The Structure and Function of the Integrated Polyoma Virus DNA in 82-Rat and 53-Rat Transformed Cells

By H. Earl Ruley, Furzana Chaudry, Moira Read, Luigi Lania and Mike Fried

Department of Tumour Genetics, Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, U.K., Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138, Department of Biochemistry, State University of New York, Stony Brook, New York 11794, U.S.A. and Istituto di Biologia, Generale e Genetica, Università di Napoli, Naples, Italy

(Accepted 14 September 1987)

SUMMARY

Integrated viral sequences and adjacent cellular sequences from the polyoma virus (Py)-transformed 53-Rat and 82-Rat cell lines which contain two and three partial early regions respectively, each in a single viral insert, have been molecularly cloned. Each of the cloned partial early regions have been subcloned and assessed with regard to their transcription, translation products (T antigens, TAgs) and biological activity including their transforming ability. The 53-Rat 5-3 kb EcoRI fragment is an intact Py EcoRI linear genome (derived from within the tandem duplicated sequences) which transforms rat cells with high efficiency and produces infectious virus when circularized and transfected into mouse cells. The 82-Rat cell line expresses three novel TAg species of 63K, 40K and 32K in addition to the Py middle and small TAgs. The 63K protein was found to be a truncated form of large TAg produced as the result of an addition/deletion in early region B sequences unique to large TAg. The 40K and 32K proteins are hybrid viral-cellular middle and large TAgs respectively, which are expressed from early region A that has been truncated by recombination with rat cellular DNA. Differences in the nuclear and cytoplasmic location of the different 82-Rat early region RNAs are due to RNA stability and/or transport from the nucleus to the cytoplasm most likely as a result of different cellular sequences at their 3' ends. Finally no common structural feature or sequence specificity was observed at the virus-host DNA joins of the two cell lines.

INTRODUCTION

Cells transformed with polyoma virus (Py) invariably contain viral DNA inserted into the cellular genome (Tooze, 1981). The viral DNA is often present in a head-to-tail tandem arrangement, containing more than one copy of the viral genome (Birg et al., 1979; Basilico et al., 1979; Hayday et al., 1982; Lania et al., 1980, 1981, 1982). As no sequence specificity or common structural features have been detected at the virus-host DNA joins in Py-transformed cells (Hayday et al., 1982; Ruley et al., 1982) integration of Py sequences appears to involve random non-specific illegitimate recombination events. Integration generally causes rearrangements of host sequences at the integration site (Hayday et al., 1982; Williams & Fried, 1986). Sequences encoding the viral 56K middle and 22K small TAgs (TAgs) are always retained and expressed, whereas sequences required for expression of the 100K large TAg are frequently missing from the viral insert (Hayday et al., 1982; Lania et al., 1980, 1981, 1982; Ruley & Fried, 1983). The inability to produce a functional large TAg can result from point mutations (Hayday et al., 1983) or deletions (Ruley & Fried, 1983) within the large TAg coding sequence, but more...
frequently results from the loss of sequence encoding the carboxy end of the large T Ag following recombination with cellular DNA (Hayday et al., 1982; Lania et al., 1980, 1981, 1982; Ruley et al., 1982). In the latter case, the major viral polyadenylation [poly(A)] signal at the end of the early region is lost. This can result in the production of virus–cell hybrid transcripts as well as hybrid proteins (Ruley et al., 1982). In other cases, such as 53-Rat viral cytoplasmic transcripts may utilize an alternative poly(A) signal located in the middle of the viral early region (Kamen et al., 1980).

In the Py-transformed cell line 82-Rat the single viral insert contains a number of rearrangements (Ruley & Fried, 1983), and three Py T Ag-related proteins of 65K, 40K and 32K are produced in addition to the Py middle and small T Ags (Lania et al., 1980, 1981). In the Py-transformed cell line 53-Rat the single viral insert contains a partial duplication of viral sequences and only the Py middle and small T Ags are produced (Lania et al., 1980, 1982). In order to correlate the virus-specified proteins and transcripts with integrated viral DNA it is advantageous to clone molecularly and analyse the integrated viral and adjacent cellular DNA sequences. In this report cloned integrated viral and adjacent cellular DNA sequences of 82-Rat and 53-Rat cells have been analysed with respect to translation products, transcription, biological activity and sequences at some of the virus–host DNA joins.

