JOURNAL OF VIROLOGY, Sept. 1981, p. 733-745
0022-538X/81/090733-13$02.00/0

Vol. 39, No. 3

Extension of the Transcriptional and Translational Map of the Left End of the Vaccinia Virus Genome to 21 Kilobase Pairs

JONATHAN A. COOPER,† RICCARDO WITTEK,‡ AND BERNARD MOSS*

Laboratory of Biology of Viruses, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20205

Received 3 April 1981/Accepted 20 May 1981

Physical, transcriptional, and translational maps of an *Eco*RI fragment located between 15,800 and 20,600 base pairs from the left end of the vaccinia virus genome were prepared. Major polypeptides with molecular weights of 14,000 (14K polypeptide), 32,000 and 38,000 were synthesized in a reticulocyte cell-free system programed with immediate early RNA made in the presence of cycloheximide and selected by hybridization to λ recombinant DNA containing the *Eco*RI fragment. With early RNA made in the presence of cytosine arabinoside, an inhibitor of DNA replication, the polypeptide pattern was similar except for quantitative differences in which less 38K polypeptide was detected as a translation product. With late RNA, isolated 6 h after infection without inhibitors, only traces of the early translation products were found and a new 40K polypeptide was detected. The sizes of the mRNA's for the 14K, 32K, and 38K polypeptides were determined to be approximately 760, 880, and 1,150 nucleotides, respectively, by several independent procedures. Several large early RNAs not shown to code for any additional translation products were also detected. The size of the late message for the 40K polypeptide varied from 920 to 3,100 nucleotides. This heterogeneity appeared to be a general property of vaccinia virus late mRNA's. No evidence of RNA splicing was obtained by analysis of RNA-DNA hybrids after nuclease S1 treatment. Further analyses using separated recombinant DNA strands and restriction fragments indicated that all mRNA's were encoded by the leftward-reading DNA strand and at least two were overlapping. Since early and late mRNA's were encoded by the same DNA strand, the possibility of temporal regulation by transcriptional strand switching was eliminated. In conjunction with previous studies, a transcriptional map of the left 20,600 base pairs of the vaccinia virus genome was derived.

To understand the organization and expression of the 180-kilobase-pair (kbp) vaccinia virus genome, detailed transcriptional and translational mapping is required. With this in mind, we have cloned in phage λ three adjacent *Eco*RI fragments from the left end of the DNA molecule (21, 23). Together, these DNA segments comprise an uninterrupted sequence of nearly 21 kbp that represents more than 10% of the genome. The end 9-kbp *Eco*RI fragment was of special interest since it includes the major portion of the 10-kbp inverted terminal repetition (7, 24). Following a 3-kbp terminal noncoding segment containing two sets of 13 to 17 direct tandem repeats of a 70-base-pair (bp) sequence (25), there is a region that encodes messages for three early polypeptides (6). These mRNA's do not appear to be spliced (22), and their 5′ capped ends contain the ribonucleotides used to initiate transcription (19). A fourth mRNA, which begins near the end of the first *Eco*RI fragment and continues into the adjacent one, was subsequently discovered (19). The penultimate 6.6-kbp *Eco*RI fragment, which contains the junction of the inverted terminal repetition with the unique region at the left end of the genome, encodes four early polypeptides and one late polypeptide (23). Of the seven early messages identified, five are copied from the same leftward-reading DNA strand.

To continue these investigations, we have now extended the transcription map 4.7 kbp further into the genome. Three additional early mRNA's and one late mRNA were detected. All of these were transcribed from the apparently dominant leftward-reading DNA strand, and at

† Present address: Tumor Virology Laboratory, The Salk Institute, San Diego, CA 92138.

‡ Present address: Institut für Virologie der Universitat Zürich, CH-8057 Zürich, Switzerland.

least two of the early mRNA's as well as one late and one early mRNA appeared to be overlapping.

MATERIALS AND METHODS

Isolation of RNA from infected cells. Total cytoplasmic RNA was isolated by CsCl centrifugation 4 h after HeLa cells were infected with 1,500 particles (30 PFU)/cell of purified vaccinia virus (WR strain) in the presence of 100 μg of cycloheximide or 40 μg of cytosine arabinoside per ml or 6 h after infection in the absence of inhibitors (5).

Hybridization selection of RNA and cell-free translation. RNA was selected by hybridization to DNA immobilized on nitrocellulose filters and translated in a micrococcal nuclease-treated reticulocyte system as described previously (5) and in the preceding paper (23).

Size fractionation of RNA in agarose gels containing methylmercuric hydroxide. RNA samples were heated at 60°C for 5 min in buffer containing 10 mM methylmercuric hydroxide and resolved by electrophoresis in 1.5% agarose gels containing 5 mM methylmercuric hydroxide exactly as described (6). After electrophoresis, the gel was washed three times in 0.5 M ammonium acetate and then exposed to X-ray film placed against an intensifying screen for 2 h at 4°C to determine the position of internal ^{32}P-labeled DNA restriction fragment markers. Gel slices (4 mm thick) were frozen and thawed, and the RNA was extracted with phenol and ethanol precipitated with 20 μg of carrier tRNA before translation (6).

Blotting and hybridization of size-fractionated RNA. RNA was denatured in buffer containing 1 M glyoxal, 50% (vol/vol) dimethyl sulfoxide, and 10 mM NaH_2PO_4-Na_2HPO_4 (pH 6.9), fractionated on a 1.6% agarose slab gel (12), and transferred to diazotized paper (20). Hybridization to ^{32}P-labeled DNA was as described (22).

Nuclease S1 analysis. The procedure of Berk and Sharp (1) was carried out as previously described by hybridizing RNA to purified restriction fragments or separated DNA strands isolated from ^{32}P-labeled λ recombinants (22). After nuclease S1 treatment, the samples were analyzed by electrophoresis in neutral or alkaline agarose gels. The gels were then dried on DEAE-paper under vacuum and fluorographed. ^{32}P-labeled DNA restriction fragments analyzed on the same gels served as size markers.

RESULTS

Physical mapping of the cloned *Eco*RI fragment C of *Hin*dIII-C. Fifteen fragments are obtained by digestion of vaccinia virus (strain WR) DNA with restriction endonuclease *Hin*dIII (11). The left-terminal fragment, *Hin*dIII-C, contains three *Eco*RI sites (21; Fig. 1). The terminal *Eco*RI-A and penultimate *Eco*RI-B of *Hin*dIII-C were cloned in coliphage λ, and those recombinants, designated λA7/1 and λ52, respectively, have been used for transcriptional and translational mapping studies (6, 22, 23). In this report, we describe similar analyses, using a recombinant designated λ34 that contains the 4.8-kbp *Eco*RI-C of *Hin*dIII-C. This fragment mapped between 15.8 and 20.6 kbp from the left end of the vaccinia virus genome (Fig. 1). For convenience, the data establishing the identity of the cloned fragment of λ34 were included with data for λ52 in the preceding report (23).

Figure 1 also shows maps of the *Xho*I, *Hpa*II, *Hinc*II, and *Bam*HI sites. The locations of the *Xho*I, *Xba*I, and *Bam*HI sites were determined by double digestions and the *Hpa*II and *Hinc*II sites by the partial mapping procedure of Smith and Birnstiel (17). The *Hpa*II, *Hinc*II, and *Xho*I sites of the cloned DNA agreed with restriction maps of the *Hin*dIII fragment isolated from authentic vaccinia virus DNA (E. Barbosa, unpublished data). The orientation of the cloned insert within the λ recombinant was determined from the sizes of the *Xho*I recombinant DNA fragments, using the known *Xho*I sites in λ and vaccinia virus DNA for reference. In this manner, the left end of the insert was found to be attached to the left λ arm.

