JOURNAL OF VIROLOGY, July 1983, p. 233–237
0022-538X/83/070233-05$02.00/0

Vol. 47, No. 1

Sequence Repeats in a Polyoma Virus DNA Region Important for Gene Expression

H. EARL RULEY† AND MIKE FRIED*

Department of Tumour Virus Genetics, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, United Kingdom

Received 24 January 1983/Accepted 28 March 1983

The sequenced prototype strains (A2 and A3) of polyoma virus lack sequence duplications characteristic of other papovaviruses. However, we found that five polyoma virus strains (P16, Toronto large plaque, MV, Ts 48, and NG59R) contain tandemly duplicated sequences in a region near the late RNA leader. Although the duplications vary in size (31 to 84 base pairs) and location (between nucleotide [nt] 5068 and nt 5185), the sequence between nt 5114 and nt 5137 is contained within all five duplicated segments. This region is known to be important in polyoma virus early gene expression, and it contains sequences capable of enhancing the expression of nonviral genes. Inspection of the sequences at and around the ends of the repeats indicated that the duplications do not arise by homologous recombination, and there was no indication that a sequence-specific mechanism results in their formation. However, the variation in the structure of the repeats among different polyoma virus strains suggests that these sequence duplications are a recent evolutionary occurrence. The potential biological significance of this variation is discussed.

The genomes of the papovaviruses BK and simian virus 40 (SV40) contain repeated sequences located on the late side of the origin of replication (9, 22, 24, 33). Most extensively studied has been the 72-base-pair (bp) repeat of SV40. This element contains sequences important in viral early gene expression (2, 14), and it is located in a region which has the ability to enhance the expression of nonviral genes (1, 21). In contrast, wild-type strains of polyoma virus (Py) have been thought to lack similar repeats since sequence duplications are not present in the two prototype Py strains which have been sequenced (5, 12, 25). That Py would lack repeated sequences is surprising given the similarities in genome organization found among the papovaviruses.

For many years, however, it has been appreciated that the size of restriction fragments derived from the noncoding region is variable in different plaque isolates of Py (4, 11, 26, 31). This variation has been seen on the late side of the origin of replication in the size of the viral *Hpa*II-3 fragment and on the early side of the replication origin in the size of the *Hpa*II-5 fragment. In this study, we report that the variation in the viral *Hpa*II-3 fragments of five wild-type and mutant Py strains results from tandem sequence duplications located near sequences encoding the late RNA leader in a region important for early gene expression. Furthermore, the sequence duplications are located in a region containing Py sequences that enhance the expression of nonviral genes.

A representation of the Py genome drawn in relation to the *Hpa*II cleavage map is shown in Fig. 1A. Variation in different Py plaque isolates can be detected by analysis of the cleavage products generated by the *Hpa*II restriction enzyme (4, 11, 26, 31). This variation is illustrated in Fig. 1B, which shows the *Hpa*II or *Bgl*I and *Bcl*I cleavage products from different Py strains separated by polyacrylamide gel electrophoresis. The principle variation is seen in the presence of an extra *Hpa*II site in the *Hpa*II-1 fragment (P16 and Toronto large plaque [TOR]) and in the sizes of *Hpa*II-3 and *Hpa*II-5 fragments. Strains P16, TOR, Ts 48, and MV had larger *Hpa*II-3 fragments than the corresponding fragments from the A2 or A3 strain. The *Hpa*II-5 fragments of A2 and A3 have been shown to differ by 11 bp (25). In addition, Ts 48 lost the *Hpa*II cleavage site between fragments 7 and 8, giving rise to a fragment migrating just below the *Hpa*II-6 fragment, as previously reported (8). The smaller *Bcl*I-*Bgl*I fragment, containing sequences from *Hpa*II-5 and *Hpa*II-3 fragments, was larger in strain NG59R, as compared with

† Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

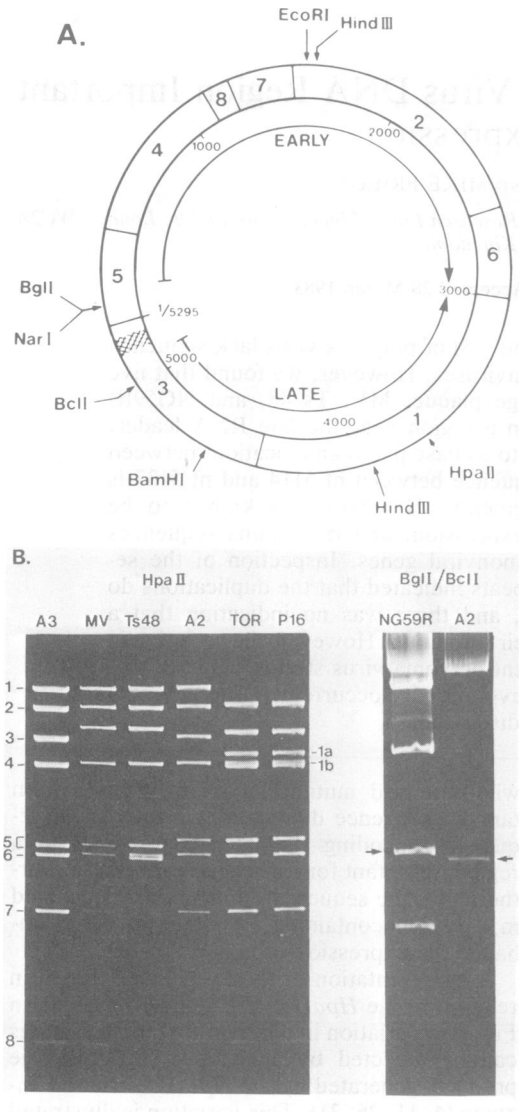

FIG. 1. (A) Physical map of the Py genome. A representation of the Py genome (A2 strain), showing the direction and extent of viral early- and late-region transcripts, is drawn in relation to the eight fragments generated by *Hpa*II cleavage. Other restriction enzyme cleavage sites, including the additional *Hpa*II site present in the TOR and P16 strains (see below). are indicated. The modified polyoma virus nucleotide numbering (29) of Soeda et al., (25) is indicated. The region in which tandem duplicated sequences occur is shown as cross-hatched. (B) Restriction enzyme cleavage analysis of polyoma virus DNAs. The A2 and A3 strains originated in this laboratory (11). The P16 and Ts 48 strains were obtained from W. Eckhart (7), whereas the MV and TOR strains were provided by M. Vogt (31) and A. Levine (23), respectively. Strain NG59R (26) was provided by T. Benjamin as a *Bam*HI-generated linear fragment cloned in plasmid DNA. Viral DNA was extracted from infected mouse embryo cells (15) and purified by ethidium bromide-cesium chloride equilibrium centrifugation as previously described (10). DNAs were cleaved with *Hpa*II or *Bgl*I and *Bcl*I as shown, and the cleavage products were separated on a 5% polyacrylamide gel. The eight *Hpa*II fragments in A2 and A3 (11) are indicated, as are the two fragments (1a and 1b) in TOR and P16 resulting from an additional *Hpa*II cleavage site in the *Hpa*II-1 fragment (see A above). The *Hpa*II-3 fragments of MV, Ts 48, TOR, and P16 contain additional sequences in comparison to the prototype A2 and A3 strains. For the differences in the *Hpa*II-5 fragment between A3, MV, Ts 48 and A2, TOR, P16, and for the loss of the *Hpa*II site in Ts 48 between fragments *Hpa*II-7 and *Hpa*II-8, see the text. The *Bcl*I-*Bgl*I fragment from NG59R pBR322 plasmid DNA containing additional sequences in the *Hpa*II-3 fragment and the corresponding *Bcl*I-*Bgl*I fragment from A2 viral DNA are indicated with arrows.

the prototype strain, A2. That this difference is contained in the *Hpa*II-3 fragment has been demonstrated for viral DNA (26).

The variation in the *Hpa*II-3 fragment suggests that many Py strains contain additional DNA sequences compared with the previously sequenced A2 or A3 strain. To understand the exact nature of this variation, strains with larger *Hpa*II-3 fragments were sequenced from the single viral *Bcl*I site located at 65.4 map units (nucleotide [nt] nt 5021) (Fig. 1). The sequences determined for the MV, TOR, P16, Ts 48, and NG59R strains are shown in Fig. 2. In all five strains, the increased size of fragment 3 was the result of tandem sequence duplications. The duplications varied in size (Fig. 2), ranging from 31 bp (NG59R) to 84 bp (MV), and in location, extending from nt 5068 to nt 5185. However, the region from nt 5114 to nt 5137 was contained in all five duplicated segments. The duplicated sequence in MV, NG59R, and P16 was exact, whereas in Ts 48 and TOR, the two repeats differed at a single nucleotide. Only strains P16 and TOR had the same duplicated sequence; however, the base change (at nt 5113) in the second copy of the TOR repeat introduced an *Eco*RI site at 67 map units in the viral genome. Analysis of the sequences at and around the duplicated segments (Fig. 3) shows no evidence of sequence homology or common sequence or structural features.

