The genome of minute virus of mice, an autonomous parvovirus, encodes two overlapping transcription units

David Pintel, Dinshaw Dadachanji, Caroline R.Astell* and David C.Ward

Departments of Human Genetics and Molecular Biophysics-Biochemistry, Yale University School of Medicine, New Haven, CT 06510, USA and *Department of Biochemistry, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, V6T 1WF, Canada

ABSTRACT
     Four virus-specific transcripts have been identified in murine cells
infected with Minute-Virus-of-Mice (MVM). These RNAs, 4.8, 3.3, 3.0 and
1.8 kilobases in length, designated R1 to R4 respectively, are all trans-
scribed from the virion (-) strand of DNA and they are all polyadenylated
and spliced. The R1 transcript is derived from sequences that reside on
the genome between 4.0 and 95 map units (mu). Transcript R2 is composed of
exon sequences derived from mu coordinates 4.0-10.0, 40-46 and 48-95. The
most abundant RNA, R3, is transcribed from sequences mapping between 40 and
95 mu. All three of these RNAs have a short intron sequence between 46-48
mu removed. The least abundant transcript, R4, has not been mapped
precisely, however it hybridizes with all three EcoRI fragments which span
the entire 5 kb genome. In vitro transcription of cloned restriction
fragments of MVM DNA confirm the existence of functional promoters at map
coordinates 4.0 and 39 and sequence analysis of these regions of the viral
DNA reveal the characteristic features of RNA polymerase II promoters.
These results indicate that MVM DNA encodes two overlapping transcription
units with separate promoters near the left end (4.0 mu) and middle (39
mu) of the genome.

INTRODUCTION
     Mammalian parvoviruses possess linear, single-stranded DNA genomes

that are only 4500-5200 nucleotides in length (1). The limited genetic

complexity of these genomes makes them attractive model systems for study-

ing features of eukaryotic gene expression. To date the majority of such

studies have been done with Adeno-Associated Virus, type 2 [AAV-2] (2-6), a

prototype of the defective parvovirus subgroup that require Adenovirus (7)

or Herpes simplex virus (8) functions to initiate their own replication.

The emerging picture of AAV-transcription (9) is that there are six species

of polyadenylated cytoplasmic RNA synthesized in infected cells These are

4.2, 3.9, 3.6, 3.3, 2.6 and 2.3 kilobases (kb) in length, respectively.

The 4.2 and 3.9 kb transcripts are initiated from a promoter at map

position (mu) 5.0 and differ only in that the 3.9 kb species has genomic

sequences between 40 and 48 mu spliced out (5,9). The same region of the

genome also has been spliced out in the 3.3 and 2.3 kb RNAs. These latter transcripts and their unspliced (3.6 and 2.6 kb) counterparts are initiated from promoters at 19.0 and 39.0 mu, respectively (5,9, E. Lusby and K. Berns, personal communication). Since each of the six RNAs share an extensive block of common sequence, derived from the genome between 48 and 96 mu, it is apparent that AAV DNA encodes three overlapping transcription units which coterminate at mu 96.

Initial studies on the transcription of the DNA from H-1 virus and Minute Virus of Mice (MVM), two members of the autonomous parvovirus sub-group that replicate without the aid of a helper virus (1), suggested that these viral genomes encode only a single transcription unit with a promoter near mu 4.0 (10,11). To further characterize the transcriptional organization of an autonomous parvovirus genome we undertook an analysis of the RNA species produced in MVM-infected mouse cells. The results presented here demonstrate the presence of four spliced viral RNAs, all transcribed from the virion strand of DNA. The three major transcripts have been shown to arise from two overlapping transcription units with separate promoters positioned near the left end (4.0 mu) and the middle (39 mu) of the genome. Additional features of the transcriptional organization of MVM DNA are discussed in the following paper, which presents the complete nucleotide sequence of the MVM genome (12). A preliminary report of some of these data has been presented (13).

MATERIALS AND METHODS
Virus and Cells

Plaque-purified MVM(p), the prototype strain of MVM, was grown in the A-9 strain of mouse L cells (14). The maintenance of cell cultures and virus stocks was as described (15).

To obtain ^{32}P-labeled viral DNA, carrier-free ^{32}P-orthophosphoric acid (New England Nuclear) was added at 80 μCi/ml to culture media, containing only 2-5% of the normal inorganic phosphate, 7-9 hours post infection. Cells were harvested 26 hours post infection, and viral DNA prepared from purified virions as described (16). This labeling condition yielded DNA of approximately 6-8x10^5 cpm/μg.

RNA extraction and cell fractionation

RNA was extracted from whole cells by the guanidinium chloride method described by Cox (17) and modified by Strohman et al. (18).

Cells were divided into nuclear and cytoplasmic fractions with Nonidet

P-40 (NP-40) as described by Borun et al. (19) and modified by Carter and
Rose (20).   RNA was extracted from the nuclear and cytoplasmic fractions
of infected and mock infected cells exactly as described by Carter and Rose
(20). Poly A$^+$ RNA was selected on an oligo(dT)-cellulose column
(Collaborative Research).

RNA-DNA Hybridization in Solution.

   a) In aqueous solution.

      RNA-DNA hybridizations were carried out in 2xSSC at 67°C for 2
hours (21). Each assay mix (100 µl), containing 2-5x10^3 cpm of ^{32}P-labeled
MVM DNA (2x10^5 cpm/µg) and varying concentrations of unlabeled RNA from
infected or mock-infected cells, was heat-denatured before annealing. At
the end of the hybridization incubation, 25 µl of each mix was TCA-precipi-
tated onto GF/A filters and counted, while 50 µl of the mix was digested
with S1 nuclease prior to TCA precipitation.

   b) In 80% formamide.

      Hybridizations under conditions favoring DNA-RNA annealing were
carried out as described (22,23). The hybridization buffer consisted of
purified (24) 80% formamide (Macalaster-Bicknell or Fisher), 0.4 M NaCl,
0.04 M PIPES-Na, pH 6.4, and 1 mM EDTA. The hybridization mix (100 µl)
usually contained between 30-100 ng MVM [^{32}P]-DNA, and between 3-36 µg
cellular RNA (as specified in the figure legends). Each assay mix was
first incubated at 80°C for 5 min to denature the nucleic acids and then
placed at the hybridization temperature. The optimum temperature for
selective hybridization was found to be 57°C, and the incubation time
ranged from 3-18 hours. At the end of each incubation, aliquots were
precipitated with TCA before and after treatment with either S1 or mung
bean nuclease.

S1 Nuclease Digestion.

      The digestion buffer consisted of 0.03 M NaOAc, pH 4.5, 0.25 M NaCl, 1 mM
ZnSO$_4$. Samples obtained from RNA-DNA hybridization assays were diluted 10-
fold in the above digestion buffer, and incubated with 65 U/ml S1 nuclease
(Sigma) in the presence of sonicated, heat-denatured calf thymus DNA (20
µg/ml) at 48°C for 75 min and the TCA-precipitable counts were then
determined in the presence of carrier RNA. (1 unit of S1 nuclease is
defined as the amount of enzyme that converts 1 µg of nucleic acid to a
TCA-soluble form in 1 min at 37°C, at pH 4.6).

