Topography of the Three Late mRNA’s of Polyoma Virus Which Encode the Virion Proteins†

ROBERT KAMEN,* JENNIFER FAVALORO, AND JACQUELINE PARKER

Transcription Laboratory, Imperial Cancer Research Fund, London WC2A 3PX, England

The three cytoplasmic polyadenylated mRNA’s which separately encode the three capsid proteins (VP1, VP2, and VP3) of polyoma virus were mapped on the viral genome by one- and two-dimensional gel electrophoreses of nuclease S1-resistant RNA-DNA hybrids. The mRNA’s, which we designated mVP1, mVP2, and mVP3 to indicate the coding functions deduced from the cosedimentation of the RNAs and the messenger activities, comprise an overlapping set of 3'-coterminial molecules which also share a heterogeneous family of noncoding 5'-terminal regions (Flavell et al., Cell 16:357–371, 1979; Legon et al., Cell 16:373–388, 1979). The three species differ in the length of the 3' colinear coding region which is spliced to the 5' leader sequences. The common polyadenylated 3' end maps at map unit 25.3. The 5' ends of the colinear bodies of mVP1, mVP3, and mVP2 map at 48.5, 59.5, and 66.5 map units, respectively. An examination of the polyoma virus DNA sequence (Arrand et al., J. Virol. 33:606–618, 1980) in the vicinities of splicing sites approximated by the S1 gel mapping data for sequences common to the ends of known intervening sequences allowed prediction of the precise splice points in polyoma virus late mRNA’s. In all three cases, the leader sequences are joined to the mRNA bodies at least 48 nucleotides before the translational initiation codon used in each particular messenger. The start signal which functions in each mRNA is the first AUG (or GUG) triplet after the splice junction.

The mRNA’s which encode the three virion structural proteins of polyoma virus (VP1, VP2, and VP3) accumulate at late times during productive infection of permissive mouse cells (1). These mRNA’s are transcribed from the L strand of the viral DNA and have been mapped (4, 27) into the region extending counterclockwise from approximately 70 to 25 map units (m.u.) on the standard physical map (20) (Fig. 1). The major structural protein, VP1, differs entirely in tryptic peptide composition from the other two capsid proteins, whereas the two minor proteins, VP2 and VP3, are related in that all tryptic peptides present in VP3 are contained within the larger protein, VP2 (14, 18, 22). Studies in which a temperature-sensitive mutant was used (19) have suggested that VP3 comprises the C-terminal portion of VP2. Earlier work had established that there were at least two distinct polyoma virus late mRNA’s, a major component which sediments at 16S and a minor component which sediments at 19S (8). These two species were shown to have a common polyadenylated 3' end mapping at about 25 m.u. and apparent 5' ends located at 48 and 67 m.u. for the 16S and 19S mRNA’s, respectively (29). In vitro translation experiments identified the 16S mRNA as the message for VP1 and the 19S mRNA as the message for VP2 (43). Capsid protein VP3 could be derived from VP2 by proteolysis, or alternatively could be translated from a discrete mRNA intermediate in size between the 19S and 16S RNAs. In vitro translation of viral nuclear RNA (30) demonstrated that the large nuclear transcripts encoded VP2 but not VP3, whereas cytoplasmic mRNA, when translated in identical cell-free systems, yielded both polypeptides. This argued against the proteolysis hypothesis. We therefore reinvestigated the structure of the polyoma virus late mRNA’s by using the powerful technique of nuclease S1 gel mapping (6) and RNase T1 oligonucleotide fingerprinting to find that there were indeed three discrete late mRNA’s (15, 26, 34). At the same time, in vitro translation of mRNA’s separated on sucrose gradients (42) or polyacrylamide gels (25) clearly established that a separate 18S species encodes VP3. A further motivation for our reinvestigation of the late viral mRNA’s was that our original mapping conclusions were based on the assumption that mRNA’s are exclusively determined by contiguous DNA sequences. The discovery of RNA splicing (5, 11, 12, 21, 24, 31, 33)

†This paper and the accompanying papers are dedicated to M. G. P. Stoker on the occasion of his retirement as Director of Research of the Imperial Cancer Research Fund.
Fig. 1. Physical map of polyoma virus (Py) DNA showing the topography of the late mRNA's, as determined in this paper. The circular DNA is oriented with the single cleavage site for restriction endonuclease EcoRI at the top and is divided into 100 m.u., which increase in the clockwise direction from this site (20). The cleavage sites for the restriction enzymes which were used in this work are indicated (3, 43a-45); it should be noted that only cleavage sites used in this study are included. The portion of the DNA which encodes the late proteins (the late region) extends counterclockwise from approximately 70 to 25 m.u. (27). The indicated topography of the three late mRNA's summarizes the results described in the text and those published elsewhere (15, 34). The abbreviations used are as follows: mVP2, the mRNA for capsid protein VP2; mVP3, the mRNA for capsid protein mVP3; mVP1, the mRNA for capsid protein VP1. These functional assignments are derived by correlating in vitro translation studies of size-fractionated mRNA's (25, 42) with data presented in the text. The short 5'-terminal leader sequences which are spliced to the bodies of mVP2 and mVP3 are schematically indicated to show the map positions of the DNA sequences from which they derive. However, it is important to note that these leaders have heterogeneous capped 5' ends (15) and are not colinear with the DNA sequence since they comprise an imperfect tandem repeat of a sequence represented uniquely in the genomic DNA (34). As discussed in the text, a similar structure also occurs at the 5' end of mVP2.

