

## Arrangement of Late RNAs Transcribed from a 7.1-Kilobase EcoRI Vaccinia Virus DNA Fragment

ANNA MAHR<sup>†\*</sup> AND BRYAN E. ROBERTS

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

Received 6 June 1983/Accepted 31 October 1983

Three late transcripts have been mapped to the vaccinia virus 7.1-kilobase (kb) *EcoRI F* fragment, which is located approximately 85 kb from the left end of the viral genome. Hybrid selection and translation of RNA have demonstrated that these mRNAs encode three polypeptides with apparent molecular weights of 32,000 (32K), 26K, and 19K. The transcripts encoding the 32K and 19K polypeptides are synthesized in a leftward direction and overlap each other as well as the RNA encoding the early 110K polypeptide. In addition, the RNA encoding the 32K polypeptide also overlaps an early transcript of 1.35 kb (see accompanying paper). The transcript encoding the 26K polypeptide overlaps the early 2.45-, 0.6- to 0.7-, and 1.7-kb RNAs, and all four mRNAs are transcribed from left to right. The protein coding sequences for the 26K and 32K polypeptides lie outside the 7.1-kb DNA fragment. Due to their heterogeneity in size, each of the three late transcripts is undetectable as a distinct size by filter hybridization. In addition, although S1 mapping has detected the 5' terminus of the late RNA which maps entirely within this fragment, it has been unable to size and locate the 3' ends of all three transcripts, and this result indicates that the heterodisperse size of the RNAs is due to heterogeneity at the 3' ends of these transcripts. The cause of this heterogeneity in termination of transcription is not known.

Crucial to elucidating the mechanisms by which vaccinia virus replicates and turns on its transcriptional program in the cytoplasm of infected cells is an understanding of the structure of its transcripts and how they are arranged and regulated. For this reason, we have determined the transcriptional organization of a 7.1-kilobase (kb) *EcoRI F* DNA fragment to which a number of early and late transcripts map.

The mapping of the four major early mRNAs and the polypeptides they encode has been described in the preceding paper (13). These four messages are all transcribed in the same direction, with their 5' ends towards the vaccinia virus *HindIII C* fragment (WR strain) on the left. A 0.6- to 0.75-kb transcript is immediately followed by a 1.7-kb transcript and then by a 3.6-kb transcript; the 1.7-kb mRNA overlaps a 1.25-kb message. These four messages code for polypeptides with apparent molecular weights of 19,000, 36,000, 110,000, and 22,000 (19K, etc.), respectively. The 19K polypeptide has been determined to be thymidine kinase and has been assigned a molecular weight of 19K by others (11, 23). The present paper deals with the organization of the late transcripts and polypeptides which map to this fragment. A late vaccinia virus gene is defined as one whose transcription is dependent on DNA synthesis (15, 20), and only two (9, 24) late vaccinia virus mRNAs have been studied.

In this study, we present the arrangement of three late vaccinia virus mRNAs transcribed from the 7.1-kb *EcoRI F* DNA fragment which is located near the central region of the viral genome. The experiments described in this paper show that the organization of three heterogeneously sized late mRNAs has features which are found in the arrangement of the early RNAs which are transcribed from this DNA fragment. The functions of the encoded late polypeptides remain to be determined.

### MATERIALS AND METHODS

**Cells and viruses.** Vaccinia virus (WR strain) was grown on L cell monolayers (13).

**Purification of nucleic acids.** Nucleic acids were isolated, purified, and labeled as described in the previous paper (13). Vaccinia virus DNA inserted into lambda or into plasmids was isolated from agarose and polyacrylamide gels after cleavage with restriction endonucleases. The DNA was labeled *in vitro* by nick translation, 5' end labeling, or by 3' end labeling, and plasmid DNA was uniformly labeled *in vivo*.

RNA was isolated by extraction with phenol or by the guanidine thiocyanate method (13).

**Filter hybridization.** RNA was fractionated on methyl mercury gels and transferred to DBM paper (2). Alternatively, RNA was fractionated on formaldehyde gels, transferred to a Gene Screen membrane, and hybridized according to the protocol supplied by New England Nuclear Corp.

**Cell-free translation.** RNA was translated in a reticulocyte cell-free system (17), using [<sup>35</sup>S]methionine (New England Nuclear), and the translation products were fractionated on sodium dodecyl sulfate (SDS)-polyacrylamide gels (12).

RNA was fractionated by size on a denaturing gel (3), extracted, and translated. Hybrid selection of RNA was carried out as described previously (13).

**Hybrid arrested cell-free translation (HART).** HARTs were done by using the procedures of Cooper et al. (8). RNA and DNA were precipitated together with 25 µg of tRNA, and the pellet was dissolved in 45 µl of 89% formamide-10 mM sodium piperazine-*N,N'*-bis(2-ethanesulfonic acid) (PIPES) (pH 6.4). The nucleic acids were heated to 100°C for 2 min to denature the DNA, and 5 µl of 4 M NaCl-0.4 M sodium PIPES (pH 6.4)-10 mM EDTA was added to the mixture, which was then incubated at 37°C for 3 h. After the addition of 350 µl of ice-cold water, half of the reaction was boiled for 2 min. Sodium acetate was then added to a final concentration of 0.2 M, both tubes containing the hybrid and the now

\* Corresponding author.

† Present address: Integrated Genetics, Framingham, MA 01701.

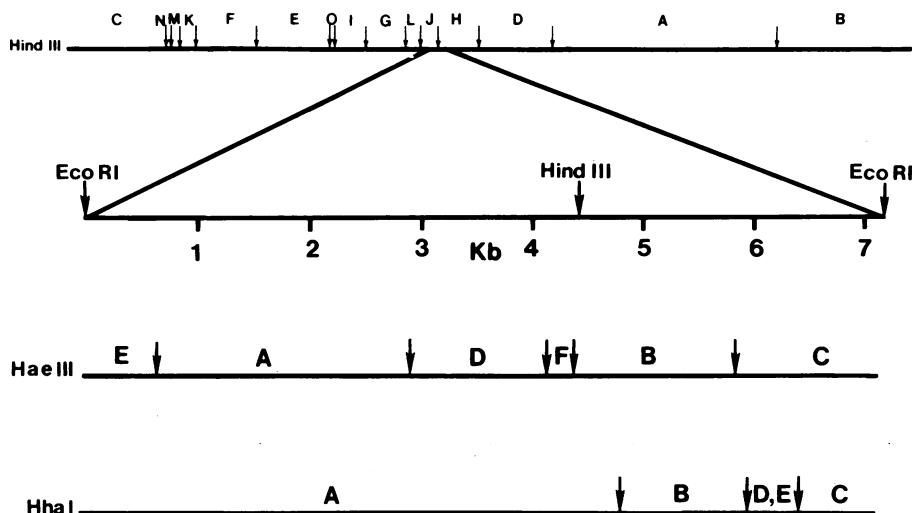


FIG. 1. Location and restriction map of the 7.1-kb DNA. The *Hind*III map has been published previously (25), and the restriction map for the 7.1-kb DNA was taken from the accompanying paper (13).

melted hybrid were precipitated with ethanol, and the RNA was translated in the reticulocyte system.

**S1 nuclease mapping.** The S1 nuclease mapping procedure of Sharp et al. (21) was followed as described in the preceding paper (13).

## RESULTS

**Location and restriction map of the 7.1-kb DNA fragment.** This DNA fragment maps to the vaccinia virus *Hind*III H

and J fragments as previously described (13), and the restriction map is reviewed in Fig. 1.

**Early and late transcripts map to this fragment.** Hybrid selection of RNA isolated at different points during the infection cycle showed that there were early transcripts present throughout the course of infection as well as other transcripts which were detectable only during late infection (13). The results of cell-free translation of early or late RNA that has been hybrid selected with 7.1-kb DNA immobilized on DBM paper show that there were three transcripts encoding proteins with apparent molecular weights of 32K, 26K, and 19K which appeared during late infection (Fig. 2A).

**Orientation of transcripts.** The direction of transcription of the late RNAs was determined by hybrid selection to the separated DNA strands. The separated DNA strands of a lambda clone containing 7.1- and 4.0-kb vaccinia virus inserts were fractionated by electrophoresis on an agarose gel, transferred to DBM paper, and used to hybrid select late mRNA. The mRNAs encoding the early 36K and 22K proteins and the late 26K protein hybridized to the faster-migrating strand (Fig. 2B), indicating that the transcript for the 26K polypeptide was synthesized in the same direction as the messages for the early 36K and 22K polypeptides, i.e., from left to right. However, the 32K and 19K polypeptides appeared to be encoded equally by the RNA hybridizing to both DNA strands, making it impossible to conclude from which DNA strand these two mRNAs were transcribed. HART and S1 nuclease mapping data to be presented later indicated that these mRNAs were transcribed from right to left. The late 95K polypeptide visible in this experiment mapped to the 4.0-kb insert in this lambda clone (13).