Mung Bean Nuclease Digestion.

      Samples obtained from RNA-DNA hybridization assays were diluted 10-fold

with buffer (0.05 M NaOAc/HOAc, pH 5.0, 0.05 M NaCl, 1 mM $ZnCl_2$). Both
native and heat-denatured, sonicated calf thymus DNA, each at a final
concentration of 10μg/ml, were added and each assay mix was incubated with
20 U/ml mung bean nuclease (PL Biochemicals), at 45°C for 45 min (unless
otherwise indicated). Samples were either tested for TCA-precipitable
counts or alcohol-precipitated, and redissolved in TE 8.3 buffer, and
analyzed by gel electrophoresis. (1 unit of mung bean nuclease is defined
as the amount of enzyme that produces 1 μg TCA-soluble material in 1 min at
37°C, at pH 4.6).

RNA transfer and hybridization.

RNA from MVM infected cells was run on 1.4% agarose, 2.2 M formaldehyde
gels as described by Rave et al. (25), transferred to nitrocellulose and
hybridized to ^{32}P-nick-translated cloned DNA or ^{32}P-labeled vDNA as
described by Thomas (26).

In vitro transcription.

Extracts were prepared, reactions performed, and products analyzed as
described by Manley et al., (27,28). ^{32}P-GTP (400-600 Ci/mmole) was obtai-
ned from Amersham or New England Nuclear. Glyoxal, 40%, was from MCB Co.

Other Methods

The preparation and purification of MVM complementary DNA was accord-
ing to Faust and Ward (29). λ exonuclease was a gift from C. Radding.

[32]P-labeled MVM RF DNA was prepared by the method of Bourguignon et al.
(16). Restriction enzymes were purchased from New England Bio-Labs or
Boehringer-Mannheim and used under the reaction conditions recommended by
the manufacturer. Assays for symmetrical transcripts were described by
Aloni (30,31). Cloned viral DNA fragments were nick translated according
to the method of Rigby et al (32).

RESULTS

Initially, unlabelled RNA was prepared from MVM infected and mock
infected A9 cells 14 hours post infection and the RNA was extracted from
nuclear, cytoplasmic or whole cell fractions. These RNA preparations were
annealed in aqueous solution to ^{32}P-vDNA over a range of RNA/DNA ratios, and
assayed for the extent of hybridization by determining the proportion of
[32]P-radioactivity found resistant to the action of S1 nuclease. The re-
sults of these analyses are shown in Figure 1. It can be seen that less
than 10% of the viral DNA is resistant to S1 nuclease when it is annealed
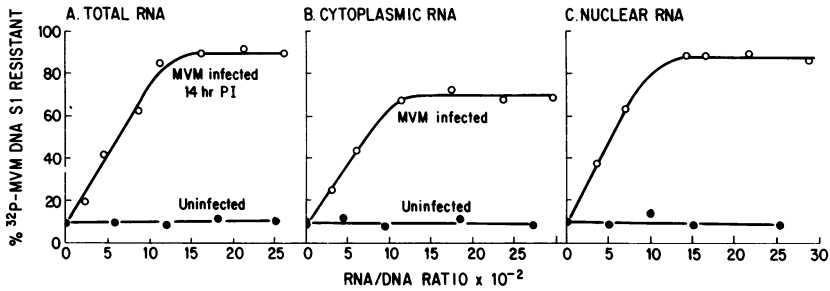to itself or to RNA from uninfected cells. This background level of annea-

Figure 1. Hybridization of MVM [^{32}P-]-vDNA to RNA extracted from MVM-infec-ted (o) and mock-infected (●) A9 cells 14 hours post infection. A) Total RNA, B) cytoplasmic RNA, C) nuclear RNA. Aqueous solution hybridization reactions were done as described in Materials and Methods.

ling can be attributed to the fact that the MVM genome possesses stable hairpin duplex structures at both the 3' and 5' termini which comprise ∿ 8% of the total viral DNA (33). Further, it can be seen that at saturation both total RNA (Fig. la) and nuclear RNA (Fig. lc) are complementary to approximately 80% (above background) of the viral genome sequences, while cytoplasmic RNA (Fig. lb) is complementary to approximately 60% of the viral genome. It can also be calculated from the slopes of these curves that at 14 hours post infection virus specific RNA constituted approximate-ly 0.065% of the pool of cellular RNA while other experiments (data not shown) have determined that at late times (20-30 hrs post infection) viral specific RNA approaches 0.3% of the total intracellular RNA.

MVM specific RNA is transcribed using only the viral strand as temp-late; we were unable to detect RNA which hybridized to viral C-strand probes. When MVM ^{32}P-cDNA was purified as described (29), hybridized in solution with cytoplasmic fractions of MVM infected A9 cells, and treated with S1 nuclease the percent of nuclease resistant cDNA did not rise above the background obtained using mock infected cell RNA, even at RNA/DNA ratios up to 60,000:1. At this ratio, had transcripts complementary to cDNA been represented as 5×10^{-5}% of the cellular RNA tested, the proportion of cDNA resistant to S1 would have risen to at least twice the background level. MVM ^{32}P-cDNA was also used to probe "Northern" blots from MVM infected cells. Exposure times 50X longer than those necessary to visualize RNA hybridizing to ^{32}P-vDNA probes detected a very weak pattern. However, the pattern obtained was identical to that seen using ^{32}P-vDNA probes, and most probably reflects hybridization with a trace amount of

contaminating vDNA in the cDNA preparations. A search for symmetrical transcripts, first described by Aloni and Attardi (30), which might have prevented detection of low abundance RNA complementary to cDNA (due to the hybridization of transcripts of opposite polarity to each other) was also negative. There was no increase of double stranded RNA species, as detected by pancreatic and Tl nuclease resistance, in MVM infected cells compared to mock infected cells, and neither pool of duplex RNA hybridized in solution with excess cDNA probes (data not shown).

To determine the number and size of viral specific transcripts, total cell RNA was extracted from MVM infected A9 cells 23 hours post infection and fractionated by electrophoresis in a formaldehyde-agarose gel. The RNA was transferred to nitrocellulose and hybridized with ^{32}P-labeled, nick-translated, MVM RF DNA. Four RNA species, 4.8, 3.3, 3.0 and 1.8 kilobases (kb) in length, can be detected by this type of analysis (Figure 2a). Similar experiments using poly A selected RNA and ^{32}P-labeled vDNA yield identical results (data not shown).