invalidated this assumption. Indeed, differences in the map position assigned to the 16S mRNA by Türl et al. (47) and by us (29) could be explained by the presence of a 5'-terminal leader sequence spliced onto the body of this messenger. Several lines of evidence now indicate that the 5'-terminal portion of all three late mRNA's is a complex family of leader sequences derived from DNA sequences in the vicinity of 67 m.u. (15, 23, 34, 35). In this paper we report the accurate mapping of late polyoma virus mRNA's by using gel electrophoresis of nuclease S1-resistant hybrids (6). This method measures the distances between ends of colinear transcripts and restriction sites in the viral DNA. Its precision depends critically on the accuracy of the restriction map. As the complete nucleotide sequence of polyoma virus DNA was being determined by others (3, 43a-45), it was efficacious to delay publication of our results until the sequence (and thus the exact restriction map) was available. Moreover, correlation of the S1 mapping results with the DNA sequence has allowed prediction of probable splicing sites by comparison with the splicing sites which occur in other known examples of viral and cellular messengers. The preliminary assignments for the map positions of the three late mRNA's, nevertheless, were the basis of subsequent work in this laboratory, and therefore a summary of these conclusions (26) and some of the data have already appeared (15, 34).

MATERIALS AND METHODS

Polyoma virus strain A2 (20) was used throughout this study except where noted. The virus was propagated in secondary whole mouse embryo cells, cytoplasmic polyadenylated RNA was purified from infected 3T6 cells late during productive infection, and viral DNA was prepared as described elsewhere (J. Favaloro, R. Treisman, and R. Kamen, Methods Enzymol., in press). Nuclease S1 gel electrophoresis map-
ping was done by a modification of the Berk-Sharp procedure as described in detail by Favaloro et al. (in press). The amounts of RNA and DNA used in particular experiments, as well as the gel fractionation system, are described in the figure legends.

RESULTS

Nuclease S1 gel electrophoresis of RNA-DNA hybrids. Polyoma virus mRNA's were analyzed by using a modification (Favaloro et al., in press) of the endonuclease S1 gel electrophoresis method originally described by Berk and Sharp (6). In this technique, unlabeled RNA containing a proportion of homologous molecules is annealed to denatured double-stranded restriction fragments of the corresponding DNA under conditions (9) which favor RNA-DNA hybridization but exclude DNA reannealing. The RNA-DNA hybrids are then digested with the single-strand-specific endonuclease S1 to hydrolyze single-stranded DNA and RNA, including loops of single-stranded DNA resulting from the annealing of a spliced RNA to genomic DNA. The S1-treated hybrids are then analyzed by neutral or alkaline agarose gel electrophoresis, and chain lengths are calculated with respect to the mobility of marker DNA restriction fragments. In experiments producing small fragments, the size determination is alternatively done on urea-polyacrylamide gels after alkaline denaturation of the hybrids. In most of the experiments described below we used polyoma virus DNA labeled either in vivo or in vitro with $^{32}$P. In some cases, however, the DNA was unlabeled, and the gel patterns were visualized after transfer to nitrocellulose sheets (48) and annealing with highly radioactive polyoma virus-specific probes.

The assumptions on which this technique is based are that nuclease S1 effectively and precisely removes single-stranded tails and loops without introducing nicks within a continuous DNA-RNA hybrid. These were tested for polyoma virus DNA by preparing high-molecular-weight complementary RNA (cRNA) (which is a continuous transcript of the DNA L strand synthesized in vitro by *Escherichia coli* RNA polymerase) from either our wild-type polyoma virus DNA (strain A2) or the DNA of a naturally occurring viable deletion mutant (strain A3) and annealing this with restriction fragments of A2 DNA, which together span the entire genome. The A3 strain contains an 11-base pair deletion extending from nucleotide 44 to nucleotide 54 in the sequence of the early region of A2 DNA (45). Thus, hybrids formed between A3 cRNA and an appropriate fragment of A2 DNA contain an 11-nucleotide single-stranded DNA deletion loop analogous to that which would occur in hybrids between spliced RNA and its template DNA. In the experiment shown in Fig. 2 mixtures of either A2 or A3 cRNA and chicken embryo fibroblast rRNA carrier or carrier RNA alone was hybridized separately to the three SstI fragments of A2 DNA. After nuclease S1 digestion, the resulting hybrids were fractionated on an alkaline agarose gel (36), the DNA bands were transferred to nitrocellulose, and the blot was annealed to $^{32}$P-labeled A2 DNA. In the absence of viral cRNA