Five wild-type and mutant Py strains were found to contain tandemly duplicated sequences in a region encoding the late RNA leader. These duplications account for much of the size variation of restriction fragments derived from the noncoding region of different Py strains. Overall, the duplications varied in size as mentioned above. However, the region from nt 5114 to nt 5137 was contained in all five duplicated segments (Fig. 2). Two other Py strains (Ts 251, ev

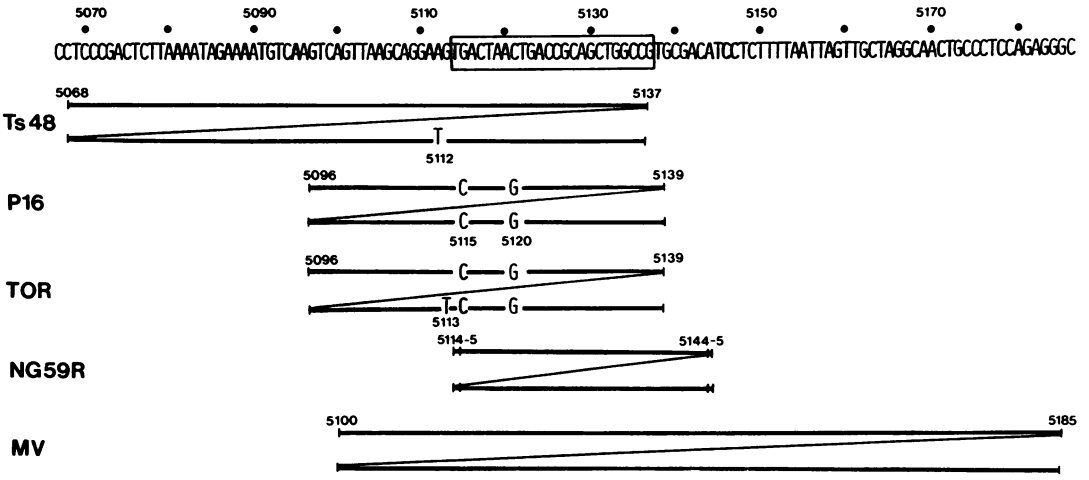

FIG. 2. Tandem repeats in polyoma strains. Before DNA sequence analysis, MV, P16, and Ts 48 viral DNAs were cleaved with *Hin*dIII, and the 3-kb fragment containing the additional sequences was cloned into the *Hin*dIII site of plasmid pAT153. NG59R DNA was cloned into pBR322 as a *Bam*HI-generated linear fragment. Plasmid DNAs were isolated from *dam Escherichia coli* (19) to allow cleavage by the endonuclease *Bcl*I. The TOR strain was sequenced by using viral DNA. DNAs were cleaved at the *Bcl*I site and 5′ end-labeled by treatment with calf intestine alkaline phosphatase and T4 polynucleotide kinase (32). After secondary cleavage with *Bgl*I or *Nar*I (Fig. 1), singly end-labeled fragments were separated by polyacrylamide gel electrophoresis, eluted, and sequenced by the chemical degradation method of Maxam and Gilbert (20). The DNA sequence of the A2 strain (29) is shown relative to the sequence repeats in the MV, Ts 48, P16, TOR, and NG59R strains, which are indicated by parallel lines connected to form a "Z." Additional nucleotide changes are also found in one or both of the repeats as indicated. The region of the sequences (nt 5114 to nt 5137) contained in all five repeats is boxed. There is a one-base ambiguity where the repeat in NG59R starts (nt 5114 or nt 5115) and ends (nt 5144 or nt 5145). The sequence beyond nt 5150 is that of strain MV only. In the TOR and the related P16 strains, there are other base changes found in the sequence outside the repeated region (data not shown) which are not observed in the related A2, A3, MV, Ts 48, and NG59R strains.

1001) (10, 18) which have large tandem repeats of about 250 bp (encompassing the Py origin of DNA replication) also contain the region from nt 5114 to nt 5137 in each of their duplicated segments (18; J. Arrand, personal communications). The sequence from nt 5114 to nt 5137 is also present in the different sequence repeats (containing the origin of DNA replication) found in different defective Py DNAs (10).