**Hybrid selections with *Hae*III and *Hha*I DNA fragments.** Digestion of the 7.1-kb DNA fragment or its *Hind*III-EcoRI subclones with the *Hae*III or *Hha*I restriction enzymes resulted in six and five fragments, respectively, which were then fractionated and transferred to DBM paper. These filters were used to hybrid select late RNA to determine the map positions of the transcripts encoding the three identified late polypeptides (Fig. 3). The 26K polypeptide was encoded by a transcript which was hybrid selected by the *Hha*-A and *Hae*III E and A DNA fragments; these fragments also hybrid selected the RNA for the early 36K polypeptides, and this

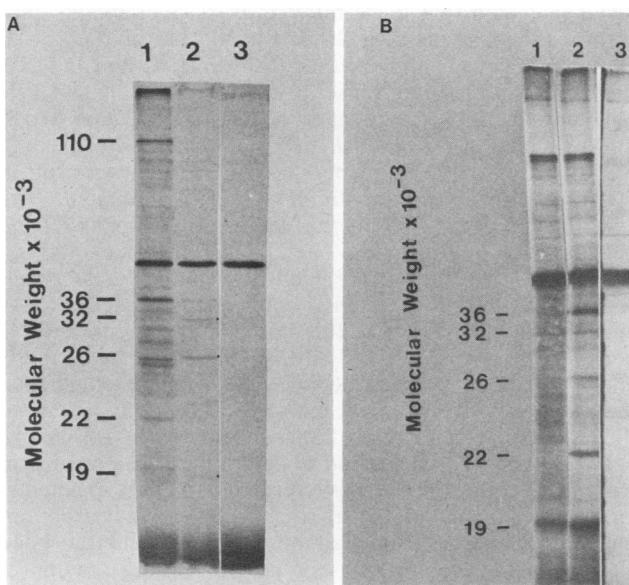


FIG. 2. Hybrid-selection of early and late RNA. (A) The 7.1-kb DNA was immobilized onto DBM paper and used to select cycloheximide or 8-h RNA. This RNA was then translated in a reticulocyte cell-free system and fractionated on an SDS gel. Lane 1, early; lane 2, late; lane 3, endogenous. (B) Determination of direction of transcription by hybrid selection. Each separated strand of the DNA of lambda clone 15, containing the 7.1-kb DNA as one of two inserts, was used to select late RNA. The orientation of these two strands had been previously determined. Lane 1, slower strand; lane 2, faster strand; lane 3, endogenous.

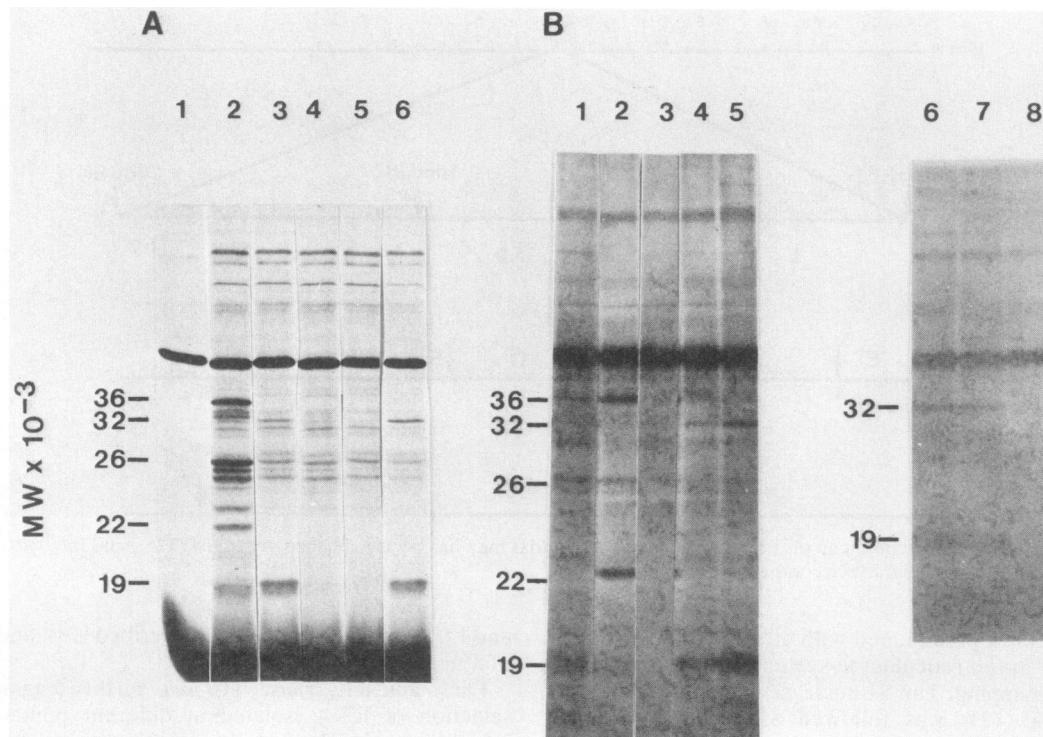


FIG. 3. Selection and cell-free translation of mRNA hybridizing to the 7.1-kb DNA fragment. The 7.1-kb DNA fragment was cleaved with *Hae*III or *Hha*I, and the fragments produced were fractionated on polyacrylamide or agarose gels and transferred to DBM paper. The filters were used to hybrid select late (8 h) RNA which was translated in the reticulocyte system. (A) Polypeptide products encoded by the RNA hybridizing to the *Hha*I DNA fragments: lane 1, endogenous; lane 2, A; lane 3, B; lane 4, D; lane 5, E; lane 6, C. (B) Polypeptide products encoded by the RNA hybridizing to the *Hae*III DNA fragments: lane 1, E; lane 2, A; lane 3, F; lane 4, B; lane 5, mixture of C and D; lane 6, B; lane 7, C; lane 8, D.

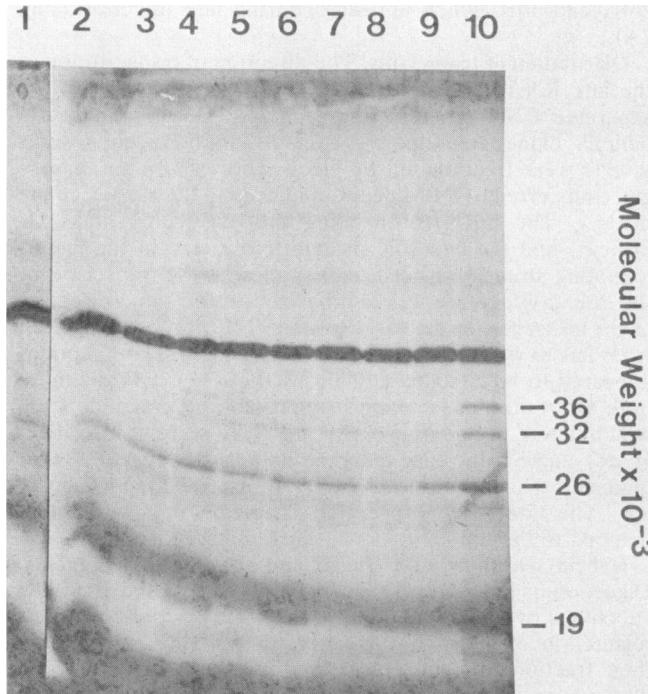
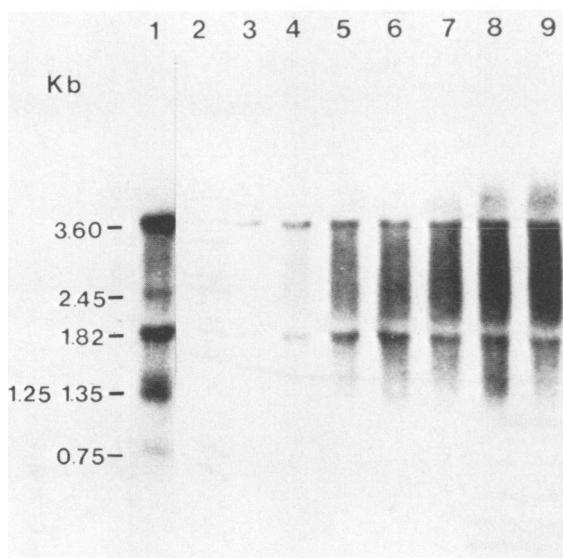


FIG. 4. Results of HART to determine the locations of protein coding regions. Late RNA was hybrid selected with 7.1-kb DNA that had been immobilized onto DBM paper. Increasing amounts of the 7.1-kb DNA were incubated with a constant amount of hybrid-selected RNA under conditions allowing the formation of DNA-

result indicated that these mRNAs were overlapping transcripts which mapped to the left-hand side of the 7.1-kb DNA fragment. The mixture of the *Hae*III C and D DNA fragments selected the transcripts for the late 32K and 19K polypeptides; subsequent experiments (Fig. 3B, lanes 6 to 8) in which the two *Hind*III-*Eco*RI fragments were used to prepare the C and D fragments showed that the transcripts in question mapped to the *Hae*III C DNA fragment. The *Hae*III B fragment also hybrid selected the mRNAs for the 19K and 32K polypeptides. The D and F fragments did not hybridize to any translatable late RNAs. Whereas the smaller *Hha* B and C fragments both selected the mRNAs encoding the late 32K and 19K polypeptides, neither the D nor the E fragment selected any late messages; since the D and E fragments mapped between the B and C fragments, their failure to hybrid select the same RNAs may be due to their small size or to an AT-rich sequence.