Cell-free translation of early and late mRNA's. Total cytoplasmic RNA from cells infected with purified vaccinia virus was used for these studies. Cytosine arabinoside, an inhibitor of DNA synthesis, or cycloheximide, an inhibitor of protein synthesis, was added before and during virus infection to obtain early classes of RNA. According to the conventional nomenclature, RNA made in the absence of protein synthesis is referred to as immediate early. Both classes of early RNAs, however, have very similar translation products (5) and were isolated 4 h after infection. Late RNA was obtained 6 h

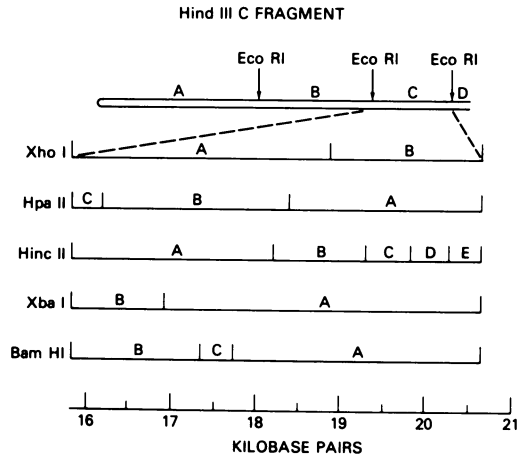

FIG. 1. *Restriction endonuclease maps of the C fragment obtained by EcoRI digestion of the left-terminal HindIII fragment C.*

after infection in the absence of any inhibitors. Under our conditions of infection, DNA replication occurs primarily between 1.5 and 4 h.

Those RNA species transcribed from the portion of the genome corresponding to the cloned *Eco*RI fragment C of *Hin*dIII-C were selected by hybridization to λ34 recombinant DNA immobilized on a nitrocellulose filter and translated in a micrococcal nuclease-treated reticulocyte cell-free system. In preliminary experiments, we noted that the amounts of translatable immediate early RNA obtained by hybridization to this 4.7-kbp portion of the vaccinia genome approached that obtained by hybridization to the terminal 9-kbp *Eco*RI-A/*Hin*dIII-C in λA7/1 (6) and were substantially greater than that obtained by hybridization to the 6.6-kbp *Eco*RI-B/*Hin*dIII-C in λ52 (23).

Three major [^{3}H]leucine-labeled 38,000-dalton (38K), 32K, and 14K polypeptides were resolved by polyacrylamide gel electrophoresis of immediate early RNA selected by hybridization to λ34 DNA (Fig. 2). These polypeptides were distinct from those made by RNA selected to either λA7/1 or λ52. The 38K polypeptide formed a broad band that sometimes appeared to separate into two components during electrophoresis. Some radioactive material also migrated with the globin front; in general, the intensity of that band correlated with the total amount of radioactive protein applied.

The lower amount of total translatable RNA obtained when cytosine arabinoside was used instead of cycloheximide was reflected in the lower amounts of specific mRNA's selected to λ34 DNA (Fig. 2). In addition, the relative amounts of the messages for the three polypeptides were different. In particular, the 38K polypeptide could barely be detected.

The three early messages were virtually absent at 6 h after infection, judging from the translation products of late RNA selected to λ34 DNA (Fig. 2). However, a new polypeptide of 40K was detected. The mobility of this polypeptide was very similar to that of the immediate early 38K polypeptide. However, these two polypeptides had distinctive tryptic peptide maps, indicating that they are different (J. A. Cooper, unpublished data).

Similar results were obtained by using [^{35}S]methionine in place of [^{3}H]leucine in the cell-free system (Fig. 3). Since all of the polypeptides encoded within the cloned DNA segment were shown to contain methionine, this amino acid was used for further experiments.

Sizes of translatable early and late mRNA's. The sizes of the mRNA's encoded within the *Eco*RI fragment C of *Hin*dIII-C were determined by agarose gel electrophoresis of methylmercuric hydroxide-denatured RNAs. Figure 4A shows the translation products of immediate early RNAs eluted from individual

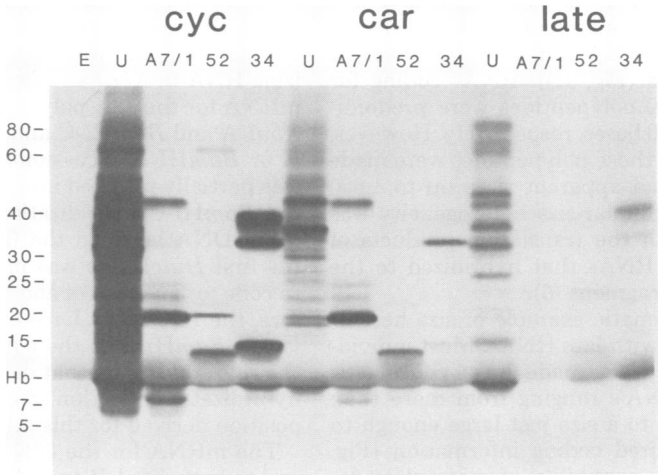

FIG. 2. *Polyacrylamide gel electrophoresis of [^{3}H]leucine-labeled cell-free translation products of RNA selected by hybridization to recombinant DNA. Cytoplasmic RNA was isolated from cells 4 h after infection in the presence of cycloheximide (cyc) or cytosine arabinoside (car) or 6 h after infection in the absence of inhibitors (late) and hybridized to λA7/1, λ52, or λ34 recombinant DNA immobilized on nitrocellulose filters. The selected RNA was translated in a micrococcal nuclease-treated reticulocyte cell-free system containing [^{3}H]leucine, and the products were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. A fluorograph is shown. Abbreviations: E, endogenous translation products of the cell-free system; U, translation products of unselected RNA; Hb, position of the globin band. Molecular weights × 10^{-3} are indicated on the left.*

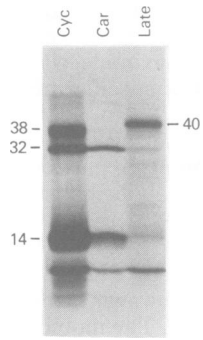

FIG. 3. *Polyacrylamide gel electrophoresis of [^{35}S]methionine-labeled translation products of RNA selected by hybridization to λ34 recombinant DNA. RNA was isolated, hybridized, and translated as in Fig. 2 except for the use of [^{35}S]methionine instead of [^{3}H]leucine. Molecular weights × 10^{-3} as indicated.*

gel slices. RNA sizes were determined by coelectrophoresis of denatured DNA restriction fragment markers. As previously shown (5, 6), there was a general relationship between RNA and polypeptide size, with the largest polypeptides of each RNA size class using most of the coding capacity. To select the messages of present interest, RNA from each fraction was hybridized to recombinant DNA before translation. mRNA coding for the 38K polypeptide was recovered mainly in two gel slices corresponding to a size of approximately 1,290 bases, but a small amount of mRNA had an apparent size of about 1,850 bases (Fig. 4B). RNAs coding for the 32K and 14K polypeptides were predominantly 860 and 680 bases, respectively. However, small amounts of those polypeptides were made by RNAs that had apparent sizes up to 1,650 bases (Fig. 4B). Similar-size heterogeneity was not observed with the translation products of immediate early RNAs that hybridized to the terminal *Eco*RI fragment (6).

A far more dramatic example of size heterogeneity occurred with late RNAs. Most individual polypeptides were made by a continuous spectrum of mRNA's ranging from more than 6,000 bases down to a size just large enough to contain the required coding information (Fig. 5A). To avoid the possibility of incomplete denaturation, these RNAs were treated with methylmercuric hydroxide, and the gel also contained this denaturing agent. Furthermore, labeled marker DNA restriction fragments mixed with the RNA sample were fully denatured. In other control experiments, reovirus double-stranded RNAs, which included species of approximately 4 kbp in length, were completely denatured to single strands under identical conditions. When vaccinia virus late RNA fractions of 2,200 to 3,400 bases were eluted, treated a second time with methylmercuric hydroxide, and electrophoresed again, much of the RNA had the same mobility as before and directed the synthesis of the same set of polypeptides.