Although the biological significance of these duplications is currently unknown, the sequence around the *Pvu*II site at nt 5128 has been shown to be important for Py early gene expression (29). Whereas Py deletion mutants lacking a sequence on either side of the *Pvu*II site (nt 5128) are viable, mutants lacking a sequence on both sides of the *Pvu*II site are either nonviable or are greatly impaired in their plaque-forming ability (29). It is interesting that the sequence common to the repeats (nt 5114 to nt 5137) also occurs on both sides of the *Pvu*II site at nt 5128. In addition, the repeats are contained within a Py DNA region (nt 5021 to nt 5268) which has been shown to enhance the expression of distant genes (6). Thus, the Py duplications may be similar to the region containing the SV40 72-bp repeat, which is known to be required for efficient SV40 early gene expression (2, 14) and which is able to enhance the expression of distant genes (1, 21).

The mechanism whereby short DNA sequences become repeated in tandem is not known. Comparison of the sequences flanking the duplicated segments rules out the possibility that the repeats arise by homologous recombination. Moreover, inspection of the sequences at and around the repeats fails to reveal common sequences or structural features to suggest that a sequence-specific mechanism resulted in their formation. The fact that the boundaries of the repeats vary in different Py strains may well argue against a sequence-specific mechanism leading to their formation.

It should be noted that the duplication in Ts 48 was not present either in the parental virus stock or in six sister *ts* mutants (*ts*3, *ts*52, *ts*59, *ts*609, *ts*616, *ts*441), as judged by the sizes of their *Hpa*II-3 fragments (data not shown). This raises the possibility that the duplication in Ts 48 arises during outgrowth of the mutant virus from the mutagenized parental stock. Ts 48 contains at least two mutations in addition to being tempera-

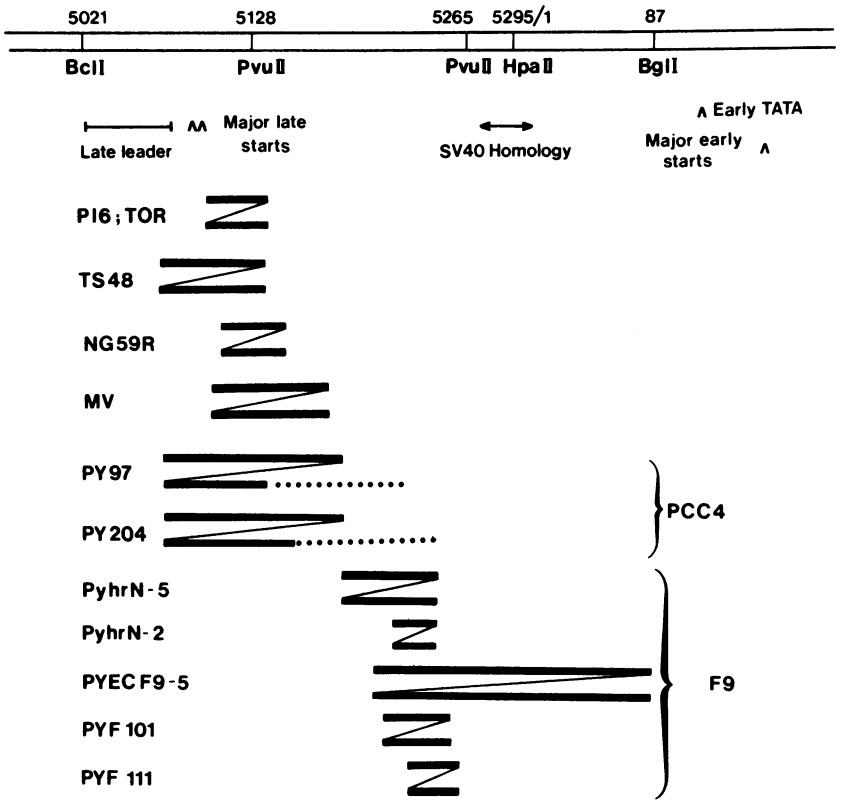

FIG. 3. Sequence duplications in Py strains. The repeats in the wild-type strains reported in this study, as well as those found in Py mutants able to grow on undifferentiated F9 and PCC4 teratocarcinoma cells (13, 16, 17, 23), are drawn in relation to the noncoding region of the Py virus. The region encoding the late RNA leader (27), the major late and early transcriptional starts (3, 28), early TATA box (28), and the region of greatest homology to SV40 (12, 25) are indicated. Sequences repeated in tandem are indicated by bold parallel lines connected to form a "Z" (Fig. 2). In the PCC4 mutants, deletions which appear to have removed a portion of the second copy of the repeat and sequences immediately adjacent to the repeat are represented by a dotted line.

