda and Joklik (1967), who concluded 20 years ago on the basis of hybridization competition experiments that late vaccinia virus mRNA contains sequences transcribed early. If early mRNAs are indeed used for the production of late mRNAs, this could explain the shutoff of early gene expression late in infection, which has been particularly well studied in the case of the viral TK gene (McAuslan, 1963a, 1963b; Jungwirth and Joklik, 1965; Zaslavsky and Yakobson, 1975; Hruby and Ball, 1981). Synthesis of this enzyme is switched off about 4 hr after infection, when late genes begin to be expressed. Nevertheless, functional TK mRNA, identified by *in vitro* translation, can be isolated from cells that have stopped making the enzyme. This discrepancy could be explained by a possible fusion of the early mRNA to late coding sequences and by assuming that late in infection the latter are translated preferentially, but that *in vitro* systems also allow translation of the upstream coding sequences.

The fact that mRNAs contain 5' leader sequences that are not derived from the immediate vicinity of the genes is not without precedent. Perhaps the best-studied example is influenza virus, which uses the first 10–13 nucleotides of cellular mRNAs, including the cap structure, to prime the synthesis of its mRNAs (Plotch et al., 1981; Herz et al., 1981; reviewed in Krug, 1985), which thus consist of cellular and viral sequences. A similar mechanism has also been reported for bunyavirus (Patterson et al., 1984). Coronaviruses use a virus-encoded leader RNA of about 70 nucleotides to prime the synthesis of the individual mRNAs on the template molecule (Lai et al., 1984; Baric et al., 1985; reviewed in Krug, 1985). A nonviral example is the mRNAs of trypanosomes, which contain an identical 35 nucleotide leader sequence (reviewed in Borst, 1986; Van der Ploeg, 1986). It is not clear how the leader

is added to the body of the mRNA, but *trans* splicing appears to be the most likely mechanism (Murphy et al., 1986; Sutton and Boothroyd, 1986).

We can only speculate on the mechanism by which the long polyadenylated RNAs are fused to the late coding sequences in vaccinia virus. Splicing of a single, very long precursor molecule can be excluded, since the sequences added to the body of the late mRNA are transcribed from either DNA strand and are located either upstream or downstream of the 11K gene. A priming mechanism can be envisaged in which the poly(A) sequence of the donor RNA interacts with sequences on the DNA coding strand. An alternative hypothesis is that the two RNAs are joined together by ligation. One would then have to explain how transcription of the late coding sequences is initiated. Although it cannot be excluded that the upstream sequences act as promoters, we prefer an alternative model in which the highly conserved 5'-TAA-ATG-3' sequences represent processing sites at the RNA level. It is well established that very long transcripts are made particularly late in infection, and from both DNA strands (reviewed in Moss, 1985). These transcripts might represent precursor molecules from which the late coding sequences are excised by a site-specific endoribonuclease. Interestingly, an enzyme that appears to possess the required properties has been isolated from vaccinia virions and partially characterized (Paoletti and Lipinskas, 1978).

Clearly, further work is needed to understand the mechanism by which donor and acceptor RNAs become joined. However, in comparison with trypanosomes, vaccinia virus represents a relatively simple system that should greatly facilitate the unraveling of this unusual mechanism of RNA production.

## Experimental Procedures

### Virus and Cells

The WR strain of vaccinia virus was obtained from Bernard Moss, National Institutes of Health, Bethesda, Md. HeLa cells were grown as monolayer cultures in Dulbecco's modified Eagle's minimal essential medium supplemented with 5% fetal calf serum.

### RNA Extraction

Total RNA from infected HeLa cells was isolated after lysis of the cells with 8 M guanidinium hydrochloride exactly as previously described (Wittek et al., 1984). Early RNA was isolated at 4 hr after infection from cells maintained in medium containing either 100 µg/ml of cycloheximide or 40 µg/ml of cytosine arabinoside. Late RNA was extracted at 7 hr after infection from cells that were not treated with either inhibitor.

