JOURNAL OF VIROLOGY, Oct. 1982, p. 67–77
0022-538X/82/100067-11$02.00/0
Copyright © 1982, American Society for Microbiology
Vol. 44, No. 1

Structural and Biological Analysis of Integrated Polyoma Virus DNA and Its Adjacent Host Sequences Cloned from Transformed Rat Cells

ADRIAN HAYDAY,† H. EARL RULEY, AND MIKE FRIED*

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

Received 5 February 1982/Accepted 22 June 1982

*Eco*RI fragments containing integrated viral and adjacent host sequences were cloned from two polyoma virus-transformed cell lines (7axT and 7axB) which each contain a single insert of polyoma virus DNA. Cloned DNA fragments which contained a complete coding capacity for the polyoma virus middle and small T-antigens were capable of transforming rat cells in vitro. Analysis of the flanking sequences indicated that rat DNA had been reorganized or deleted at the sites of polyoma virus integration, but none of the hallmarks of retroviral integration, such as the duplication of host DNA, were apparent. There was no obvious similarity of DNA sequences in the four virus-host joins. In one case the virus-host junction sequence predicted the virus-host fusion protein which was detected in the transformed cell line. DNA homologous to the flanking sequences of three out of four of the joins was present in single copy in untransformed cells. One copy of the flanking host sequences existed in an unaltered form in the two transformed cell lines, indicating that a haploid copy of the viral transforming sequences is sufficient to maintain transformation. The flanking sequences from one cell line were further used as a probe to isolate a target site (unoccupied site) for polyoma virus integration from uninfected cellular DNA. The restriction map of this DNA was in agreement with that of the flanking sequences, but the sequence of the unoccupied site indicated that viral integration did not involve a simple recombination event between viral and cellular sequences. Instead, sequence rearrangements or alterations occurred immediately adjacent to the viral insert, possibly as a consequence of the integration of viral DNA.

Polyoma virus (Py) can transform a variety of cell types to a neoplastic state. In virus-transformed cells, Py sequences are retained and expressed as part of the host cell genome (reviewed in reference 34). The study of integrated viral sequences has been complementary to other approaches in understanding the interaction of the virus with the host cell. These studies, which were largely dependent on Southern blot analysis, indicated that the retention of the coding capacities for small and middle T-antigens (T-Ags) is sufficient to maintain cell transformation (17, 19, 21, 22, 27). Retention of the large T coding capacity is not necessary to maintain transformation (20), although the large T does play a role in initiating transformation after viral infection (10–12). In transformed cells Py sequences have been observed to become integrated at different chromosomal locations (19, 22, 23). The questions which defied previous analyses were whether the chromosomal locations of integrated viral sequences fell into a subset of the host genome (e.g., highly reiterated DNA); whether the adjacent cellular DNA modified the biological activity of the viral DNA; whether there was a sequence specificity to integration manifested by similar sequences at the virus-host joins; whether homology between viral and host DNA mediated integration; and whether the integration of viral sequences perturbed (duplicated/deleted/inverted) the host cell sequences at the site of integration. These questions require cloning of the integrated viral DNA from transformed cells and of the cellular DNA at the site of viral insertion from uninfected cells. This approach has been followed here.

† Present address: Center for Cancer Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.

MATERIALS AND METHODS

Cell culture and biological assays. The isolation of cell lines 7axT and 7axB has previously been described (22, 23). Cells were grown in Dulbecco modi-

fied Eagle medium containing 5% fetal calf serum. Whole mouse embryo cells were used for plaque assays as previously described (20). For transformation assays, 0.5 μg of Py-pAT153 recombinant plasmid DNAs in the presence of 20 μg of rat carrier DNA were applied to semiconfluent Rat-1 cells (about 10^6 per dish) in calcium phosphate precipitates as previously described (13, 39). Cell sheets were stained for transformed foci 2 weeks later with Leishman stain.

Preparation, restriction, and fractionation of eucaryotic cell DNA. Total cell DNA was prepared essentially as previously described (20, 21, 23). For Southern (31) analysis of cell DNA, endonuclease digestions were performed on about 10 μg of cell DNA for 3 to 6 h, depending on the enzyme used. The enzymes were purchased from New England Biolabs, and the prescribed buffers were used. Preparative fractionation of endonuclease-cleaved cell DNA was on 10 to 40% sucrose velocity gradients, as described by Maniatis et al. (25). Fractions were assayed for the presence of Py sequences by Southern blot analysis.

Gel electrophoresis, transfer, and hybridization procedures. DNA was electrophoresed in agarose gels of 0.6 to 2% with E buffer (18). The DNA was denatured in situ, transferred to nitrocellulose filters (Schleicher and Schuell, BA85), and hybridized to virion-purified [^{32}P]DNA nick-translated to a specific activity of 1×10^8 to 2×10^8 cpm/μg (20, 21, 23). Blots were autoradiographed on presensitized Fuji Rx film at −70°C with intensity screens (Dupont).

Cloning eucaryotic DNA into procaryotic vectors. Fractions of *Eco*RI-cleaved 7axB and 7axT cell DNA which contained Py sequences were prepared as described above and ligated to the purified arms of λgt.WES.λb. In the standard reaction, between 0.15 and 0.25 μg of cell DNA was ligated to 1.5 μg of λ arms. T4 DNA ligase was supplied by New England Biolabs. Packaging of λ DNA in vitro was performed as previously described (15); between 2×10^5 and 1×10^6 plaques per μg of λ arms were obtained. The phage were plated on *Escherichia coli* LE392 (33) and screened by the method of Benton and Davis (3).

TABLE 1. Summary of Southern blot data derived from analysis of 7axB cell DNA with the enzymes indicated and a probe of virion-purified Py DNA

Enzyme	Py fragments present in digest of 7axB DNA	Putative virus-host linker fragments (kb)
*Bgl*II		16
*Eco*RI	5.3 kb, linear	4.9, 3.65
*Bam*HI	5.3 kb, linear	10, 2.2
*Xba*I	5.3 kb, linear[a]	7.5, 3.7
*Bam*HI + *Eco*RI	3.1 kb, 2.2 kb	3.65, 1.9
*Bam*HI + *Xba*I	3.2 kb, 2.1 kb	3.7, 2.2
*Bgl*II + *Eco*RI	5.3 kb, linear	4.9, 3
*Bgl*II + *Xba*I	5.3 kb, linear	7.8, 3.7
*Pvu*II	*Pvu*II fragments 1–4	7.5, 3.9
*Pst*I	*Pst*I fragments 1–4	4.6, 3.5
*Msp*I	*Msp*I fragments 1–8	>6

[a] *Xba*I cuts Py DNA twice, but since one of the fragments is 45 bp long, *Xba*I is regarded as a single-cut enzyme when used to analyze genomic blots.