To determine the proper annealing conditions for analyzing hybrids by the nuclease procedure of Berk and Sharp, ^{32}P-labelled vDNA was annealed, in 80% formamide, at various temperatures, to total intracellular RNA taken from MVM infected A9 cells 26 hours post infection. The hybridization mix was digested with mung bean nuclease, and run on a 1.4% agarose gel under neutral pH conditions. For each temperature tested both DNA/DNA hybridization of separate viral restriction fragments and DNA hybridization to total RNA from mock infected cells were run in parallel. The results shown in Fig. 2b indicate that the optimum temperature for hybridization is $56^{\circ}C$-$58^{\circ}C$ (lanes 2-7). The gels also demonstrate the existence of four major hybrid bands when ^{32}P-labelled vDNA was hybridized with RNA extracted from MVM infected cells. The three largest of these bands 4.8, 3.3 and 3.0 kb correspond both in size and relative intensity to the three largest transcripts seen by Northern analysis. In addition, a fourth band appears at approximately 2.6 kb. This band is smaller than any of the major transcripts seen by Northern analysis and, as discussed below, is the size of the major shared exon of the three largest and most abundant RNA species. We believe that this band represents an artifact of the nuclease digestion, resulting from RNA strand scission at the point of a splice junction. Similar artifactual cleavages have been reported previously by others (10,34).

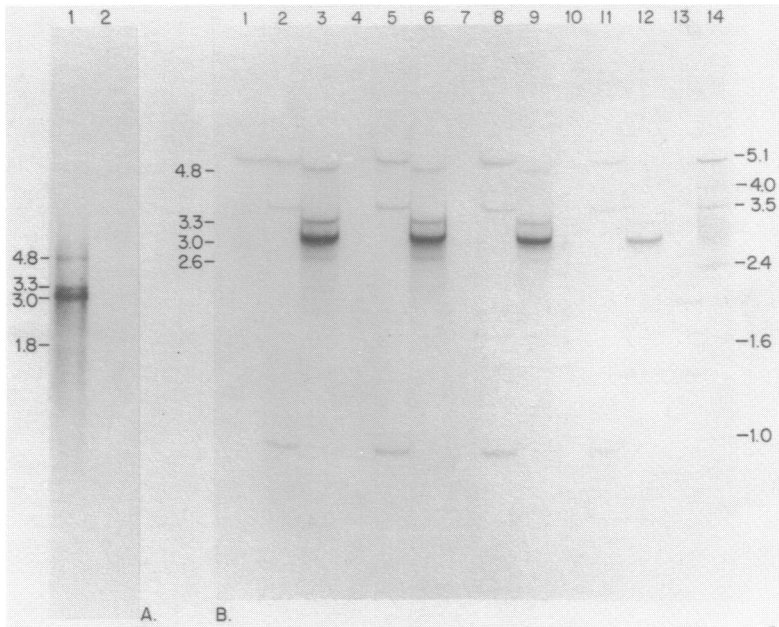Since the three largest transcripts represent over 95% of the viral RNA

Figure 2A. Hybridization of cloned MVM RF DNA ^{32}P-labeled by nick-translation to RNA extracted from MVM-infected (lane 1) and uninfected (lane 2) A9 cells 26 hours post infection. The methods of RNA extraction, blotting and hybridization are described in Materials and Methods. The relative abundance of the four viral transcripts are: 4.8 kb band (10-15%), 3.3 kb band (15-20%), 3.0 kb band (65-70%), and 1.8 kb band (2-3%). B. Effect of temperature on the specific formation of DNA/RNA hybrids. Total RNA (17 µg), prepared from infected A9 cells 26 hours post infection, was hybridized in 80% formamide buffer to MVM [32P]-vDNA (12,000 cpm, 21 ng) at 56° (lane 3), 58° (lane 6), 60° (lane 9) or 62° (lane 12), then treated with mung bean nuclease and electrophoresed on a 1.4% agarose gel for 4 hours at 100V. (At this RNA/DNA ratio, less than 20% of the DNA probe becomes S1 resistant). RNA (18 µg) from uninfected A9 cells was hybridized at these temperatures with the ^{32}P-MVM DNA probe (21 ng), nuclease treated and electrophoresed in lane 4 (56°), lane 7 (58°), lane 10 (60°) and lane 13 (62°). To control for DNA:DNA reannealing, ^{32}P-MVM RF DNA, synthesized in vitro (16), was subjected to partial digestion with EcoRI, heat denatured and annealed at the indicated temperatures for 2 hours in 80% formamide buffer. The three bands observed represent fragments which spontaneously renature since the V and C strands are covalently linked. The EcoRI A,B and AB fragments (see Figure 4), in which V and C strands are not covalently associated, do not reanneal at any of the four hybridization temperatures used.

in infected cells, total cytoplasmic RNA (extracted 26 hr post infection) was hybridized to ^{32}P-labelled vDNA using the optimal hybridization conditions described above, treated with mung bean nuclease, and run on
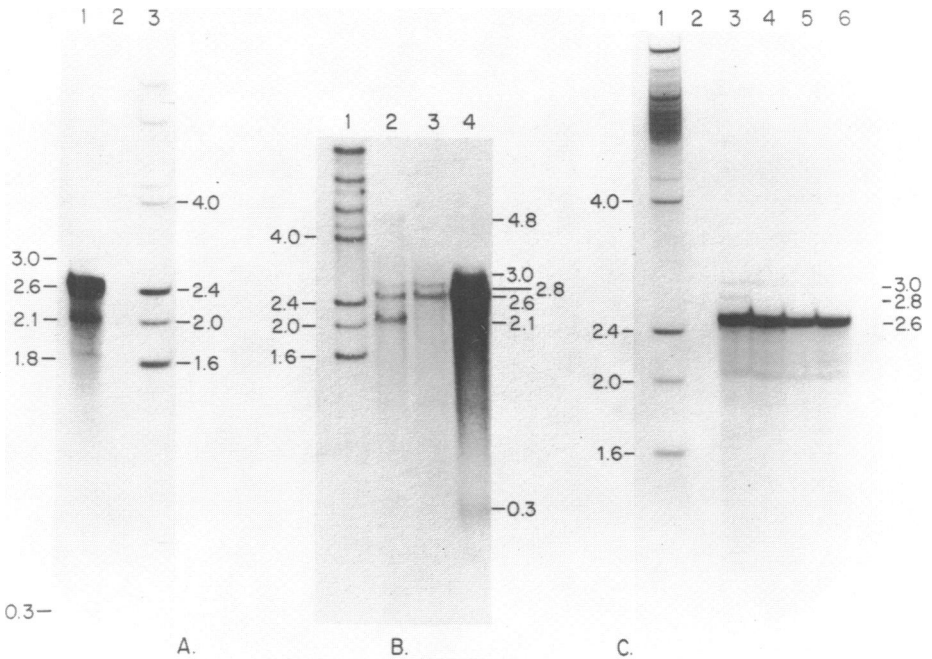
Figure 3. Alkaline gel analysis of MVM RNA-DNA hybrids obtained by the "Berk and Sharp" technique. A. Equal aliquots of MVM [^{32}P]-vDNA (20,000 cpm, 0.03 µg) were annealed with 36 µg of cytoplasmic RNA from MVM-infected A9 cells taken 24 hours post infection (lane 1) or from mock-infected cells (lane 2). Hybridization were performed in 80% formamide hybridization buffer, at 57°C for 4 hours. Samples were then digested with mung bean nuclease and run on an alkaline 1.4% agarose gel. Lane 3: partial digest of MVM [^{32}P]-RF DNA. B. Autoradiogram of an alkaline 1.4% agarose gel, analyzing the three major RNA-DNA hybrid bands individually isolated from a neutral gel. MVM [^{32}P]-vDNA was hybridized to total RNA from MVM-infected A9 cells (extracted at 25 h p.i.), in 80% formamide hybridization buffer at 57°C for 20 h (at an RNA/DNA ratio of 515:1). The sample was digested with mung bean nuclease, electrophoresed through a preparative 1.4% agarose gel, pH 8.3, and each of the three largest hybrid bands (see text) was sliced out, dissolved in saturated KI solution, and recovered by chromatography on hydroxylapatite. The RNA-DNA hybrids derived from the 4.8, 3.3 and 3.0 kb transcripts are shown in lanes 2 through 4, respectively. Lane 1: partial EcoRI digest of MVM [^{32}P]-RF DNA. C. The effect of varying mung bean nuclease concentrations on RNA/DNA hybrids. An aliquot of [^{32}P]-vDNA (72,000 cpm, 168 ng) was annealed with total RNA from MVM infected A9 cells extracted 26 hours post infection, at an RNA/DNA ratio of 200:1, in 80% formamide hybridization buffer, at 57°C for 13 h. The sample was then divided into 4 equal aliquots, digested with increasing concentrations of mung bean nuclease: 3.6 U/ml (track 3), 7.2 U/ml (track 4), 14.4 U/ml (track 5) and 28.8 U/ml (track 6), and electrophoresed on an alkaline 1.4% agarose gel. Lane 1: partial EcoRI digest of MVM [^{32}P]-RF DNA. Lane 2: blank.