**Hybrid-arrested translation to determine locations of protein coding regions.** RNA was hybridized to DNA fragments, and the resulting hybrids were translated. Before translation, half of each hybridization reaction was melted and translated as a control. The locations of protein coding regions of

RNA hybrids but minimizing DNA renaturation. Half of each reaction served as a control and was boiled to melt the hybrids. This control, and the other half of the reaction which was still in hybrid form, were both translated in the reticulocyte cell-free system. Lane 1, endogenous; lanes 2 and 3, melt, hybrid with 0.5 µg of DNA; lanes 4 and 5, melt, hybrid with 1.0 µg of DNA; lanes 6 and 7, melt, hybrid with 2.0 µg of DNA; lanes 8 and 9, melt, hybrid with 5.0 µg of DNA; lane 10, hybrid selection.



transcripts, relative to DNA restriction sites, can be determined because translation of an RNA is prevented when a given DNA fragment hybridizes to the coding sequence of a message.

HART experiment were carried out by using the entire 7.1-kb DNA fragment (Fig. 4) as well as the *Hae*III fragments (data not shown). To locate the coding regions of the late RNAs mapping to the 7.1-kb DNA, late mRNA was first hybrid selected with the 7.1-kb DNA. When samples containing the highest concentration of linearized plasmid DNA were melted, there was still no inhibition of translation and no change in translation of RNA (Fig. 4). The early 36K and

FIG. 5. Hybridization of labeled DNA to RNA immobilized on filters. A 20- $\mu$ g amount of total RNA isolated from different times after infection was fractionated on an agarose gel containing methyl mercury hydroxide, transferred to DBM paper, and hybridized to labeled 7.1-kb DNA. Lane 1, longer exposure of lane 3, 2-h RNA; lane 2, 1 h; lane 3, 2 h; lane 4, 3 h; lane 5, 4 h; lane 6, 5 h; lane 7, 6 h; lane 8, 7 h; lane 9, 8 h.

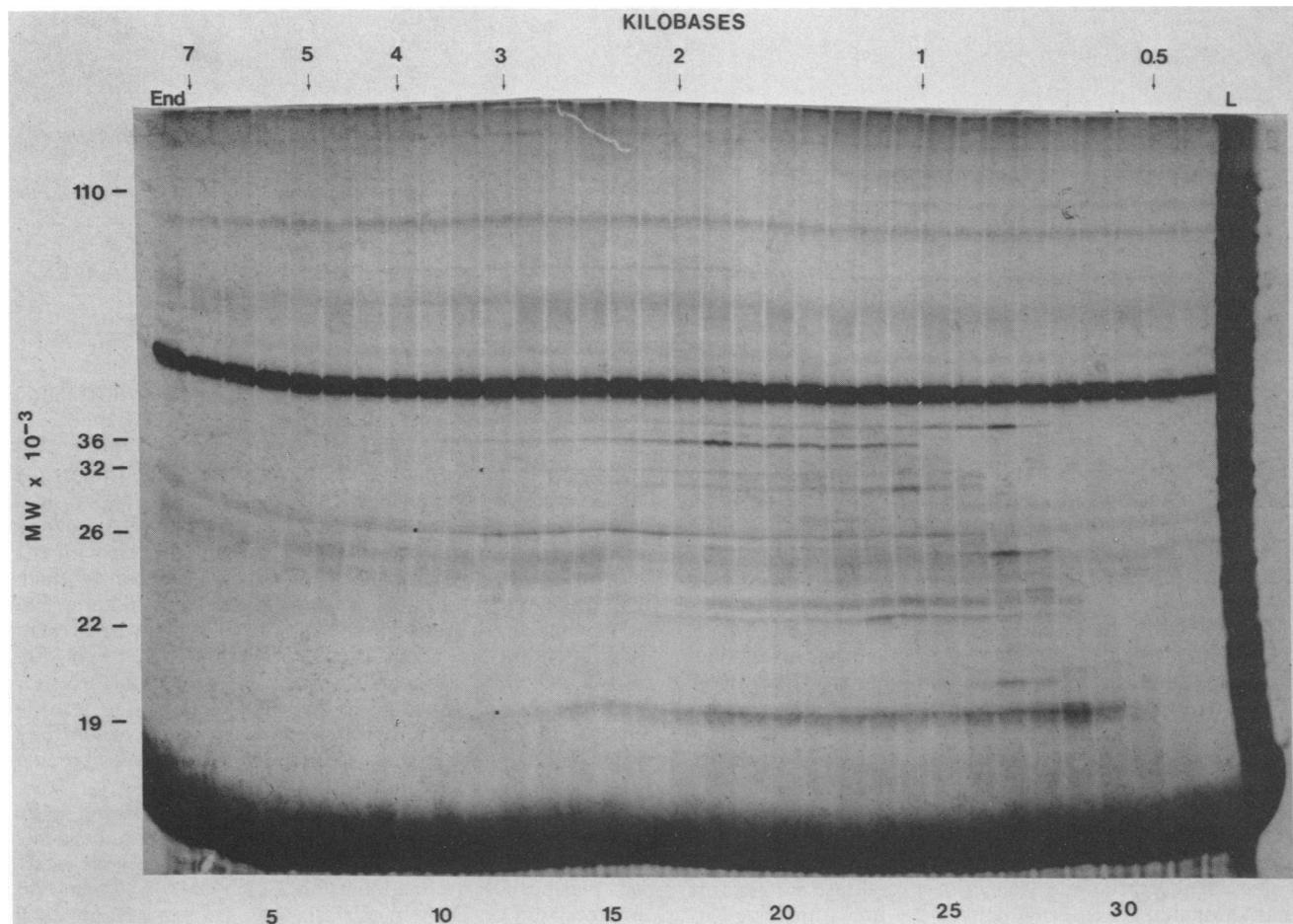


FIG. 6. Size fractionation and translation of hybrid-selected late (8 h) RNA. RNA was hybrid selected by using the 7.1-kb DNA immobilized on DBM paper. The RNA was then fractionated on a 1.1% low-melt agarose gel containing 12.5 mM methyl mercury hydroxide. The lane containing the RNA was then cut into 1-mm slices; every three consecutive slices were pooled, and the RNA was extracted and translated. The polypeptide products were fractionated on an SDS-polyacrylamide gel. Lane 1, endogenous; lanes 2 to 33, fractionated, hybrid-selected RNA; lane 34, unfractionated late RNA.