![Fig. 2. Alkaline agarose gel electrophoresis of nuclease S1-resistant DNA in hybrids between restriction fragments of polyoma virus strain A2 DNA and cRNA transcribed from strain A2 or strain A3 DNA. High-molecular-weight cRNA was prepared under conditions yielding predominantly complete transcripts of the DNA L strand, as described previously (28). Each hybridization contained the indicated restriction endonuclease SstI fragment of strain A2 DNA (4 ng of SstI-1, 4 ng of SstI-2, or 2 ng of SstI-3) and either no cRNA (tracks A, D, and G), A2 cRNA (150 ng; tracks B, E, and H), or A3 cRNA (110 ng; tracks C, F, and I). A mixture of polyoma virus DNA restriction fragments (EcoRI linear segments, the two Hsu fragments, and the eight HpaII fragments) in tracks labeled M served as size markers; the numbers on the left indicate their lengths in nucleotides, as deduced from the nucleotide sequence of polyoma virus DNA (3, 43a–45). The numbers on the right are the lengths (in nucleotides) of the indicated nuclease S1-resistant DNA fragments, as calculated from their mobilities relative to the marker fragments. The DNA used in this experiment was nonradioactive, and the bands were visualized by transfer to nitrocellulose followed by annealing to viral DNA labeled with $^{32}$P by nick translation, as described by Favaloro et al. (in press). Exposure time was 14 h, and Fuji Rx film without intensification was used.](http://jvi.asm.org/Downloaded from http://jvi.asm.org)
(Fig. 2, tracks A, D, and G), no S1-resistant DNA was detected, demonstrating the low background level of the Berk-Sharp technique. Both crNA preparations protected the entire lengths of fragments SstI-1 (tracks B and C) and SstI-3 (tracks H and I) from nuclease digestion, although a faint series of small bands was detectable with the larger fragments; this may reflect a very minor degree of internal nicking at sites slightly sensitive to the nuclease, but alternatively it may have been due to the locations of the crNA ends. The A2 crNA also entirely protected fragment SstI-2 (track E), which contains the deletion present in strain A3, whereas the DNA strand in the hybrid between A3 crNA and fragment SstI-2 (track F) was completely cleaved at the site of the deletion loop. Table 1 compares the lengths of the nuclease S1-resistant DNA bands obtained from the alkaline gel shown in Fig. 2 with the lengths expected from the DNA sequence and demonstrates excellent agreement between observed and expected values.

**Mapping the colinear regions of three late mRNA’s**. A physical map of polyoma virus DNA indicating the restriction enzyme cleavage sites used in this study is shown in Fig. 1. Initial experiments to detect late mRNA’s were done by hybridizing polyadenylated cytoplasmic RNA extracted from 3T6 mouse cells at a late time (30 h) after polyoma virus infection to 32P-labeled viral DNA cleaved outside the late region of the genome. Figure 3 shows a saturation curve in which 60 to 600 ng of unfractionated late mRNA was hybridized to a fixed amount (approximately 1 ng) of the 3,820-nucleotide DNA fragment extending clockwise from 0 (EcoRI) to 72.2 m.u. (BglII). With the smallest amount of mRNA (Fig. 3, track A) three major bands approximately 2,200, 1,850, and 1,250 nucleotides long were resolved by alkaline agarose gel electrophoresis of the nuclease S1-resistant hybrids. The intensity of the bands decreased with increasing length. Since the viral DNA was in excess in this case, the band intensities were roughly proportional to the relative abundances of the three RNA species which hybridized to the DNA. With increasing amounts of added

![Image: Alkaline agarose gel of nuclease S1-resistant DNA in hybrids between restriction fragment EcoRI-BglI-1 (0 to 72.2 m.u.) and different amounts of late mRNA. Each annealing contained the 32P-labeled restriction fragment (2,500 cpm; ca. 2 x 10^6 cpm/μg) and polyadenylated cytoplasmic RNA extracted from 3T6 mouse cells 30 h after infection and growth at 37°C (track A, 0.06 μg; track B, 0.145 μg; track C, 0.290 μg; track D, 0.580 μg). The first marker track (track M on left) contained the eight HpaI fragments of viral DNA, and the second marker track (track M, on right) contained the three SstI fragments. Chain lengths (in nucleotides) of marker fragments are shown on the right, and the calculated lengths (in nucleotides) of the S1-resistant DNAs are shown on the left. The photograph shows an overexposed autoradiograph (1-week exposure with intensifying screen) to facilitate reproduction. The measurements for the chain length calculations in this case and for the other figures were taken from short exposures, where the bands were far sharper.
mRNA (Fig. 3, tracks B through D), the relative intensities changed as the available DNA was saturated. Note that the 1,250-nucleotide band actually decreased in darkness as the 1,850- and 2,200-nucleotide bands became prominent. This is because (see below) the three bands are due to a set of overlapping mRNA’s and the larger ones displace the smaller one from the hybrid as conditions of RNA excess are reached; when there is excess RNA, the relative band intensities reflect the RNA lengths rather than their abundances. At the two highest RNA concentrations (Fig. 3, tracks C and D) a 1,380-nucleotide band was detected in addition. This band is the 3’ end of the early mRNA’s (extending from the end of the DNA fragment at 0 m.u. clockwise to 26 m.u.), as will be discussed elsewhere (R. Kamen, J. Favaloro, J. Parker, M. Fried, B. E. Griffin, and Y. Ito, manuscript in preparation). We show below that the 2,200-, 1,850- and 1,250-nucleotide bands are the colinear portions of the messengers for capsid proteins VP2, VP3, and VP1, respectively. To simplify the discussion below, we refer to the three mRNA species as mVP2, mVP3, and mVP1 to indicate their coding functions.

To position the ends of the three colinear mRNA transcripts on the physical map, we hybridized late cytoplasmic polyadenylated RNA to various restriction fragments which end within the late region. The chain lengths of the nuclease-resistant DNA fragments obtained can be used to determine the distance between an endpoint and the restriction cleavage site, the precise location of which is determined from the known DNA sequence. Unambiguous assignments can be made by comparing the results obtained with a variety of restriction fragments. Because the chemical polarity of the late mRNA’s with respect to the physical map is known (28), endpoints can readily be identified as either 5’ or 3’ termini.