### Primer Extension and RNA Sequencing

For primer extension, 20 µg of total RNA was hybridized to 2–4 ng of 5' end-labeled DNA fragments in 40 µl containing 80% formamide, 0.4 M NaCl, 0.04 M PIPES (pH 6.4), and 0.001 M EDTA, at 42°C for 3–4 hr. Reverse transcription was performed for 1 hr at 40°C in 20 µl reactions containing 50 mM Tris-HCl (pH 8.3), 50 mM KCl, 6 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, 0.25 mM each of dATP, dTTP, dCTP, and dGTP, and 12 units of reverse transcriptase (Life Sciences, St. Petersburg, Fla.). The reaction was stopped by the addition of NaOH to a final concentration of 0.166 M. After incubation for 10 min at 37°C, the solution was neutralized and nucleic acids were precipitated with ethanol. For RNA sequencing, reverse transcription was performed for 10 min in the presence of 12 µM of each dNTP and 12.5 µM ddATP, 50 µM

ddCTP, 50  $\mu$ M ddGTP, or 12  $\mu$ M ddTTP. After the concentration of all dNTPs was raised to 125  $\mu$ M, synthesis was continued for 1 hr.

#### Nuclease S1 Analysis

The 5' ends of RNA were mapped by S1 analysis (Berk and Sharp, 1977) using 5' end-labeled DNA fragments as hybridization probes (Weaver and Weissman, 1979). The experimental details were as previously described (Wittek et al., 1984).

#### cDNA Cloning and Identification of cDNA Clones

Poly(A)-containing cytoplasmic RNA from infected cells isolated late in infection was annealed to oligo(dT)-tailed plasmid pcDV-1 vector DNA (Okayama and Berg, 1983), purchased from Pharmacia. All subsequent steps were performed exactly as described in the detailed protocol supplied by Pharmacia, except that JM109 host bacteria were used for the initial plasmid amplification. Desired cDNA clones were identified by colony hybridization using an EcoRI-HindIII DNA fragment located at the beginning of the 11K coding sequences (Bertholet et al., 1985); the fragment was  $^{32}$ P-labeled by repair synthesis (Maniatis et al., 1982). DNA from positive colonies was isolated, and used to transform competent DH5.1 host cells (Vector Cloning Systems, San Diego, Cal.).

#### Slot Blot Hybridization

DNA or RNA was immobilized on nitrocellulose membranes using a minifold apparatus (Schleicher and Schuell). The filters were dried at 80°C for 2 hr and prehybridized in a solution consisting of 150  $\mu$ g/ml of denatured herring sperm DNA, 5 $\times$  SSC (SSC is 0.15 M NaCl, 0.015 M sodium citrate), 25 mM sodium pyrophosphate, 1 $\times$  Denhardt's solution (Denhardt, 1966), and 50% formamide, at 40°C for 4–12 hr. Hybridizations were carried out at 40°C for 16–90 hr with  $^{32}$ P-labeled DNA probes (Rigby et al., 1977) that were denatured and diluted into prehybridization solution. For RNA–DNA hybridizations (Figure 3), parallel filters were washed at 38°C in 50% formamide, 0.1% SDS, and either 2 $\times$  SSC, 0.5 $\times$  SSC, or 0.2 $\times$  SSC. For DNA–DNA hybridizations (Figure 4), the filters were washed at 38°C in 50% formamide, 0.1% SDS, 0.2 $\times$  SSC.

#### Northern Blot Analysis

For Northern blot analysis, 10  $\mu$ g of total early or late RNA was denatured with glyoxal as previously described (McMaster and Carmichael, 1977) and then size-fractionated by electrophoresis in a 1% agarose gel. The RNA was then transferred to nitrocellulose membranes (Thomas, 1980) and hybridized to DNA fragments that had been  $^{32}$ P-labeled by the random-primer procedure (Feinberg and Vogelstein, 1983). Prehybridization, hybridization, and washing of the membranes in 0.5 $\times$  SSC were performed as described for slot blot hybridization.

