JOURNAL OF VIROLOGY, Oct. 1979, p. 293-303
0022-538X/79/10-0293/11$02.00/0

Vol. 32, No. 1

Electron Microscopic Mapping of RNA Transcribed from the Late Region of Polyoma Virus DNA†

HAIM MANOR,[1]‡ MADELINE WU,[1] NAVA BARAN,[2] AND NORMAN DAVIDSON[1]*

Department of Chemistry, California Institute of Technology, Pasadena, California 91125,[1] and Department of Biology, Technion-Israel Institute of Technology, Haifa, Israel[2]

Received for publication 8 January 1979

The polyoma virus (Py) RNA species transcribed from the L DNA strand of the "late" region of the Py genome in Py-infected mouse cells have been mapped by hybridization with specific fragments of Py DNA followed by electron microscopic visualization of the hybrids. Total cellular polyadenylated Py-specific RNA molecules having an S value in the range of 16S to 20S were purified by oligodeoxythymidylic acid-cellulose column chromatography, preparative hybridization with Py DNA, and sucrose gradient centrifugation. Cytoplasmic Py-specific RNA was similarily purified, except that it was not fractionated by sucrose gradient centrifugation. Hybrids of these RNA molecules and Py DNA fragments were spread for electron microscopy by either the cytochrome *c* technique or the bacteriophage T4 gene 32 protein method. The polyadenylic acid at the 3'-end of the RNA in the hybrids was identified by labeling with simian virus 40 DNA circles to which polybromodeoxyuridylic acid tails had been covalently attached. These experiments revealed the presence of three L DNA strand transcripts in both RNA preparations. Two of these RNA molecules were found to be spliced from chains transcribed from two noncontiguous parts of the late region. The third molecule either is a continuous transcript of the entire late region or contains a splicing feature which is too small to be reliably observed by the electron microscope methods used. The 5'-ends of the three RNA species map within a region extending from 68 to 70 map units on the Py restriction endonuclease map. Each of the two spliced molecules contains a 5'-terminal leader sequence transcribed from a DNA segment with an estimated length of 60 to 110 nucleotides. The 3'-ends of the leaders map at 66.7 ± 1.0 and 66.4 ± 0.50 map units. In these molecules the 5'-ends of the other part (the main body) map at 59.4 ± 0.90 and 49.4 ± 2.0 map units, respectively. The 3'-termini of all three RNA species map at 24 to 25 map units.

At the late phase of polyoma virus (Py) infection in mouse cells viral RNA is predominantly transcribed from one of the strands of the Py DNA designated as the L DNA strand (12, 17). The major products of viral transcription at this stage are giant RNA molecules longer than the Py genome (1, 7, 23, 28). These giant molecules are confined to the nuclei. The L DNA strand transcripts found in the cytoplasm of Py-infected cells are smaller than the Py genome. The most abundant cytoplasmic L DNA strand transcript has a sedimentation coefficient of 16S and serves as the mRNA for the major Py capsid protein VP1 (30). A second Py L DNA strand transcript found in the cytoplasm has a sedimentation coefficient of 19S and codes for the capsid protein VP2 (30). Recently, the existence of a third cytoplasmic L DNA strand transcript which codes for the third Py capsid protein VP3 has been postulated on the basis of in vitro translation experiments. The predicted sedimentation coefficient of this RNA species is 18S (29).

The giant nuclear RNA molecules and the cytoplasmic 16S and 19S mRNA species have been mapped on the Py restriction endonuclease map by annealing the RNA with radioactively labeled separated strands of various Py DNA fragments and by measuring the fraction of DNA which formed RNA-DNA hybrids. Hybrid formation was assayed in these experiments by digesting the DNA with the single-strand-specific endonuclease S1 and by measuring the radioactivity which remained acid insoluble. The giant RNA was found by this technique to consist of sequences complementary to the entire

† Contribution no. 5925 from the California Institute of Technology.

‡ Permanent address: Department of Biology and the Joseph and Arlene Taub Institute of Biological Research, Technion-Israel Institute of Technology, Haifa, Israel.

Py L DNA strand (17, 22). The 19S mRNA was shown to contain sequences complementary to the L DNA strand of the region extending between 68 and 25 map units (m.u.) on the Py physical map, defined as the "late" region (17, 18). The 16S mRNA was found to be transcribed from the L DNA strand of a region extending between 48 and 25 m.u. (18).

Recently, a number of viral mRNA's were found to be spliced from sequences transcribed from noncontiguous regions of DNA (2, 5, 6, 9, 11, 16, 19, 20). In view of this discovery, it was interesting to re-examine the mapping of the Py RNA species by more sensitive techniques, with the aim of identifying splicing patterns not detectable by the previous method in these RNA molecules. In the present report we describe an attempt to map the Py RNA species transcribed from the L DNA strand of the late region by electron microscopy. Our approach was to anneal the RNA with specific Py DNA fragments and to determine by electron microscopy the map coordinates of the DNA regions which formed stable hybrids, as well as the coordinates of the regions which remained single stranded. This method is suitable for mapping RNA molecules transcribed from noncontiguous regions of DNA, because the DNA sequences found between the template regions are expected to form clearly visible single-stranded loops flanked by hybrid regions (5).

MATERIALS AND METHODS

Cells and virus growth. Mouse kidney cells were grown and infected with Py, as described by Lev and Manor (23). The two strains of virus used for these experiments were the Py A2 strain supplied by M. Fried and a large-plaque strain obtained from E. Winocour. Virus stocks were prepared as previously described (26).

Preparation of Py-specific RNA. (i) Purification of RNA from whole cells. Py-infected cells were harvested 28 h postinfection. The method developed by Parish and Kirby (27) was used for RNA extraction, as described by Lev and Manor (23). The RNA was chromatographed on oligodeoxythymidylic acid-cellulose columns to select for polyadenylated molecules (3, 23), and the Py-specific fraction among these molecules was isolated by preparative hybridization with 15 μg of Py DNA fixed on a nitrocellulose filter. The hybridization was carried out as follows. The filter containing the DNA and a blank filter were preincubated for 2 h at 37°C in 0.70 ml of a solution consisting of 50% formamide, 0.05 M PIPES [piperazine-*N*,*N*'-bis(2-ethanesulfonic acid)] buffer (pH 7.0), 0.75 M NaCl, 10 mM EDTA, 120 μg of polyadenylic acid [poly(A)] per ml, 85 μg of *Escherichia coli* tRNA per ml, and 0.10% of sodium dodecyl sulfate. The filters were removed and inserted into 1 ml of the same solution, which also contained 40 to 50 μg of ^{3}H-labeled polyadenylated RNA molecules (specific radioactivity, 4,000 cpm/μg) purified from Py-infected cells as described above. At the end of the incubation period the filters were removed from the hybridization buffer and incubated for 2 h at 37°C in 5 ml of the above solution from which the [^{3}H]RNA, poly(A), and tRNA were omitted. The filters were then soaked successively in three beakers, each containing 10 ml of 0.50 M NaCl, 1 mM EDTA, 0.10% sodium dodecyl sulfate, and 10 mM PIPES (pH 7.0). To elute the RNA each filter was washed with 0.20 ml of a solution containing 90% formamide, 1 mM EDTA, and 2 mM sodium acetate (pH 5.5) and then incubated for 1 h at 45°C in 1 ml of the same solution. The filter was further incubated for 15 min at 45°C in 0.30 ml of this solution. The three eluants were combined, and the RNA was precipitated with ethanol. It was then dissolved in 0.20 ml of 10 mM PIPES buffer (pH 7.0) and centrifuged in a 5 to 20% sucrose gradient also containing 0.10 M NaCl, 10 mM PIPES (pH 7.0), and 1 mM EDTA. Centrifugation in the Spinco SW50.1 rotor was performed at 40,000 rpm and 22°C for 2 h. The fractions sedimenting between 16S and 20S were pooled and used for the electron microscopic studies.