Figure 4A (tracks A, B, and C) shows the nuclease S1-resistant DNA fragments obtained by annealing late mRNA to the restriction fragment extending from 36.0 to 72.2 m.u. (BglI–HincII-2 [Fig. 1]). Three bands with lengths of 1,620, 1,240, and 670 nucleotides were detected. These lengths were tentatively identified as the distances between the HincII site at 36.0 m.u. and the 5’ ends of the colinear portions of mVP2, mVP3, and mVP1, respectively, as indicated diagrammatically in Fig. 5. Hybridization with the fragment extending from 36.0 to 58.1 m.u. (BamHI–HincII-2) showed complete protection of the 1,188-nucleotide restriction fragment and a smaller 680 nucleotide band (Fig. 4B), whereas hybridization with the fragment extending from 26.5 to 36.0 m.u. (between the two HincII sites [Fig. 1]) yielded only complete protection of the entire fragment from S1 digestion (Fig. 4C). As shown in the diagram in Fig. 5, these three experiments together yield a provisional map for the 5’ ends of the colinear portions of mVP2, mVP3, and mVP1. This map was confirmed with other restriction fragments (SstI-2, HpaII-1, HpaII-3, BamHI–EcoRI-2), and all of the results are summarized in Table 2. The data comprise an internally consistent set and map the 5’ ends of the colinear bodies of mVP2, mVP3, and mVP1 at 66.5, 59.5 and 48.5 m.u., respectively. To illustrate the internal consistency, it was useful to compare the observed lengths of S1-resistant DNA with those expected from a model specifying the 5’ ends of the colinear transcripts. In constructing this model (which is discussed in more detail below), we first estimated the positions of the 5’ ends of the three colinear transcripts by averaging the distances measured from various restriction sites. We then inspected the sequence of polyoma virus DNA within 20 nucleotides of mean endpoints and predicted splice sites by using the GT/AG rule (7) and the requirement for a pyrimidine cluster 5’ to the AG (10, 17, 39–41). These predictions (Table 2) were then used to calculate expected lengths of S1-resistant DNA fragments. Table 2 shows that there is excellent agreement between the observed and predicted lengths.

The common 3’ end of the three mRNA’s was mapped in the first instance by hybridization to the fragment extending clockwise from 96.5 to 52.6 m.u. (SstI-1, [Fig. 1]). With a small amount of late mRNA from cells infected with wild-type strain A2 virus, two bands 1,450 and 1,250 nucleotides long were obtained (Fig. 6A, track F). With larger amounts of wild-type late mRNA under conditions of RNA excess, the intensity of the 1,250-nucleotide band decreased, whereas that of the 1,450-nucleotide band increased (Fig. 6A, tracks G and H). In addition, a further band 1,550 nucleotides long was detected (Fig. 6A, tracks G and H). As Fig. 5 shows, the 1,250-nucleotide band is the full length of the colinear portion of mVP1, whereas the 1,450-nucleotide band is the distance between the restriction site at 52.6 m.u. and the common 3’ end of mVP2 and mVP3. The 1,550-nucleotide band, which is minor in wild-type-infected cells, is overproduced in cells infected by the large-T antigen mutant tsA104 after a temperature shift from the permissive to the nonpermissive temperature (Fig. 6A, tracks C through E). Since early mRNA’s are coordinately overproduced under these conditions (13), we identified the 1,550-nucleotide band as the 3’ end of the early mRNA’s (Fig. 5). These assignments were confirmed by hybridization with the fragment ex-
FIG. 4. Alkaline agarose gels of nuclease S1-resistant DNA in hybrids between late mRNA's and various restriction fragments which end within the late region. (A) Hybridization with fragment BglII to HincII-2 (36.0 to 72.2 m.u.). Each annealing contained 20 ng of nonradioactive restriction fragment. Track A, 0.15 μg of tsA mRNA; track B, 0.22 μg of a different preparation of tsA mRNA; track C, 0.15 μg of wild-type mRNA. The tracks labeled M contained polyoma virus restriction fragments (see Fig. 2), and the numbers on the left are their chain lengths in nucleotides. The numbers on the right are the calculated chain lengths (in nucleotides) of the three observed nuclease S1-resistant bands. Tracks 1 through 15 are hybrids formed with samples of sucrose gradient fractions separating the late mRNA's, with the position of the 18S RNA marker indicated; 100 pg of polyadenylated tsA RNA was sedimented through a 5 to 20% sucrose gradient containing 50% formamide (29) for 22 h at 35,000 rpm in a Beckman SW40 rotor, and 5-μl samples of the 500-μl fractions were hybridized to the restriction fragment. The slowest migrating band visible in all of the tracks, as well as in the blank (track B1, no viral RNA added), was due to a small amount of self-annealing of the restriction fragment and was not seen in other experiments in which this fragment was used. The DNA bands were detected after transfer to nitrocellulose and annealing to viral DNA labeled with 32P by nick translation. (B) Hybridization with fragment BamHI–HincII-2 (36.0 to 58.1 m.u.). Each annealing contained approximately 500 cpm of DNA (specific activity, ca. 2 × 10^6 cpm/μg) and either 0.5 μg of A2 cRNA (track A), 0.05, 0.1, or 0.5 μg of A2 wild-type polyadenylated mRNA (tracks B, C, and D, respectively), or 0.5 μg of tsA polyadenylated mRNA (track E). The positions of polyoma virus DNA marker fragments (see Fig. 2) are shown on the left, and the calculated lengths (in nucleotides) of the observed bands are shown on the right. Note that the cRNA protected the entire length of the fragment. No bands were visible when the viral mRNA was omitted (data not shown). (C) Hybridization with 32P-labeled fragments BamHI–HincII-2 (tracks A through D) and BamHI–HincII-3 (tracks F through H). Tracks labeled M contained markers, the lengths (in nucleotides) of which are shown on the left. Tracks A and E were blanks with no added mRNA. Tracks B and F were annealings with 0.08 μg of A2 cRNA. Tracks C and G were annealings containing 1.0 μg of A2 wild-type polyadenylated mRNA, whereas tracks D and H were annealings containing 1.25 μg of tsA polyadenylated mRNA.
We also conclude that the data from F through H) produced a mutant which contained, in addition, mRNA which produced a 500-nucleotide band (Fig. 6B, tracks F through H). This is the 3' end of the early mRNA's. The data from Fig. 6A and B are also summarized in Table 2. We conclude that the late mRNA's have a common 3' terminus at 25.3 m.u. We also conclude that the common 3' end of the early mRNA's lies at 25.8 m.u. Note that these conclusions suggest that the 3' ends of early and late mRNA's overlap by about 30 nucleotides.