By hybridizing the size-fractionated RNA to recombinant DNA, the mRNA encoding the late 40K polypeptide was shown to be heterogeneous in size and to vary from 920 to 3,100 bases (Fig. 5B). The smaller size is barely sufficient to code for the 40K polypeptide. Although there appeared to be discontinuities in the size of the mRNA, tracks with lighter amounts of translation products may reflect variable recovery of RNA during the gel elution and hybridization steps.

Mapping of early polypeptides. The three immediate early polypeptides of 38K, 32K, and 14K were mapped by cell-free translation of mRNA's purified by hybridization to the separated strands and restriction fragments of the recombinant DNA. Each of the mRNA's hybridized to the light or leftward-reading DNA strand (Fig. 6, track L). In addition, all three hybridized to *Hinc*II-A but not to *Hinc*II-B, -C, -D, or -E or to *Hpa*II-C (Fig. 6), indicating that all three polypeptides were encoded within a contiguous 2,000-bp region. However, this region was somewhat smaller than that necessary to code for the three mRNA's with a combined size of 2,800 bases without their overlapping.

Finer mapping was accomplished by hybridizing RNA to *Xba*I and *Bam*HI fragments. The mRNA for the 38K polypeptide was selected on *Xba*I-A and *Bam*HI-C and -A but not on *Xba*I-B or *Bam*HI-B. Occasionally, this polypeptide was partially resolved into two bands, as seen in the *Bam*HI-C hybridization of Fig. 6. The 840 bp of DNA between the first *Bam*HI site and the first *Hinc*II site was not quite long enough to code for an RNA of about 1,200 bases. Therefore, the RNA may have extended into the adjacent *Bam*HI-B to the left or into *Hinc*II-B to the right to extents that did not permit efficient hybridization selection. Figure 7 shows the map position derived for this mRNA.

The mRNA for the 32K polypeptide hybridized to both *Xba*I-B and -A but only to *Bam*HI-B (Fig. 6). Since the mRNA was 860 bases long and did not hybridize to *Hpa*II-C, it could lie mostly to the left of the *Xba*I site. Nevertheless, *Xba*I-A effectively selected this RNA, indicating that a significant portion must be transcribed to the right of the *Xba*I site as well. In Fig. 7 we have shown the 5′ end close to the *Bam*HI site to minimize overlap of this RNA with that cod-

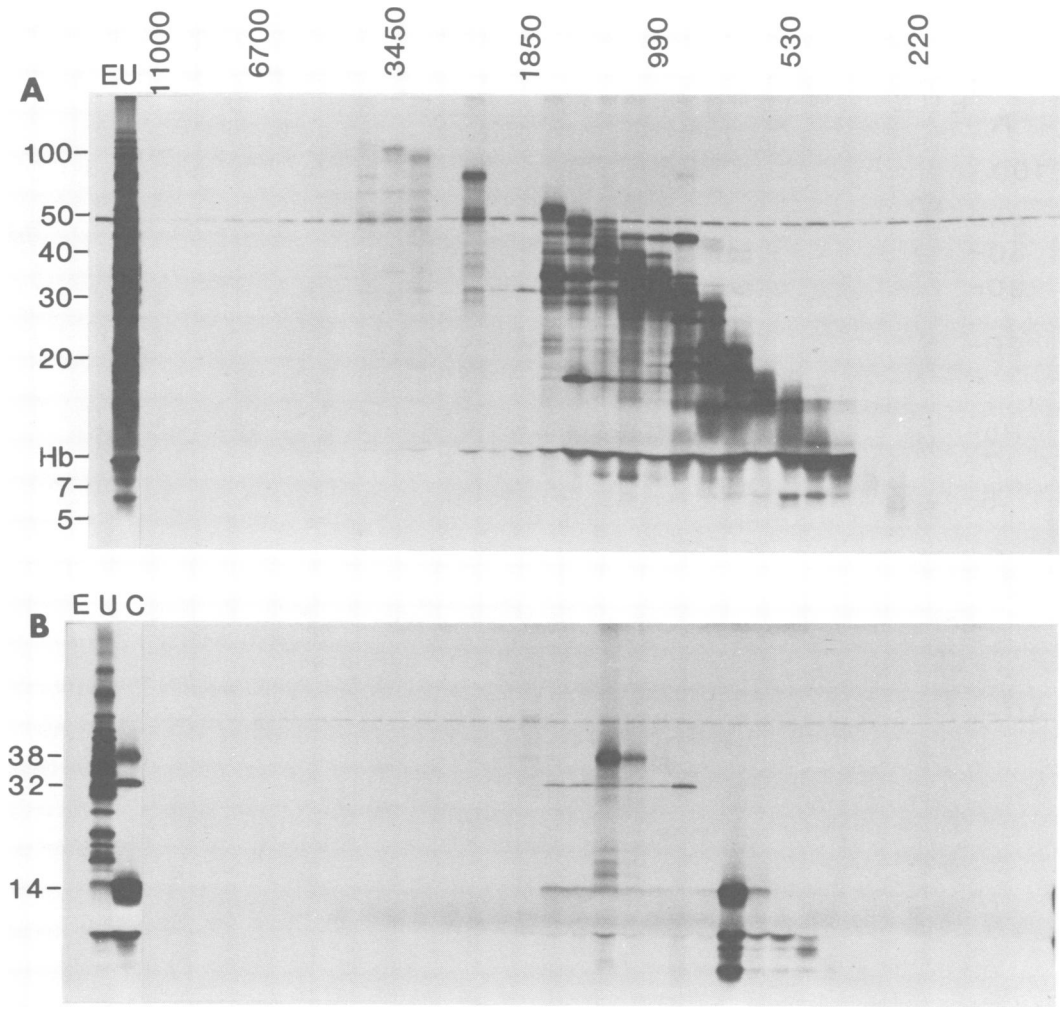

FIG. 4. *Translation products of size-fractionated immediate early RNAs. RNA eluted from each slice of agarose gel containing methylmercuric hydroxide was translated either directly (A) or after hybridization to recombinant DNA (B). [^{35}S]methionine-labeled products were analyzed by electrophoresis on a polyacrylamide slab gel. Abbreviations: E, endogenous translation products; U, translation products of unselected RNA; C, translation products of total RNA selected by hybridization to recombinant DNA; Hb, globin band. The numbers on the left are the molecular weights × 10^{-3} of the polypeptides. The numbers on the top represent estimates of the size in nucleotides of RNA eluted from that position in the methylmercuric hydroxide gel.*

ing for the 14K polypeptide.

The mRNA for the 14K polypeptide hybridized to *Xba*I-B and *Bam*HI-B but not to *Xba*I-A, *Bam*HI-C or -A, or *Hpa*II-C (Fig. 6). The DNA segment defined by this hybridization pattern was just large enough to code for an RNA of the observed size, although a small overlap into *Hpa*II-C cannot be ruled out. With the mRNA's positioned as shown in Fig. 7, several hundred bases of the 3′ end of the 32K polypeptide mRNA overlapped the 5′ end of the 14K polypeptide mRNA.