(ii) Purification of cytoplasmic RNA. The cytoplasm was separated from the nuclei by the method of Lindberg and Darnell (24), and the RNA was extracted from the purified cytoplasm, as previously described (25). Purification of polyadenylated Py-specific RNA molecules was carried out by oligodeoxythymidylic acid-cellulose column chromatography, followed by preparative hybridization with Py DNA fixed on nitrocellulose filters, as described in (i) above. However, no further size selection was performed in this case, and the RNA eluted from the filters was used for the electron microscopic studies.

Preparation of restriction fragments of Py DNA. Supercoiled Py DNA was prepared as described by Manor and Neer (26). Digestion with the enzyme *Eco*RI was carried out for 30 min at 37°C in a solution containing 10 mM Tris (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$, and 1 mM dithiothreitol. The concentrations of DNA and enzyme were 50 to 70 μg/ml and 100 to 140 U/ml, respectively. Digestion with the enzyme *Hha*I was carried out for 60 min at 37°C in a solution containing 10 mM Tris (pH 7.9), 50 mM NaCl, 7 mM $MgCl_2$, and 6.5 mM β-mercaptoethanol. The concentrations of DNA and enzyme were 50 to 70 μg/ml and 50 to 70 U/ml, respectively. The enzymes were obtained from R. Kamen. DNA digests were electrophoresed for 18 h at 1.2 to 1.5 V/cm on 1.4% agarose (Sigma) gels. Bands containing DNA fragments were cut from the gels, and the DNA was extracted from the agarose as described by Kamen and Shure (18). Alternatively, the DNA was eluted from the agarose by electrophoresis into dialysis tubes.

RNA-DNA hybridization. RNA at a concentration of 0.2 to 0.6 μg/ml was annealed for 2 h at 50°C with 2 to 6 μg of denatured DNA per ml in 10 μl of a solution which also contained 80% recrystallized formamide, 0.40 M NaCl, 1 mM EDTA, and 10 mM PIPES buffer at pH 7.1. These conditions were found by Casey and Davidson (8) to allow RNA-DNA hybridization with no concomitant DNA self-annealing.

Electron microscopic techniques. (i) Spreading by the cytochrome *c*-formamide technique.

The RNA-DNA hybrids were spread by the cytochrome *c*-formamide method (10), using a hyperphase containing 58% recrystallized formamide, 40 mM NaCl, 0.10 mM EDTA, 1 mM PIPES, 60 mM TES buffer [*N*-tris(hydroxymethyl)methyl-2-aminomethane-sulfonic acid] at pH 7.0, and 30 μg of cytochrome *c* per ml. The hypophase contained 17% formamide, 1 mM EDTA, and 10 mM TES buffer at pH 7.0. Alternatively, the hybrids were first incubated at 25°C with oligobromodeoxyuridylic acid (dBrU)-simian virus 40 (SV40) DNA circles (4) in a solution containing 20% formamide, 0.10 M NaCl, 0.17 M TES (pH 7.0), and 0.25 mM EDTA and then spread, using the conditions described above. Grids were rotary-shadowed with platinum palladium and examined with a Philips 300 electron microscope. Hybrid molecules were photographed, and the lengths of various parts were measured using a Hewlett-Packard digitizer. The SV40 circles were used as an internal-length standard in most spreadings. In the spreadings with no SV40 circles, the lengths were determined using as a standard the known length of the Py *Hha*I DNA fragment 1. Under the conditions used for these preparations, the length of RNA-DNA hybrids regions was within 5% of the length of the corresponding single-stranded DNA regions.

(ii) Spreading by the phage T4 gene 32 protein technique. The conditions for gene 32 protein spreading were as previously described (31), except that pure carbon grids were used instead of carbon-coated Parlodion grids. The stock solution of ethidium bromide was prepared weekly.

RESULTS

Preparation of RNA-DNA hybrids. Polyadenylated Py-specific RNA molecules having a sedimentation coefficient in the range of 16S to 20S were isolated from whole Py-infected cells 28 h postinfection. Alternatively, the cytoplasm of these cells was separated from the nuclei and Py-specific RNA was isolated from the purified cytoplasm. Each of these RNA preparations was annealed with fully denatured linear Py DNA produced by cleavage of the circular genome with the enzyme *Eco*RI or with *Hha*I DNA fragment 1 isolated after cleavage of the Py genome with the enzyme *Hha*I (13). This fragment includes almost the entire late region (Fig. 1). The hybridization reaction was carried out under conditions previously found to minimize DNA renaturation (8).

Analysis of RNA-DNA hybrids prepared for microscopy by the cytochrome *c*-formamide technique. In these experiments Py-specific RNA prepared by either of the two methods described above was annealed with Py *Hha*I DNA fragment 1. In some cases the hybrids were further incubated with circular SV40 DNA molecules to which poly(dBrU) was enzymatically attached. This incubation was carried out under conditions which favored hybridization of the poly(A) at the 3'-end of the RNA with the poly(dBrU) attached to the SV40 circles (4).