alkaline 1.4% agarose gels. As seen in Fig. 3a, five bands are generated
under denaturing conditions, whereas four bands were observed under the
neutral gel conditions discussed above and shown in Fig. 2b. The very
prominent band at 2.6 kb in addition to the bands at 2.1 and 0.3 kb
represent exons which constitute the major RNA species as detailed below.
The origin of the minor bands at 3.0 and 1.8 kb are also discussed below.

When the three largest hybrid bands (4.8, 3.3 and 3.0 kb) illustrated
in Fig. 2b were individually eluted, and separately run on alkaline 1.4%
agarose gels the bands shown in Fig. 3b were observed. It can be seen
immediately that all three of the major transcripts have an exon fragment
2.6 kb in length. The 4.8 kb transcript (Fig. 3b, lane 2) can be seen to
contain the 2.1 kb exon fragment, previously noted in Fig. 3a. The 4.8 kb
transcript also protects DNA sequences on the alkaline gel which migrate at
4.8 kb and 2.8 kb. These bands could represent either incomplete digestion
products or minor unspliced species (see below). The 3.3 kb transcript can
be seen to contain the 2.6 kb exon, the 2.8 kb band, and, upon longer
exposure, a 0.3K exon band. The major 3.0 kb transcript generates the 2.6
kb exon band, a 0.3 kb exon band, the 2.8 kb band, the 3.0 kb band
(previously noted in Fig. 3a) and a minor band at 3.2 kb - larger than the
original 3.0 kb transcript.

The presence of a 2.8 kb band in each of these samples was unexpected
since no such band was observed in Fig. 3a, which displays the hybrid pro-
tected DNA fragments generated by the total viral RNA pool after alkalai
treatment. To account for this apparent discrepancy we next analyzed the
effect of varying mung bean nuclease concentration on the DNA/RNA hybrids
generated during such analysis. Fig. 3c shows that at low nuclease concen-
tration bands at 2.8 kb as well as at 4.9 and 5.1 kb are seen; these bands
disappear as the nuclease concentration increases. In contrast the bands
at 2.1, 2.6 and 3.0 kb in these digests remain unaffected by increasing the
nuclease level. (The 0.3 kb exon was run off the gel in this experiment in
order to get better resolution of higher molecular weight bands). We
interpret the 2.8 kb band, seen in Fig. 3b, as a partial digestion product
which is not present in the sample illustrated in Fig. 3a. Restriction
fragment mapping data presented below shows that the major 2.6 kb exon RNA
extends very close to the 5' hairpin duplex. We conclude that the partial
digestion band at 2.8 kb represents incomplete cutting at this site, near
the 5'-termini, which results in fragments approximately 250 nucleotides
longer than expected. The minor 3.2 kb fragment observed in the analysis

of the 3.0 kb transcripts (lane 4, Fig. 3b) is also interpreted to be due
to incomplete removal of the 5' terminal hairpin sequence. The 5.1 and 4.9
kb bands obtained at low nuclease concentration (lane 1 of Fig. 3c) most
likely results from incomplete nuclease digestion at one or both of the
terminal hairpin sequences.

   Although no band of 1.8 kb was observed in the nuclease protection
studies illustrated in Fig. 2b, where the protected DNA fragments were
resolved on neutral agarose gels, a minor 1.8 kb band is observed on the
alkaline gel shown in Fig. 3a. The band could reflect an unspliced form of
the 1.8 kb transcript seen by Northern analysis (Fig. 2a).   Unfortunately,
since this RNA species appears in such low abundance in MVM infected cells we
have not been able to analyze this RNA further. However, the 1.8 kb
transcript does hybridize to all three EcoRI fragments of MVM RF DNA, which
span the entire 5 kb genome of MVM, suggesting that the majority of the
molecules in the RNA species are spliced.

   Although Exonuclease VII has been used widely in nuclease protection
experiments to determine the size of intron sequences, such studies, using
whole MVM genomic DNA as probe, have proven extremely difficult in our hands
since the hairpin duplexes at the genomic termini render the MVM-vDNA probe
extremely resistant to this enzyme. Therefore, the sites and sizes of
intron sequences and the map coordinates of the splice junctions were
determined directly by the method of Berk and Sharp (23) using mung bean
nuclease. Total cellular RNA from MVM infected A9 cells, taken at 26 hours
post infection, was hybridized in solution with isolated restriction
fragments of ^{32}P-labelled MVM RF DNA. After treatment with mung bean
nuclease the RNA protected DNA fragments were analyzed under neutral and
denaturing gel conditions. The Hind III B and Hind III C fragments, which
map between 54-80.5 (nucleotide 2650-3993) and 80.5-100 (nucleotide
3993-5081) respectively on the viral genome, are wholly protected from
nuclease in the above analysis. Nucleotide numbers are taken from the
complete nucleotide sequence of MVM, (12). The alkaline agarose gel profile
shown in Figure 4c indicates that the rightward half of the genome is
transcribed without interruption, and since the Hind III B and C fragments
span a region approximately 2440 nucleotides in length and the two largest
exons have been determined to be 2.6 and 2.1 kb, this region must encode
the large exon of 2.6 kb, common to the three most abundant transcripts,
demonstrated in Figure 3a. There is also a diffuse band generated from the
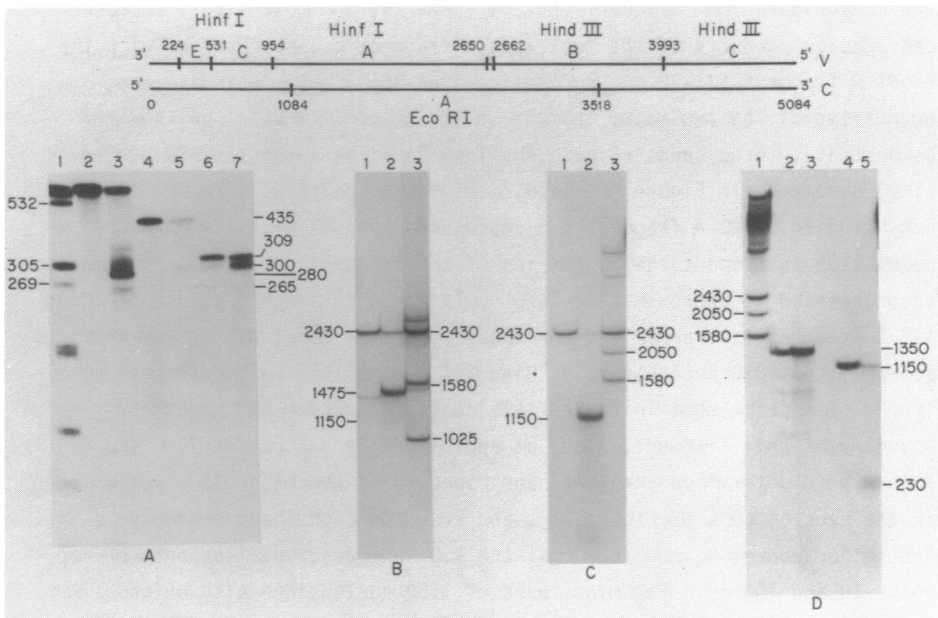Hind III C hybrid on the alkaline gel which migrates at approximately