#### DNA Sequencing

Sequencing of DNA was performed by the chain-termination method (Sanger et al., 1977) either on linearized, denatured plasmid DNA or after appropriate DNA fragments were subcloned on single-stranded recombinant phage WB238 DNA (Barnes and Bevan, 1983; Sahli et al., 1985).

#### Electron Microscopy

For electron microscopy, 11K-specific RNA was selected by hybridization to a single-stranded HindIII–C1a1 fragment containing the 11K coding sequences. The DNA was covalently bound to Sephacryl S-1000 (Pharmacia) by the procedure of Seed (1982) with modifications (Bünemann et al., 1982; Bünemann, 1982). A C1a1 DNA fragment spanning the 11K gene (Wittek et al., 1984) was purified from an agarose gel. DNA–RNA hybrids were formed as described by Brack et al. (1978). RNA and DNA were dissolved at 10  $\mu$ g/ml, and 1  $\mu$ l of each was diluted into 14  $\mu$ l of 76% formamide, 100 mM PIPES–NaOH (pH 7.8), 20 mM Tris–HCl (pH 7.8), 4 mM EDTA, 500 mM NaCl. Samples were heated to 80°C for 10 min and then incubated at 49°C for 1 hr. The mixture was then diluted with 10 volumes of 70% formamide, 100 mM Tris–HCl (pH 8.5), 10 mM EDTA, 50  $\mu$ g/ml of cytochrome C, and spread on distilled water as hypophase. The samples were processed for electron microscopy as previously described (Brack et al., 1978). Plasmid pBR322 and phage  $\phi$ X174 DNA were included as double- and single-stranded length standards, respectively.

#### Acknowledgments

We are grateful to Anne Seiler-Tuyns and Philippe Walker for the stimulating discussions that initiated this work, to Daniel Schümperli for suggesting to us the S1 protection experiment, and to Walter Wahli for advice concerning electron microscopy. We would also like to thank Claude Mancini for technical assistance, Bob Hipskind, Deoki Tripathy, and Walter Wahli for comments on the manuscript, Len Archard for materials, and Nelly Buechi and Hannelore Pagel for secretarial help. This work was supported by the Swiss National Science Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 22, 1986; revised April 13, 1987.

#### References

- Bajsar, G., Wittek, R., Weir, J. P., and Moss, B. (1983). Vaccinia virus thymidine kinase and neighboring genes: mRNAs and polypeptides of wild-type virus and putative nonsense mutants. *J. Virol.* 45, 62–72.
- Baric, R. S., Stohlman, S. A., Razavi, M. K., and Lai, M. M. C. (1985). Characterization of leader-related small RNAs in coronavirus-infected cells: further evidence for leader-primed mechanism of transcription. *Virus Res.* 3, 19–33.
- Barnes, W. M., and Bevan, M. (1983). Kilo-sequencing: an ordered strategy for rapid DNA sequence data acquisition. *Nucl. Acids Res.* 11, 349–368.
- Berk, A. J., and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* 12, 721–732.
- Bertholet, C., Drillien, R., and Wittek, R. (1985). One hundred base pairs of 5' flanking sequence of a vaccinia virus late gene are sufficient to temporally regulate late transcription. *Proc. Natl. Acad. Sci. USA* 82, 2096–2100.
- Bertholet, C., Stocco, P., Van Meir, E., and Wittek, R. (1986). Functional analysis of the 5' flanking sequence of a vaccinia virus late gene. *EMBO J.* 5, 1951–1957.
- Boone, R. F., Parr, R. P., and Moss, B. (1979). Intermolecular duplexes formed from polyadenylated vaccinia virus RNA. *J. Virol.* 30, 365–374.
- Borst, P. (1986). Discontinuous transcription and antigenic variation in trypanosomes. *Ann. Rev. Biochem.* 55, 701–732.
- Brack, C., Hiram, M., Lenhard-Schuller, R., and Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. *Cell* 15, 1–14.
- Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50, 349–383.
- Bünemann, H. (1982). Immobilization of denatured DNA to macroporous supports: II. Steric and kinetic parameters of heterogeneous hybridization reactions. *Nucl. Acids Res.* 10, 7181–7196.
- Bünemann, H., Westhoff, P., and Herrmann, R. G. (1982). Immobilization of denatured DNA to macroporous supports: I. Efficiency of different coupling procedures. *Nucl. Acids Res.* 10, 7163–7180.
- Cochran, M. A., Puckett, C., and Moss, B. (1985). In vitro mutagenesis of the promoter region for a vaccinia virus gene: evidence for tandem early and late regulatory signals. *J. Virol.* 54, 30–37.
- Colby, C., and Duesberg, P. H. (1969). Double-stranded RNA in vaccinia virus infected cells. *Nature* 222, 940–944.
- Colby, C., Jurale, C., and Kates, J. R. (1971). Mechanism of synthesis of vaccinia virus double-stranded ribonucleic acid in vivo and in vitro. *J. Virol.* 7, 71–76.
- Cooper, J., Wittek, R., and Moss, B. (1981). Extension of the transcriptional and translational map of the left end of the vaccinia virus genome to 21 kilobase pairs. *J. Virol.* 39, 733–745.
- Denhardt, A. T. (1966). A member filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Comm.* 23, 641–646.
- Duesberg, P. H., and Colby, C. (1969). On the biosynthesis and struc-