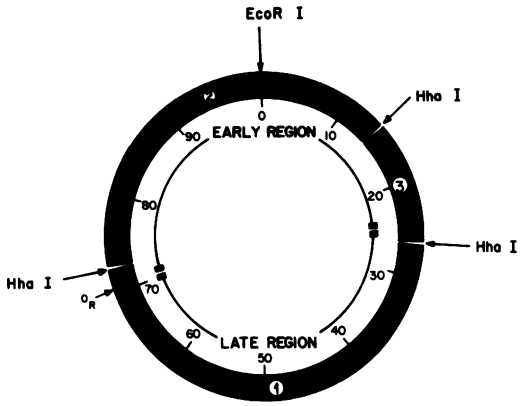

FIG. 1. *Physical map of the Py genome and the cleavage sites of the enzymes EcoRI and HhaI. The circular map of the Py genome is divided into 100 units, point 0 on the map being the single site of cleavage by the restriction endonuclease EcoRI (14). The early and late regions were defined by Kamen et al. (17). The three fragments produced by cleaving the Py DNA with the restriction endonuclease HhaI (13) are marked 1, 2, and 3.*

Figure 2 presents examples of the hybrid molecules found in the cytochrome *c* spreads. These molecules must be hybrids between the L DNA strand and the L DNA strand transcripts, because even though the presence of some molecules transcribed from the E DNA strand of the late region in the 16S-20S Py RNA could not be entirely excluded, their abundance must have been too low to be scored in these spreads (17). In the cytoplasmic RNA no such E DNA strand transcripts were previously found (17, 18).

The hybrids observed most frequently in the cytochrome *c* spreads are shown in Fig. 2a and b. These molecules contain a single-stranded DNA loop whose length constitutes about 35% of the total length. The loop is asymmetrically located relative to the two ends and is flanked by hybrid regions. We shall refer to the distance between the loop and the closer end of these molecules as the short arm and to the distance between the loop and the other end as the long arm. The presence of the loop indicates that the RNA chain in each of these hybrid molecules is spliced from two parts which are transcribed from the regions of DNA flanking the loop. Figure 2c and d shows examples of a second type of hybrid molecules observed in the cytochrome *c* spreads. These molecules contain a smaller single-stranded DNA loop flanked by hybrid regions. Therefore, the RNA in these hybrids is also spliced from two chains transcribed from the DNA flanking the loop.

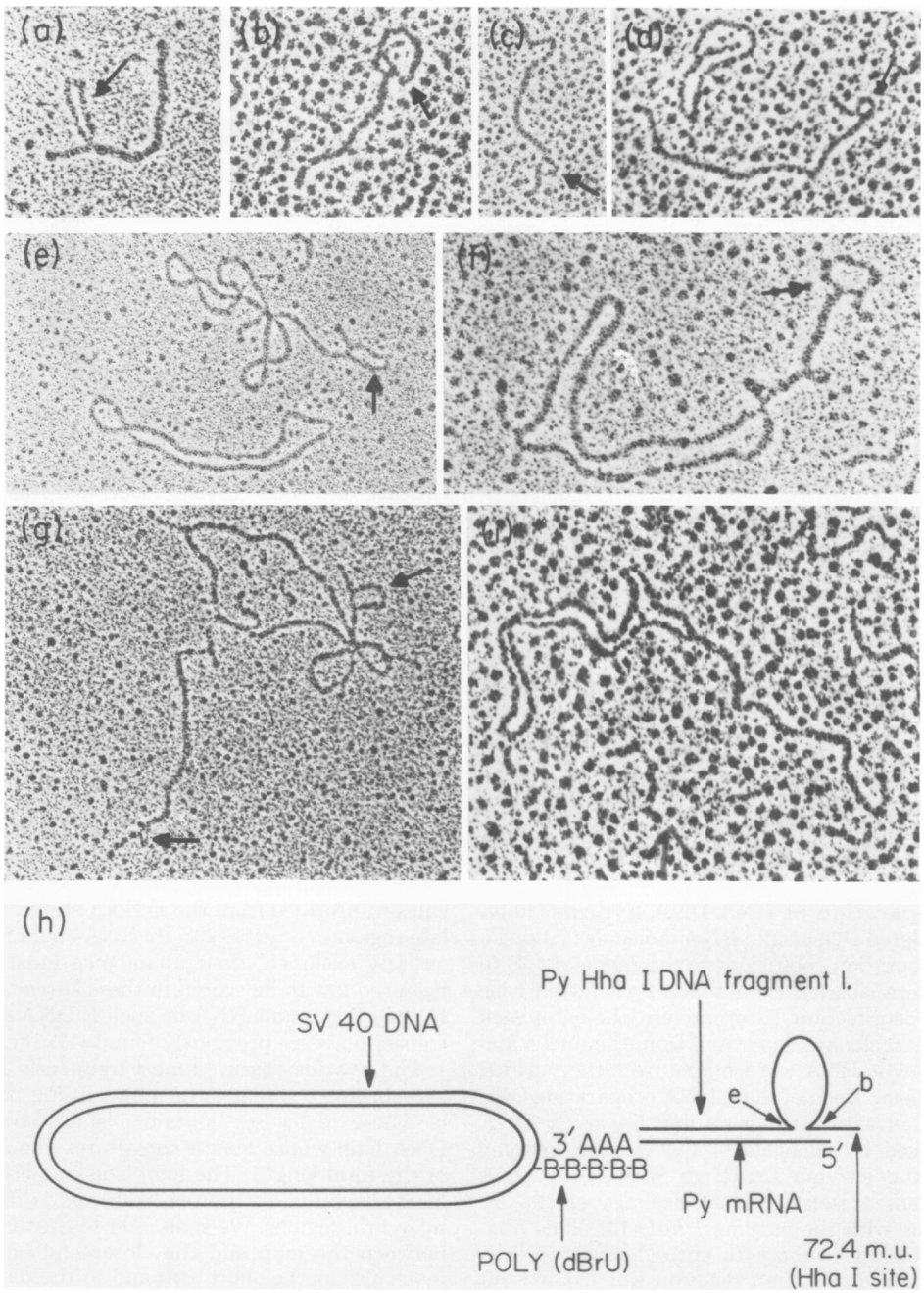

FIG. 2. *Electron micrographs of Py RNA-DNA hybrids spread by the cytochrome c-formamide technique. Polyadenylated Py RNA having an S value in the range of 16S to 20S was annealed with HhaI DNA fragment 1, as described in the text. The RNA was extracted from whole Py-infected cells (a, c, e, f, g) or from a purified cytoplasm (b, d, i). The hybrids were either directly spread for microscopy or annealed with poly(dBrU)-SV40 DNA circles before being spread for microscopy. (a, b) RNA-DNA hybrids having a large single-stranded loop. (c, d) RNA-DNA hybrids having a small single-stranded DNA loop. (e, f) RNA-DNA hybrids having a large single-stranded DNA loop and with the poly(A) tails of the RNA molecules labeled with poly(dBrU)-SV40 DNA circles. (g) A hybrid having a small single-stranded DNA loop and a hybrid having a large single-stranded DNA loop bound to the same poly(dBrU)-SV40 circle. (h) A schematic drawing of a hybrid having a large single-stranded DNA loop bound to a poly(dBrU)-SV40 DNA circle. Points b and e mark the beginning and the end of the spliced region. (i) RNA-DNA hybrid having no visible loop bound to a poly(dBrU)-SV40 DNA circle. Each of the single-stranded DNA loops is marked by an arrow.*

Figure 2e, f, and g presents several hybrids of the two types bound to the SV40 circles through the end of the long arm, which is therefore the site of the polyadenylated 3'-end of the RNA. Clearly, the 5'-end of the RNA must be on the short arm. It is to be noted that the boundary between an RNA-DNA hybrid region and a single-stranded DNA region in the same molecule cannot always be easily identified in cytochrome *c* spreads. Thus, it is difficult to map the position of the 5'-end of the RNA on the short arm. However, it is shown below that the 5'-terminal portion of the RNA, which will be referred to as the leader sequence, is shorter than the short arm; therefore, the end of the short arm must be one of the two ends of *Hha*I DNA fragment 1. On the basis of the previously determined polarity of the L DNA strand transcripts relative to the Py physical map (17), we conclude that this end must be the cleavage site of the enzyme *Hha*I at 72.4 m.u. A schematic representation of a hybrid molecule bound to an SV40 DNA circle is drawn in Fig. 2h on the basis of these inferences.