Positive plaques were picked and purified, and their DNA was isolated by the method of Maniatis et al. (25). Cloned Py DNA fragments were ligated into *Eco*RI-cleaved pAT153 (36) DNA and introduced into *E. coli* HB101.

Preparation of plasmid DNA. The method of Birnboim and Doly (4) was used to obtain yields of recombinant plasmid DNA of >1 mg/liter of culture. After CsCl density centrifugation, plasmid DNAs were further purified on sucrose gradients (5 to 20%) containing ethidium bromide.

DNA sequence analysis. Restriction endonuclease cleavage sites were 5′-end labeled after treatment with calf intestine alkaline phosphatase and T4 polynucleotide kinase (38). After secondary enzyme cleavage, labeled fragments were fractionated on polyacrylamide gels and eluted (26). The chemical degradation method of Maxam and Gilbert was employed as described (26), except the G + A reaction was stopped with 200 μl of hydrazine stop buffer and ethanol precipitated. The degradation products were resolved on thin urea-acrylamide gels as described by Sanger and Coulson (28). DNA sequences were analyzed with the SEQ computer program (6).

RESULTS

State of the viral DNA in cell lines 7axB and 7axT. Cell lines 7axB and 7axT were isolated from independent monoclonal tumors induced in rats by the injection of Py-transformed cells of the line 7ax (23). Although 7ax contained multiple copies of both free and integrated Py DNA, both 7axB and 7axT contained only a single insert of Py DNA and lacked free viral forms (23). The structure of the integrated viral sequences in the 7axT cells as determined by Southern blot analysis has previously been reported to comprise a single insert of continuous viral sequences (0.8 to 0.85 genomes in length) extending from 26 to 32 map units (MU) at one end to between 12 and 14 MU at the other end (22). A preliminary description of the 7axB insert indicated the presence of head-to-tail tandemly duplicated Py sequences (23). By an analysis of DNA fragments from a combination of cleavages with enzymes which cut Py DNA once or several times and enzymes which do not cut Py DNA which used the strategy employed by Lania et al. (21) for mapping tandem head-to-tail integrated viral sequences, the structure of the 7axB insert was derived. Briefly, enzymes which cleave the Py genome once (*Eco*RI, *Bam*HI) generated a 5.3-kilobase (kb) linear Py DNA fragment. A 5.3-kb fragment was also generated by the cleavage of 7axB DNA with *Xba*I. Cleavage of 7axB DNA with enzymes that cut Py DNA at multiple sites (*Pst*I, *Pvu*II, *Hpa*II) generated all the fragments expected from the cleavage of the circular Py genome. From this data the 7axB insert was judged to contain a head-to-tail tandem duplication extending from 59 to 70 MU at the left end to 79 to

90 MU at the right end. The results of some of these digestions have been previously presented (23) and are listed along with additional cleavage data in Table 1. The extent of viral sequences at each end was judged as follows. At the left-hand end, the *Eco*RI virus-host linker fragment (3.65 kb) hybridized to a purified probe of Py *Hpa*II-3 (54 to 70 MU) but did not contain the Py *Bam*HI site (59 MU). At the other end, the *Eco*RI virus-host linker fragment hybridized to a purified probe of Py *Hpa*II-4 (78 to 91 MU) but did not contain the Py *Ava*I site at 90 MU. The resulting maps of Py DNA in 7axT and 7axB cells are shown (Fig. 1). 7axT contains an intact coding region for small and middle T-Ags, whereas 7axB contains two complete coding regions for large, middle, and small T-Ags. The sequence arrangement of the two inserts was studied in more detail after molecular cloning had been performed.

Isolation and characterization of the 7axB and 7axT viral inserts and adjacent host sequences. The integrated viral sequences of 7axT and 7axB were contained in two and three *Eco*RI fragments, respectively (Fig. 1). Size-fractionated *Eco*RI-cleaved DNA from both cell lines was cloned into λgt. WES.λb as described above. The *Eco*RI fragments were then subcloned into the plasmid pAT153 and analyzed. The cloned 3.6-, 5.3-, and 4.9-kb *Eco*RI fragments from 7axB are referred to as 7B-L (left), 7B-M (middle), and 7B-R (right), respectively (Fig. 1). The 4.7- and 5.1-kb cloned *Eco*RI fragments of 7axT are referred to as 7T-L (left) and 7T-R (right) (Fig. 1). These cloned fragments were indistinguishable in electrophoretic mobility from the *Eco*RI fragments generated from total cellular DNA. In addition, analysis of the cloned fragments with a number of restriction enzymes both confirmed their derivation from the integrated viral DNA structures mapped in Fig. 1 and allowed better resolution of the virus-host join regions. In several cases putative virus-host linker fragments detected by Southern analysis of cellular DNA were verified as such by restriction cleavage of cloned DNA. For 7T-L, the extent of viral sequences at the left-hand end of the viral insert was 100 base pairs (bp) greater than had previously been determined by Southern blot analysis of cell DNA (22). This discrepancy resulted from the inability to cleave 7axT DNA at either the *Hha*I site at nucleotide (NT) 2,962 or the *Hpa*II site at NT 2,991. Since *Msp*I (37) cleaves both 7T-L and 7axT DNA at NT 2,991, the inability of *Hpa*II to cleave 7axT DNA was presumably due to methylation at that site.

Nature of the adjacent host sequences and cloning of an unoccupied site DNA from Rat-1 cellular DNA. The clones 7B-R, 7B-L, 7T-R, and 7T-L were used to probe untransformed Rat-1 cellular DNA to estimate the copy number of the adjacent host sequences in the cellular genome (Fig. 2). The host sequences in three of the clones were found to be represented only once in *Eco*RI-cleaved Rat-1 DNA (7B-L in a fragment of 6.2 kb; 7B-R, 2 kb; 7T-L, 0.75 kb). The host sequences in 7T-R were found in at least two *Eco*RI fragments of Rat-1 DNA of 5 and 1.3 kb, respectively. The clones containing flanking cellular DNA were also used to probe *Eco*RI- and *Xba*I-cleaved DNA from 7axB and 7axT cells. In each case the probes detected two classes of fragments: (i) the fragment(s) due to viral sequences in common (which represent all or part of the proviral inserts), and (ii) the fragment(s) detected by that probe in untransformed cell DNA. The presence of class ii fragments suggested that the integration of viral DNA is haploid in the respective transformed cell lines. The observation that probes from 7axB cells hybridized to different fragments than did probes from 7axT cells confirms that the Py DNA in these cell lines is integrated into different sites in the host genome. That the right-hand flanking sequences of either 7axT or 7axB hybridized to fragments in Rat-1 DNA distinct from those detected by the left-hand flanking sequences shows that the viral DNA has not merely been inserted into the host DNA, but that cellular sequences have been rearranged (e.g., deleted) at the site of integration.