Figure 4. Analysis of hybrids between [^{32}P]-labeled restriction fragments of MVM RF DNA and total cell RNA which had been extracted from MVM infected A9 cells 26 h post infection. Hybridization were performed in 80% forma-mide hybridization buffer at 57°C for 15-18 h at an RNA/DNA ratio of 100:1 (at which ratio only 5% [above background] of the genome is S1 resistant). The samples were then digested to completion with mung bean nuclease and run on either an alkaline 1.4% agarose gel, Fig. 4D, 4C; neutral 1.4% agarose gel, Fig. 4B; or an 8M urea - 8% polyacrylamide gel, Fig. 4A. A map of the restriction fragments appears on top. A: analysis of nuclease treated hybrids obtained with HinfI A (956-2664), lane 3; Hinf I C (534-956), lane 5; and Hinf I E (225-534), lane 7. Lanes 2,4,6: unhybridized, purified ^{32}P-labeled HinfI A, C and E fragments respectively. Lane 1: HhaI digest of φX174 [^{32}P]-RF DNA. B. Neutral agarose gel analysis of nuclease treated hybrids obtained with the EcoRI A (1086-3521) fragment (lane 2). Lane 1: [^{32}P]-labeled purified EcoRI A fragment before hybridization. Lane 3: EcoRI digest of MVM [^{32}P]-RF DNA. C. Alkaline agarose gel analysis of the same samples as in B. D. Alkaline agarose analysis of nuclease treated hybrids with the Hind III B (2652-3996) fragment (lane 3) and Hind III C (3996-5081) fragment (lane 5). Lanes 2 and 4 are [^{32}P]-labeled purified Hind III B and C fragments respectively, before hybridization. Lane 1: EcoRI digest of MVM [^{32}P]-RF DNA.

220-240 nucleotides (Figure 4d, lane 5). This is the size of the 5'

terminal hairpin which is most likely released intact after nuclease

digestion. Although four additional minor bands are seen in the HindIII B

protection experiment, (Fig. 4d, lane 3), these bands are also present in

the unhybridized DNA fragment control (lane 2), and most likely represent
degradation products of the restriction fragment probe. Hybrids with the
EcoRI A fragment (21-70 mu, nucleotide 1084-3518) help delineate the
boundaries of the two major introns in the genome. Neutral gel analysis
demonstrates three bands (Figure 4b, lane 2). The largest (2430 nucleo-
tide) band seen in Figure 4b, lane 2, migrates identically to the
unhybridized EcoRI A fragment and represents hybrids with the RNA sub-
population that spans the boundaries of the fragment. However, the major
band observed is approximately 1475 nucleotides in length. Since the Hind
III B fragment is transcribed co-linearly and the EcoRI A fragment overlaps
approximately 870 nucleotides of Hind III B, the 1475 bp hybrid must be
from a transcript that initiates 1475 nucleotides from the right end
boundary of this fragment, i.e., at approximately nucleotide 2000 (mu 40).
This hybrid determines the left hand boundary of the major 3.0 hybrid (seen
in the neutral gels of Figure 2b), and from the right hand boundary
determined above, we conclude that the 3.0 kb transcript lies between map
units 40 and 100 mu. The minor band of 1150 nucleotides also on the
gel is the size of the largest exon segment protected in the hybrid (see
below) and constitutes another example of the nuclease digestion artifact
described above. On alkaline gels (Figure 4c) the major band migrates at
1150 nucleotides and, in addition, there is a less intense band at 0.3 kb.
This indicates that the transcript which initiates at mu 40 continues for
approximately 300 nucleotides, contains a small splice between mu 46-48,
and then continues uninterrupted through the remainder of the EcoRI A
fragment and through the Hind III B and C fragments as well. Together
these results demonstrate that the common 2.6 kb exon maps between mu 48
and 96 on the genome.

The Hinf I A fragment (mu 17.1 to 49.4, nucleotide 954-2662) overlaps
the majority of the EcoRI A fragment and is useful in confirming the re-
sults described above. Neutral gel analysis shows that approximately
650-690 nucleotides of the Hinf I A fragment are protected by MVM RNA (data
not shown); this provides further evidence that the 3.0 kb transcript
starts at mu 40. The hybrids protected by the Hinf I A fragment were also
run on alkaline acrylamide gels as shown in Figure 4a. It can be seen that
the 650-690 hybrid forms multiple bands centering around 280 nucleotides
(Figure 4a, lane 3). The sizes of these fragments corroborate the
existence of two exons of approximately 265-330 nucleotides, which are
separated by a small splice at mu 46-48. The multiple band pattern seen

may indicate heterogeneity in these RNA species. As expected, there is also a minor hybrid fragment that spans the entire Hinf I fragment, indicative of the large 4.8 kb transcript that is unspliced. Alkaline gel analysis of hybrids with HinfI C fragment (mu 9.6-171, nucleotide 531-954) shows no interruption of the coding sequence in this region (Fig. 4a, lane 5). Analysis of hybrids with the Hinf I E fragment (mu 4.3-9.6, nucleotide 224-531), however, gives evidence for two differentially spliced molecules coded for by this region. Neutral gel analysis shows that all RNA hybridizing with this fragment span the entire fragment (data not shown) while the alkaline acrylamide gel displayed in Fig. 4a, lane 7 demonstrates that while one species remains unspliced through the region, another species is truncated approximately 25 nucleotides from the 3' end of the fragment, or approximately nucleotide 510, map position 10. This leaves a leader sequence 300 nucleotides long between mu 4.5-10. Since the 3.3 kb message appears to be comprised of two 0.3 kb leader sequences attached to the common 2.6 kb exon, the smaller region of HinfI E which is protected encodes the leftmost of these two leaders segments.