Mapping of the late polypeptide. Mapping of the 40K polypeptide proved difficult because of persistent nonspecific polypeptide background encountered when this minor late mRNA was hybridized to immobilized restriction fragments. Whether this was partly due to heterogeneity of RNA size or to formation of double-stranded branched RNA structures (3) is not known. Nevertheless, the data were consistent and provided useful map information (Fig. 8). Most significantly, the mRNA for the 40K polypeptide hybridized selectively to the light or

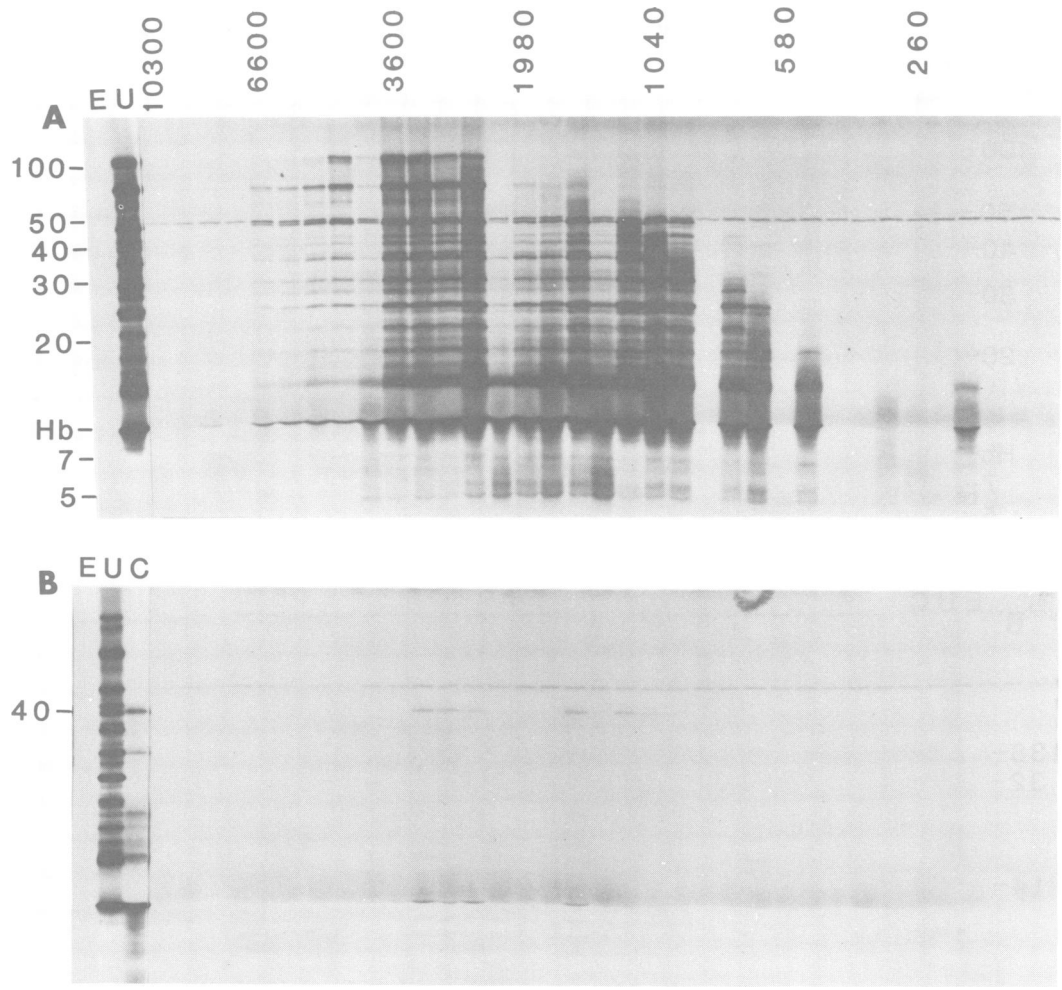

FIG. 5. *Translation products of size-fractionated late RNAs. RNA eluted from each slice of a gel containing methylmercuric hydroxide was translated either directly (A) or after hybridization selection to recombinant DNA (B). The key to symbols is in the legend to Fig. 4. Decreased amounts of mRNA may have been present in fractions of approximately 2,100 bases because of overloading the gel with rRNA.*

leftward-reading DNA strand, *Xba*I-A and *Bam*HI-A, a mixture of *Hpa*II-A and -B, and *Hinc*II-A and -B. No hybridization to *Hpa*II-C, to *Hinc*II-C, -D, or -E, or to the second-largest fragment of combined *Hpa*II and *Hinc*II digestion was detected. The relatively weak hybridization to *Hinc*II-B and the failure to hybridize to the *Hpa*II/*Hinc*II double-digestion B fragment, which was slightly shorter than *Hinc*II-B, suggested that the 5′ end of the late message lies between the *Hinc*II-A/B junction and the *Hpa*II-B/A junction. From the heterogeneity in size of this late RNA, the 3′ ends could hybridize to the right and left of the first *Xba*I and *Bam*HI sites. In support of this, weak hybridization to *Bam*HI-B and *Xba*I-B was detected. We have considered that any degradation near the 3′ ends of these large late transcripts during hybridization might actually sharpen the mapping of the translated regions. Because of the heterogeneity in size, we did not include the map position of the late RNA in Fig. 7. However, it appeared to overlap extensively with the early mRNA for the 38K polypeptide.

Mapping of immediate early mRNA's by hybridization to labeled DNA. In the previous sections of this report, mRNA's were identified by their cell-free translation products. An independent approach was used to confirm the map positions and sizes as well as to identify RNAs without detectable translation products. Total cytoplasmic RNA from vaccinia virus-in-

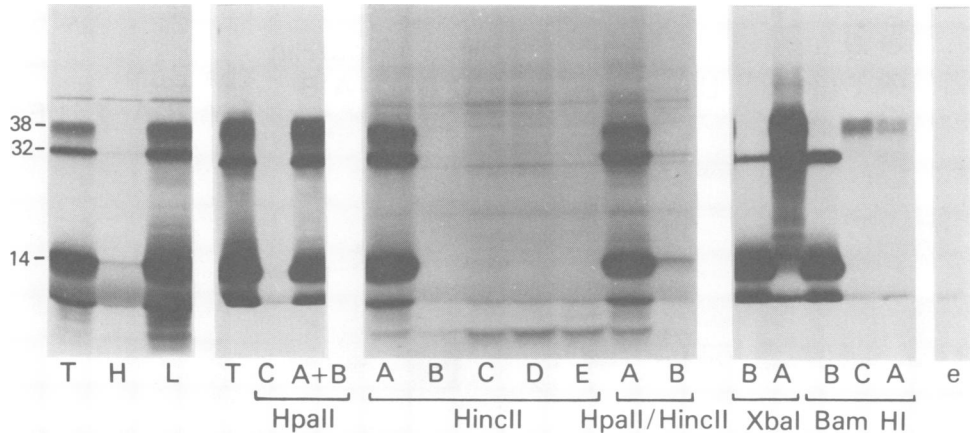

FIG. 6. *Cell-free translation of immediate early transcripts selected by hybridization to separated DNA strands and purified restriction fragments. Cytoplasmic RNA from cells infected in the presence of cycloheximide was hybridized to (T) total λ34 recombinant DNA, (H) the heavy or rightward-reading strand, (L) the light or leftward-reading strand or to specified restriction fragments of the cloned insert, and (e) endogenous translation products. The molecular weights of polypeptides × 10^{-3} are indicated. Fluorographs of polyacrylamide gels are shown. Since HpaII-A and -B were too similar in size to be preparatively separated, the two largest fragments produced by combined HpaII and HincII digestion were used. "HpaII/HincII A" is contained within HpaII-B and "HpaII/HincII B" is within HpaII-A, as can be seen in Fig. 1.*

fected cells treated with cycloheximide was denatured with dimethyl sulfoxide and glyoxal and fractionated according to size on an agarose gel (12). The RNA was then transferred by blotting to diazotized paper, and individual strips were hybridized to various ^{32}P-labeled probes. When the probe was prepared from the total cloned fragment (Fig. 9, strip T), five major RNA bands calculated to be 758, 879, 1,152, 1,333, and 1,606 bases, respectively, were resolved. These bands were more clearly seen by using probes made with smaller DNA fragments. In addition, four larger RNAs of 2,121, 2,485, 2,909, and 3,394 bases, respectively, were detected on longer autoradiographic exposure. A probe prepared from the heavy strand of the recombinant did not hybridize to any of the RNAs (strip H), but all RNA bands were labeled when the light strand was used as a probe (strip L). Therefore, all RNAs encoded within the cloned DNA segment were transcribed from the leftward-reading DNA strand.