The *Eco*RI fragment homologous to 7B-L was cloned from untransformed Rat-1 cells into λgt. WES.λb and subcloned into pAT153 (pAU-L). A restriction map for the unoccupied site, pAU-L, was determined by complete and partial digestion with more than a dozen restriction enzymes (Fig. 3). For about 1.3 kb from the left-hand *Eco*RI site, pAU-L had a restriction map identical to 7B-L. At about that point, the flanking sequences in 7B-L became joined to viral DNA.

DNA sequences of the virus-host joins and the unoccupied-site DNA. The DNA sequences around the virus-host joins of 7B-L, 7B-R, 7T-L, and 7T-R are shown in Fig. 4. Continuous viral sequences were detected in 7B-L up to NT 4,682, in 7B-R up to NT 916, in 7T-L up to NT 2,838, and in 7T-R up to NT 2,252. At each of these points there was an abrupt discontinuity in the viral sequences. This is a feature of other Py virus-host joins that have been studied (27a; H. E. Ruley, unpublished data). A closer analysis of the sequence of 7T-L revealed that the discontinuity in viral sequences could have been due to a virus-virus rather than a virus-host join. The eleven nucleotides which abut the join could be derived from Py DNA (NTs 2,943 to 2,932). This arrangement of the viral sequence could have resulted from deletion of viral DNA from a

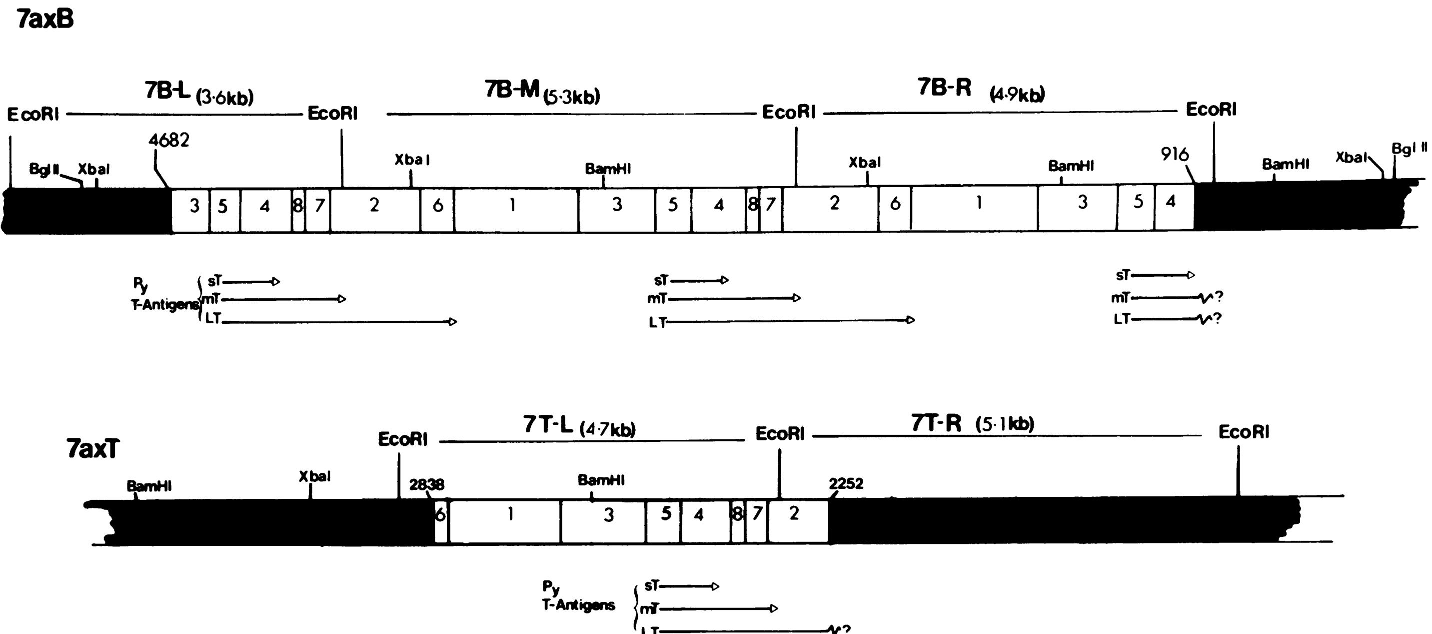

FIG. 1. Maps of the integrated viral sequences in the 7axB and 7axT cell lines. The maps were determined by Southern blot hybridization (see Table 1; references 22, 23), molecular cloning, and DNA sequencing (see the text). Flanking DNA sequences (shaded) and viral sequences are shown relative to the *Hpa*II/*Mps*I physical map (14). The nucleotide numbers (30) indicate the first discontinuities in the viral DNA. The coding region for the Py small T-Ag (sT), middle T-Ag (mT), and large T-Ag (LT) are indicated. The size of each cloned *Eco*RI fragment from 7axB (7B-L, 7B-M, 7B-R) and 7axT (7T-L, 7T-R) is shown.