Correlation of the three late mRNA's with coding function. Siddell and Smith (42) identified three functional late mRNA's by translating in vitro viral mRNA's fractionated on sucrose gradients. They found that mVP2, mVP3, and mVP1 sedimented at approximately 19S, 18S, and 16S, respectively. Hunter and Gibson (25), using polyacrylamide gel electrophoresis to fractionate the mRNA's, confirmed these conclusions. To correlate the three mRNA species characterized above with the in vitro translation results, we sedimented late mRNA through a sucrose gradient and hybridized the RNA in each fraction to the segment of viral DNA extending from 36.0 to 72.2 m.u. (Fig. 4A). The data obtained clearly show that the S1 nuclease-resistant DNA bands tentatively identified above as the 5'-terminal portions of the colinear regions of mVP2, mVP3, and mVP1 were produced by mRNA's with the expected sedimentation coefficients. We also analyzed the same sucrose gradient fractions used by Siddell and Smith (data not shown) and confirmed that the colinear transcripts cosedimented with their respective messenger activities.

Leader sequences on the late mRNA's. The mapping experiments described above did not reveal leader sequences at the 5' ends of late polyoma virus mRNA's, as reported for adeno-virus and simian virus 40 late messengers (5, 11, 12, 21, 24, 31, 33); no bands other than the colinear mRNA bodies were detected on the gels. However, other experiments demonstrated that the actual 5' ends of all of the mRNA's mapped near the 5' end of the mVP2 body at 66 to 67 m.u. Flavell et al. (15) showed that all three viral mRNA's have 5'-terminal cap struc-
TABLE 2. Summary of the lengths of S1-resistant DNA obtained with various restriction fragments compared with expected lengths derived from an examination of the DNA sequence for tracts characteristic of the ends of intervening sequences
d

<table>
<thead>
<tr>
<th>Restriction fragment</th>
<th>Map coordinates</th>
<th>Length of S1-resistant DNA (nucleotides)</th>
<th>Expected length (nucleotides)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI-BglI-1 or BglII</td>
<td>0-72.2 (full length, cut at 72.2)</td>
<td>2,200 ± 44</td>
<td>2,175</td>
</tr>
<tr>
<td>BglI-HincII-2</td>
<td>36.0-72.2</td>
<td>1,620 ± 24</td>
<td>1,607</td>
</tr>
<tr>
<td>HincII-HincII</td>
<td>36.0-36.0</td>
<td>505 ± 14</td>
<td>504</td>
</tr>
<tr>
<td>BamHI-HincII-2</td>
<td>36.0-58.1</td>
<td>1,170 ± 10</td>
<td>1,166</td>
</tr>
<tr>
<td>BamHI-EcoRI</td>
<td>58.1-100</td>
<td>445 ± 30</td>
<td>443</td>
</tr>
<tr>
<td>BamHI</td>
<td>Full length, cut at 58.1</td>
<td>1,750</td>
<td>1,731</td>
</tr>
<tr>
<td>Hpall-3</td>
<td>53.9-70.5</td>
<td>658</td>
<td>666</td>
</tr>
<tr>
<td>Hpall-1</td>
<td>27.1-53.9</td>
<td>300</td>
<td>297</td>
</tr>
<tr>
<td>SstI-1</td>
<td>96.5-52.6</td>
<td>1,450 ± 24</td>
<td>1,444</td>
</tr>
<tr>
<td>Sst-2</td>
<td>52.6-81.3</td>
<td>720 ± 5</td>
<td>730</td>
</tr>
<tr>
<td>MboII-1</td>
<td>16.4-36.7</td>
<td>595 ± 5</td>
<td>603</td>
</tr>
</tbody>
</table>

* See reference 3 for the DNA sequence.