ture sensitive in the large T antigen (8). The first is a base change which abolishes the *Hpa*II cleavage site between fragments 7 and 8, resulting in single amino acid changes in the middle and large-T antigens. The second mutation abolishes the termination codon for VP2 and VP3, resulting in the addition of 21 amino acids to the end of each capsid protein. It is interesting to speculate that the lesion(s) in Ts 48 may exert a selective pressure favoring the outgrowth of a strain containing the duplication. Such might be the case if the duplication results in an increased synthesis of poorly functioning mutant proteins.

The duplications reported in this study recall the tandem repeats found in Py host range mutants which are able to grow in undifferentiated teratocarcinoma cells (13, 16, 17, 23). The duplications in mutants selected to grow on F9 and PCC4 teratocarcinoma cells varied in extent and location with common sequences contained in the repeats of each type (Fig. 3). We speculate that the sequence arrangement in the PCC4 mutants arose from a tandem duplication followed by the deletion of part of the second copy which extended into the adjacent viral DNA (Fig. 3), rather than from a transposition or deletion as originally suggested (17). Seen as such, portions of the region contained within the PCC4 duplications are also present in the repeats reported in this study, whereas the region repeated among the F9 mutants lies closer to the early region genes (Fig. 3). The duplications in the PCC4 and F9 mutants arose after selection for growth on undifferentiated teratocarcinoma cells. Likewise, the finding of sequence duplications in many wild-type strains of Py virus suggests a biological advantage for having two copies since the second copy can be easily lost by homologous recombination.

The function of tandem sequence repeats as a structural feature is not understood; thus, it is possible that changes in the spacing of different

sequences is the important consequence of some duplications as opposed to having two copies of a particular sequence. In either event, the formation of a repeated sequence appears to be a dynamic process resulting in a high degree of structural variability. This variability is seen in other papovaviruses as well. SV40 strains lacking the 72-bp repeat (30) or containing a 91-bp repeat (22) have been reported. In addition, the structures of the repeats (present as partial sequence triplications) found in the two sequenced strains of BKV differ greatly (24, 33). It is remarkable that the arrangement of sequences important in viral gene expression would show such extreme variability. This may reflect the fluctuating nature of the virus-host relationship in natural populations and adaptation of the virus to grow on differentiated cells of different tissue types.

We thank Brendan Davies and Moira Griffiths for their excellent technical assistance and Mary Ann Hollis and Gina Yiangou for help in the preparation of this manuscript.