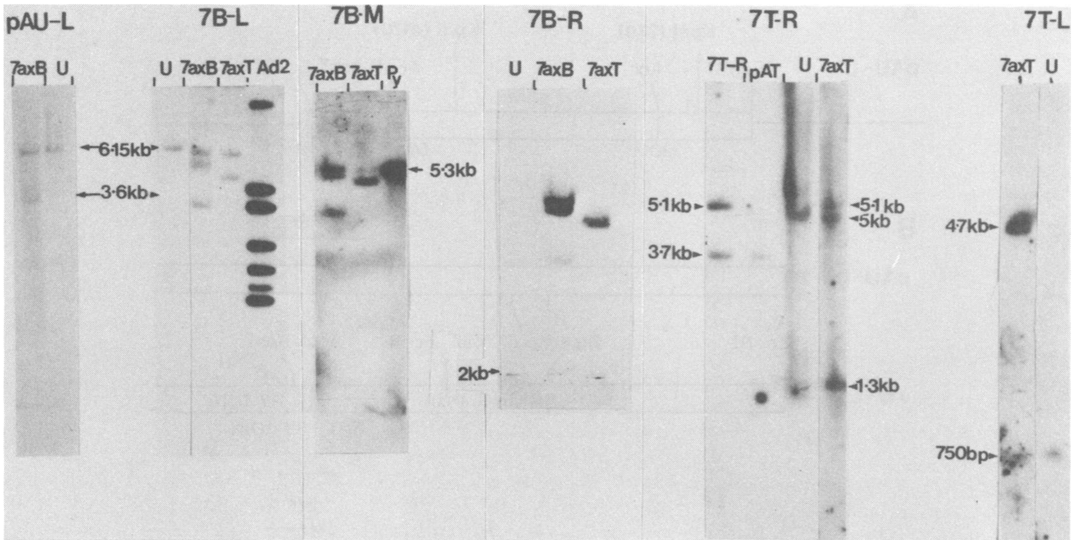

FIG. 2. Detection of flanking host DNA sequences in normal and rat cell DNA, indicating the haploid nature of the insertions of Py sequences and rearrangements at the site of integration in 7axT and 7axB cells. Total cellular DNA from 7axT, 7axB, and untransformed cells (U) was cleaved with *Eco*RI, separated on a 1% agarose gel, transferred to nitrocellulose paper, and probed with the pAT153 recombinant DNAs. The caption above each panel indicates the probe used, and the notation above each lane indicates the *Eco*RI-cleaved DNA probed (pAT, pAT153 DNA; Py, Py DNA). The bands detected with the 7B-M probe (contains only Py sequences) identified fragments of 7axT (4.7 and 5.1 kb) and 7axB (3.6, 5.3, and 4.9 kb) cellular DNAs containing viral sequences (see Fig. 1). Bands in addition to these fragments detected in untransformed, 7axB, and 7axT cell DNA identified host DNA sequences homologous to rat cellular sequences in pAU-L or the flanking sequences of 7B-L, 7B-R, 7T-L, or 7T-R. The 6.15-kb rat cell DNA fragment (sequences contained in the unoccupied site rat cellular DNA clone pAU-L) was detected in untransformed, 7axB, and 7axT cell DNAs when either pAU-L or 7B-L was used as probe. In 7axB DNA this host 6.15-kb fragment was detected in addition to the 3.6-kb 7B-L fragment with the pAU-L and 7B-L probes. When 7B-R was used as a probe, the 2-kb host DNA found in untransformed cell DNA was detected in 7axB DNA in addition to the 4.9-kb 7B-R fragment. When 7T-R was used as a probe, the 5- and 1.3-kb rat DNA fragments found in untransformed cell DNA were detected in 7axT DNA in addition to the 5.1-kb 7T-R fragment. When 7T-L was used as a probe, the 750 bp found in untransformed cell DNA was detected in 7axT DNA in addition to the 4.7-kb 7T-L fragment. Ad2, *Hin*dIII-cleaved adenovirus 2 DNA used as a marker.

tandem duplication of Py sequences. Because the probability that 11 out of 13 nucleotides around the virus-host join would occur by chance once on either strand of the Py genome is less than 1 in 300 (6), we favor the representation of the 7T-L virus-host join shown in Fig. 5.

When the flanking DNA sequences were compared to the displaced viral sequences (viral sequences adjacent to those at the join but no longer present), various amounts of patchy homology were observed (Fig. 4 and 5). Inspection of the 7T-R sequence revealed an open reading frame of 24 nucleotides (eight amino acids) continuous with the large T-Ag reading frame. This could encode a hybrid protein of about 76 kilodaltons (kd) (565 virus-encoded amino acids plus 8 flanking DNA-encoded amino acids [Fig. 6]). This is in good agreement with the 75-kd protein detected in 7axT cells (22). Inspection of 7B-R sequences predicted the formation of both large and middle T-Ag-related hybrid proteins. The sequence results indicated that there are 2 flanking DNA-coded amino acids in the middle T reading frame and 34 flanking DNA-coded amino acids in the large T reading frame (data not shown). In 7axB cells, only the normal-size large, middle, and small T-Ags are detected (23).

DNA sequences for pAU-L were determined in either direction from the *Acc*I site located 1.3 kb from the left-hand *Eco*RI site (Fig. 3). The sequence through the *Acc*I site was determined from an *Eco*RI linker introduced approximately 30 bp beyond the *Acc*I site in the flanking cellular DNA (Fig. 3). This was accomplished by controlled digestion with exonuclease *Bal*31 from the *Xba*I site located approximately 500 bp from the *Acc*I site and the insertion of an *Eco*RI linker. Approximately 180 bp of the sequence was determined to lie in the region where Py DNA recombined with Rat-1 DNA in pAU-L. Part of the sequence was compared to that of the homologous region of 7B-L and also to that of Py DNA (NTs 4,696 to 4,630) (Fig. 7). In the case of a simple recombination event between

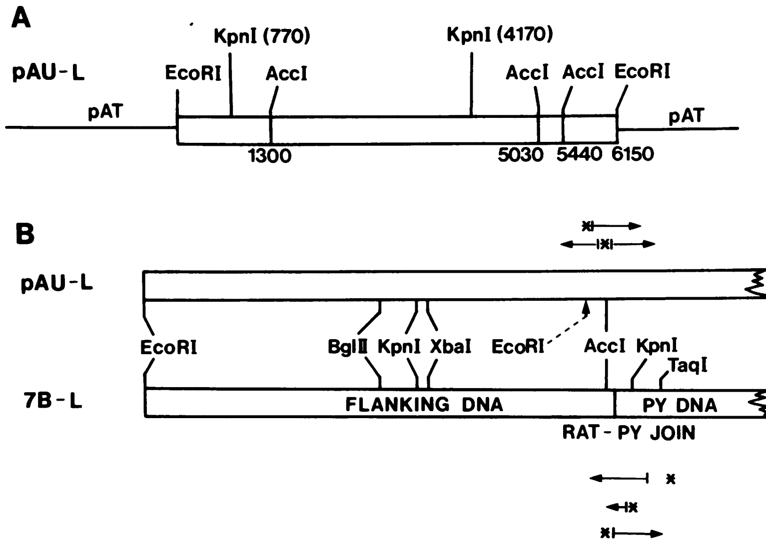

FIG. 3. (A) Restriction endonuclease map of pAU-L. The *Eco*RI fragment containing the cellular sequences flanking the left end of the 7axB viral insert (see Fig. 1) is drawn in relation to the *Acc*I restriction endonuclease cleavage sites. (B) Comparison of pAU-L and 7B-L DNAs and strategy for DNA sequencing. For over 1.3 kb from the leftward *Eco*RI site, the map of pAU-L was the same as 7B-L. A comparison of this region in the two DNAs is shown until the site at which 7B-L becomes joined to Py DNA. Restriction endonuclease cleavage sites were 5'-end-labeled, and the arrows refer to the direction and extent of the sequence determined from each labeled site. The sequence through the *Acc*I site in pAU-L was determined from an *Eco*RI linker inserted within approximately 30 bp.

viral or cellular DNAs, the transition from viral to non-viral sequence in 7B-L would be expected to coincide with the beginning of homology between the flanking DNA sequence in 7B-L and pAU-L. Instead, homology did not begin until 37 nucleotides beyond the last nucleotide (NT 4,682) colinear with the viral insert, after which perfect homology between the two sequences began and continued for more than 100 bp (the extent to which both sequences were determined). The origin of the 37 bp immediately adjacent to the viral insert ending at NT 4,682 is not known. The patchy homology noted (Fig. 7) may be significant, in which case the first 16 nucleotides beyond the join at NT 4,682 may be largely virus derived; likewise, some of the remaining sequence may be host derived. However, there still remain some 7 to 11 nucleotides which appeared to be derived from neither Py or pAU-L sequences.