To corroborate the emerging map of viral transcripts, we next probed four identical samples of total RNA from MVM infected cells which had been electrophoresed through formaldehyde-agarose and transferred to nitrocellulose with four ^{32}P-labeled cloned MVM DNA fragments from different areas across the genome. As expected the cloned Hind III C fragment (nucleotide 3993-5081) and the cloned EcoRI A (nucleotide 1084-3518) fragment hybridized to each of the 4.8, 3.3, and 3.0 kb messages. Cloned DNA from nucleotide 1084-1661, which falls totally within the large intron hybridizes with only the 4.8 kb message, and the cloned PvuII C fragment (nucleotide 1-760) hybridizes with only the 4.8 and 3.3 kb messages (data not shown).

The map coordinates of the 3 major transcripts, deduced from the data presented above, are summarized in Figure 5. The largest message, R1, is initiated at the left end of the genome at ∿ mu 4 and extends the length of the genome to ∿ mu 96. Most of these molecules have a small region between mu 46-48 spliced out, although there apparently does exist a subpopulation that lack the small splice. The next largest transcript, R2, which also begins at ∿ mu 4.0, extends to ∿ mu 10 at which point it is spliced to ∿ mu 40 on the genome. The RNA continues ∿ 300 more nucleotides downstream at which point a small region between mu 46-48 is spliced out; it then continues uninterrupted to ∿ mu 96. This message lacks two blocks of
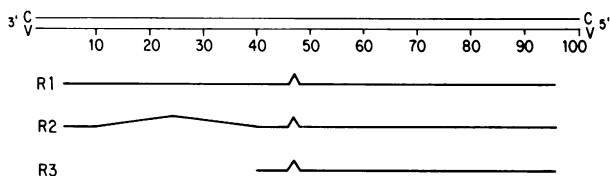
Figure 5. Map displaying the organization of the three major transcripts of MVM (see text for details).

genomic sequences, one large block between 10-40 mu and one small block between 46-48 mu. The map coordinates of the R2 transcript deduced by the nuclease protection study agree extremely well with those determined by Tal et al. (11) using electron microscopic measurements of DNA-RNA hybrids. However, the small splice in this transcript at 46-48 mu was not observed by Tal et al., presumably because the resultant intron loop was too small to be detected by electron microscopy. Since only hydrid molecules with detectable intron loops were analyzed in the EM study, the map coordinates of the most abundant 3.0 kb transcript were not reported. The data presented above indicates that the major R3 transcript is initiated from a promoter near mu 40 and possesses sequences complementary to genomic DNA mapping between 40-46 and 48-96 mu. It remained possible however, that the above analyses might have missed a very smaller leader sequence on the R3 transcript encoded by the left end of the genome around mu 4.

To corroborate the existence of two promoters in the MVM genome we have examined restriction fragments from viral RF DNA and cloned MVM DNA segments for promoter activity in the standard in vitro transcription system developed by Manley et al.,(27). The cloning of the complete molecule of MVM or subgenomic fragments of the viral DNA into pBR322 and the characterization of the resultant clones will be presented elsewhere (Merchlinsky et al., manuscript in preparation).

In our hands the in vitro HeLa cell transcription system neither splices nor polyadenylates transcripts, however it is very useful for identifying and mapping promoter sites onto isolated DNA. When full length MVM RF DNA cloned into pBR322 is used to prime the in vitro transcription system two transcripts approximately 4.8 and 3.2 kb in length are produced (Fig. 6, part A). The 3.2 kb band is a doublet because the template pool contains a minor population of molecules which contain a deletion of approximately 85 base pairs within the 5' hairpin sequences of MVM which is generated during propagation in bacteria (Merchlinsky et al., unpublished
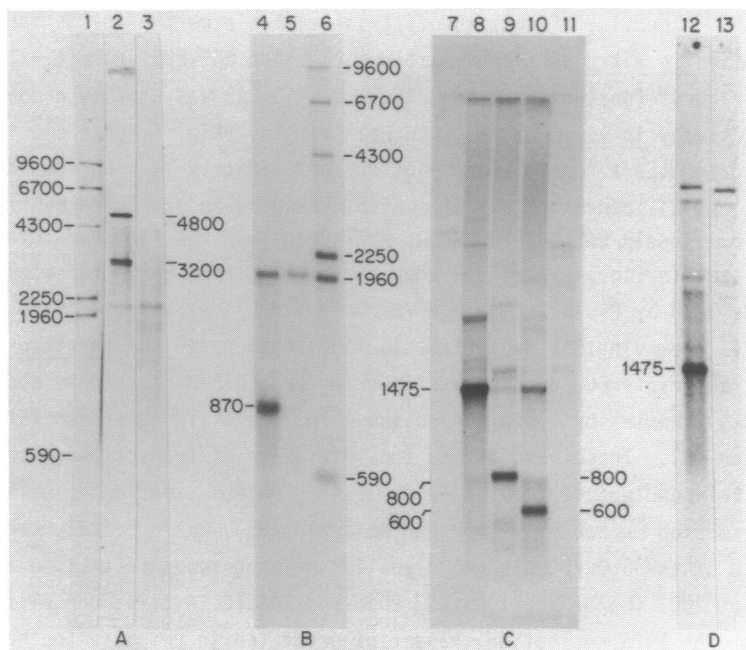
Figure 6. In vitro transcription of cloned MVM RF DNA in Hela cell lysates.
Incubation conditions were exactly as described by Manley et al. (27). RNA
products were treated with glyoxal at 50°C for 1 hour, electrophoresed on
1.4% agarose gels and autoradiographed. Lanes 1 and 6 are ³²P-labeled
lamba DNA cut with Hind III. A. Transcription of cloned MVM RF DNA.
Templates are: lane 2, cloned total RF DNA excised from pBR322 vector; lane
3, linearized pBR322 vector alone. B. Transcription of EcoRI C fragment
(nucleotide 0-1086). Templates are: lane 4, purified EcoRI C fragment;
lane 5, purified EcoRI C fragment plus 2 µg/ml of α-amanitin. C. Trans-
scription of cloned EcoRI A (nucleotide 1086-3521) and EcoRI B (nucleotide
3521-5081) fragments. Templates are: lane 7, linearized pBR322 alone; lane
8, EcoRI A fragment purified from the plasmid vector; lane 9, purified
EcoRI A fragment cut with PVUII (nuc. 2842) truncating 679 bp from the
right hand end; land 10, purified EcoRI A fragment cut with Hind III (nuc.
2652) truncating 869 bp from the right hand end; lane 11, EcoRI B fragment
excised from the plasmid vector. D. Templates are: lane 12, clone EcoRI A
fragment excised from pBR322 vector; lane 13, the same as lane 12 above
plus 2 µg/ml α-amanitin.