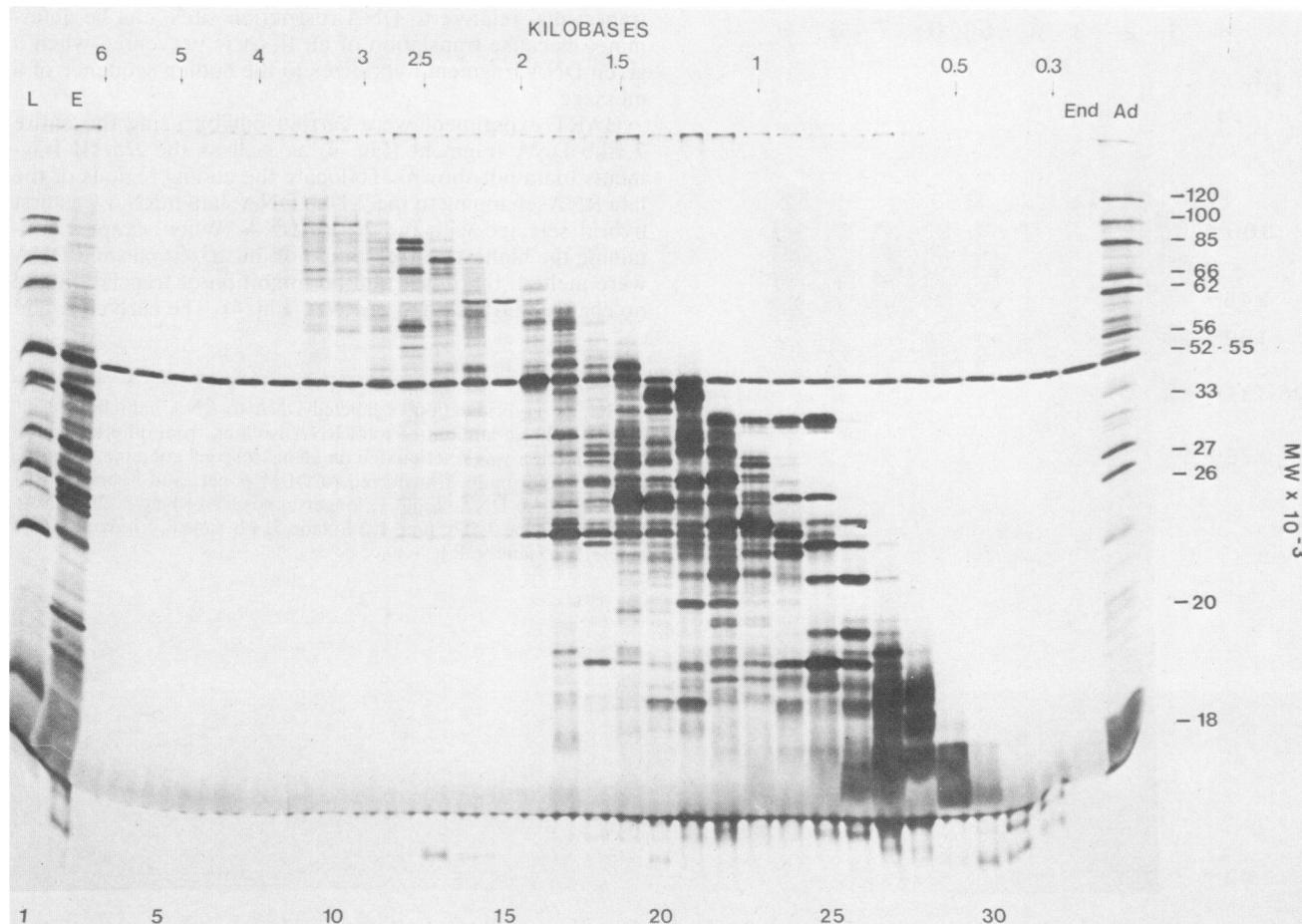


FIG. 7. Fractionation of 20 µg of cycloheximide RNA on an agarose gel containing methyl mercury. Lane 1, late RNA; lane 2, early RNA; lanes 3 to 32, translation of fractionated early RNA; lane 33, endogenous; lane 34, 27-h adenovirus RNA.

late 19K proteins were not synthesized, demonstrating that their coding regions were within this DNA fragment. More specifically, the *Hae*III C fragment hybrid-arrested production of the late 19K polypeptide (data not shown). At the highest concentration of DNA, the 26K and 32K proteins were unchanged, leading to the conclusion that their coding regions were outside this DNA fragment. Although the coding region of the mRNA encoding the 32K polypeptide was outside the 7.1-kb DNA fragment, the mRNA was hybrid selected by both the *Hae*III B and C and the *Hha* B and C DNA fragments. The simplest assumption was that, as is the case for most RNAs, the coding region of this RNA was at the 5' end of the transcript, and the noncoding portion of the message was at the 3' end. This 3' noncoding sequence was transcribed from the right end of the 7.1-kb DNA fragment, and the DNA sequence for the 5' coding region was found to the right of the 7.1-kb fragment; this message was thus transcribed in a leftward direction.

**Analysis of transcripts by filter hybridization.** To determine the sizes of the three late transcripts, total late RNA (8 h) was fractionated on a formaldehyde gel and transferred to Gene Screen paper. The resulting filter was hybridized to the 7.1-kb DNA fragment or with its *Hae*III digestion products labeled in vitro, and the results (data not shown) demonstrate there were no exclusively late bands which could be the candidates for these transcripts; instead,

there was a large amount of "background" hybridization which was due to the presence of RNA heterogeneous in size. We also failed to detect late transcripts when we used conditions which enabled us to visualize early transcripts that were still present in late infection. In this experiment (Fig. 5), a DBM filter containing fractionated RNA isolated at different times during infection was hybridized with the labeled 7.1-kb DNA fragment. Although RNA bands which first appeared early could still be detected late, no new late RNAs appeared. Although the early transcripts were present in a much lower concentration (visible on a longer exposure of Fig. 5) by late infection, they were still detectable by filter hybridization of RNA and by hybrid selection. Because the search for late transcripts was made at a time when they were optimally present, these methods should have detected late RNAs if they existed as a distinct size.

**Translation of mRNA size fractionated on denaturing methyl mercury hydroxide gels.** To determine the sizes of the late transcripts encoding specific polypeptides by another method, plasmid DNA containing the 7.1-kb DNA fragment was immobilized on DBM paper and was used to hybrid select late RNA which had been purified by the guanidine thiocyanate procedure. RNA was fractionated by size on a methyl mercury hydroxide gel, extracted, and translated in a cell-free system.

Translation of this RNA (Fig. 6) showed that the RNAs

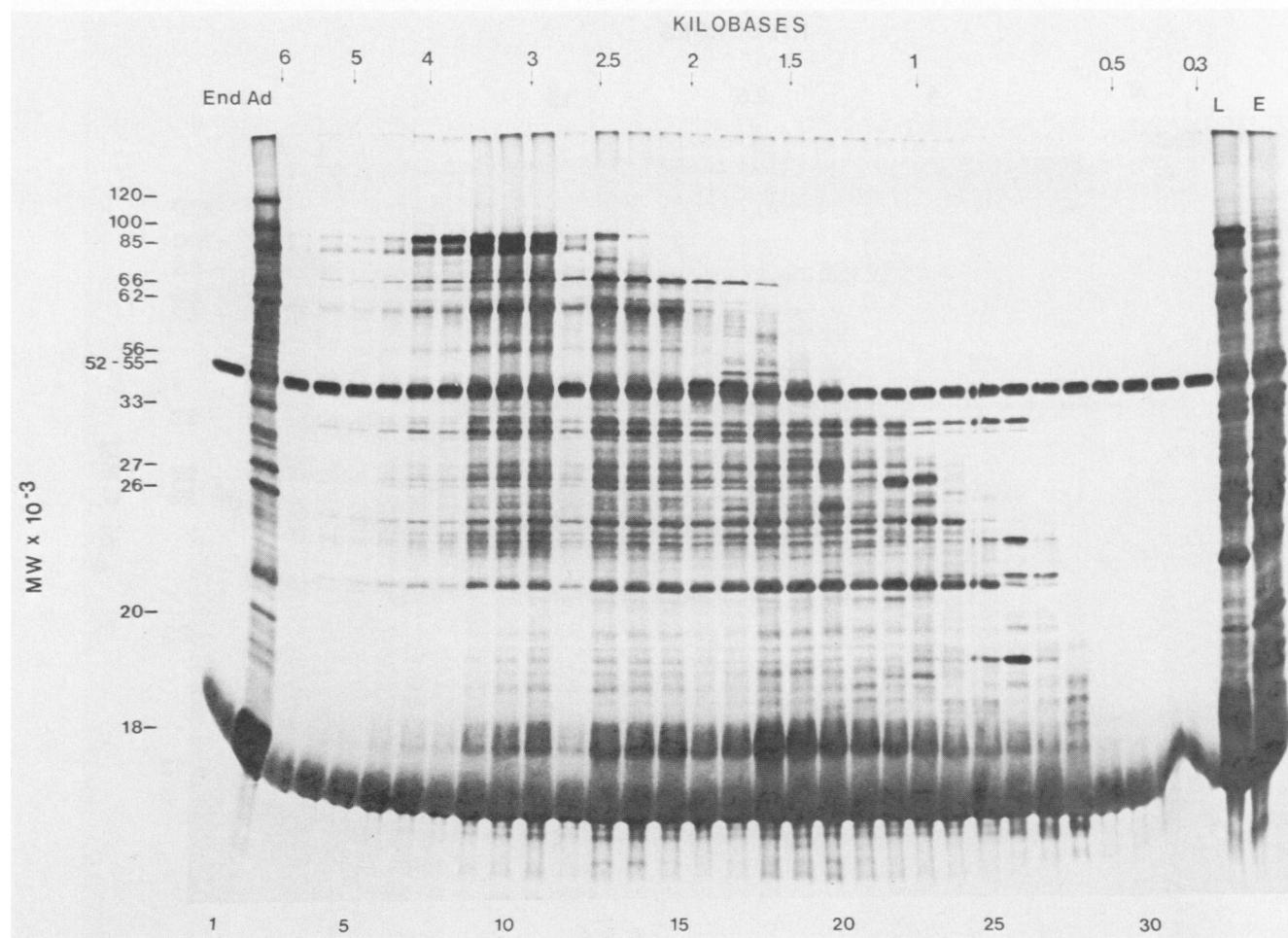


FIG. 8. Fractionation of 20  $\mu$ g of total late (8 h) RNA on a denaturing gel. Lane 1, endogenous; lane 2, 27-h adenovirus RNA; lanes 3 to 32, translation of fractionated RNA; lane 33, unfractionated late RNA; lane 34, unfractionated early RNA.

which first appeared during early infection were still visible during late infection, and their molecular weights peaked at the same sizes (3.6, 1.7, 1.2, and 0.6 kb) as when isolated from cells in early infection (13). In contrast, the RNAs encoding the three late polypeptides showed no peak of translational activity, demonstrating that these polypeptides were encoded by transcripts which were extremely heterogeneous in size. The 32K polypeptide was encoded by transcripts ranging in size from 1 to 3 kb; the message encoding the 26K polypeptide was 1.0 to 3.0 kb, and the late 19K polypeptide was encoded by an RNA which was 0.7 to 3.0 kb. The finding that the functional RNAs spanned a wide range of molecular weights which did not peak at any specific size correlated with the inability to find discrete late transcripts by filter hybridization of RNA to the 7.1-kb DNA or to its *Hae*III digestion products.