The expected lengths (see text) were obtained as follows. The distances of the mRNA 5' ends from the different restriction sites were used to calculate mean estimates for their map positions. The DNA sequence (3) within 20 nucleotides of these mean values was then searched for an AG dinucleotide preceded by a pyrimidine-rich tract because such regions occur at the ends of all known intervening sequences (7, 10, 16, 39-41). Such sequences were found in all three cases and the nucleotides after the AGs (positions 216, 585, and 1,170 in the DNA sequence of the late region [3]) were used as the 5' ends of the mRNA bodies in calculating the expected lengths. The location for the 3' end of the mRNA used in the expected length calculations was the mean distance from the SstI and MboII sites, which localized the polyadenylated tract just after an AAUAAA hexanucleotide predicted by the DNA sequence (45).

Two-dimensional analysis of late mRNA-DNA hybrids. We have developed a two-dimensional gel fractionation technique (Favaloro et al., in press) which is a general diagonal assay for characterizing spliced mRNA's. Berk and Sharp (6) originally showed that S1 nuclease removes the single-stranded DNA loop from a hybrid between a spliced RNA and its template DNA without cutting the RNA strand in the majority of molecules. Such an S1-digested hybrid migrates as a single band on nondenaturing gels, but dissociates into its component single-stranded DNA fragments on alkaline gels. In the two-dimensional technique, native S1-resistant hybrids are resolved by electrophoresis along one edge of a square neutral agarose gel. The gel is then soaked in alkali to denature the hybrids, and electrophoresis is continued at 90° to the original dimension in alkaline buffer. After neutralization, the single-stranded DNA is transferred to nitrocellulose by the Southern procedure (46) and visualized by annealing to 32P-labeled total or strand-specific nucleic acid probes. The DNA present in a continuous DNA-RNA hybrid has the same relative mobility in both neutral and alkaline dimensions and therefore is detected as a dark radioactive spot along the faintly visible diagonal which is the result of the background of nonspecifically nicked DNA always present in S1 gel analyses. The component DNA fragments from a hybrid between a
spliced RNA and its template, however, have greater relative mobilities in the alkaline dimension and form radioactive spots below the diagonal. Moreover, the two or more component DNA fragments of a spliced RNA-DNA hybrid appear as vertically aligned spots because they have a common mobility in the neutral dimension but different mobilities in the alkaline dimension.

Although the 5'-terminal leader sequences of the polyoma virus late mRNA's are too short and heterogeneous for direct detection on agarose gels, we reasoned that their presence would yield a characteristic displacement from the diagonal when the colinear bodies of the mRNA's were resolved on two-dimensional gels. Figure 8 shows the two-dimensional gel analysis of hybrids formed between late cytoplasmic mRNA and viral DNA cleaved once at 58.1 m.u. with BamHI. The RNA in this experiment was purified from cells infected with the tsA early mutant and contains almost equal quantities of early and late messengers. Duplicate blots were annealed to either nick-translation viral DNA to detect all species or to 32P-labeled asymmetric cRNA to detect specifically those DNA fragments protected by mRNA's transcribed from the DNA L strand (late mRNA's). Eight spots were resolved on the gels. Only those identified as 1, 2, 2', and 3 were detected with the late mRNA-specific probe (Fig. 8C). Spots 1 and 3 lie along the diagonal and correspond to the colinear portions of MVP1 and MVP3, respectively. Spot 2 lies below the diagonal and is vertically above spot 2', which is clearly visible in the longer exposure shown in Fig. 8B and is also detectable in Fig. 8C. Spots 2 and 2' both correspond to the colinear portion of MVP2. Because the BamHI site lies within the region determining the mRNA, the messenger circularized the DNA and after S1 nuclease digestion yielded a hybrid containing a specific nick at the BamHI site. Spot 2 is the 1,750-nucleotide fragment from the 3' end of MVP2 to the BamHI...
site, whereas spot 2' is the 400-nucleotide fragment from the BamHI site to the 5' end of the mRNA at 66.5 m.u. The other spots on the two-dimensional gel are hybrids between the DNA and early mRNA's. Since these messengers are transcribed from the DNA E strand (27), the DNA they protect from nuclease digestion is not detected with the L-strand-specific probe used in Fig. 8C. The early mRNA's provide an im-

*Fig. 7. Polyacrylamide gel fractionation of small DNA fragments protected from S1 nuclease digestion by late mRNA's. Polyadenylated cytoplasmic RNA from wild-type polyoma virus-infected cells (5 μg) was annealed to the 32P-labeled fragment extending from 0 to 72.2 m.u. (BglII-EcoRI-1; 12,000 cpm) (track A) and to the fragment extending from 36.0 to 72.2 m.u. (BglII-HincII-2; 6,600 cpm) (track B). The S1-resistant hybrids were denatured in alkali and fractionated on a 10% urea-polyacrylamide gel as described by Favaloro et al. (in press). The marker tracks (M) contained HaeIII fragments of polyoma virus DNA, whose chain lengths (in nucleotides) are shown on the left. The sizes (in nucleotides) of the protected DNA fragments are shown on the right. The autoradiogram was overexposed to visualize the smaller bands, thus obscuring the larger bands due to the mRNA bodies near the top of the gel.*