^{32}P-labeled *Hpa*II, *Hin*cII, *Xba*I and *Bam*HI fragments were used to map the RNAs. RNA with a size of 758 bases hybridized to *Hpa*II-C and -B, *Hin*cII-A, *Xba*I-B, and *Bam*HI-B, indicating transcription from the left side of the cloned *Eco*RI DNA segment (Fig. 9). Both the size and map position of this RNA suggest that it was the message for the 14K polypeptide (Fig. 7). One discrepancy, however, was the hybridization to *Hpa*II-C. An RNA translated to form the 14K polypeptide was not detected by hybridization to this fragment (Fig. 6), possibly because the extent of overlap into *Hpa*II-C was too small for efficient hybridization selection of the message for the 14K polypeptide in the cell-free translation experiments. However, a similar-size message for a 19K polypeptide that was previously mapped to the end of *Eco*RI-B/*Hin*dIII-C (23) is known from nuclease S1 protection experiments to extend into *Eco*RI-C (R. Wittek and B. Moss, manuscript in preparation). Therefore, the RNA of 760 bases detected by hybridization to *Hpa*II-C probably represents

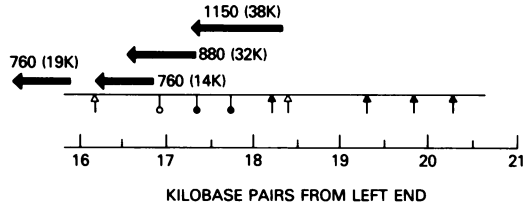

FIG. 7. *Transcriptional and translational map of the 4.8-kbp EcoRI fragment. The arrows represent immediate early mRNA's pointing in the directions of their synthesis. The numbers by the arrows indicate their sizes in bases determined by blot hybridization, and the numbers in parentheses represent the polypeptide sizes. The following symbols stand for sites of cleavage with the indicated restriction endonuclease: ↑, HpaII; ♁, XbaI; ♦, BamHI; ↟, HincII. The late RNA coding for the 40K polypeptide is not shown because of its size heterogeneity, but appears to overlap considerably with the 38K polypeptide mRNA.*

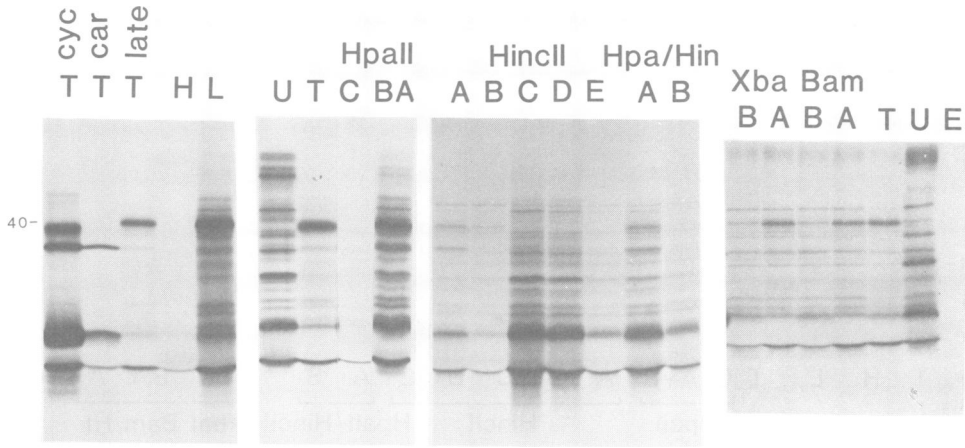

FIG. 8. *Cell-free translation products of late transcripts selected by hybridization to separated DNA strands and purified restriction fragments. Cytoplasmic RNA from cells infected in the presence of cycloheximide (cyc) or cytosine arabinoside (car) or 6 h after infection in the absence of inhibitors (late) was selected by hybridization to total (T) recombinant DNA. Late RNA was selected by hybridization to the heavy (H) or light (L) recombinant DNA strand or to the specified restriction fragments of the cloned inserts. Symbols: (E) endogenous translation products; (U) translation products of unselected late RNA. The molecular weight of the late polypeptide* $\times 10^{-3}$ *is indicated.*

the message for this 19K polypeptide, as indicated in Fig. 7.

The next larger RNA, of 879 bases, hybridized to *Hpa*II-B, *Hinc*II-A, *Bam*HI-B, and *Xba*I-A and -B (Fig. 9). Its hybridization to *Xba*I-A distinguished it from the 758-base RNA and suggested that it was the message for the 32K polypeptide (Fig. 7).

The RNA of 1,152 bases was the best candidate for the 38K message. It hybridized to *Hpa*II-B, *Hinc*II-A, *Xba*I-A, and *Bam*HI-A (Fig. 9). In some cases, the band appeared broad enough to be a doublet.

The RNA of 1,333 bases hybridized to the same fragments as the 758-base RNA, suggesting that it could also code for the 14K polypeptide. In this respect, small amounts of that polypeptide were made by RNAs up to 1,650 bases long.

The RNA of 1,606 bases hybridized primarily to the right side of the cloned segment, as indicated by the strong labeling with *Hinc*II-D and -E, *Xba*I-A, and *Bam*HI-A (Fig. 9). However, no message hybridizing to this region was detected by cell-free translation experiments.

The large RNAs of 2,121 to 3,394 bases were too faint to map, and no messages of that size range were detected by cell-free translation.

When late RNA was fractionated by electrophoresis under identical denaturing conditions, only radioactive smears were obtained by hybridization to ^{32}P-labeled restriction fragments (not shown). This result was consistent with the size heterogeneity detected by cell-free translation experiments.

Analysis of nuclease S1-resistant RNA-DNA duplexes. Additional information regarding the size, structure, and map positions of RNA molecules was obtained by analysis of nuclease-resistant RNA-DNA duplexes (1). The cloned *Eco*RI fragment from the uniformly ^{32}P-labeled recombinant was hybridized to unlabeled cytoplasmic RNA from cells infected with vaccinia virus in the presence of cycloheximide. This was done at a temperature and formamide concentration which prevented DNA-DNA reannealing. After hybridization, single-stranded DNA was digested with nuclease S1 and resistant DNA-RNA hybrids were separated by electrophoresis in a neutral agarose gel and then visualized by autoradiography or fluorography. The sizes of the duplexes were determined from the positions of radioactively labeled DNA restriction fragment markers resolved on the same gel.

In initial experiments, a constant amount of RNA was hybridized to increasing amounts of DNA to ensure that we were working in DNA excess (Fig. 10, track 1 through 4). A complex array of bands was seen. The lowest three, with sizes of 604, 747, and 1,057 bases, were each approximately 150 nucleotides shorter than the mRNA's encoding the 14K, 32K, and 38K polypeptides, respectively. This difference could have resulted from elimination of the polyadenylate [poly(A)] tail by nuclease S1 digestion. The higher-molecular-weight RNAs, for which no messenger function has been demonstrated, were more clearly revealed by this procedure than by blot hybridizations.