Because no late vaccinia virus transcripts of discrete size were detected, the possibility existed that late RNAs were very large and discrete in size and that the results obtained here were due to RNA degradation during cell lysis before the addition of SDS and phenol or during the purification of RNA homologous to the 7.1-kb DNA fragment. The end product of this degradation would be RNA which was very heterogeneous in size. To test this possibility, RNA was

prepared by a procedure which minimized degradation. Total RNA was prepared by lysing infected L cell monolayers with the highly chaotropic agent guanidine thiocyanate and then pelleting the RNA through a cushion of CsCl. The RNA was extracted once with phenol and then fractionated on a 1.1% low-melt agarose gel containing methyl mercury hydroxide. The results of size fractionation and cell-free translation of RNA isolated from vaccinia virus-infected L cells are shown in Fig. 7, 8, and 9.

Translation of size-fractionated cycloheximide RNA (Fig. 7) demonstrated that most of the messages encoding early polypeptides were small and discrete in size. The size of many early mRNAs were either a little larger or approximately the same as the minimum coding length necessary for a given polypeptide. However, a small fraction of the mRNA for most polypeptides did demonstrate heterogeneity in size, and this was especially clear upon longer exposure of this film.

Because individual late proteins were encoded by a population of mRNAs which were evenly distributed along the length of the denaturing gel, it can be concluded that mRNA for a given late protein was very heterogeneous in size, ranging over several kilobases in length (Fig. 8). Those RNAs that were discrete in size and were restricted to a few

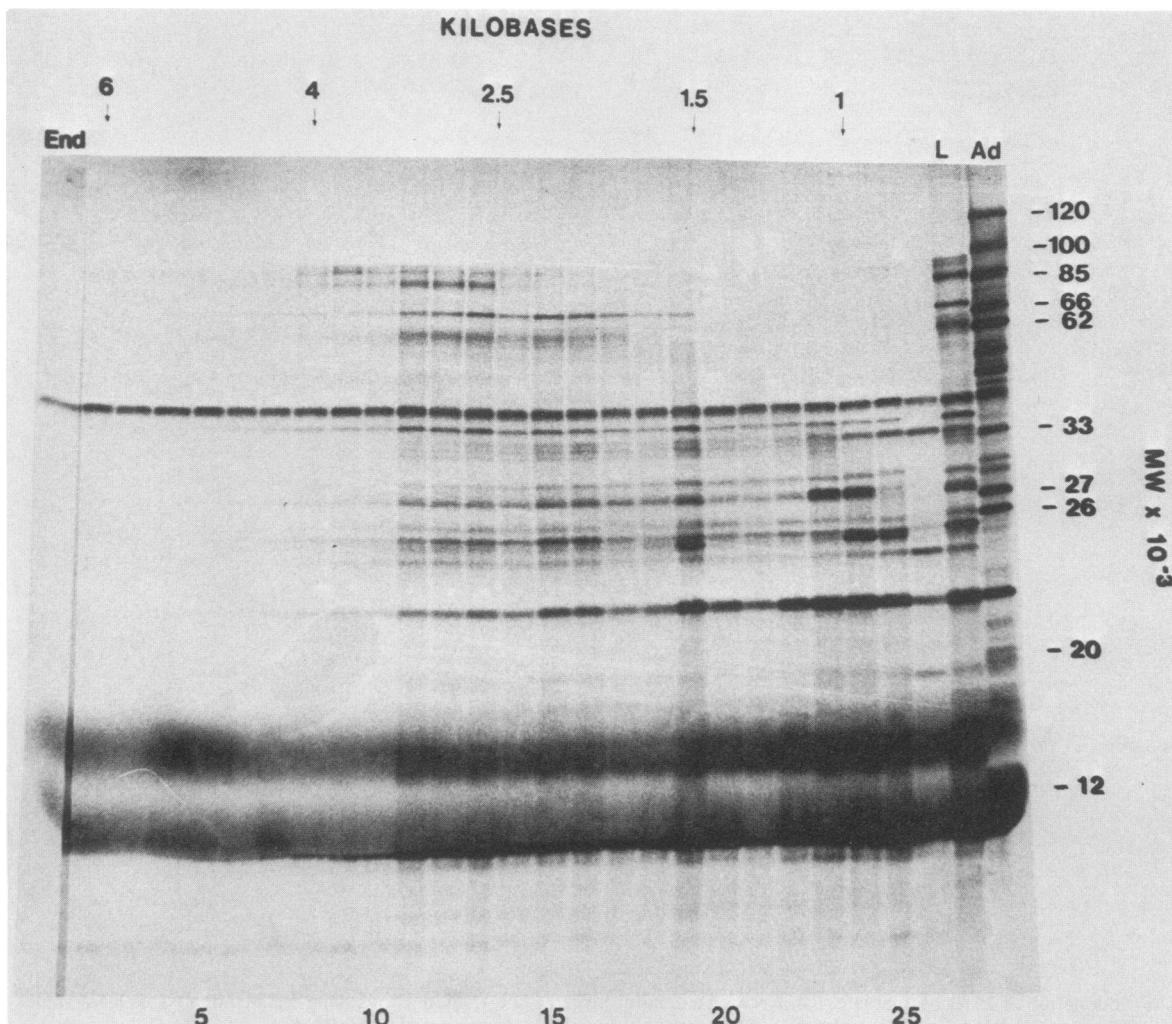


FIG. 9. Size fractionation and translation of late polysomal RNA. At 8 h after infection, 40 µg of polysomal RNA was isolated from cells as follows: cells were washed in 10 mM Tris-hydrochloride (pH 7.8)-150 mM KCl-2 mM MgCl<sub>2</sub> and lysed in the same buffer containing 0.05% Triton X-100. The lysate was centrifuged at 12,000 rpm in a Sorvall SS34 rotor, and the supernatant was layered over 3 ml of 33% sucrose cushions and centrifuged for 75 min at 4°C in a 50 Ti rotor at 49,000 rpm. The resulting polysomal pellet was suspended in 100 mM NaCl-10 mM EDTA-10 mM Tris-hydrochloride (pH 7.4)-0.5% SDS, extracted with phenol, and translated. Lane 1, endogenous; lanes 2 to 26, fractionated late polysomal RNA; lane 27, unfractionated late polysomal RNA; lane 28, 27-h adenovirus RNA.

gel fractions were actually RNAs that first appeared during early infection. Unlike most early RNAs, the mRNA for a late polypeptide could be much larger than the calculated minimum coding length. The late heterogeneous mRNAs were also found on polysomes (Fig. 9) and were thus actively directing the synthesis of proteins *in vivo*.

**S1 nuclease protection of labeled DNA by hybridization to RNA.** The Berk-Sharp S1 nuclease and exonuclease VII procedures for RNA mapping were used in an attempt to locate more precisely the 5' and 3' ends of the late transcripts. Size heterogeneity of the late transcripts could be due to heterogeneous synthesis or processing at the 5' or 3' termini or at internal regions of the RNA, and S1 nuclease mapping could localize a transcript or a fraction of a transcript and also demonstrate that it was of a discrete size. The 7.1-kb DNA fragment was cleaved with *Hae*III or *Hha*I, and the resulting fragments were then labeled at the 3' or 5' ends (13). These fragments were then hybridized to late RNA, the

resulting DNA-RNA hybrids were digested with S1 nuclease or exonuclease VII, and the nuclease-resistant DNA was sized on alkaline agarose gels. Differences in the results with these two enzymes enabled us to determine whether splicing played a role in the synthesis of these transcripts.