*Fig. 8. Two-dimensional (neutral/alkaline) gels of S1 nuclease-resistant DNA in hybrids between tsA mRNA and viral DNA cleaved with BamHI (58.1 m.u.). The gels are oriented with the neutral dimension from left to right and the alkaline dimension from top to bottom. The spots along the left edge are marker polyoma virus (Py) DNA fragments loaded before the alkaline dimension. (A) and (B) Shorter and longer exposures of the blots after annealing with nick-translated polyoma virus cRNA (Favaloro et al., in press), which detected specifically DNA fragments from hybrids containing late mRNA's.*
portant internal control. For example, as described in another publication (Kamen et al., in preparation), spots 5 and 5' are produced by the colinear segments of an early mRNA which contains a 400-nucleotide splice. Spot 5 is the colinear segment from 85.8 m.u. to the 3' end at 25.8 m.u., and spot 5' is the 260-nucleotide fragment from the 5' end at 73.3 m.u. to the splice at 78.3 m.u. The obvious displacement of spot 5 from the diagonal because of the 260-nucleotide fragment 5' shows that leaders 100 to 200 nucleotides long should have displaced mVP3 and mVP2 spots 3 and 1 from the diagonal if they were present. This was not the case. We inferred that the late leader sequences, as suggested by the data in Fig. 7, are mostly shorter than 100 nucleotides. We therefore repeated the two-dimensional analysis with a restriction fragment extending from 36.0 to 72.2 m.u. This produces shorter S1-resistant fragments corresponding to the 5' ends of the late mRNA bodies and thus increases the fractional alteration in hybrid chain length caused by very short leaders. We also lowered the temperature of the S1 nuclease digestion from 37°C used in the experiment shown in Fig. 8 to 11°C (Fig. 9B and C) to stabilize hybrids containing the short leader sequences. Figure 9A demonstrates that minor spots displaced from the diagonal were obtained for mVP3 and mVP1. The difference in mobility between the major and minor spots in the neutral dimension suggests that the hybrids generating the minor spots contained a second DNA fragment about 50 nucleotides long. Figures 9B and C confirm this result (in this case size, fractionated mVP2, mVP3, and mVP1 were used) and demonstrates that even at 11°C, only a fraction of mVP3 and mVP1 hybrids contained a continuous RNA chain joining the DNA fragment protected by the mRNA body to that protected by the leader. These data are consistent with the repeated structure proposed for the leader sequences spliced to polyoma virus late mRNA's (34). The DNA sequence complementary to the leader sequence can hybridize to any of the three or four repeat units in the mRNA. Only those hybrids in which the DNA has hybridized to the body-proximal repeat unit would lack a nuclease-sensitive loop in their RNA strands.

DISCUSSION

We mapped three different late mRNA's on the polyoma virus genome (Fig. 1). These comprise an overlapping set of messengers with 5' ends of colinear sequences at 66.5, 55.5, and 46.5 m.u. and a common polyadenylated 3' end at 25.3 m.u. Correlation of the sedimentation rates of the mRNA's with those of the messenger activities identified the three components as mVP2, mVP3, and mVP1, respectively. Additional evidence demonstrating that the three mRNA's share a complex family of 5'-terminal leader sequences derived from the DNA sequence in the vicinity of 67 m.u. may be summarized as follows. (i) All three mRNA's have at least seven alternative capped 5' ends, all of which map within the region between 70.5 and 65.4 m.u. (15). (ii) Precise localization of three of the alternative 5' ends by sequencing capped RNase T1 oligonucleotides showed them to occur at positions 198, 199, and 201 in the DNA sequence (3), corresponding to 66.7 m.u. (14a). (iii) Each of the three mRNA's contains four large RNase T1 oligonucleotides uniquely derived from a continuous region of the DNA sequence between 66.28 and 65.38 m.u. (from nucleotide 225 to nucleotide 273). Although this sequence occurs only once in the genomic DNA (3, 44, 45), the four T1 oligonucleotides are present in multimeric amounts in the mRNA's, leading to the conclusion that the leader sequences are tandem repeats of a basic sequence 50 to 60 nucleotides long (34). The S1 mapping data presented in this paper are consistent with the structures predicted for the 5' ends of the late mRNA's. We showed that some S1-resistant hybrids between late mRNA's and DNA from the late region contained, in addition to the long colinear bodies, a heterogeneous set of smaller DNA fragments, most of which were about 55 nucleotides long. This is strikingly different from the results obtained when similar techniques were used in studies of simian virus 40 late mRNA's (32, 48). Although the leader sequences of the simian virus 40 late mRNA's are also heterogeneous, S1-protected DNA fragments more than 200 nucleotides long were readily detected. Analysis of the polyoma virus late mRNA's with the two-dimensional variation of the S1 gel mapping technique provided evidence consistent with the existence in mVP3 and mVP1 of a 5'-terminal leader containing repeats of a sequence 50 nucleotides long. It will be necessary, however, to sequence the 5' ends of a representative set of late mRNA's to rigorously establish their structures.

Although all of the available data clearly indicate that mVP3 and mVP1 are spliced, we are less certain about the largest species, mVP2. The S1 mapping experiments demonstrated that mVP2 is colinear with the DNA sequence up to 66.5 m.u. and thus its body extends through the region determining the four large T1 oligonucleotides (66.28 to 65.38 m.u.) which define the leader sequence. We find no evidence for a discontinuity between 66.5 m.u. and the initiation codon for VP2, which is at 65.06 m.u. (3, 22a).
Legon et al. (34), however, found that all four of the leader sequence T1 oligonucleotides were over-represented in mVP2 molecules purified by sucrose gradient centrifugation. The two electron microscopic studies of polyoma virus late mRNA's (23, 35) found that mVP2 hybridized continuously with the late region DNA, but Ho-rowitz et al. (23) suggested that a short tail of single-stranded RNA may exist at the 5' end. The most probable interpretation of these data is that mVP2 does have the repeated leader but lacks a splice between the last copy of the repeat and its coding region.