LITERATURE CITED

1. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
2. **Benoist, C., and P. Chambon.** 1981. In vivo sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
3. **Cowie, A., C. Tyndall, and R. Kamen.** 1981. Sequences of the capped 5′-ends of polyoma virus late region mRNAs: an example of extreme terminal heterogeneity. Nucleic Acids Res. **9**:6305–6322.
4. **Crawford, L. B., A. K. Robbins, P. M. Nicklin, and K. Osborn.** 1974. Polyoma DNA replication: location of the origin in different virus strains. Cold Spring Harbor Symp. Quant. Biol. **39**:219–225.
5. **Deininger, P., A. Esty, P. LaPorte, and T. Friedmann.** 1979. Nucleotide sequence and genetic organisation of the polyoma late region: features common to the polyoma early region and SV40. Cell **18**:771–779.
6. **de Villiers, J., and W. Schaffner.** 1981. A small segment of polyoma virus DNA enhances the expression of a cloned β-globin gene over a distance of 1400 base pairs. Nucleic Acids Res. **9**:6251–6262.
7. **Eckhart, W.** 1974. Properties of temperature-sensitive mutants of polyoma virus. Cold Spring Harbor Symp. Quant. Biol. **39**:37–40.
8. **Eckhart, W., S. Delbruck, P. Deininger, T. Friedman, and T. Hunter.** 1981. A mutation increasing the size of the polyoma virion-proteins, VP2 and VP3. Virology **109**:35–46.
9. **Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. Vande Voorde, H. Van Heuverswyn, J. Van Herreweghe, G. Volckaert, and M. Ysebaert.** 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) **273**:113–120.
10. **Fried, M., and B. E. Griffin.** 1977. Organisation of the genomes of polyoma and SV40. Adv. Cancer Res. **24**:67–113.
11. **Fried, M., B. E. Griffin, E. Lund, and D. L. Robberson.** 1974. Polyoma virus—a study of wild-type, mutant and defective DNAs. Cold Spring Harbor Symp. Quant. Biol. **39**:45–52.
12. **Friedmann, T., A. Esty, P. LaPorte, and P. Deininger.** 1979. Nucleotide sequence and genetic organisation of the polyoma late region: features common to the polyoma early region and SV40. Cell **18**:715–724.
13. **Fujimura, F. K., P. L. Deininger, T. Friedmann, and E. Linney.** 1981. Mutation near the polyoma DNA replication origin permits productive infection of F9 embryonal carcinoma cells. Cell **23**:809–814.
14. **Gruss, P., R. Dhar, and G. Khoury.** 1981. Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. U.S.A. **78**:943–947.
15. **Hirt, B.** 1967. Selective extraction in polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
16. **Katinka, M., M. Vasseur, N. Montreau, M. Yaniv, and D. Blangy.** 1981. Polyoma DNA sequences involved in control of viral gene expression in murine embryonal carcinoma cells. Nature (London) **290**:720–722.
17. **Katinka, M., and M. Yaniv.** 1980. Expression of polyoma early functions in mouse embryonal carcinoma cells depends on sequence rearrangements in the beginning of the late region. Cell **20**:393–399.
18. **Magnusson, G., and M.-G. Nillson.** 1982. Viable polyoma virus variant with two origins of DNA replication. Virology **119**:12–21.
19. **Marinus, M. G.** 1980. Influence of *uvrD3*, *uvrE502*, and *recL152* mutations on the phenotypes of *Escherichia coli* K-12 *dam* mutants. J. Bacteriol **141**:223–226.
20. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
21. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, and P. Chambon.** 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. **9**:6047–6068.
22. **Reddy, V. B., B. Thimmapaya, R. Dahr, K. N. Subramanian, B. S. Zain, J. Pan, P. K. Ghosh, M. I. Celma, and S. M. Weissmann.** 1978. The genome of simian virus 40. Science **200**:494–502.
23. **Sekikawa, K., and A. J. Levine.** 1981. Isolation and characterisation of polyoma host range mutants that replicate in nullipotential embryonal carcinoma cells. Proc. Natl. Acad. Sci. U.S.A. **78**:1100–1104.
24. **Sief, I., G. Khoury, and R. Dhar.** 1979. The genome of human papovavirus BKV. Cell **18**:963–977.
25. **Soeda, E., J. R. Arrand, N. Smolar, J. E. Walsh, and B. E. Griffin.** 1980. Coding potential and regulatory signals of the polyoma virus genome. Nature (London) **283**:445–453.
26. **Staneloni, R. J., M. M. Fluck, and T. L. Benjamin.** 1977. Host range selection of transformation-defective hr-t mutants of polyoma virus. Virology **77**:598–609.
27. **Treisman, R.** 1980. Characterisation of polyoma late mRNA leader sequences by molecular cloning and DNA sequence analysis. Nucleic Acids Res. **8**:4867–4888.
28. **Treisman, R., A. Cowie, J. Favaloro, P. Jat, and R. Kamen.** 1981. The structures of the spliced mRNAs encoding polyoma virus early region proteins. J. Mol. Appl. Genet. **1**:83–92.
29. **Tyndall, C., G. La Mantia, C. Thacker, J. Favaloro, and R. Kamen.** 1981. A region of the polyoma virus genome between the replication origin and late protein coding sequences is required in *cis* for both early gene expression and viral DNA replication. Nucleic Acids Res. **9**:6231–6250.
30. **van Heuverswyn, H., and W. Fiers.** 1979. Nucleotide sequence of the Hind-C fragment of simian virus 40 DNA. Eur. J. Biochem. **100**:51–60.
31. **Vogt, M., L. T. Bacheler, and L. Boice.** 1976. Proposed structure of two defective viral DNA oligomers produced in 3T3 cells transformed by the ts-a mutant of polyoma virus. J. Virol. **17**:1009–1026.
32. **Weaver, R. F., and C. Weissman.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15 S β-globin mRNA precursor and mature 10 S β-globin mRNA have identical map coordinates. Nucleic Acids Res. **7**:1175–1193.
33. **Yang, R. C. A., and R. Wu** 1979. BK virus DNA: complete nucleotide sequence of human tumour virus. Science **206**:456–462.