The biological activity of the sequences cloned from 7axB and 7axT cells. The presence and expression of integrated viral sequences in Py-transformed cells is assumed to be responsible for cell transformation (34). Cloning the fragments containing integrated viral DNA allows their biological activity to be tested by transfection of Rat-1 cells (see above). Clones containing the complete coding regions for the Py middle and small T-Ags (7B-L, 7B-M, 7T-L) were capable of transforming rat cells with efficiencies equal to that of a complete Py genome cloned via the *Eco*RI site into pAT153 DNA (Table 2). This result provided assurance that viral sequences were faithfully maintained during their isolation from cell DNA. Clones 7B-R and 7T-R did not transform cells. Although 7B-R contained the complete coding region for Py small T-Ag, it may be unable to express this protein because the viral polyadenylation signals for this mRNA have been removed by fusion to host sequences. However, within 7B-R the major Py early polyadenylation signal occurs 5' to the small T coding region. By recloning the Py insert through the *Bam*HI site in pAT153 DNA it was possible to bring this polyadenylation signal 3' to the small T coding region. This reoriented plasmid did not transform rat cells.

The *Eco*RI fragment of 7B-M (the full-size Py genome cloned from the tandem Py DNA insert in 7axB cells) was recircularized with T4 DNA ligase and tested for its ability to produce plaques on mouse cells. Transfection with 1 μg of this fragment produced no plaques. This compares with the 10^5 plaques per μg formed by Py DNA after cleavage and recircularization of the Py genome cloned in the *Eco*RI site of pAT153. It was subsequently found that 7B-M was defective for DNA replication because of a single base change in the large T coding region (A. C. Hayday, F. Chaudry, and M. Fried, submitted for publication). This lesion was tan-

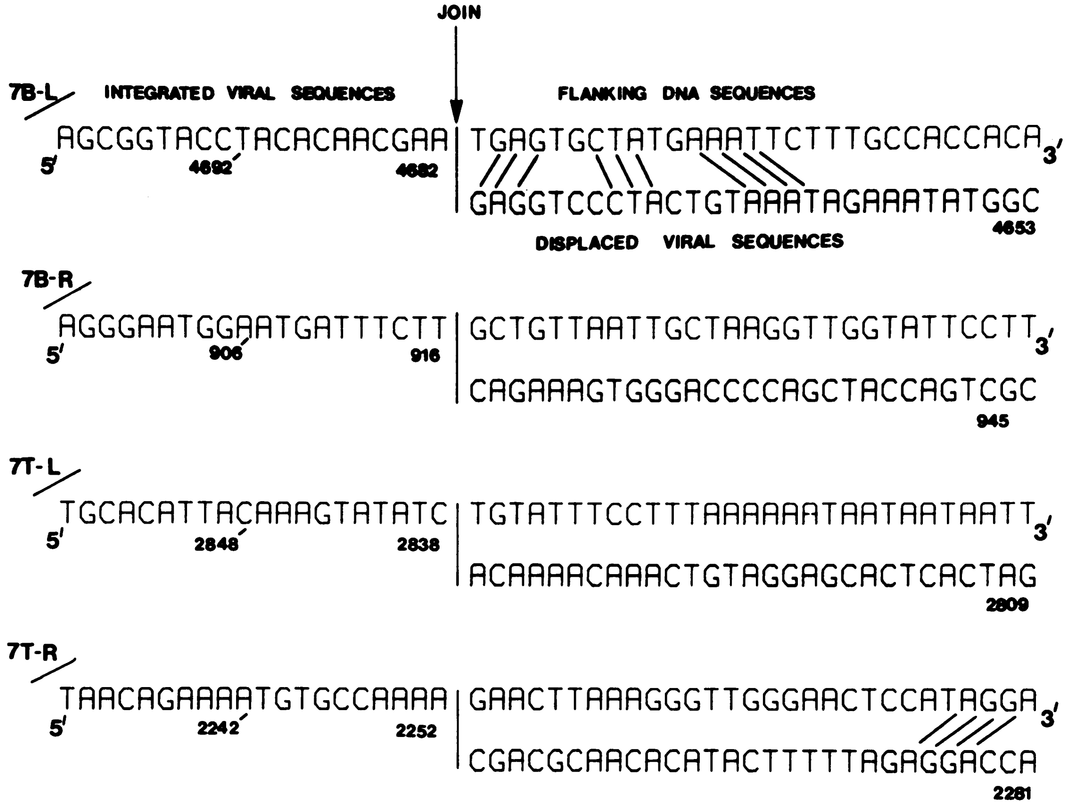

FIG. 4. Sequences of virus-host joins in 7B-L, 7B-R, 7T-L, and 7T-R DNAs. The sequences flanking the integrated viral DNAs are shown above the displaced viral sequences (i.e., viral sequences which are no longer present as a result of recombination between viral and cellular DNA). Sloping lines indicate homologies of ≥3 bp (within a distance of 4 bp) between the flanking DNA sequences and displaced viral sequences. The nucleotide numbers of the Py genome are indicated (30). The sequences of 7B-L and 7B-R were determined from analysis of both strands, and of 7T-L and 7T-R from overlapping fragments on the same strand.

demly repeated in the Py DNA in 7axB, i.e., it also occurred in the large T coding region of 7B-R.

DISCUSSION

In this report we describe experiments aimed at understanding the recombination events between viral and host sequences which accompany cell transformation by Py. The integrated viral and adjacent host DNA sequences from two Py-transformed cell lines were cloned into procaryotic vectors. The structures of the cloned DNA sequences were apparently unaltered by this process. The flanking host DNA sequences at three of the four virus-host joins were found to exist in a single copy in untransformed rat cells. An *Eco*RI fragment containing one of these unoccupied sites was also cloned. The primary DNA sequence was determined for this clone and for each of the virus-host joins. In

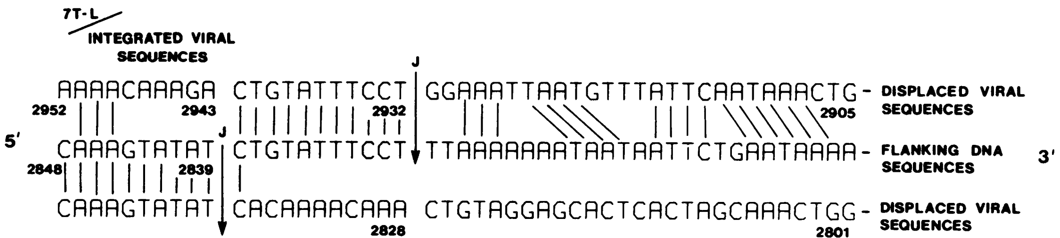

FIG. 5. Alternative representation of the virus-host join in clone 7T-L. In this representation (see the text), a virus-virus join (J), NTs 2,838 to 2,842, precedes the virus-host join (J) at NT 2,932. Vertical and sloping lines indicate homologies between Py and 7T-L DNA.