Hybridization of uniformly labeled and 3' end-labeled *Hha*-A or *Hae*III-E DNA to late RNA (Fig. 10) followed by S1 nuclease digestion produced a band of 430 nucleotides. This was due to the 3' end of an RNA which was transcribed from left to right, and this RNA did not correspond to the map positions of the late RNAs which had been mapped by translation of hybrid-selection RNA. The 3' end-labeled *Hha* A fragment was not protected by late RNA to produce any exclusively late band which was large enough to reach from the left end of the *Hha* A fragment into the *Hae*III A fragment; this protection was expected for the 3' end of the mRNA encoding the late 26K polypeptide. The 3' end label of the *Hha* A DNA fragment as well as the uniformly labeled

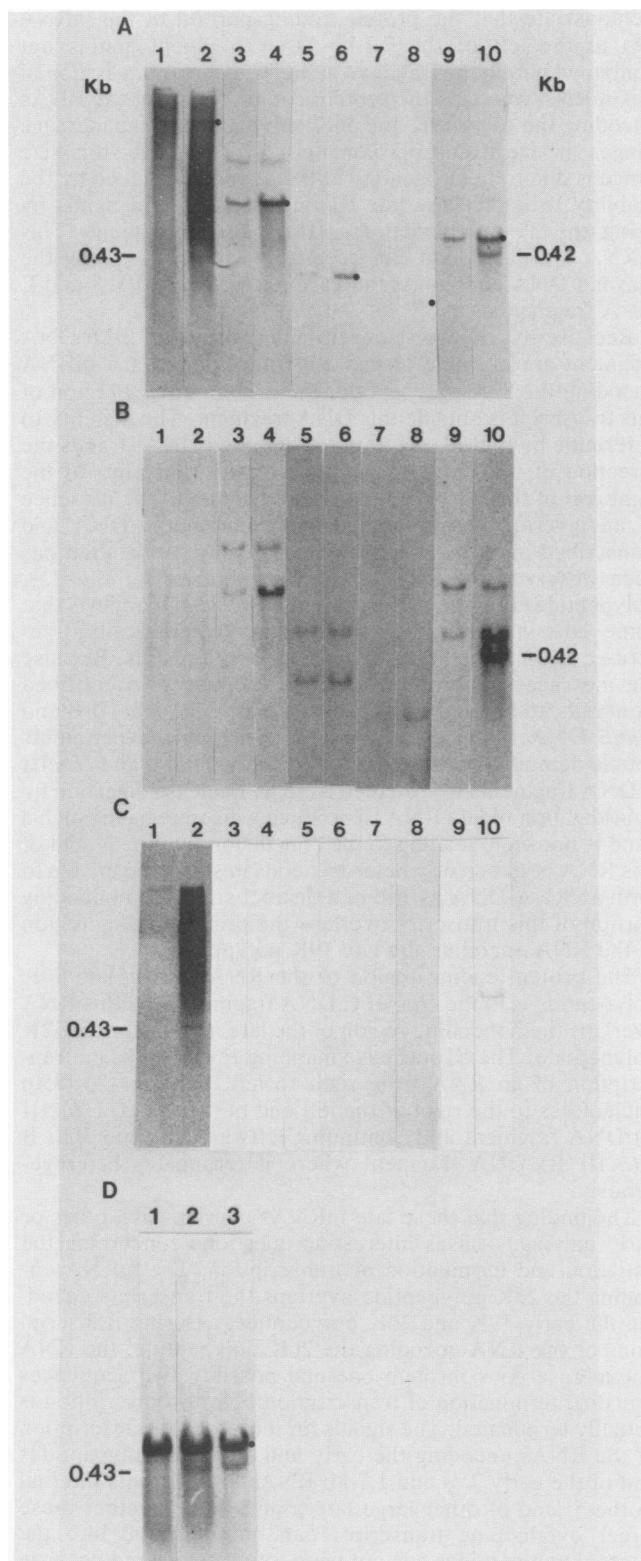


FIG. 10. S1 nuclease protection of labeled *Hha*I DNA fragments by hybridization to late RNA. Small closed circles in (A) and (D) indicate the position of renatured DNA. (A) Hybridization of uniformly labeled DNA obtained by labeling plasmid DNA *in vivo*. Lanes 1, 3, 5, 7, 9: hybridization of labeled *Hha*I fragments A, B, D, E, C in the absence of RNA. Lanes 2, 4, 6, 8, 10: hybridization of labeled *Hha*I fragments A, B, D, E, C in the presence of 25 µg of

and 3' end-labeled *Hae*III E DNA fragment were both protected by late RNA to produce the same 280-nucleotide band seen with early RNA. Failure to detect other constitutive mRNAs was attributed to their much lower levels during late infection.

The uniformly labeled and 3' or 5' end-labeled *Hha* B, D, and E fragments and the *Hae*III B, D, and F fragments did not generate any S1 nuclease-resistant bands when hybridized to late RNA (Fig. 10). Either this DNA was not transcribed at all during late infection, or the entire fragment was transcribed and detection of the bands due to this transcription was difficult to distinguish from DNA renaturation. Alternatively, transcription did not start or end at unique sites on the DNA. This last explanation may account for the lack of protection of the 3' or uniformly labeled *Hha* B DNA fragment from S1 nuclease, insofar as the mRNAs encoding the 32K and 19K polypeptides were concerned. The *Hae*III B fragment produced a result which was due to contamination by the C fragment.

Hybridization of the uniformly labeled and 5' end-labeled *Hha* C DNA fragment to late RNA resulted in an S1 nuclease-resistant, 420-nucleotide band, whereas hybridization with the uniformly labeled and 5' end-labeled *Hae*III C DNA resulted in an S1 nuclease-resistant band of 850 nucleotides. The 420-nucleotide band was not the same 410-nucleotide band produced by hybridization to early RNA (13), and both were due to RNAs transcribed from right to left. The late band is in the correct region of the 7.1-kb DNA fragment to be due to the 5' end of the mRNA encoding the late 19K polypeptide.

Comparison of S1 nuclease and exonuclease VII digestion of RNA-DNA hybrids demonstrated no detectable differences in the sizes of the nuclease-resistant bands when the uniformly labeled *Hha* A (Fig. 10) or the *Hae*III C DNA fragments (Fig. 11) were hybridized to late RNA. This result indicated that there was no splicing step in the processing of these portions of these late transcripts.

## DISCUSSION

Three late transcripts map to the 7.1-kb DNA fragment (Fig. 12; Table 1); one of these is transcribed from left to right, in the same direction as the message for the vaccinia virus thymidine kinase, and the other two are transcribed from right to left. The most striking feature of these three RNAs is that they are heterogeneous in size and are therefore difficult to detect by S1 nuclease analysis and impossible to size or map by filter hybridization; their organization has been determined primarily by translational assays. This heterogeneity in size is characteristic of total late RNA as well as polysomal RNA. The S1 nuclease mapping studies were unable to detect the 3' ends of these three late transcripts, indicating that the heterogeneity of these three RNAs is at their 3' ends. The functions of the three polypeptides encoded by these three late RNAs is not known.

The late 26K polypeptide is encoded by a message which is transcribed from left to right, the same direction as the early transcripts mapping to this DNA (13). HART results

---

late (8 h) RNA. (B) S1 nuclease protection of 5' end-labeled *Hha*I DNA fragments. All lanes are the same as described in (A), except that the DNA is labeled at both 5' ends. (C) S1 nuclease protection of 3' end-labeled *Hha*I DNA fragments. All lanes are as described in (A), except that the DNA is labeled at both 3' ends. (D) Comparison of exonuclease VII and S1 digestion of RNA-DNA hybrids. Uniformly labeled *Hha*-C: lane 1, S1, no RNA; lane 2, S1, RNA; lane 3, exonuclease VII, RNA.