Less viral RNA was present when cytosine

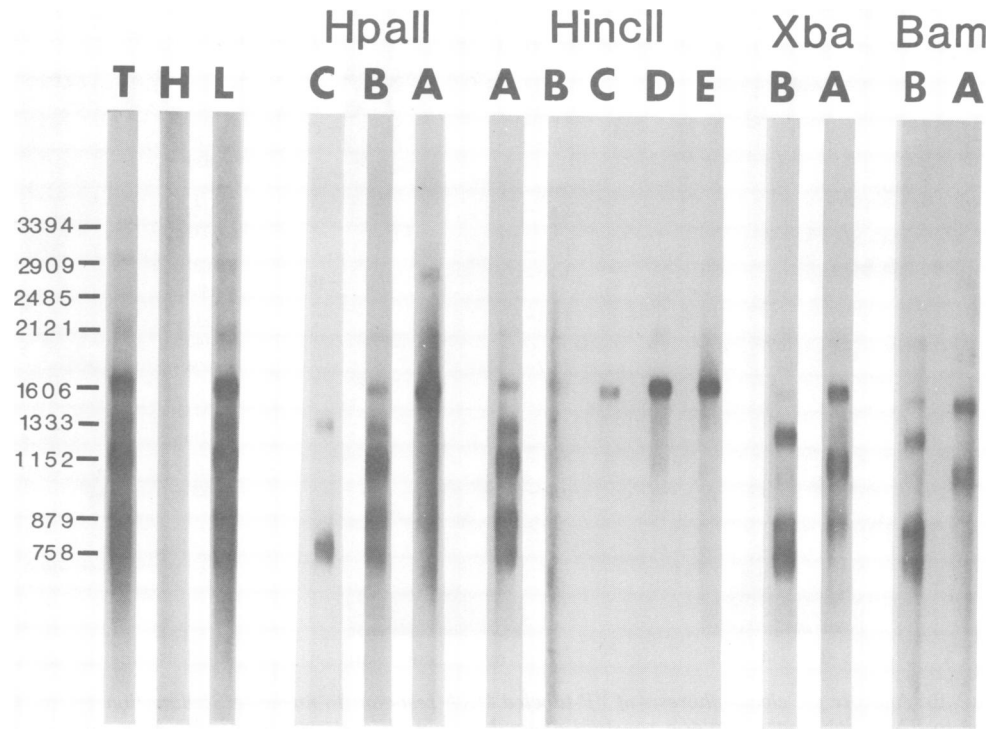

FIG. 9. *Autoradiograph of size-fractionated RNAs hybridized to ^{32}P-labeled DNA probes. Total cytoplasmic RNA was denatured with dimethyl sulfoxide and treated with glyoxal before agarose gel electrophoresis. After this, the RNA was transferred by blotting to diazotized paper and hybridized to ^{32}P-labeled total recombinant DNA (T), heavy DNA strand (H), light DNA strand (L), or the indicated restriction fragment of the cloned insert. The sizes in bases of the RNAs detected are shown at the left.*

arabinoside was used instead of cycloheximide. However, except for their lighter intensity, the pattern of nuclease S1-resistant bands obtained by using RNA made in the presence of cytosine arabinoside and cycloheximide were very similar (Fig. 10). One curious difference was the replacement of the 1,057-bp band with one very slightly larger (marked by a dot). This difference may be significant, since the 1,057-base RNA appeared to code for the 38K polypeptide, which appeared to be made in very low amounts, in cell-free systems programed with RNA from cytosine arabinoside-treated cells.

When late RNA was examined by this technique, discrete bands were more difficult to see against a smeared high-molecular-weight background (Fig. 10), as might be expected from the heterogeneity indicated by cell-free translations of size-fractionated RNA. However, apparent bands of about 3,352 and 1,512 bases fell within the size range giving maximal amounts of 40K polypeptide (Fig. 5).

Additional experiments were designed to evaluate the possibility of splicing. Berk and Sharp (2) demonstrated that after hybridizing a spliced mRNA to genomic DNA, the nonhybridized intron DNA is susceptible to S1 nuclease degradation. As a result, the RNA-DNA duplexes contain interrupted segments of DNA which remain associated under neutral conditions but dissociate under alkaline conditions. Accordingly, splicing can be revealed by comparing the sizes of the S1 nuclease-protected fragments by electrophoresis under neutral and alkaline conditions. When we analyzed the fragments of the cloned vaccinia DNA protected by immediate early or early RNAs from S1 nuclease under alkaline conditions, however, the pattern of bands and the sizes of the individual DNA fragments were very similar to those observed under neutral conditions (Fig. 10). Thus, there was no evidence for splicing of either class of early vaccinia virus RNAs encoded within this region of the genome. The neutral and alkaline electrophoresis patterns obtained by using late RNA for hybridization were also similar to each other, although the resolution appeared slightly greater in the alkaline gel.

In Table 1, we have compared the sizes of the polypeptides, the estimated minimal coding

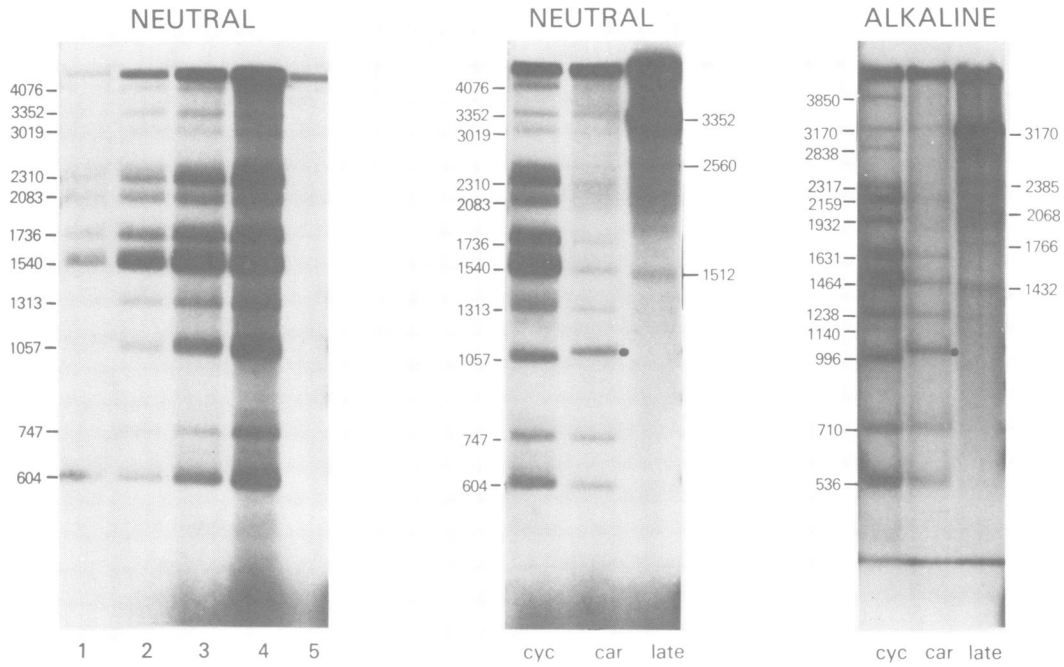

FIG. 10. *Agarose gel electrophoresis of ^{32}P-labeled DNA resistant to nuclease S1 digestion after hybridization to RNA. For tracks 1 to 4, a constant amount (25 μg) of total cytoplasmic RNA was hybridized to increasing amounts of purified insert from ^{32}P-labeled recombinant DNA before nuclease S1 digestion. For sample 5, no RNA was added. In the other tracks, a constant amount of ^{32}P-labeled DNA equivalent to that in track 3 was hybridized with 25 μg of RNA from cells infected in the presence of cycloheximide (cyc), cytosine arabinoside (car), or no inhibitors (late) before nuclease S1 digestion. The use of neutral and alkaline gel conditions and sizes in bases of the DNA segments are indicated.*