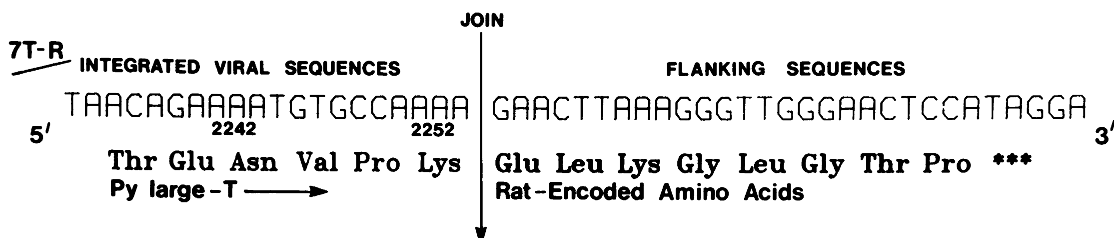

FIG. 6. A large T-Ag-related virus-host fusion protein predicted by the DNA sequence across the Py-host join in clone 7T-R. The virus-host join in 7T-R interrupts the coding region for the viral large T-Ag. The cellular sequences remain open in the large T-Ag reading frame for an additional eight amino acids, as shown.

addition, the clones derived from the transformed cells were assayed for their biological activity.

Expression of integrated viral sequences. Cells transformed by Py express both the middle and small T-Ags but not necessarily the full-size Py large T-Ag (19, 21, 23, 27). Recently it has been shown that expression of only the middle T-Ag is sufficient for most of the properties of the Py-transformed phenotype (35). Both 7axT and 7axB cells contain at least one unaltered Py coding region for both the small and middle T-Ags, and both of these cell lines produce these proteins (23). The cloned fragments 7B-M, 7B-L, and 7T-L, which each encode both small and middle T-Ags, transformed rat cells efficiently. Neither the 7B-R clone, which contains the complete coding region for small T-Ag, nor a modified clone with the major viral polyadenylation signal brought 5′ to the small T-Ag coding region transformed rat cells. That clones 7B-R and 7T-R did not transform cells indicates that the host sequences which are adjacent to the viral inserts in these cells are non-transforming.

7axT cells do not contain sufficient viral sequences to encode Py large T-ag. Rather, the sequence data for 7T-R suggests that a large T-related protein of 76 kd could be expressed. An anti-T immunoreactive protein of 75 kd has been detected in 7axT cells (22). The 100-kd anti-T immunoreactive protein detected in 7axB cells also could not support viral DNA replication due to a mutation in the coding region for the carboxy-terminal portion of the protein (Hayday et al., submitted for publication). The absence of functional Py large T-Ag in these cell lines correlates with the stability of the inserts of viral sequences both upon further passage in vivo and upon selection for morphological revertants of transformation (2, 23; A. C. Hayday, unpublished data). The DNA sequence of 7B-R predicts large and middle T-Ag-related virus-host fusion proteins. No such proteins are detectable in 7axB cells (23). If such proteins are produced, they may be very unstable, non-immunoreactive, or both. Alternatively, their synthesis may be blocked at the level of transcription or processing. By using the cloned proviral fragments as probes, it was shown that one copy of the flanking host sequences exists unaltered in both 7axB and 7axT cells (Fig. 2). Thus, a haploid copy of the viral transforming gene(s) is sufficient to induce transformation, and any theory of transformation by inactivation of host sequences after Py insertion becomes less tenable.

Site and specificity of integration. It has previously been shown by restriction enzyme analysis that Py DNA is integrated in different chromosomal sites in different transformed cell lines (19, 21–23). The question arises whether there is

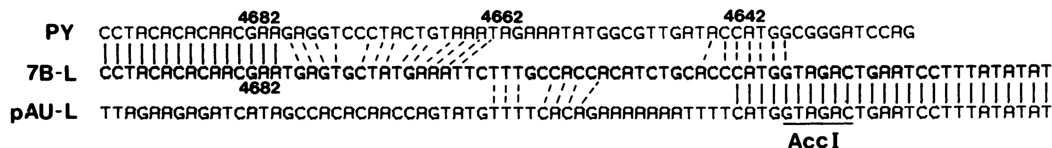

FIG. 7. Analysis of the integration site in pAU-L. The site of recombinant Py and cellular DNA which generated the left-hand join of the viral insert in 7axB is shown by comparing the DNA sequences of Py, 7B-L, and pAU-L. The sequence of the viral insert in 7B-L is colinear with the Py sequences until NT 4,682, as shown. Sequences flanking the viral insert in 7B-L are not homologous with cellular sequences around the site of integration until 4 bp before the *Acc*I site, at which point homology between the two sequences abruptly begins and continues over the extent that both sequences were determined (100 bp; only 26 bp of this homology is shown). The origin of the 37 nucleotides immediately adjacent to the colinear extent of the viral insert in 7B-L is not known, but some of the sequences may have been derived from small patches of Py and cellular DNAs, as shown by dotted lines between the sequences.

TABLE 2. Transforming activities of cloned DNAs[a]

DNA	No. of foci per 0.5 μg of plasmid DNA	Average no. of foci per ng of Py-transforming gene[b]
7B-L	682	5.6
7B-M	743	7.5
7B-R		
7T-L	727	6.9
7T-R		
Py-pAT153[c]	796	8

[a] The DNAs were applied to Rat-1 cells and stained for transformed foci 2 weeks later as described in the text.

[b] Calculated by assuming that the oncogenic capacity of each clone lies 1.75 kb from the *BclI* site on the Py genome until the end of the middle T-Ag coding region.

[c] Py-pAT153 is an *EcoRI*-cleaved genome of pAT153 into which *EcoRI*-linearized Py DNA has been ligated.

any sequence specificity or structural feature common to sites of recombination between viral and host DNA. For example, the integration of retroviral genomes or bacterial transposable elements results in duplication of adjacent host DNA, and transposable elements apparently have preferred sites of integration (7, 24, 29). In the cases presented here, the DNA sequences flanking the integrated Py DNA were different and shared no apparent common structural feature. This is a consistent finding for other Py-host join sequences (27a; Ruley, unpublished data). Some of the flanking sequences showed homology to displaced viral sequences (7T-L, 7B-L); others did not (Fig. 4 and 5). Some of the flanking sequences were A+T rich (e.g., 7T-L, 90% over 20 bp); others were not. Some of the adjacent displaced viral sequences were purine-rich (e.g., 7B-R, 80% over 10 bp); others were not. An extensive hairpin loop structure could be postulated for viral sequences which surrounded and included the site of recombination of viral and host DNA in 7B-R; for other joins this was not possible. In short, no common sequences or structural features were observed at the sites of integrated Py sequences studied here. Similar conclusions were reached in studies on integrated simian virus 40 (SV40) DNA (5, 32).