METHODS

Cell culture and biological activity of cloned DNAs. The isolation of the 82-Rat and 53-Rat cell lines has been previously described (Lania et al., 1980). Cells were grown in Dulbecco's modified Eagle's medium supplemented with a 5% foetal calf serum. Infectivity of cloned viral DNA was monitored by transfecting whole mouse embryo cells in the presence of DEAE-dextran and the transforming ability of cloned DNA was determined by transfecting Rat-1 cells using a modification of the calcium phosphate co-precipitation method (Hayday et al., 1982). Foci of transformed cells were stained (Leishman's stain) after 14 to 21 days.

Cloning. The strategy for cloning cellular DNAs enriched for integrated viral DNA sequences in 2gtWes has been previously described (Hayday et al., 1982; Ruley et al., 1982). Briefly, EcoRI-cleaved DNA from 82-Rat and 53-Rat cells was enriched for Py sequences by velocity sedimentation in 10 to 40% sucrose gradients. Enriched fractions were ligated to the purified arms of 2gtWes B, packaged in vitro and plated. Phages were transferred to nitrocellulose filters and screened for Py DNA sequences by hybridization to nick-translated (Rigby et al., 1977) Py DNA as described by Benton & Davis (1977). Positive clones were checked to verify that they contained EcoRI fragments derived from integrated Py DNA. Subsequently, the EcoRI fragments and fragments derived from them by secondary restriction enzyme cleavages were subcloned into the plasmid pAT153 (Twigg & Sherratt, 1980) for DNA sequence analysis and for studies testing the biological activity of integrated viral DNAs.

DNA sequencing. Restriction enzyme cleavage sites were 5' end-labelled following treatment with calf intestine alkaline phosphatase and T4 polynucleotide kinase. Following secondary enzyme cleavage singly end-labelled fragments were fractionated by polyacrylamide gel electrophoresis, eluted and sequenced by the chemical degradation method of Maxam & Gilbert (1981). Sequences were analysed by the SEQ computer program (Brutlag et al., 1982).

SI analysis of 82-Rat cell RNA. Methods for the isolation of poly(A)-containing nuclear and cytoplasmic RNAs and the analysis of SI nuclease-resistant RNA–DNA hybrids have been thoroughly described (Favaloro et al., 1980). Briefly, 5 g of poly(A)-containing RNA was annealed to 10 ng of agarose gel-purified 82-Rat early region A 5.2 kb PstI–EcoRI fragments for 3 h at 52 °C and treated with SI nuclease as described. SI-resistant hybrids were analysed following alkaline agarose gel electrophoresis, blotting to nitrocellulose, hybridization to nick-translated probes and autoradiography (Berk & Sharp, 1978; Favaloro et al., 1980). The filter was probed first with a nick-translated 1 kb PstI–EcoRI Py early region fragment from plasmid R82J and the autoradiogram developed. The filter was then washed twice for 15 min with 100 mM NaOH to remove all the hybridized Py probe. The filter was then probed with nick-translated 5.2 kb PstI–EcoRI from clone 82JF3 containing 780 bp of early region A Py sequence and 4.4 kb of flanking cellular DNA.