TABLE 1. *Size of mRNA's and polypeptides*

Mol wt of polypeptide[a]	No. of amino acids[a]	Minimal coding length[b]	RNA size			
			Methylmercury[c]	Glyoxal[c]	S1, neutral	S1, alkaline
14,000	122	366	680	758	604	536
32,000	278	834	860	879	747	710
38,000	330	990	1,290	1,152	1,057	996
40,000	347	1,041	920–3,100			

[a] Molecular weight divided by 115.
[b] No. of amino acids × 3.
[c] Includes poly(A).

lengths, and the actual RNA sizes determined by the several different methods used. As already mentioned, a correction of 100 to 150 residues for the poly(A) tail would make the results obtained by denaturing gels and nuclease S1 procedures comparable. The estimated sizes of the nuclease S1-resistant RNA-DNA hybrids in neutral gels were about 50 bases longer than the sizes of DNA segments in alkaline gels. Although the same DNA restriction fragments were used as markers for both, it is possible that RNA-DNA hybrids migrated very slightly slower than DNA-DNA hybrids, causing overestimation of size. Alternatively, the nuclease S1 treatment might not have removed all of the adenylate residues at the 3′ end of the RNA, leading to slight electrophoretic retardation. The minimal coding lengths suggest that the sizes of the 32K and 38K polypeptides may have been slightly overestimated by sodium dodecyl sulfate-gel electrophoresis.

Mapping the 5′ ends of RNAs by in vitro capping. Specific labeling of the 5′ ends of RNA molecules was done by in vitro capping. This procedure involved the removal of the original unlabeled cap by periodate oxidation and β-

elimination followed by addition of a new cap by incubating the RNA with [α-^{32}P]GTP and vaccinia virus RNA guanylyltransferase (22). The specificity of the enzyme for di- or triphosphate termini (18) ensured that labeling did not occur within the body of the message. In addition, a portion of the labeled RNA was digested with nuclease P1 and alkaline phosphatase and analyzed by paper electrophoresis to prove that the label was exclusively in cap structures.

Cap-labeled immediate early RNA was then hybridized to immobilized *Hpa*II and *Hinc*II fragments of the cloned DNA. Nonhybridized RNA was removed by washing and by RNase treatment. The fluorograph shown indicates that the capped ends of the RNAs hybridized mainly to *Hpa*II-B and *Hinc*II-A, although some hybridization to *Hinc*II-B was also detected (Fig. 11). These results were consistent with the map positions of the immediate early mRNA's. When similar experiments were attempted with late RNA, insufficient labeled material was obtained to detect hybrids. Subsequent experiments with capped late RNA, however, showed hybridization primarily to *Bam*HI-A (M. Haffey, personal communication).

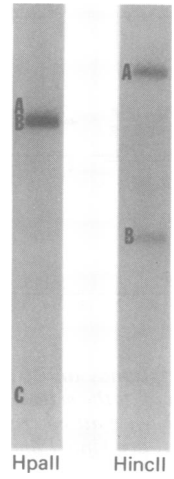

FIG. 11. *Hybridization of cap-labeled RNA to immobilized DNA restriction fragments. Recombinant DNA was cleaved with EcoRI, to separate the cloned DNA from the λ vector, and then with HpaII or HincII. After agarose gel electrophoresis, the DNA fragments were transferred to a nitrocellulose sheet and hybridized to ^{32}P-cap-labeled immediate early RNA. After washing and RNase A treatment, a fluorograph was made. DNA fragments were located by probing parallel strips with ^{32}P-labeled vaccinia virus DNA.*

DISCUSSION

The present report, together with previous ones (6, 21–23), provides a preliminary transcriptional and translational map of the left 21 kbp of the vaccinia virus genome. A total of 11 early mRNA's encoding polypeptides of 6K and 60K and two late mRNA's encoding polypeptides of 22K and 40K were identified. Inspection of the map of the early RNAs (Fig. 12) reveals several interesting features. For example, the majority of mRNA's were transcribed from the leftward-reading DNA strand. Indeed, the only two mRNA's that hybridized to the rightward-reading DNA strand were encoded within the inverted terminal repetition. Therefore, at the right end of the genome, those two mRNA's were also encoded by the leftward-reading DNA strand. Thus, all of the early genes mapped so far would be expressed if only one DNA strand were transcribed. Late mRNA's also hybridized to the leftward-reading DNA strand, indicating that temporal regulation of expression did not occur by transcriptional strand switching. To what extent this predominance for one strand will hold for mRNA's transcribed from other regions of the genome remains to be seen. Certainly, some transcripts, although not necessarily with messenger function, must be copied from the rightward-reading DNA strand because of the large amounts of complementary RNA made late in infection (3, 4).

Given the very large size of the vaccinia virus genome, it is interesting that some mRNA's appeared to be transcribed in an overlapping fashion. This is a tentative conclusion, however, that is not yet based on sequence data. Although certain regions of the genome appeared to be crowded with mRNA's, other portions appeared unexpressed (Fig. 12). This is somewhat misleading, since additional high-molecular-weight RNAs were detected both by hybridization of ^{32}P-labeled DNA fragments to size-fractionated RNA and by nuclease S1 protection experiments. The nature of these RNAs is uncertain, however, since their translation products have not been identified. Although technical problems cannot be excluded, it is possible that these RNAs were not mature messages but were either high-molecular-weight precursors of mRNA's (15) or read-through transcripts.

Perhaps the most puzzling data resulted from attempts to size fractionate late mRNA by electrophoresis. Despite the use of methylmercuric hydroxide under conditions that denatured double-stranded DNA and double-stranded RNA, late mRNA's migrated as heterogeneous-size molecules up to and exceeding 6,000 bases. Similar observations have been made independently by Brian Roberts and co-workers (personal communication) during investigations of vaccinia virus mRNA made in L cells. This heterogeneity is not due to variability in poly(A) length, since

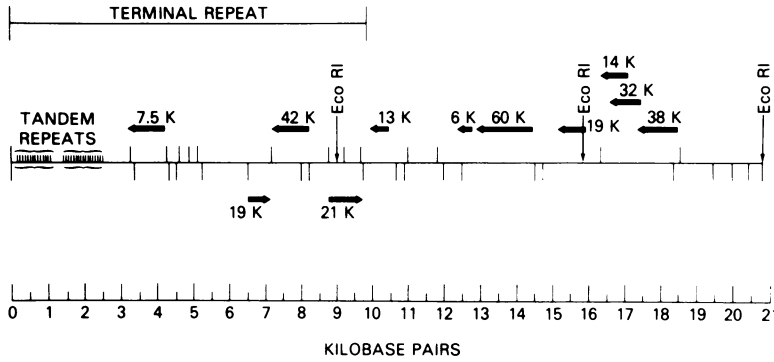

FIG. 12. *Transcriptional and translational map of the left 21 kbp of the vaccinia virus genome. Early mRNA's are shown with the sizes of the polypeptides that they encode. The late mRNA's for the 22K and 40K polypeptides have been omitted. Except for the tandem repeats at the far left, the vertical lines above and below the horizontal are HpaII and HincII sites, respectively.*

late RNA has been reported to contain uniform poly(A) tails of about 100 nucleotides (13). We suspect either that late genes do not have strong termination signals or that late in infection the termination mechanism becomes less effective. If the latter were true, then one might expect residual early mRNA's made late to also be heterogeneous. We have preliminary evidence that RNA as large as 3,100 bases encoded small amounts of the 14K polypeptide late in infection. Indeed, Oda and Joklik (14) reported many years ago that early sequences which are represented in late RNA are in larger RNA species. Failure to properly terminate transcription on both DNA strands would also explain the large amounts of double-stranded RNA and the branched RNA structures seen by electron microscopy after incubating late RNA under hybridization conditions (3, 4). It is interesting that large heterogeneous RNA is synthesized in vitro by vaccinia virus cores when ATP is depleted (16), ATP analogs are added (8), or the virions are heat treated (9). We have speculated previously that the DNA- or RNA-dependent nucleoside triphosphate phosphohydrolases (8) might function as a rho factor which mediates ATP-dependent RNA termination in procaryotes (10). If ATP is depleted late in infection, failure of termination could occur.