A general pattern has emerged from the known DNA sequences spanning corresponding splicing sites in many mRNA's (7, 10, 16, 40),
including those of simian virus 40 (17, 39). Virtually all splicing sequences obey the so-called GT/AG rule (7), which specifies that the first two nucleotides at the 5’ end of intervening sequences are GT and that the last two nucleotides at the 3’ end are AG. Furthermore, a pyrimidine-rich tract precedes the 3’-terminal AG dinucleotide. Seif et al. (41) recently proposed consensus sequences for splice donors and splice acceptors, which extend the GT/AG rule. The proposed optimal donor sequence is AG↓GTAAGT (↓ indicates the splice point). Variations from this preferred sequence belonged to one of four less specific sequences: R↓GTANG, R↓GTANNT, R↓GTNNGT, and R↓GTNAG (R = purine; N = any nucleotide). Their preferred sequence for acceptor sites was YYNYAG↓ (Y = pyrimidine), with the additional restriction that no other AG dinucleotide occur within the 13 nucleotides before the splicing point. An examination of the polyoma virus DNA sequence (3) in the regions identified by the present data for such donor and acceptor sites allows prediction of the splice junctions in the late mRNA’s. These predictions were used to deduce the theoretical lengths listed in Table 2; it is apparent that the agreement between the expected and the observed lengths of S1-resistant DNA is very good. Although RNA sequencing (or sequencing of cloned complementary DNA) is obviously required to confirm the splicing predictions, they are of sufficient interest to mention at this time.

The leader-to-leader splice postulated in the amplified leader sequence (34) is predicted to occur between nucleotides 272 and 216 in the sequence of the late region of polyoma virus DNA (3). We presume that this “backwards” splice in fact occurs during the processing of giant nuclear RNA, which comprises tandemly repeated complete transcripts (1, 27; R. Treisman, personal communication) of the polyoma virus DNA L strand. Thus, the donor nucleotide would in fact be joined to the acceptor in the next copy of the viral genome within a giant transcript. As discussed above, the acceptor site for the leader-to-leader splice may also form the 5’ end of the collinear portion of mVP2. The proposed donor sequence (GATCAA↓-GTAAGT) is only one nucleotide different from the optimal sequence (41) and is the only oligonucleotide within 300 nucleotides preceding the initiation codon for VP2 which fits the general pattern for splice donors (3). The proposed acceptor sequence (TTTTCTATTTAAG↓AG) is also the only oligonucleotide near the 5’ end of the late region which fits the general pattern (one of the ovalbumin splice acceptors has an AAG instead of YAG). Moreover, the acceptor sequence occurs immediately before the first T1 oligonucleotide found within the tandemly repeated leader sequence (34) and includes the end of the large capped T1 oligonucleotide identified by Flavell et al. (14a). The strongest support for the donor and acceptor assignments is that the sequence across the splice junction uniquely predicts that the T1 oligonucleotide AUCCAAAG is present in the repeated leaders. This oligonucleotide was found by Legon et al. (34), and these authors suggested that it spans a splice junction. Some of the heterogeneous family of 5’-leader sequences may, however, be more complex than perfect tandem repeats (34), and it seems possible that other splices may occasionally occur or that sequences from other regions of the large precursor molecules may be incorporated into the leader sequences.

The S1 mapping data locate the 5’ end of the mVP3 body at 59.5 m.u., and a sequence (CCTCTACTATTTCCCTAG↓GA) is found in the DNA at this point (nucleotides 567 to 586) (3), which fits the acceptor pattern for a splice before nucleotide 585. The 5’ end of the mVP1 body at 48.5 m.u. is near the sequence CTTCCCTTAAATCTAG↓GG (nucleotides 1,154 to 1,171) (3), and we thus predict that the mVP1 splice occurs before nucleotide 1,170. The homology between the polyoma virus and simian virus 40 DNA sequences is most striking in this region; 27 of 31 nucleotides surrounding the splice junction are identical. This is one of the two longest regions of DNA sequence conservation observed between the two viruses (3, 44, 45).

Two final points of interest arise from comparing the S1 mapping results with the DNA sequence. First, the 5’ ends of the collinear portions of all three mRNA’s are at least 48 nucleotides before the initiation codons (3, 22a). Thus, the leader sequences are noncoding and are not adjacent to the AUGs after splicing. The initiation codons used are in each case the first AUG (or GUG) triplets after the splice junctions (3). Second, the polyadenylated 3’ ends of the late mRNA’s overlap the 3’ ends of the early mRNA’s which are encoded by the other DNA strand. Both DNA strands contain AATAAA hexanucleotides shortly before the points corresponding to the mRNA 3’ ends (45), and thus the polyoma virus mRNA’s extend the general observation that all messengers have the sequence AUUAAA (38) near their polyadenyllic acid tails.

ACKNOWLEDGMENTS

We are grateful to J. R. Arrand, B. E. Griffin, N. Smolar,
and E. Soeda for providing DNA sequence information before publication and to J. R. Arrand and B. E. Griffin for help in the preparation of this manuscript.

LITERATURE CITED