An interesting feature of the hybrid molecules shown in Fig. 2c, f, and g is that not only are there loops of single-stranded DNA in these molecules, but also there appear to be short unhybridized RNA sequences across the loops. The significance of this finding will be discussed below.

A third type of hybrid molecule bound to a poly(dBrU)-SV40 DNA circle is shown in Fig. 2i. This molecule does not contain a visible loop, and therefore the RNA in this case may be transcribed from a continuous DNA template (see below and Discussion).

Figure 3a shows the size distribution of the single-stranded DNA loops in the hybrid molecules prepared by annealing *Hha*I DNA fragment 1 with the 16S–20S Py RNA. Figure 3c presents similar data for the hybrid molecules prepared by annealing this DNA fragment with cytoplasmic RNA. Each histogram clearly re-

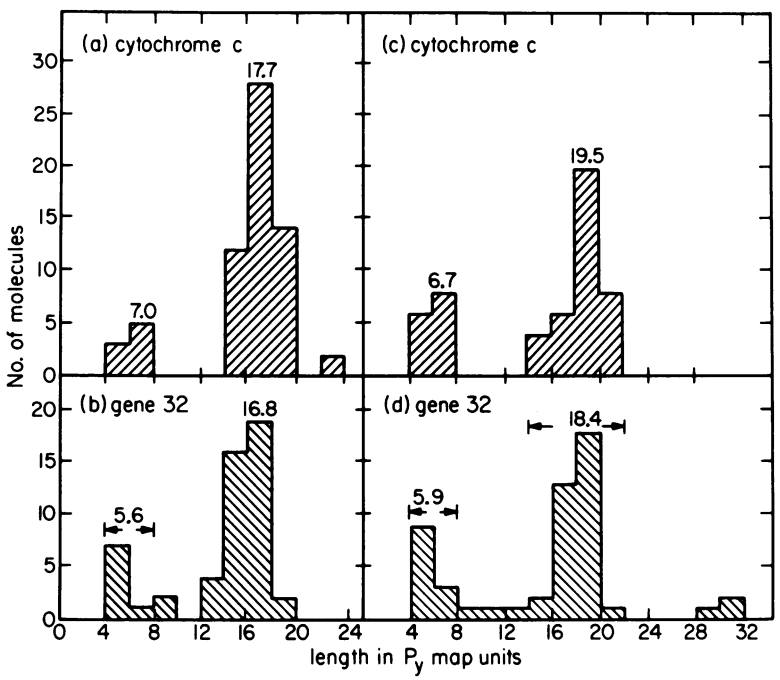

FIG. 3. *Size distribution of single-stranded DNA loops in RNA-DNA hybrids. (a) Hybrids between polyadenylated 16S-20S Py RNA prepared from whole Py-infected cells and HhaI DNA fragment 1 (cytochrome c-formamide spreads). The lengths of the large loop and the small loop were found to be 17.7 ± 1.7 and 7.0 ± 0.9 m.u., respectively. (b) Hybrids between polyadenylated 16S-20S Py RNA prepared from whole Py-infected cells and EcoRI-cleaved linear Py DNA (gene 32 protein method). The lengths of the large loop and the small loop were found to be 16.8 ± 1.2 and 5.6 ± 0.7 m.u., respectively. (c) Hybrids between cytoplasmic Py RNA and HhaI DNA fragment 1 (cytochrome c-formamide spreads). The lengths of the large loop and the small loop were found to be 19.5 ± 1.8 and 6.7 ± 1.1 m.u., respectively. (d) Hybrids between cytoplasmic Py RNA and EcoRI-cleaved linear Py DNA (gene 32 protein method). The lengths of the large loop and the small loop were found to be 18.4 ± 1.7 and 5.9 ± 0.8 m.u., respectively.*

veals the presence of the two distinct hybrids differing in the size of the loop, of which examples are presented in Fig. 2. The sizes of the larger and the smaller loops are, within experimental error, the same in the two preparations of hybrids. However, the ratio of the large loops to small loops is significantly greater in Fig. 3a than in Fig. 3c.

Figure 4 shows the map coordinates of the two sides of the loop in the hybrids prepared by annealing *Hha*I DNA fragment 1 with the 16S–20S Py RNA. These coordinates were determined relative to the end of the short arm at 72.4 m.u. The length of the short arm gives the coordinate of the 3′-end of the leader sequence (point b in Fig. 2h). The length of the short arm plus the length of the loop gives the coordinate (point e in Fig. 2h) of the 5′-end of the second part of the RNA chain to which we shall refer as its main body. The data are plotted separately for the molecules having a large loop and for those having a smaller loop. These histograms show that the position of the 3′-end of the leader sequence is the same for the two spliced RNA species. Similar data were obtained in the experiments conducted with cytoplasmic RNA (not shown).

It was more difficult, however, to determine the coordinates of the 3′-termini of the RNAs in these hybrids because their main bodies presumably extend beyond the end of *Hha*I DNA fragment 1 at 26.4 m.u. (see reference 18). Thus, in the hybrids which were not bound to the SV40 circles, the unhybridized 3′-terminal portions of the RNAs could collapse under the conditions used for spreading, and hence the length of the long arm of these molecules (plotted in Fig. 5a)