The presence of RNA from opposing DNA strands was used to explain our previous finding that self-annealed late RNA was untranslatable unless denatured (5). Similarly, late RNA selected by hybridization to the λ34 recombinant was also untranslatable after self-annealing, suggesting the presence of RNA from the opposite DNA strand in this region of the genome. The finding of only one late polypeptide, however, does not indicate that the complement is translatable.

Results obtained using the S1 nuclease procedure (1, 2), blot hybridization of cap-labeled RNA, and the mapping of each RNA to a single or two contiguous fragments of DNA were all consistent with the likelihood that the mRNA's encoded within the DNA segment investigated here were unspliced. However, further work is necessary to rigorously exclude this possibility of splicing, since it would be difficult to detect short intervening sequences located near the ends of genes. Similar negative results were also obtained upon analysis of the mRNA's encoded within the terminal 9 kbp of the genome (22). Anomalous nuclease S1 results were obtained with RNAs encoded near the junction of the terminal repetition and unique region (Wittek and Moss, manuscript in preparation). However, this was shown to result from a small deletion or rearrangement within the genome and not from RNA processing.

In some experiments, very little of the 38K polypeptide was made. This was particularly noticeable when RNA from cytosine arabinoside-treated cells was used and might be correlated with a slight difference in the size of the putative message for the 38K polypeptide, determined by nuclease S1 digestion of DNA hybrids formed with RNA from cycloheximide- and cytosine arabinoside-treated cells. Clearly, further work will be required to sort out the multiple, apparently overlapping, RNAs from this region of the genome and to determine whether the proteins share amino acid sequences.

It is interesting to note that a pattern of polypeptides, apparently identical to those made with vaccinia virus mRNA's, were made with rabbitpox virus immediate early RNA purified by hybridization to λ34 recombinant DNA (Cooper, unpublished data). By contrast, a sim-

ilar comparison of vaccinia virus and rabbitpox virus polypeptides made with RNA that hybridized to the 9-kbp terminal *Eco*RI fragment of vaccinia virus DNA revealed some differences (6). Thus, the genes present in the λ34 recombinant appear to have been highly conserved.

The few late mRNA's encoded within the terminal 21 kbp is consistent with a survey recently made of the left half of the genome (Belle Isle et al., in press). That study, which consisted of translating RNA that hybridized to 13 cloned *Hin*dIII fragments, indicated that the major late mRNA's are encoded within the central portion of the genome. Attempts to prepare detailed transcriptional and translational maps within the predominantly late regions are now in progress.

ACKNOWLEDGMENTS

We thank Ernest Barbosa and Alan Gershowitz for hybridizing cap-labeled RNA-to-DNA restriction fragments, Norman Cooper for RNA preparations, Mary Haffey for carefully reading the manuscript, and Jean Carolan and members of the National Institute of Allergy and Infectious Diseases editorial office for assistance in preparing the manuscript.

LITERATURE CITED

1. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNA's by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
2. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNA's of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **75**:1274–1278.
3. **Boone, R. F., R. P. Parr, and B. Moss.** 1979. Intermolecular duplexes formed from polyadenylylated vaccinia virus RNA. J. Virol. **30**:365–374.
4. **Colby, C., C. Jurale, and J. R. Kates.** 1971. Mechanism of synthesis of vaccinia virus double-stranded ribonucleic acid in vivo and in vitro. J. Virol. **7**:71–76.
5. **Cooper, J. A., and B. Moss.** 1979. *In vitro* translation of immediate early, early, and late classes of RNA from vaccinia virus-infected cells. Virology **96**:368–380.
6. **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.
7. **Garon, C. F., E. Barbosa, and B. Moss.** 1978. Visualization of an inverted terminal repetition in vaccinia virus DNA. Proc. Natl. Acad. Sci. U.S.A. **75**:4863–4867.
8. **Gershowitz, A., R. F. Boone, and B. Moss.** 1978. Multiple roles for ATP in the synthesis and processing of mRNA by vaccinia virus: specific inhibitory effects of adenosine (β,γ-imido)triphosphate. J. Virol. **27**:399–408.
9. **Harper, J. M. M., M. T. Parsonage, H. R. B. Pelham, and G. Darby.** 1978. Heat inactivation of vaccinia virus particle-associated functions: properties of heated particles in vivo and in vitro. J. Virol. **26**:646–659.
10. **Lowry-Goldhammer, C., and J. P. Richardson.** 1974. An RNA-dependent nucleoside triphosphate phosphohydrolase (ATPase) associated with rho termination factor. Proc. Natl. Acad. Sci. U.S.A. **71**:2003–2007.
11. **Mackett, M., and L. C. Archard.** 1979. Conservation and variation in Orthopoxvirus genome structure. J. Gen. Virol. **45**:683–701.
12. **McMaster, G. K., and G. G. Carmichael.** 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. U.S.A. **74**:4835–4838.
13. **Nevins, J. R., and W. K. Joklik.** 1975. Poly(A) sequences of vaccinia virus messenger RNA: nature, mode of addition and function during translation *in vitro* and *in vivo*. Virology **63**:1–14.
14. **Oda, K., and W. K. Joklik.** 1967. Hybridization and sedimentation studies on "early" and "late" vaccinia messenger RNA. J. Mol. Biol. **27**:395–419.
15. **Paoletti, E.** 1977. High molecular weight virion-associated RNA of vaccinia. A possible precursor to 8 to 12S mRNA. J. Biol. Chem. **252**:872–877.
16. **Paoletti, E., and B. R. Lipinskas.** 1978. The role of ATP in the biogenesis of vaccinia virus mRNA *in vitro*. Virology **87**:317–325.
17. **Smith, H. O., and M. L. Birnstiel.** 1976. A simple method for DNA restriction site mapping. Nucleic Acids Res. **3**:2387–2398.
18. **Venkatesan, S., A. Gershowitz, and B. Moss.** 1980. Modification of the 5′ end of mRNA: association of RNA triphosphatase with the RNA guanylyltransferase·RNA(guanine-7-)methyltransferase complex from vaccinia virus. J. Biol. Chem. **255**:903–908.
19. **Venkatesan, S., and B. Moss.** 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.
20. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization using dextran sulfate. Proc. Natl. Acad. Sci. U.S.A. **76**:3683–3687.
21. **Wittek, R., E. Barbosa, J. A. Cooper, C. F. Garon, H. Chan, and B. Moss.** 1980. The inverted terminal repetition in vaccinia virus DNA encodes early mRNAs. Nature (London) **285**:21–25.
22. **Wittek, R., J. A. Cooper, E. Barbosa, and B. Moss.** 1980. Expression of the vaccinia virus genome: analysis and mapping of mRNA's encoded within the inverted terminal repetition. Cell **21**:487–493.
23. **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.
24. **Wittek, R., A. Menna, H. K. Muller, D. Schumperli, P. G. Boseley, and R. Wyler.** 1978. Inverted terminal repeats in rabbit poxvirus and vaccinia virus DNA. J. Virol. **28**:171–181.
25. **Wittek, R., and B. Moss.** 1980. Tandem repeats within the inverted terminal repetition of vaccinia virus DNA. Cell **21**:277–284.