- ture of double-stranded RNA in vaccinia virus-infected cells. Proc. Natl. Acad. Sci. USA 64, 393-403.
- Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6-13.
- Hänggi, M., Bannwarth, W., and Stunnenberg, H. G. (1986). Conserved TAAAT motif in vaccinia virus late promoters: overlapping TATA box and site of transcription initiation. EMBO J. 5, 1071-1076.
- Herz, C., Stavnezer, E., Krug, R. M., and Gurney, T., Jr. (1981). Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells. Cell 26, 391-400.
- Hirt, P., Hiller, G., and Wittek, R. (1986). Localization and fine structure of a vaccinia virus gene encoding an envelope antigen. J. Virol. 58, 757-764.
- Hruby, D. E., and Ball, L. A. (1981). Control of expression of the vaccinia virus thymidine kinase gene. J. Virol. 40, 456-464.
- Jungwirth, C., and Joklik, W. K. (1965). Studies on "early" enzymes in HeLa cells infected with vaccinia virus. Virology 27, 80-93.
- Kates, J., and Beeson, J. (1970). Ribonucleic acid synthesis in vaccinia virus. II. Synthesis of polyriboadenylic acid. J. Mol. Biol. 50, 19-23.
- Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell 15, 1109-1123.
- Kozak, M. (1986). Bifunctional messenger RNAs in eukaryotes. Cell 47, 481-483.
- Krug, R. M. (1985). The role of RNA priming in viral and trypanosomal mRNA synthesis. Cell 47, 651-652.
- Lai, M. M. C., Baric, R. S., Brayton, P. R., and Stohman, S. A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. Proc. Natl. Acad. Sci. USA 81, 3626-3630.
- Mahr, A., and Roberts, B. (1984). Arrangement of late mRNAs transcribed from a 7.1-kilobase EcoRI vaccinia virus DNA fragment. J. Virol. 49, 510-520.
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
- McAuslan, B. R. (1963a). Control of induced thymidine kinase activity in the poxvirus infected cell. Virology 20, 162-168.
- McAuslan, B. R. (1963b). The induction and repression of thymidine kinase in the poxvirus-infected HeLa cell. Virology 21, 383-389.
- McMaster, G. K., and Carmichael, G. G. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. USA 74, 4835-4838.
- Moss, B. (1974). Reproduction of poxviruses. In Comprehensive Virology, vol. 3, H. Fraenkel-Conrat and R. R. Wagner, eds. (New York: Plenum Publishing Corp.), pp. 405-473.
- Moss, B. (1985). Replication of poxviruses. In Virology, B. N. Fields, R. M. Chanock, and B. Roizman, eds. (New York: Raven Press), pp. 685-703.
- Murphy, W. J., Watkins, K. P., and Agabian, N. (1986). Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. Cell 47, 517-525.
- Oda, K., and Joklik, W. K. (1967). Hybridization and sedimentation studies on "early" and "late" vaccinia messenger RNA. J. Mol. Biol. 27, 395-419.
- Okayama, H., and Berg, P. (1983). A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell. Biol. 3, 280-289.
- Paoletti, E., and Lipinskas, B. R. (1978). Soluble endoribonuclease activity from vaccinia virus: specific cleavage of virion-associated high-molecular weight RNA. J. Virol. 26, 822-824.