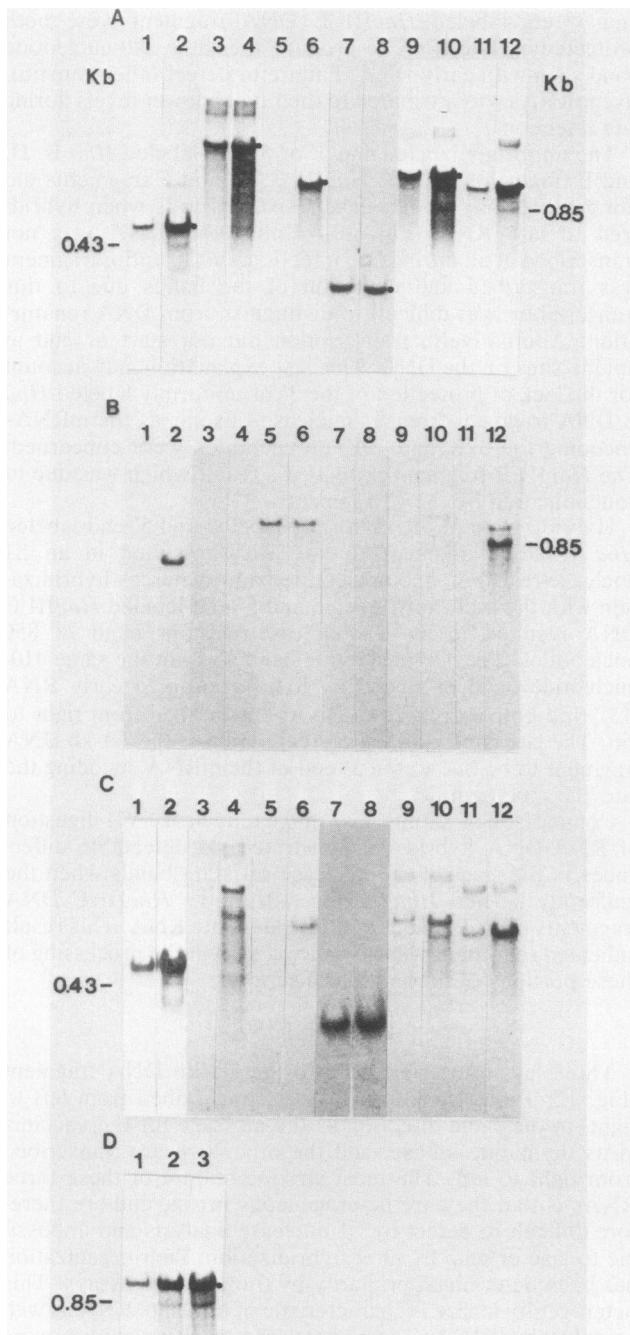


FIG. 11. S1 nuclease protection of labeled *Hae*III DNA fragments by hybridization to late RNA. Small closed circles in (A) and (D) indicate the position of renatured DNA. (A) Uniformly labeled DNA obtained by labeling *in vivo*. Lanes 1, 3, 5, 7, 9, 11: hybridization of labeled *Hae*III DNA fragments E, A, D, F, B, C in the absence of RNA. Lanes 2, 4, 6, 8, 10, 12: hybridization of labeled *Hae*III DNA fragments E, A, D, F, B, C in the presence of 25  $\mu$ g of late RNA. (B) S1 nuclease protection of 5' end-labeled *Hae*III DNA fragments. All lanes are the same as described in (A), except that the DNA fragments are labeled at both 5' ends. (C) S1

demonstrate that the protein coding portion of the mRNA lies to the left of the 7.1-kb DNA fragment and is not contained within the fragment at all. The noncoding region of this mRNA overlaps the protein coding region of the RNAs encoding the early 19K and 36K polypeptides. The message ranges in size from approximately 1.0 to 3.0 kb; this size range is due to heterogeneity at the 3' end as evinced by the inability to detect any late S1 nuclease-resistant bands by using the 3' end-labeled *Hae*III A DNA fragment. This mRNA and its 5' end can be further studied by using the plasmid DNA containing the vaccinia virus *Hind*III J and L DNA fragments.

Because hybrid arrest experiments with the 7.1-kb DNA fragment are not able to prevent translation of the mRNA encoding the 32K polypeptide, the protein coding region of this transcript is outside this DNA fragment. The inability to determine by hybrid selection to separated DNA strands the direction of transcription of the two late transcripts at the right end of the 7.1-kb fragment may be due to the presence of an inverted repeat sequence present in the DNA and transcribed in both mRNAs. This type of hybridization has been observed for the adenovirus transcript for the 72K polypeptide (14), and in the present case, despite reports that some late vaccinia virus RNAs are symmetrically transcribed (6), it is the best explanation for our results. Because the message encoding the 32K polypeptide is transcribed from right to left and is hybrid selected by the *Hae*III B and *Hha*B DNA fragments, S1 nuclease mapping experiments should demonstrate that the 3' label of the *Hha*B and *Hae*III B DNA fragments is protected from S1 nuclease digestion by hybridization to late RNA to produce a discrete band; such a band is not seen, leading to the conclusion that the 3' end of this RNA is sufficiently heterogeneous in size to be unable to form an RNA-DNA hybrid of a distinct size. The noncoding portion of this transcript overlaps the protein coding region of the RNA encoding the late 19K polypeptide.

The protein coding region of the RNA for the late 19K polypeptide is in the *Hae*III C DNA fragment, and this RNA overlaps the noncoding region of the late mRNA for the 32K polypeptide. The S1 nuclease mapping results indicate transcription of an RNA from right to left, starting 420 (850) nucleotides to the right of the left end of the *Hha*C (*Hae*III C) DNA fragment and continuing leftward into the *Hha*B (*Hae*III B) DNA fragment where it terminates heterogeneously.

The finding that these late mRNAs overlap each other or early messages raises interesting questions concerning the initiation and termination of transcription. The mRNA encoding the 26K polypeptide overlaps the transcripts encoding the early 19K and 36K polypeptides. During transcriptions of the RNA encoding the 26K polypeptide, the RNA polymerase goes through one and possibly two sequences signaling termination of transcription before transcription is actually terminated. The signals for initiation of transcription of the RNAs encoding the early and late 19K polypeptides and of the early 2.3- and 1.7-kb RNAs are at points internal to the 5' end of other large late transcripts. Whether these larger overlapping transcripts can be translated into the internal polypeptides has not been totally answered because

nuclease protection of 3' end-labeled *Hae*III DNA fragments. All lanes are the same as described in (A), except that the DNA fragments are labeled at both 3' ends. (D) Comparison of exonuclease VII and S1 nuclease digestion of RNA-DNA hybrids. Uniformly labeled *Hae*III C: lane 1, RNA, no S1; lane 2, RNA, S1; lane 3, RNA, exonuclease VII.

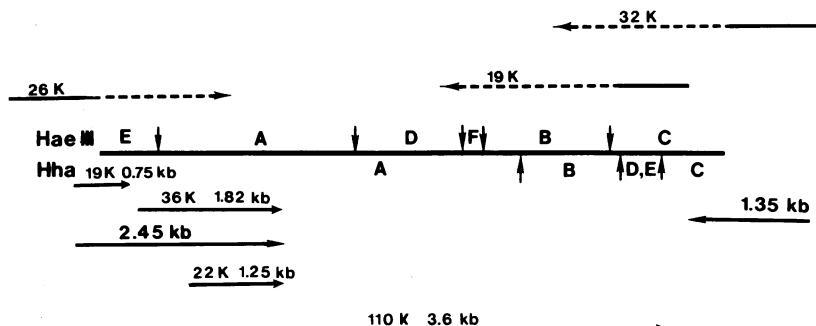


FIG. 12. Summary of the arrangement of transcripts mapping to the 7.1-kb *EcoRI F* vaccinia virus DNA fragment. Arrows are at the 3' ends of transcripts. The late transcripts are diagrammed above the DNA restriction map; the solid lines indicate the minimum size transcripts as determined from translation of size-fractionated RNA, and the dashed lines indicate the extent of the heterogeneity, as determined by translation of hybrid-selected, size-fractionated RNA, at the 3' end. Early transcripts are depicted under the restriction map, and their sizes are those obtained by filter hybridization (13).

of the heterogeneity and overlapping sizes of the late transcripts (Fig. 6). Hybrid selection and translation of RNA using the *HaeIII A* fragment does not produce the early 19K polypeptide; this demonstrates that the translation of this large late transcript encoding the 26K polypeptide is not occurring at an internal AUG (Fig. 3).

The heterogeneity of total late RNA as compared with the discrete sizes of early RNAs was first observed by A. Rice (Ph.D. thesis, Brandeis University, Waltham, Mass., 1982) in vaccinia virus-infected L cells and reported by Cooper et al. (8) for a specific transcript in infected HeLa cells. The data presented in the present paper agree with these previous studies that during infection of two different cell lines, early RNAs are discrete in size, whereas late RNAs are heterogeneous. The results in Fig. 7, 8, and 9 have verified and extended these results by showing the following: (i) due to the method of RNA purification, the observed heterogeneity of the late RNAs is unlikely to be the result of breakdown of large discrete RNAs; (ii) heterogeneous late RNAs are actively translated in vivo; and (iii) although a discrete size can be assigned to early RNAs, a small fraction of the RNA for most polypeptides is heterogeneous in size and can range up to at least 5 to 6 kb. The evidence for heterogeneity of the

late RNAs mapping to the 7.1-kb DNA has been accumulated by filter hybridization, the translation of size-fractionated RNA, hybrid selection of mRNAs, and by S1 nuclease mapping. Due to the inability to detect the 3' ends of these three late transcripts by using conditions which have allowed the detection of early transcripts and the 5' end of one late transcript, it appears that the heterogeneous size of these late RNAs is due to heterogeneous termination at their 3' ends. This conclusion is also supported by other studies (26). However, the detection of a late 3' end with the *HaeIII E* and *HhaA* DNA fragments shows that not all late transcripts exhibit 3' heterogeneity; this late 3' end is a transcript which has not been detected as a functional mRNA or by filter hybridization. It is also unclear whether this RNA is a mature message and whether it codes for one of the seven late polypeptides that have been mapped to the *HindIII J* fragment (5) by others.