observation). To demonstrate the activity of the left end promoter in
vitro we chose to transcribe the EcoRI C fragment, purified from MVM RF
DNA, which terminates at nucleotide 1084 (c.f. Fig. 4). This fragment
primes the synthesis of a discreet runoff transcript approximately 870
nucleotides in length (Fig. 6, lane 4) which is not produced in the
presence of 2 µg/ml α-amanitin, characteristic of authentic polymerase II

transcripts. The size of this runoff transcript places the initiation of
transcription in vitro at approximately nucleotide 205-225 or mu 4. The
EcoRI A fragment (nucleotide 1084-3518; c.f. Fig. 4) was used to demonstr-
ate the activity in vitro of the internal promoter (Fig. 6, part C). This
fragment generates a runoff transcript of approximately 1475 nucleotides
(Fig. 6, lane 8), consistent in size with the nuclease protected hybrid
obtained previously between the EcoRI A fragment and RNA from MVM infected
cells as seen in Fig. 4, part B. When the EcoRI A fragment, truncated at
nucleotide 2840 by PvuII (removing 679 nucleotides), or at nucleotide 2650
by Hind III (removing 869 nucleotides), is used to prime the in vitro
system, transcripts 800 nucleotides (shorter by ∿ 675 nucleotides) and 600
nucleotides (shorter by ∿ 875 nucleotides) in length are generated (Fig. 6,
lanes 9 and 10). These results map the initiation of transcription to
approximately nucleotide 2000, i.e. mu 39-40. Recent experiments using the
primer extension method of Weaver and Weissman (35), has precisely mapped
the 5' end of the 3.0 kb message, revealing heterogeneous initiation at
nucleotides 2003 (major start site), 2006 and 2010(C. Astell, unpublished
results). The 1475 nucleotide transcript generated in response to the
EcoRI A fragment is also sensitive to low levels of α-amanitin as shown in
Fig. 6, part D. In vitro transcription of cloned EcoRI B fragment
(nucleotide 3518-5081, c.f. Fig. 4) produced no specific transcript (Fig.
6, lane 11). Our in vitro system generates numerous minor bands which are
DNA dependent but α-amanitin insensitive. Such bands, presumed not to be
RNA polymerase II products, have been widely observed by others (27,28) and
they were not investigated further.

The identification of the regions at 4 and 39 map units as functional
RNA promoter regions both in vivo and in vitro led us to an examination of
the nucleotide sequences in these areas. At 3.5 mu, nucleotide 175-181,
there is a canonical TATA box (36,37), "TATATAA", on the viral comple-
mentary strand (Fig. 7a). These sequences reside approximately 60 nucleo-
tides downstream from the 3' hairpin of MVM. At 39 mu, nucleotide 1976-
1982, there is also a 7 nucleotide canonical TATA box, "TATAAAT", which is
preceeded 87 nucleotides upstream by a canonical CAT box (37,38),
"CCAATCT", at nucleotides 1889-1895 (Fig. 7b).

DISCUSSION
We have identified and mapped the three major RNA transcripts, which
together constitute greater than 97% of the total viral specific RNA,

A.

```
3' C GATAAGCGGTTCAGGGAGTTTAAACCAAGGCGCGAAAAGGAAGTGGGCGT
   V CTATTCGCCAAGTCCCTCAAATTTGGTTCCGCGCTTTTCCTTCACCCGCA
        |                |                |
       120              140              160

     GGTTTAAAGTATATAAGCAACTACTGAAGTCAGTTACTTATCTTTTCTTT
     CCAAATTTCATATATTCGTTGATGACTTCAGTCAATGAATAGAAAAGAAA
              |                |
             180              200

     CATTCTGTGAGTCGAGACGCACAGAAAGAGAGTAACCAACTAACCATGGC C 5'
     GTAAGACACTCAGCTCTGCGTGTCTTTCTCTCATTGGTTGATTGGTACCG V
         |                |                |
        220              240              260

     TGGAAATGCTTACTCTGATGAAGTTTTGGG
     ACCTTTACGAATGAGACTACTTCAAAACCC
              |                |
             280              295
```

B.

```
3' C TTTGTGCTTGGTTGGTAAAGAATGGTTACCAATCTACCATGGCAAGCTAC
   V AAACACGAACCAACCATTTCTTACCAATGGTTAGATGGTACCGTTCGATG
        |                |                |
      1861             1880             1900

     TGTGCTAAATGGGGCAAAGTTCCTGATTGGTCAGAAAACTGGGCGGAGCC
     ACACGATTTACCCCGTTTCAAGGACTAACCAGTCTTTTGACCCGCCTCGG
              |                |                |
            1920             1940             1960

     AAAGGTGCCAACTCCTATAAATTTACTAGGTTCGGCACGCTCACCATTCA C 5'
     TTTCCACGGTTGAGGATATTTAAATGATCCAAGCCGTGCGAGTGGTAAGT V
              |                |
            1980             2000

     CGACACCGAAAAGTACGCCTCTCAGCCAGA
     GCTGTGGCTTTTCATGCGGAGAGTCGGTCT
              |                |
            2020             2040
```
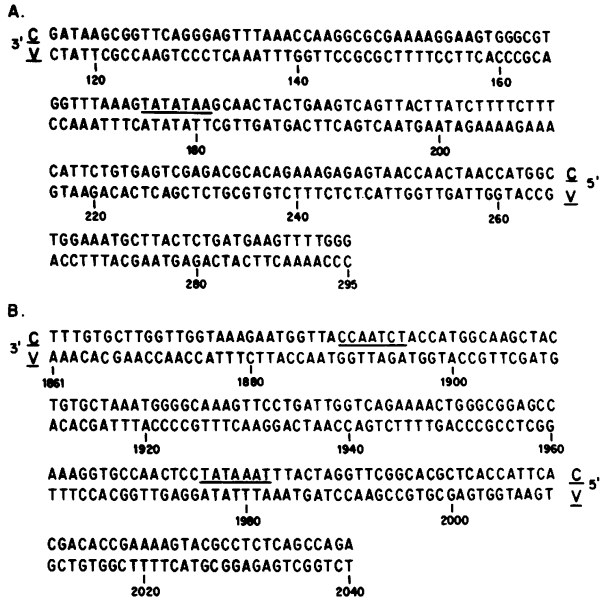
Figure 7. Nucleotide sequence of MVM RF in the vicinity of map unit 4 (7A), and 39 (7B). The complete nucleotide sequence of MVM DNA is described in the accompanying paper (12).

produced by the autonomous parvovirus MVM during productive infection of murine A9 cells. These transcripts, which are 4.8,3.3 and 3.0 kb respectively, are organized into two overlapping transcriptional units as illustrated in Fig. 5. Each of these transcripts is spliced in the region of mu 46 and each contains a similar exon of 2.6 kb between mu 48-96. Minor unspliced species of R1 and R3 have been detected at very low levels in total RNA preparations. These results suggest a more complex pattern of autonomous parvovirus transcription than has been previously described (10,11) and resembles in many ways the transcriptional topography described for the defective parvoviruses. The transcription organization of the defective parvovirus AAV-2 is that of three overlapping transcriptional units, encoding six RNA species, with each promoter (located at mu 5,19 and 39) generating both a spliced and unspliced species of RNA (9). All the spliced species of AAV-2 RNA, although initiated at three different sites, share a similar exon which comprises the right half of the genome. The spliced RNAs also share a similarly sized and placed splice at mu 40-48. The size of this intron (∿ 350 nucleotides) is, however, considerably larger than

for the analogous common intron at mu 46 that is spliced out of the MVM transcripts. In addition, while AAV-2 encodes two significant transcripts initiated at mu 19, we found no evidence for a third such promoter in MVM. No in vivo RNAs that mapped to that initiation site were detected, and in vitro transcription experiments with cloned DNA representing 6-25 mu of MVM RF DNA failed to demonstrate efficient in vitro promotion from that region of the genome (data not shown). The emerging picture of MVM transcription, while resembling that of the defective parvoviruses, is still distinctly different.