Although no common specificity was observed among the different integrated Py sequences, a unique sequence arrangement was observed for each complete provirus. Oligonucleotides found at or around the join at one end of the viral insert were observed in about the corresponding position at the other end (Fig. 8). Symmetrically placed oligonucleotides have been observed at either end of another Py insert (27a) and at either end of two out of three SV40 proviruses for which the sequence from both ends is available (32). Given the short lengths of sequences involved, the biological significance of these symmetries is unclear. However, their occurrence in five out of the six Py and SV40 proviruses for which sequence data from both ends is available suggests that future virus-host joins should be examined for their occurrence.

Mechanism of integration. When the flanking sequences from each cell line were used to probe untransformed cell DNA, no common fragments were detected (Fig. 2). Thus, for both 7axB and 7axT, the host sequences which flank the viral inserts were not adjacent in rat cell DNA. This indicates that host DNA sequences have been deleted or rearranged at the site of viral integration. If sequences have been deleted, the minimum extent of this deletion is 6.5 kb in 7axB cell DNA and 0.65 kb in 7axT cell DNA. Deletions

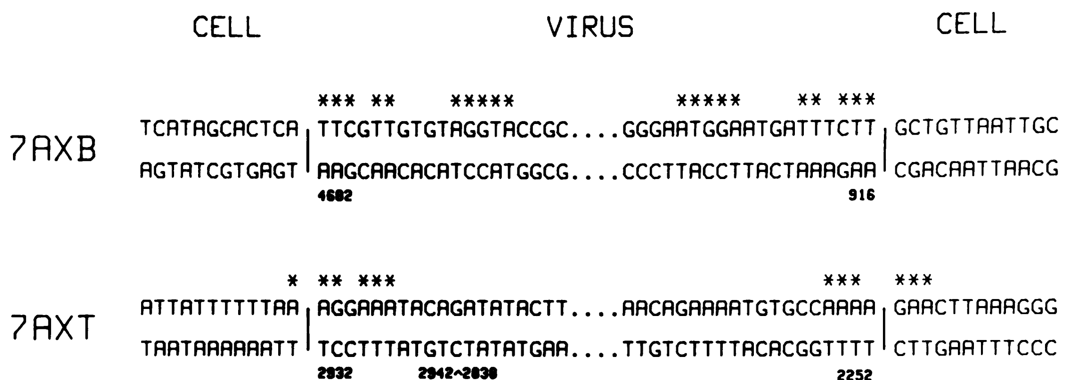

FIG. 8. The virus-host joins at either end of the Py sequences in 7axB and 7axT cells, respectively. Asterisks indicate the symmetrical sequence at the ends of each viral insert (i.e., inwards from flanking to viral sequences). Numbers are Py nucleotide numbers (30).

have also been observed around the sites of recombination between SV40 and rat cell DNAs (5); SV40 and adenovirus DNAs (40); SV40 and monkey DNAs in an SV40-defective variant (9); and different sequences of integrated Py DNA in Py-transformed cells (Ruley, unpublished data). Thus, deletions may be a general feature accompanying recombination between two largely dissimilar sequences in eucaryotic cells.

The Py DNA in 7axB cells was flanked by single-copy rat DNA. An *Eco*RI fragment of cell DNA, homologous to the flanking DNA in 7B-L, was cloned from syngeneic, untransformed rat cells. This cloned DNA (pAU-L) was compared to 7B-L. For 1.3 kb from one of the *Eco*RI sites, 7B-L and pAU-L had identical cleavage patterns as determined by analysis with more than a dozen different restriction enzymes (Fig. 3). Within this region, 7B-L and pAU-L DNA showed identical sequences for at least 100 bp. For 37 bp adjacent to the virus-host join, the sequence of 7B-L differed from that of pAU-L (Fig. 7). This dissimilarity may be explained in one of four ways. (i) pAU-L has been cloned from the allele opposite that to which viral DNA integrated. If so, the sequence divergence just at the virus-host join would be remarkably fortuitous. (ii) Py DNA transduced a stretch of sequence which is too small for its true homolog to be detected by hybridization. (iii) The sequences flanking the viral sequences (7B-L) represent a patchwork of viral and host sequences, possibly constructed by template switching of DNA polymerase(s) filling a gapped duplex at the site of integration. Some form of template switching for bacterial DNA polymerase has previously been invoked for aberrant replication events (1, 8). According to this model, parts of the flanking DNA are viral in origin, and because of this the flanking DNA shows homology to viral sequences. (iv) Recombination at the site of integration has involved other host sequences as well as the viral sequences.

As yet, the data are insufficient to distinguish between these four alternatives. The flanking DNA in 7B-L did show some homology to viral sequences, and this homology was greater in the 40 bp around the virus-host join that it was elsewhere. However, there were regions of flanking DNA which were not easily shown to be derived from either viral or cellular sequences. Around the sites of other virus-host joins, the flanking DNA showed even greater homology to viral DNA than was the case for 7B-L: e.g., Py DNA-rat DNA 7T-L, this study (Fig. 5); SV40 DNA-rat DNA (32); SV40 DNA-monkey DNA (9, 16). Only an analysis of unoccupied sites from cases such as these can reveal whether virus-host patchworks or viral transduction may be involved in integration events. When there is no detectable homology between viral and flanking DNA (e.g., 7B-R, this study), the virus-host join may have been formed by a direct recombination event, or there may have been negligible template switching in filling a gapped duplex to produce a clean transition from viral to cellular sequences.

From the results presented here it would appear that Py single-stranded regions may pair to single-stranded regions of host DNA as a prerequisite to integration. The virus-host mismatched duplex could then be converted to a homoduplex by the repair systems of the host cell, which in some cases could result in regions of patchwork between viral and host sequences. If the two joins of the viral sequences to the host sequences took place as separate events, the joins could be in nonadjacent regions of the host DNA, causing a rearrangement of host sequences. A possible candidate for stimulating the opening of the host DNA duplex for entry of the viral strands would be the Py *A* gene product (large T-Ag) since, after virus infection, a functional *A* gene is required for efficient transformation (10–12). The study of more virus-host joins and unoccupied-site DNAs should lead to a greater understanding of the mechanism of Py integration.