- Patterson, J. L., Holloway, B., and Kolakofsky, D. (1984). La crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease. J. Virol. 52, 215-222.
- Plotch, S. J., Bouloy, M., Ulmanen, I., and Krug, R. M. (1981). A unique cap ( $m^7GpppXm$ )-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell 23, 847-858.
- Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113, 237-251.
- Sahli, R., McMaster, G. K., and Hirt, B. (1985). DNA sequence comparison between two tissue-specific variants of the autonomous parvovirus, minute virus of mice. Nucl. Acids Res. 13, 3617-3633.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Sarov, I., and Joklik, W. K. (1972). Studies on the nature and location of the capsid polypeptides of vaccinia virions. Virology 50, 579-592.
- Seed, B. (1982). Diazotizable arylamine cellulose papers for the coupling and hybridization of nucleic acids. Nucl. Acids Res. 10, 1799-1810.
- Smith, A. J. H. (1980). DNA sequence analysis by primed synthesis. Methods Enzymol. 65, 560-580.
- Sutton, R. E., and Boothroyd, J. C. (1986). Evidence for *trans* splicing in trypanosomes. Cell 47, 527-535.
- Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA 77, 5201-5205.
- Van der Ploeg, L. H. T. (1986). Discontinuous transcription and splicing in trypanosomes. Cell 47, 479-480.
- Venkatesan, S., and Moss, B. (1981). In vitro transcription of the inverted terminal repetition of the vaccinia virus genome: correspondence of initiation and cap sites. J. Virol. 37, 738-747.
- Weaver, R. F., and Weissman, C. (1979). Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of  $\beta$ -globin mRNA have identical map coordinates. Nucl. Acids Res. 7, 1175-1193.
- Wei, C. M., and Moss, B. (1975). Methylated nucleotides block 5'-terminus of vaccinia virus mRNA. Proc. Natl. Acad. Sci. USA 72, 318-322.
- Weinrich, S. L., and Hruby, D. E. (1986). A tandemly-oriented late gene cluster within the vaccinia virus genome. Nucl. Acids Res. 14, 3003-3016.
- Weir, J. P., and Moss, B. (1983). Nucleotide sequence of the vaccinia virus thymidine kinase gene and the nature of spontaneous frameshift mutations. J. Virol. 46, 530-537.
- Weir, J. P., and Moss, B. (1984). Regulation of expression and nucleotide sequence of a late vaccinia virus gene. J. Virol. 51, 662-669.
- Wittek, R., Cooper, J., Barbosa, E., and Moss, B. (1980). Expression of the vaccinia virus genome: analysis and mapping of mRNAs encoded within the inverted terminal repetition. Cell 27, 487-493.
- Wittek, R., Hänggi, M., and Hiller, G. (1984). Mapping of a gene coding for a major late structural polypeptide on the vaccinia virus genome. J. Virol. 49, 371-378.
- Zaslavsky, V., and Jakobson, E. (1975). Control of thymidine kinase synthesis in IHD vaccinia virus-infected thymidine kinase-deficient LM cells. J. Virol. 16, 210-213.