There is precedence for heterogeneously sized mRNAs of viral and nonviral origin. The adenovirus mRNAs for the 5' proximal proteins in a late family contain long 3' noncoding regions and are also heterogeneous in size (14). This result was obtained by translation of size-fractionated RNA, and it simultaneously demonstrated that the other adenovirus

TABLE 1. Summary of the arrangement of late transcripts

Protein <sup>a</sup>	Calculated minimum coding size (nucleotides)	Selected by DNA fragments			mRNA size (kb) <sup>c</sup>	HART by 7.1-kb DNA	Sequence overlap of late RNAs	Sequence overlap of early RNAs (kb)	S1 nuclease-resistant bands
		<i>HaeIII</i>	<i>Hha</i>	<i>BglII</i> <sup>b</sup>					
32K	960	C,B	C,B <sup>d</sup>	A	1-3	No	RNA for 19K	1.35 3.6	No bands <sup>e</sup>
26K	780	E,A	A	B	1-3	No	0.43-kb S1 nuclease-resistant band	0.6-0.7 1.82 2.45 1.25?	No bands <sup>e</sup>
19K	570	C,B	C,B <sup>d</sup>	A	0.7-3	Yes	RNA for 32K protein	1.35 3.6	<i>Hae-C</i> 0.85-kb band, 5' end; <i>Hha-C</i> 0.42-kb band, 5' end
?								0.6-0.75 2.45	<i>Hae-E</i> 0.43-kb band, 3' end

<sup>a</sup> Molecular weights were determined by size markers of 92.5K, 68K, 43K, 25.7K, 18.4K, and 12.3K obtained from Bethesda Research Laboratories. The size of the 110K polypeptide was taken from Belle Isle et al. (5).

<sup>b</sup> Data not shown.

<sup>c</sup> From translation of size-fractionated RNA.

<sup>d</sup> The E and D fragments may fail to select due to their size or sequence.

<sup>e</sup> No protection of 5'-end-labeled DNA expected with any part of the 7.1-kb DNA fragment.

mRNAs were discrete in size. Differential regulation can produce differentially sized mRNAs from a single nonviral gene, and this heterogeneity can be at the 5' end of a message (7, 18, 27) or at the 3' end (1, 10, 19). In the case of cellular genes, there are usually two different forms of the mRNA, and the extent of heterogeneity is measured in nucleotides or tens of nucleotides. The viral mRNAs exhibit a much greater heterogeneity in size and extend from minimum coding size to several kilobases larger.

The origin of the size heterogeneity of late vaccinia virus RNAs remains to be determined. The diversity in size might arise from the inability of the RNA polymerase to terminate at the "correct" site because of multiple weak or nonexistent polyadenylic acid addition sites. Unless the sequences at the 3' ends are all different from the corresponding regions of early mRNAs, this would indicate that the RNA polymerase or any factors required for its activity are different from those present during early infection. The RNA polymerase present during late vaccinia virus infection and that present in the virion have been isolated, and one difference has been reported (4, 16, 22). In contrast to weak stop sites for RNA polymerase, one can also postulate a strong stop signal, as may be the case for the 3' end of the transcript encoding the early 36K polypeptide.

#### ACKNOWLEDGMENTS

We thank J. Morgan for the use of an RNA filter.

A.M. acknowledges postdoctoral support from American Cancer Society grant PF-1595 and Public Health Service grant F32-AI05904 from the National Institutes of Health.

#### LITERATURE CITED

- Alt, F. W., A. L. M. Bothwell, M. Knapp, E. Siden, E. Mather, M. Koshland, and D. Baltimore. 1980. Syntheses of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* **20**:293-301.
- Alwine, J. C., D. J. Kemp, B. A. Parker, J. Reiser, J. Renant, G. R. Stark, and G. M. Wahl. 1980. Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper. *Methods Enzymol.* **68**:220-242.
- Bailey, J. M., and N. Davidson. 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. *Anal. Biochem.* **70**:75-85.
- Baroudy, B. M., and B. Moss. 1980. Purification and characterization of a DNA-dependent RNA polymerase from vaccinia virion. *J. Biol. Chem.* **255**:4372-4380.
- Belle Isle, H., S. Venkatesan, and B. Moss. 1981. Cell-free translation of early and late mRNAs selected by hybridization to cloned DNA fragments derived from the left 14 million to 72 million daltons of the vaccinia virus genome. *Virology* **112**:306-317.
- Boone, R. F., R. P. Parr, and B. Moss. 1979. Intermolecular duplexes formed from polyadenylated vaccinia virus RNA. *J. Virol.* **30**:365-374.
- Carlson, M., and D. Botstein. 1982. Two differentially regulated mRNAs with different 5' ends encode secreted and intracellular forms of yeast invertase. *Cell* **28**:145-154.
- Cooper, J. A., R. Wittek, and B. Moss. 1981. Hybridization selection and cell-free translation of mRNA's encoded within the inverted terminal repetition of the vaccinia virus genome. *J. Virol.* **37**:284-294.
- Cooper, J. A., R. Wittek, and B. Moss. 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.
- Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood. 1980. Two mRNAs can be produced from a single immunoglobulin gene by alternative RNA processing pathways. *Cell* **20**:313-319.
- Hruby, D. E., and L. A. Ball. 1982. Mapping and identification of the vaccinia virus thymidine kinase gene. *J. Virol.* **43**:403-409.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680-685.
- Mahr, A., and B. E. Roberts. 1984. Organization of six early transcripts synthesized from a vaccinia virus EcoRI DNA fragment. *J. Virol.* **49**:497-509.
- Miller, J. S., R. P. Ricciardi, B. E. Roberts, B. M. Paterson, and M. B. Mathews. 1980. Arrangement of messenger RNAs and protein coding sequences in the major late transcription unit of adenovirus 2. *J. Mol. Biol.* **142**:455-488.
- Moss, B., and N. P. Salzman. 1968. Sequential protein synthesis following vaccinia virus infection. *J. Virol.* **2**:1016-1027.
- Nevins, J. R., and W. K. Joklik. 1977. Isolation and properties of the vaccinia virus DNA-dependent RNA polymerase. *J. Biol. Chem.* **252**:6930-6938.
- Pelham, H. R. B., and R. L. Jackson. 1976. An efficient mRNA-dependent translation system from reticulocyte lysates. *Eur. J. Biochem.* **67**:247-256.
- Reichardt, L. F., and A. D. Kaiser. 1971. Control of lambda repressor synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **68**:2185-2189.
- Rogers, J., P. Early, C. Carter, K. Calame, M. Bond, L. Hood, and R. Wall. 1980. Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin chain. *Cell* **20**:303-312.
- Salzman, N. P., and E. D. Sebring. 1967. Sequential formation of vaccinia virus proteins and viral deoxyribonucleic acid replication. *J. Virol.* **1**:16-23.
- Sharp, P. A., A. J. Berk, and S. M. Berget. 1980. Transcription maps of adenovirus. *Methods Enzymol.* **65**:750-768.
- Spencer, E., S. Shuman, and J. Hurwitz. 1980. Purification and properties of vaccinia virus DNA-dependent RNA polymerase. *J. Biol. Chem.* **255**:5388-5395.
- Weir, J. P., G. Bajszar, and B. Moss. 1982. Mapping of the vaccinia virus thymidine kinase gene by marker rescue and by cell-free translation of selected mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **79**:1210-1214.
- Wittek, R., J. A. Cooper, and B. Moss. 1981. Transcriptional and translational mapping of a 6.6-kilobase-pair DNA fragment containing the junction of the terminal repetition and unique sequence at the left end of the vaccinia virus genome. *J. Virol.* **39**:722-732.
- Wittek, R., A. Menna, D. Schumperli, S. Stoffel, H. K. Muller, and R. Wyler. 1977. *Hind*III and *Sst* I restriction sites mapped on rabbit poxvirus and vaccinia virus DNA. *J. Virol.* **23**:669-678.
- Wittek, R., and B. Moss. 1982. Colinearity of RNAs with the vaccinia virus genome: anomalies with two complementary early and late RNAs result from a small deletion or rearrangement within the inverted terminal repetition. *J. Virol.* **42**:447-455.
- Young, R. A., O. Hagenbuchle, and U. Schibler. 1981. A single mouse amylase gene specifies two different tissue-specific mRNAs. *Cell* **23**:451-458.