ACKNOWLEDGMENTS

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

LITERATURE CITED

1. **Backman, K., M. Betlach, H. W. Boyer, and S. Yanofsky.** 1978. Genetic and physical studies on the replication of ColE1-type plasmids. Cold Spring Harbor Symp. Quant. Biol. **43**:69–76.
2. **Basilico, C., S. Gattoni, D. Zouzias, and G. Della Valle.** 1979. Loss of integrated viral DNA sequences in Py-transformed cells is associated with an active viral A-function. Cell **17**:645–659.
3. **Benton, W. D., and R. W. Davies.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
4. **Birnboim, H., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
5. **Botchan, M., J. Stringer, T. Mitchison, and J. Sambrook.** 1980. Integration and excision of SV40 DNA from the chromosome of a transformed cell. Cell **20**:143–152.
6. **Brutlag, D. L., J. Clayton, P. Friedland, and L. H. Kedes.** 1982. SEQ: a nucleotide sequence analysis and recombination system. Nucleic Acids Res. **10**:279–294.
7. **Calos, M., and J. Miller.** 1980. Transposable elements. Cell **20**:579–595.
8. **Clark, A. J., and M. R. Volkert.** 1978. A new classification of pathways repairing pyrimidine dimer damage in DNA, p. 57–72. *In* P. C. Hanawalt, E. C. Friedberg, and C. F. Fox (ed.), DNA repair mechanisms. Academic Press, Inc., New York.
9. **Dhruva, B. R., T. Shenk, and K. N. Subramanian.** 1980. Integration *in vivo* into SV40 DNA of a sequence that resembles a certain family of genomic, interspersed, re-

peated sequences. Proc. Natl. Acad. Sci. U.S.A. 77:4514–4518.
10. **DiMayorca, G., J. Callender, G. Marin, and R. Giordano.** 1969. Temperature sensitive mutants of polyoma virus. Virology 38:126–133.
11. **Eckhart, W.** 1969. Complementation and transformation by temperature-sensitive mutants of polyoma virus. Virology 38:120–125.
12. **Fried, M.** 1965. Cell-transforming ability of a temperature sensitive mutant of polyoma virus. Proc. Natl. Acad. Sci. U.S.A. 53:486–491.
13. **Graham, F. L., and A. J. van der Eb.** 1973. Transformation of rat cells by DNA of human adenovirus 5. Virology 52:456–467.
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. **Grosveld, F. G., H. H.-M. Dahl, E. de Boer, and R. A. Flavell.** 1981. Isolation of β-globin-related genes from a human cosmid library. Gene 13:227–237.
16. **Gutai, M. W., and D. Nathans.** 1978. Evolutionary variants of SV40: cellular DNA sequences and sequences at recombinant joints of substituted variants. J. Mol. Biol. 126:275–288.
17. **Hassell, J. A., W. Topp, D. Rifkin, and P. Moreau.** 1980. Transformation of rat embryo fibroblasts by cloned polyoma virus DNA fragments containing only part of the early region. Proc. Natl. Acad. Sci. U.S.A. 77:3978–3982.
18. **Hayward, G. S., and M. G. Smith.** 1972. The chromosome of bacteriophage T5. Analysis of the single-stranded DNA fragments by agarose gel electrophoresis. J. Mol. Biol. 63:383–395.
19. **Israel, M. A., D. F. Vanderryn, M. L. Meltzer, and M. A. Martin.** 1980. Characterisation of polyoma viral DNA sequences in polyoma-induced hamster tumour cell lines. J. Biol. Chem. 255:3798–3805.
20. **Lania, L., M. Griffiths, B. Cooke, Y. Ito, and M. Fried.** 1979. Untransformed rat cells containing free and integrated DNA of a polyoma non-transforming (hr-t) mutant. Cell 18:793–802.
21. **Lania, L., D. Gandini-Attardi, M. Griffiths, B. Cooke, D. de Cicco, and M. Fried.** 1980. The polyoma virus 100K large-T antigen is not required for the maintenance of transformation. Virology 101:217–232.
22. **Lania, L., A. Hayday, G. Bjursell, D. Gandini-Attardi, and M. Fried.** 1980. Organization and expression of integrated polyoma virus DNA sequences in transformed rodent cells. Cold Spring Harbor Symp. Quant. Biol. 44:597–603.
23. **Lania, L., A. Hayday, and M. Fried.** 1981. Loss of functional large T antigen and free viral geno--- from cells transformed in vitro by polyoma virus after passage in vivo as tumor cells. J. Virol. 39:422–431.
24. **Majors, J. E., and H. E. Varmus.** 1981. Nucleotide sequences at host-proviral junctions for mouse mammary tumour virus. Nature (London) 289:253–258.
25. **Maniatis, T., R. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, and A. Efstradiadis.** 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell 15:687–701.
26. **Maxam, A. M., and W. Gilbert.** 1977. Sequencing end-labelled DNA with base-specific chemical cleavages. Proc. Natl. Acad. Sci. U.S.A. 74:560–563.
27. **Novak, U., S. Dilworth, and B. E. Griffin.** 1980. Coding capacity of a 35% fragment of the polyoma virus genome is sufficient to initiate and maintain cellular transformation. Proc. Natl. Acad. Sci. U.S.A. 77:3278–3282.
27a. **Ruley, H. E., L. Lania, F. Chaudry, and M. Fried.** 1982. Use of a cellular polyadenylation signal by viral transcripts in polyoma virus transformed cells. Nucleic Acids Res. 10:4515–4524.
28. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. 87:107–110.
29. **Shimotohno, K., S. Mizutani, and H. M. Temin.** 1980. Sequences of a retrovirus provirus resemble those of b cterial transposable elements. Nature (London) 28 :550–554.
30. **Soeda, E., J. Arrand, N. Smolar, J. Walsh, and B. E. Griffin.** 1980. Coding potential and regulatory signals of the polyoma virus genome. Nature (London) 283:445–463.
31. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.
32. **Stringer, J.** 1981. Integrated simian virus 40 DNA: nucleotide sequences at cellular virus recombinant junctions. J. Virol. 38:671–679.
33. **Tiemeier, D. C., S. M. Tilghman, and P. Leder.** 1977. Purification and cloning of a mouse ribosomal gene fragment in coliphage lambda. Gene 2:173–191.
34. **Tooze, J. (ed.)** 1980. DNA tumor viruses. Molecular biology of tumor viruses, part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Treisman, R., U. Novak, J. Favaloro, and R. Kamen.** 1981. Transformation of rat cells by an altered polyoma virus genome expressing only the middle-T antigen. Nature (London) 292:595–600.
36. **Twigg, A. J., and D. Sherratt.** 1980. Trans-amplifiable copy number mutants of plasmid ColE1. Nature (London) 283:216–218.
37. **Waalwijk, C., and R. A. Flavell.** 1978. An isoschizomer of HpaII which cleaves both unmethylated and methylated HpaII sites. Nucleic Acids Res. 5:3231–3236.
38. **Weaver, R. F., and C. Weissmann.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map co-ordinates. Nucleic Acids Res. 7:1175–1193.
39. **Wigler, M., R. Sweet, G. K. Sim, B. Wold, A. Pellicer, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel.** 1979. Transformation of mammalian cells with genes from prokaryotes and eukaryotes. Cell 18:777–785.
40. **Zain, B., and R. Roberts.** 1978. Characterisation and sequence analysis of a recombination site in the hybrid virus $Ad2^{+}ND_{1}$. J. Mol. Biol. 120:13–31.