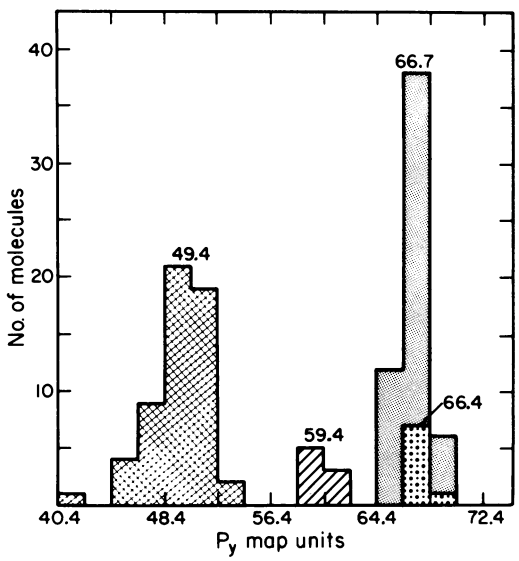

FIG. 4. *Coordinates of the 3′-ends of the leader sequences and the 5′-ends of the main bodies in the spliced L DNA strand transcripts. These coordinates were determined relative to the HhaI cleavage site at 72.4 m.u. in the same population of hybrids whose loop size distribution is shown in Fig. 3a. These measurements are described in the text. The positions of the 3′-ends of the leaders (point b in Fig. 2h) were calculated as 66.7 ± 1.0 m.u. (n = 56) and 66.4 ± 0.50 m.u. (n = 8), respectively, in the molecules with the large loop and the small loop. The 5′-termini (point e in Fig. 2h) of the main bodies of these molecules were mapped at 49.4 ± 2.0 and 59.4 ± 0.90 m.u. The errors represent the standard deviations. Symbols: (dotted bar) hybrids with a small loop, 3′-end of leader; (striped bar) hybrids with a small loop, 5′-end of main body; (stippled bar) hybrids with a large loop, 3′-end of leader; (cross-hatched bar) hybrids with a large loop, 5′-end of main body.*

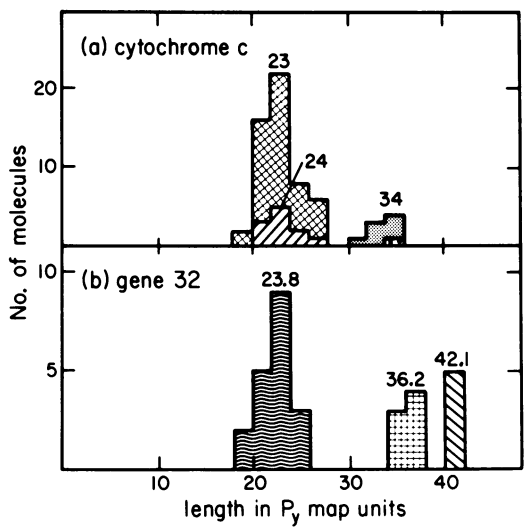

FIG. 5. *Lengths of the hybrid regions in the RNA-DNA hybrid molecules. (a) Size distribution of the long arm was determined in the hybrid molecules whose loop size distribution is shown in Fig. 3a (see also Fig. 2h). The average length of the long arm for the hybrids not attached to poly(dBrU)-SV40 circles, and with large (cross-hatched bar) and small (stippled bar) loops, respectively, was 23.5 ± 2.0 m.u. (n = 43) and 34.0 ± 1.4 m.u. (n = 8). The average length for the hybrids with a large loop attached to poly(dBrU) SV40 circles (striped bar) was 23.7 ± 2.1 m.u. (n = 11). The length for one hybrid with a small loop attached to a poly(dBrU)-SV40 circle was 34.5 m.u. (dotted bar). (b) Size distribution of hybrid regions observed by the gene 32 protein method extending from the poly(A) tail to the large (wavy bar) and small (checked bar) loops and of the total lengths of long hybrids with no loops (striped bar), as discussed in the text. The average lengths of the three types of hybrids are 23.8 ± 2.1 m.u. (n = 18), 36.2 ± 2.7 m.u. (n = 7), and 42.1 ± 3.8 m.u. (n = 5), respectively.*

may be an underestimation of the length of the RNAs. In the hybrids which were bound to the SV40 circles, on the other hand, the length of the poly(dBrU) chains is not known. Therefore, determinations of the distance between the circle and the loop (also plotted in Fig. 5a) may lead to an overestimation of the lengths of the RNAs. However, as Fig. 5a shows, the values of these two parameters are, within experimental error, the same. Thus, the 3'-termini of the two spliced RNA species map within 1 to 2 m.u. of the end of the *Hha*I DNA fragment 1 at 26.4 m.u.

Analysis of RNA-DNA hybrids prepared for microscopy by the phage T4 gene 32 protein spreading technique. The gene 32 protein spreading technique is based on the fact that this protein has a much higher affinity for single-stranded DNA than for double-stranded DNA, or for RNA-DNA hybrids. Therefore, in a molecule consisting of both single-stranded and hybrid regions, single-stranded DNA segments appear in the electron microscope as thick lines, whereas the hybrid regions are considerably thinner and more difficult to recognize at lower magnifications (31). Preliminary experiments showed that hybrids between the *Hha*I DNA fragment 1 and Py-specific RNA cannot be efficiently scored in gene 32 protein spreads due to the absence of sufficiently long single-stranded DNA regions flanking the DNA-RNA hybrid segment of interest. Therefore, we used for these experiments the whole linear Py DNA produced by cleavage of the circular genome with the enzyme *Eco*RI.

Figure 6 shows several examples of the hybrid molecules found in the gene 32 protein spreads. Three types of molecules were observed corresponding to those found in the cytochrome *c* spreads: (i) a molecule with a large single-stranded DNA loop (Fig. 6a and b), (ii) a molecule with a smaller single-stranded DNA loop (Fig. 6c), and (iii) a molecule with no visible loop (Fig. 6d). It can be seen that the boundaries between the single-stranded DNA and the hybrid regions are more clearly defined in the gene 32 protein spreads than in the cytochrome *c* spreads. In many of the molecules observed by the gene 32 protein method, the end of the RNA-DNA hybrid region corresponding to the 3'-end of the RNA appears as a fork with a long arm of single-stranded DNA and a short arm of approximate length 100 to 300 nucleotides (Fig. 6a, b, and d). We believe that this short arm is the poly(A) segment of the RNA. In the molecules containing loops, the leader sequences are also clearly visible (L in the tracings of Fig. 6a, b, and c).

Figure 3b and d shows the size distribution of the loops in the hybrid molecules observed by the gene 32 protein method. The sizes of the loops in the hybrids between the *Eco*RI linear DNA and the 16S–20S RNA (Fig. 3b), or the cytoplasmic RNA (Fig. 3d), were found to be the same, within experimental error, as those found in the corresponding molecules scored in the cytochrome *c* spreads. Three molecules with larger loops were observed among the hybrids with cytoplasmic RNA (Fig. 3d).

Figure 5b shows histograms of the lengths of the main bodies of the duplex regions, i.e., the segments extending from the poly(A) tail to the loop, in the hybrid molecules containing a loop observed by the gene 32 protein method. Measurements of the length of the main body of the hybrid region for the two types of molecules with spliced RNA chains, as estimated in cytochrome *c* spreads, have been discussed above and are presented in Fig. 5a; they are in agreement with the measurements presented in Fig. 5b.