There are a number of still unanswered questions raised by the data presented here. Foremost of these is the map position(s) of the low abundance 1.8 kb transcript R4. In addition, the results of Fig. 4A, lane 3 indicates that the small leader sequence between mu 40-46 is hetero- geneous. Although this may be an artifact of the procedures used, it may reflect, as in SV40 (39,40) and polyoma (41) transcription, subtle differences in splice junctions either among individual members of the same transcript class or differences between the different class species. Addi- tional discussion of the splice junctions in this area based on nucleotide analysis of MVM DNA is presented in the accompanying paper (12).

Previous work on the autonomous parvovirus H1 (10) had revealed 5 viral RNA species, 4.7, 3.0, 2.8, 1.45 and 1.3 kb in length. The 4.7 kb transcript, although spliced, was found only in the nucleus, leading to the speculation that it was a nuclear precursor to the smaller transcripts found in the cytoplasm. We have detected the MVM 4.8 kb transcript both in the nuclear and cytoplasmic fractions. In addition, in vitro translation of R1, which had been selected by hybridization to cloned MVM DNA from nucleotide 1086-1661 (totally within the large intron), has recently been shown to generate at least one previously unidentified non structural protein (S. Cotmore and P. Tattersall, personal communication). We conclude from this that the 4.8 kb message appears in a functional form in the cytoplasm of infected cells.

We have presented here a more comprehensive map of MVM transcription products than previously available, however, precise definition of the map coordinates of all RNA species require further analysis. Such studies are currently in progress.

REFERENCES
1. Berns, K.I. and Hauswirth, W.W. (1978) in Replication of Mammalian
   Parvoviruses, D.C. Ward and P. Tattersall, eds. Cold Spring
   Laboratory, Cold Spring Harbor, N.Y., p.13-32.
2. Carter, B.J. (1978) in Replication of Mammalian Parvoviruses, D.C.
   Ward and P. Tattersall, eds. Cold Spring Harbor Laboratory, Cold
   Spring Harbor, N.Y., p. 33-52.
3. Laughlin, C.A., Westphal, H. and Carter, B.J. (1979) Proc. Natl. Acad.
   Sci. USA 76, 5567-5571.
4. Jay, F.T., de la Maza, L.M. and Carter, B.J. (1979) Proc. Natl. Acad.
   Sci. USA 76, 625-629.
5. Green, M.R. and Roeder, R.G. (1980) J. Virol. 36, 79-92.
6. Green, M.R. and Roeder, R.G. (1980) Cell 22, 231-242.
7. Berns, K.I. and Hauswirth, W.W. (1979) Adv. Virus Res. 25, 407-449.
8. Buller, R.M.L., Jamik, J.E., Sebring, E.D. and Rose, J.A. (1981) J.
   Virol. 40, 241-247.
9. Marcus, C.J., Laughlin, C.A. and Carter, B.J. (1981) Eur. J. Biochem.
   121, 147-154.
10. Green, M.R., Lebovitz, R.M. and Roeder, R.G. (1979) Cell 17, 967-977.
11. Tal, J., Ron, D., Tattersall, P., Bratosin, S. and Aloni, Y. (1979)
    Nature 279, 649-651.
12. Astell, C.R., Thomson, M., Merchlinsky, M. and Ward, D.C. (198),
    submitted to Nucl. Acids Res.
13. Astell, C.R., Dadachanji, D., Merchlinsky, M., Pintel, D. and Ward,
    D.C. (1981) Abstracts Vth Intl. Congress Virology, p. 365.
14. Littlefield, J.W. (1964) Nature 203, 1142-1144.
15. Tattersall, P., Cawte, P.J. Shatkin, A.J. and Ward, D.C. (1976) J.
    Virol. 20, 273-289.
16. Bourguignon, G.J., Tattersall, P.J. and Ward, D.C. (1976) J. Virol.
    20, 290-306.
17. Cox, R.A. (1968) in Methods in Enzymology, Vol. 12B, L. Grossman and
    M. Moldave, eds. Academic Press, New York, N.Y. p. 120-129.
18. Strohman, R.C., Moss, P.S., Micou-Eastwood, J., Spector, D., Przybyla,
    A. and Paterson, B. (1977) Cell 10, 265-273.
19. Borun, R.W., Sharff, M.D. and Robbins, E. (1967) Biochim. Biophys.
    Acta 149, 302-304.
20. Carter, B.J. and Rose, J.A. (1974) Virol. 61, 182-199.
21. Carter, B.J., Fife, K.H., de la Maza, L.M. and Berns, K.I. (1976) J.
    Virol. 19, 1044-1053.
22. Casey, J. and Davidson, N. (1977) Nucl. Acids Res. 4, 1539-1552.
23. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
24. Tibbetts, C., Johansson, K. and Philipson, L. (1973) J. Virol. 12,
    218-225.
25. Rave, N., Crkvenjakov, R. and Boedtker, H. (1979) Nucl. Acids Res. 6,
    3559-3567.
26. Thomas, P. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.

27. Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gefter, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 3855-3859.
28. Fire, A., Baker, C.C., Manley, J.L., Ziff, E.B. and Sharp, P.A. (1981) J. Virol. 40, 703-719.
29. Faust, E.A. and Ward, D.C. (1979) J. Virol. 32, 276-292.
30. Aloni, Y. and Attardi, G. (1971) Proc. Natl. Acad. Sci. USA 68, 1757-1761.
31. Aloni, Y. and Locker, H. (1973) Virol. 54, 495-505.
32. Rigby, P.W.J., Rhodes, C., Dieckman, M. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
33. Chow, M.B. and Ward, D.C. (1978) in Replication of Mammalian Parvoviruses, D.C. Ward and P. Tattersall, eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., p. 205-217.
34. Berk, A.J. and Sharp, P.A. (1978) Proc. Natl. Acad. Sci. USA 75, 1274-1278.
35. Weaver, R.F. and Weissmann, C. (1979) Nucl. Acids Res. 7, 1175-1193.
36. Goldberg, M.L. (1979) Ph.D. Thesis, Stanford University.
37. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
38. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C. Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C., Proudfoot, N.J. (1980) Cell 21, 653-668.
39. Reddy, V.B., Ghosh, P.K., Lebowitz, P., Weissman, S.M. (1978) Nucl. Acids Res. 5, 4195-4214.
40. Haegman, G. and Fiers, W. (1980) J. Virol. 35, 955-961.
41. Flavell, A.J., Legon, S., Cowie, A., Kamen, R. (1979) Cell 16, 357-371.