In the gene 32 protein spreads, several hybrid molecules with no single-strand loop larger than 100 nucleotides, with a duplex region longer than that corresponding to the RNA-DNA duplex of the molecules with a small splicing loop, and with a poly(A) fork at the same position as for the molecules with loops, were observed. The observed lengths of these duplex regions are also plotted in Fig. 5b. The discrete distribution suggests the presence of a third kind of RNA which is either a continuous unspliced transcript or contains a splicing loop too small to be reliably observed. From previous experience with the gene 32 protein method (32), we estimate that any DNA loop 100 nucleotides or greater in length would be recognized. A loop or frayed end in the RNA would be more difficult to recognize because single-stranded RNA is usually somewhat collapsed in gene 32 protein spreads. In three of these five molecules (including the one in Fig. 6d), there was a small bump about 70 nucleotides from the 5'-end. In view of the limited data, these observations are suggestive of a sequence nonhomology feature in the hybrid, but are not decisive.

Figure 7 shows histograms of the distance between the poly(A) tails in the three types of hybrid molecules and the *Eco*RI cleavage site. Since the *Eco*RI site is defined as 0 on the Py physical map, the numbers shown on the abcissa are actually the coordinates of the 3'-termini of the RNA molecules. There is some ambiguity in measuring the length of the short RNA-DNA hybrid segment, L, due to the 5'-leader sequence (Fig. 6a, b, and c), because we do not know whether it extends to the nearer or further crossover point with the loop, or to an intermediate

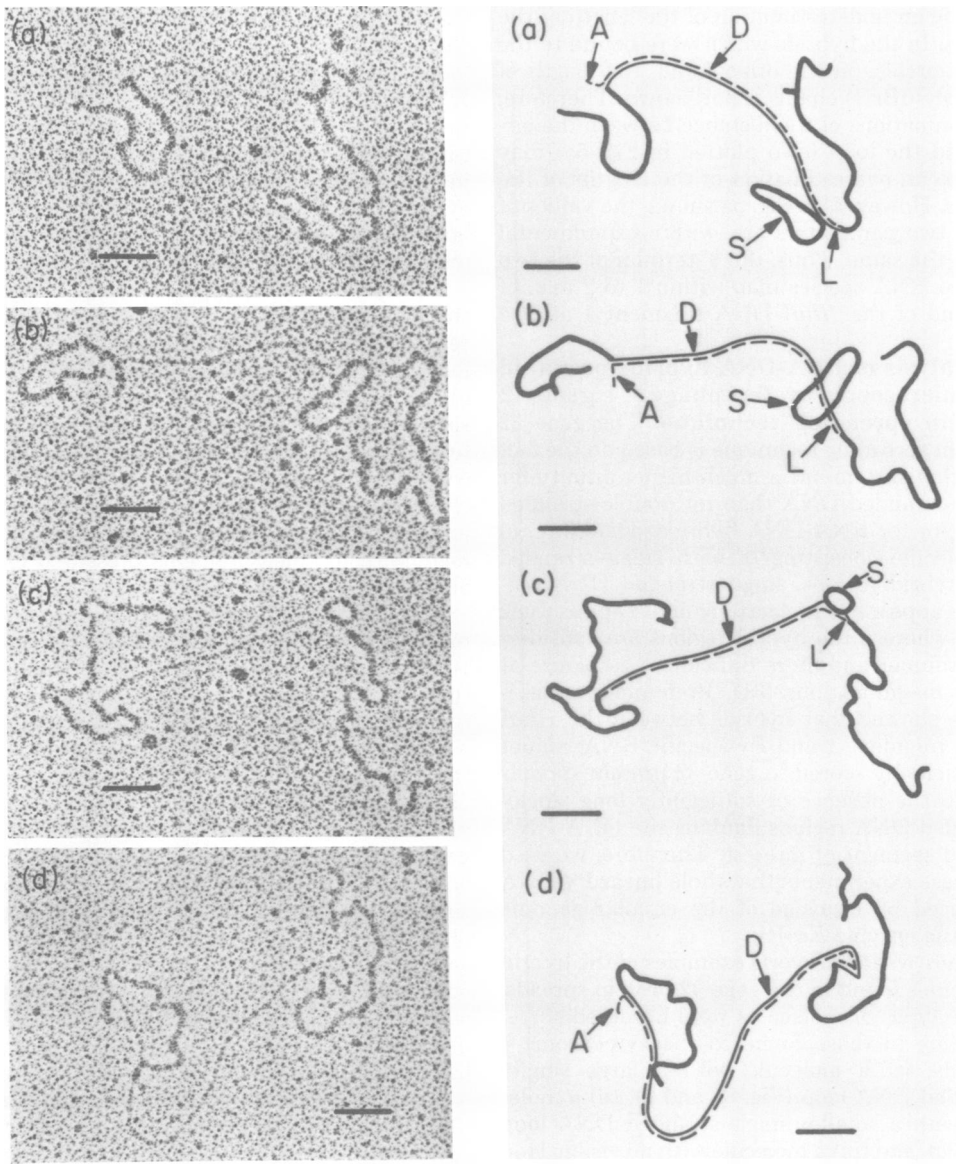

FIG. 6. *Electron micrographs of Py RNA-DNA hybrids spread for microscopy by the gene 32 protein method. Whole linear Py DNA, produced by cleavage of the circular genome with the restriction endonuclease EcoRI, and polyadenylated Py RNA prepared from whole Py-infected cells (a, c), or from a purified cytoplasm (b, d), were annealed under the same conditions used for preparing the hybrids studied by the cytochrome c-formamide spreading technique. The conditions for spreading the hybrids in the presence of gene 32 protein are described in the text. (a, b) Hybrid having a large single-stranded DNA loop; (c, d) hybrid having a small single-stranded DNA loop and a hybrid having no loop, respectively. In the interpretive tracings, A is the single-stranded segment interpreted as poly(A) tail; D, the long RNA-DNA hybrid region; S, the single-stranded DNA loop; and L, the short RNA-DNA region corresponding to the RNA leader sequence. Bar, 5 m.u. or 0.265 kilobases.*

point. By taking the two extreme limits we measure 1.2 ± 0.2 ($n = 30$) and 2.2 ± 0.4 ($n = 30$) m.u. as lower and upper limits for the length of the leader. We surmise that the lower limit is more likely to be close to the true value.

DISCUSSION

Figure 8 presents the coordinates of the three Py L DNA strand transcripts mapped in this study by electron microscopy. Two of these mol-

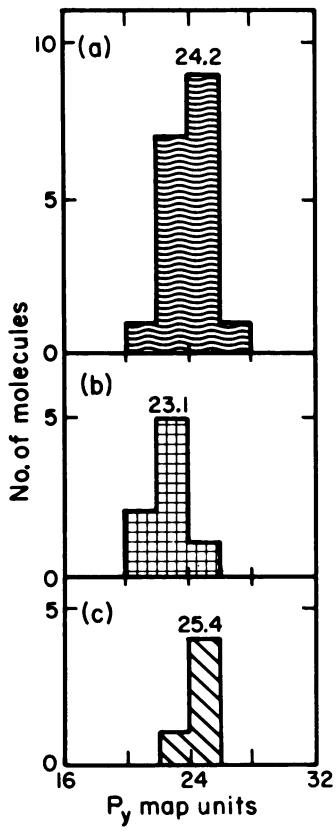

FIG. 7. *Coordinates of the 3'-termini of the Py L DNA strand transcripts. These coordinates were determined by measuring the distance between the poly(A) tail and the EcoRI cleavage site (map coordinate 0) in the hybrids having a large loop (a), in the hybrids having a small loop (b), and in the hybrids having no loop (c), in molecules prepared for microscopy by the gene 32 protein method (see Fig. 6). The coordinates are 24.2 ± 2.1 m.u. (n = 18), 23.1 ± 2.30 m.u. (n = 8), and 25.4 ± 2.7 m.u. (n = 5), respectively.*

ecules are spliced and consist of leaders whose 3'-ends map at 66.7 and 66.4 m.u. The 5'-ends of the main bodies of these molecules map at 49.4 and 59.4 m.u., respectively, and the corresponding 3'-termini map at 24.2 and 23.1 m.u. The 5'-end and the 3'-end of the third RNA molecule were found to map at 67.5 and 25.4 m.u., respectively. As noted above, this molecule either is unspliced or contains a splicing loop too small to be reliably observed by the gene 32 protein method.

A comparison of these data with the information gained in previous studies on the Py L DNA strand transcripts indicates that the three molecules shown in Fig. 8 are the three late mRNA species which code for the Py capsid proteins VP1, VP2, and VP3. The molecule with the large gap corresponds to the 16S mRNA shown previously by in vitro translation experiments to direct the synthesis of the VP1 coat protein (30). This inference is based on the observation that the coordinates of the main body of this spliced RNA molecule are identical to the previously determined coordinates of the 16S mRNA (18). The fact that the leader sequence was not detected in the previous study is probably due to the lower sensitivity of the technique used for mapping the RNA. Partial sequencing of the DNA and the VP1 coat protein has also indicated that the main body of the spliced RNA contains the genetic information for this protein (J. Arrand and B. Griffin, personal communication). A similar spliced 16S RNA molecule has been recently shown to be the mRNA for the major coat protein of the closely related SV40 (2, 15, 21).

The unspliced RNA molecule shown in Fig. 8 has approximately the same coordinates as those of the 19S mRNA, and this mRNA molecule has been previously found to code for the coat protein VP2 (30). The spliced RNA molecule with the small gap is presumably the mRNA which directs the synthesis of the coat protein VP3, since the probable location of the sequences coding for VP3 is within the main body of this molecule (29). The fact that its length, as determined by microscopy, is intermediate between that of the 16S and 19S RNA species is consistent with its predicted sedimentation coefficient of 18S (29).

It was pointed out in Results that many hybrid molecules containing a single-stranded DNA loop also contain a short unhybridized RNA across the loop. The length of the unhybridized RNA varies between 50 and 250 nucleotides. One possible explanation of this finding is that the hydrogen bonds at the single-stranded DNA-hybrid junction are less stable than those found within the hybrid region. As a result, the hybrid adjacent to the single-stranded DNA may sometimes "peel back." Another possible explanation is that the RNA at or around the leader contains sequences which are absent from the DNA. It is interesting to note, in this connection, that the leader sequences in the late Py mRNA's were found by fingerprinting analysis to consist of several repeats of a sequence which appears only once in the Py DNA (21; J. Arrand, personal communication). Our observation of unhybridized RNA sequences across the loop is consistent with these findings. The minimal and maximal estimates of the length of the leader sequences, 1.2 ± 0.2 m.u. (63 ± 10 nucleotides) and 2.2 ± 0.4 m.u. (116 ± 21 nucleotides), obtained by the gene 32 protein method, would be, according to this interpretation, minimal and maximal esti-

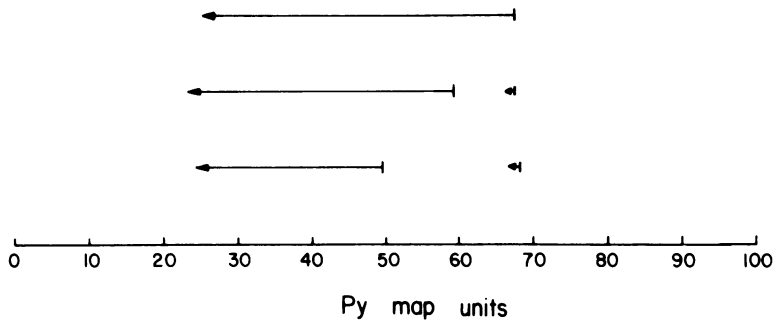

FIG. 8. *Topography of the three Py L DNA strand transcripts mapped by microscopy. The coordinates of the 3'-ends of the leader sequences in the spliced RNA molecules and the 5'-ends of their main bodies were calculated from the data shown in Fig. 4. The map positions of the 3'-ends of the three RNA species were calculated from the data shown in Fig. 7. The 5'-terminus of the unspliced RNA molecule was located on the physical map by measuring the length of the hybrid region from the poly(A) tail (Fig. 5b).*

mates of the length of one repeating unit in the leader. Recently, Manor et al. (H. Manor, M. Zuckerman, J. Parker, and R. Kamen, manuscript in preparation) hybridized the 16S Py mRNA with an *E. coli* plasmid containing two Py genomes inserted in tandem. They identified by electron microscopy hybrid molecules containing two characteristic loops which are expected if, and only if, the 16S RNA leader is in fact reiterated.

Three cytoplasmic L DNA strand transcripts having approximately the same coordinates as those shown in Fig. 8 have been identified and mapped by different methods and described in two recent articles published after completion of the present work (15, 21). These data indicate that not only the 16S and the 18S RNAs, but also the 19S RNA, contain a reiterated leader at the 5'-end. However, in contrast to the 16S and 18S RNAs, the 19S RNA appears to be a continuous transcript of the late region to which a reiterated leader with a length that is short but not precisely known is spliced at the 5'-end (21). A hybrid between such a molecule and Py DNA would not be expected to contain a DNA splicing loop, but would contain a small single-stranded RNA loop or a short chain of unhybridized RNA at the 5'-end (15). As discussed above, we have observed small bumps in the 19S RNA-Py DNA hybrids that could conceivably be related to these reiterated structures, but we do not have decisive data.

Finally, we would like to point out that the use of the gene 32 protein spreading technique, in addition to the more commonly used cytochrome *c* spreading technique, was essential for mapping the Py RNA species by electron microscopy, for the gene 32 protein spreads provided data which could not be otherwise obtained (e.g., the length of the leader sequence). Other gene 32 protein data lent further support to the results obtained by cytochrome *c* spreading. We recommend the use of both methods for similar mapping projects.

ACKNOWLEDGMENTS

This research was supported by grants from the United States-Israel Binational Science Foundation, Jersusalem, Israel, and the Leukemia Research Foundation, and by Public Health Service grant GM 20927 from the National Institutes of Health. H.M. was supported by an Eleanor Roosevelt-International Cancer Fellowship during the course of this work.

We thank Yaffa Bot for able technical assistance.

LITERATURE CITED

1. **Acheson, N. H., E. Buetti, K. Scherrer, and R. Weil.** 1971. Transcription of the polyoma virus genome: synthesis and cleavage of giant late polyoma-specific RNA. Proc. Natl. Acad. Sci. U.S.A. **68**:2231-2235.
2. **Aloni, Y., R. Dhar, O. Laub, M. Horowitz, and G. Khoury.** 1977. Novel mechanism of RNA maturation: the leader sequences of simian virus 40 mRNA are not transcribed adjacent to the coding sequences. Proc. Natl. Acad. Sci. U.S.A. **74**:3686-3690.
3. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. U.S.A. **69**:1408-1412.
4. **Bender, W., N. Davidson, K. L. Kindle, W. C. Taylor, M. Silverman, and R. A. Firtel.** 1978. The structure of M6, a recombinant plasmid containing Dictyostelium DNA homologous to actin messenger RNA. Cell **15**: 779-788.
5. **Berget, S. M., C. Moore, and P. A. Sharp.** 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. U.S.A. **74**:3171-3175.
6. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early messenger RNA's of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **75**:1274-1278.
7. **Birg, F., J. Favaloro, and R. Kamen.** 1977. Analysis of polyoma virus nuclear RNA by mini-blot hybridization. Proc. Natl. Acad. Sci. U.S.A. **74**:3138-3142.
8. **Casey, J., and N. Davidson.** 1977. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucleic Acids Res. **4**:1539-1552.
9. **Chow, L. T., R. E. Gelinas, T. R. Broker, and R. J. Roberts.** 1977. An amazing sequence arrangement at the 5'-ends of adenovirus 2 messenger RNA. Cell **12**:1-8.

10. **Davis, R. W., M. Simon, and N. Davidson.** 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. Methods Enzymol. **21**:413–428.
11. **Dunn, A. K., and J. A. Hassell.** 1977. A novel method to map transcripts: evidence for homology between an adenovirus mRNA and discrete multiple regions of the viral genome. Cell **12**:23–36.
12. **Flavell, A. J., and R. Kamen.** 1977. Strand-specific transcription of polyoma virus DNA late during productive infection. J. Mol. Biol. **115**:237–242.
13. **Griffin, B. E., and M. Fried.** 1975. Amplification of a specific region of the polyoma virus genome. Nature (London) **256**:175–179.
14. **Griffin, B. E., M. Fried, and A. Cowie.** 1974. Polyoma DNA: a physical map. Proc. Natl. Acad. Sci. U.S.A. **71**: 2077–2081.
15. **Horowitz, M., S. Bratosin, and Y. Aloni.** 1978. Polyoma infected cells contain at least three spliced late RNAs. Nucleic Acids Res. **5**:4663–4675.
16. **Hsu, M.-T., and J. Ford.** 1977. Sequence arrangement at the 5'-ends of SV40 16S and 19S mRNA's. Proc. Natl. Acad. Sci. U.S.A. **74**:4982–4985.
17. **Kamen, R., D. M. Lindstrom, H. Shure, and R. W. Old.** 1974. Virus specific RNA in cells productively infected or transformed by polyoma virus. Cold Spring Harbor Symp. Quant. Biol. **39**:187–198.
18. **Kamen, R., and H. Shure.** 1976. Topography of polyoma virus messenger RNA molecules. Cell **7**:361–371.
19. **Klessig, D. F.** 1977. Two adenovirus mRNA's have a common 5' terminal leader sequence encoded at least 10kb upstream from their main coding regions. Cell **12**: 9–21.
20. **Lavi, S., and Y. Groner.** 1977. The 5'-terminal sequences and coding regions of late SV40 mRNA's are derived from non-contiguous segments of the viral genome. Proc. Natl. Acad. Sci. U.S.A. **74**:5323–5327.
21. **Leion, S., A. J. Flavell, A. Cowie, and R. Kamen.** 1979. Amplification in the leader sequence of late polyoma virus mRNAs. Cell **16**:373–388.
22. **Lev, Z., R. Kamen, and H. Manor.** 1979. Topography of polyoma virus-specific giant nuclear RNA molecules containing poly(A) sequences. Virology **93**:445–452.
23. **Lev, Z., and H. Manor.** 1977. Amount and distribution of virus-specific sequences in giant RNA molecules isolated from polyoma-infected mouse kidney cells. J. Virol. **21**:831–842.
24. **Lindberg, U., and J. E. Darnell.** 1970. SV40-specific RNA in the nucleus and polyribosomes of transformed cells. Proc. Natl. Acad. Sci. U.S.A. **65**:1089–1096.
25. **Manor, H., and R. Kamen.** 1978. Polyoma virus-specific RNA synthesis in an inducible line of polyoma virus-transformed rat cells. J. Virol. **25**:719–729.
26. **Manor, H., and A. Neer.** 1975. Effects of cycloheximide on virus DNA replication in an inducible line of polyoma-transformed rat cells. Cell **5**:311–318.
27. **Parish, J. H., and K. S. Kirby.** 1966. Reagents which reduce interactions between ribosomal RNA and rapidly-labeled RNA from rat liver. Biochim. Biophys. Acta **129**:554–562.
28. **Rosenthal, L. J., C. Solomon, and R. Weil.** 1976. Isolation and characterization of poly (A)-containing intranuclear polyoma-specific "giant" RNAs. Nucleic Acids Res. **3**:1167–1183.
29. **Siddel, S. G., and A. E. Smith.** 1978. Polyoma virus has three late mRNA's: one for each virion protein. J. Virol. **27**:427–431.
30. **Smith, A. E., R. Kamen, W. F. Mangel, H. Shure, and T. Wheeler.** 1976. Location of the sequences coding for capsid proteins VP1 and VP2 on polyoma virus DNA. Cell **9**:481–487.
31. **Wu, M., and N. Davidson.** 1975. Use of gene 32 protein staining of single-strand polynucleotides for gene mapping by electron microscopy: application to the ϕ80 d3 ilvsu$^+$7 system. Proc. Natl. Acad. Sci. U.S.A. **72**:4506–4510.
32. **Wu, M., R. J. Roberts, and N. Davidson.** 1977. Structure of the inverted terminal repetition of adenovirus type 2 DNA. J. Virol. **21**:766–777.