Analysis of T Ag species. The conditions for cell labelling, extraction and immunoprecipitation of Py-infected and -transformed cells have been described (Lania et al., 1980, 1981). Py-transformed cells were grown to semi-confluence in a 5 cm dish and labelled for 3 h at 37 °C with [35S]methionine added at 300 μCi/dish in 4 ml Dulbecco's modified Eagle's medium without methionine. After labelling, the cells were detached with EDTA and washed twice with ice-cold Tris–saline. Subsequent extraction, immunoprecipitation and acrylamide gel analysis of T Ags from Py-transformed cells was identical to that of lytically infected cells.
Nature of integrated polyoma virus DNA

Fig. 1. The structure of the integrated viral sequences (white) and flanking cellular DNA (black) in 82-Rat and 53-Rat cells as deduced by Southern blotting, molecular cloning and DNA sequencing. The hatched areas indicate those regions in which the virus-host DNA joins have not yet been determined precisely. The viral sequences are shown in relation to the HpaII cleavage map (HpaII fragments 1 to 8) and the direction and extent of the viral early and late transcription units are indicated (Tooze, 1981). The numbers above and below the HpaII map indicate Py map units (Tooze, 1981) and the locations of a number of restriction enzyme cleavage sites are indicated. The EcoRI fragments which have been molecularly cloned from 82-Rat cells (82J and 82L) and 53-Rat cells (53L and 53J) are shown. The rearrangements in the 82-Rat insert where late region Z has recombined with early region B, where early region Y has recombined with early region X and late region B, and where early region Y has recombined with early region X and late region A (Ruley & Fried, 1983) are indicated.

RESULTS

Structure of the viral insert in 82-Rat and 53-Rat cell lines

The Py-transformed 82-Rat and 53-Rat cell lines contain single viral inserts. The arrangement of integrated viral sequences as deduced by Southern blot analysis has been previously reported (Lania et al., 1980; Ruley & Fried, 1983). EcoRI fragments containing portions of the viral inserts and flanking DNAs corresponding to the regions indicated in Fig. 1 (82J, 82L, 53J and 53L) were cloned in a prokaryotic vector as described in Methods. The cloned EcoRI fragments were analysed to determine the structures of the integrated viral DNAs more precisely by restriction enzyme mapping and by DNA sequencing. Viral sequences are present mostly in a head-to-tail tandem arrangement. Approximately 1.9 and 1.2 genome equivalents are contained in the 82-Rat and 53-Rat lines respectively. The 82-Rat cell line contains three early transcription units (early A, B and C; Fig. 1), whereas the 53-Rat line contains two partial early regions (A and B) of which only early B is transcribed (Kamen et al., 1980). The viral insert in the 82-Rat cell line contains two sites at which segments of viral DNA have recombined. The sequences at these viral joins have been previously reported and show small amounts of homology (3 or 4 bp) at the joins (Ruley & Fried, 1983).

Biological activity of cloned DNA fragments

The biological activities of different segments of the viral inserts were assessed by DNA transfection. Plasmid clones (Fig. 2) containing different segments of viral DNA were constructed and tested for their ability to transform cells. In the case of 82L, viral sequences were brought into proper alignment in plasmid C82L by cleavage with EcoRI, circularization, cleavage with BamHI and recloning into pAT153 (Fig. 2b). The two early regions (A and B) of 82J were separated after cleavage with EcoRI and BamHI and recloned into pAT153 to generate plasmids 82JF3 and 82JE6 (Fig. 2c). In order to test whether the transcriptional control signals associated with early region A were sufficient for gene expression, plasmid R82J was constructed in which a Py fragment from PstI to EcoRI was incorporated to form a complete coding region for the middle T Ag, the Py oncogene (Fig. 2c). The results (Table 1) show that the
Fig. 2. Structure of plasmids containing different Py early region sequences from 82-Rat and 53-Rat (see Fig. 1). Py sequences are shown in white, cellular sequences in black and pAT153 sequences are stippled. (a) The EcoRI fragments 53L and 53J (see Fig. 1) were subcloned into the EcoRI site of plasmid pAT153 to form plasmids 53L and 53J. (b) The 82L EcoRI fragment (see Fig. 1) was cloned into the EcoRI site of pAT153 to form plasmid 82L. The 82L EcoRI fragment was circularized, cleaved with BamHI and cloned into the BamHI site of pAT153 to form the C82L plasmid in which the 5' end of early region C was linked to the 3' end of early region B. (c) The EcoRI fragment 82J (see Fig. 1) was cloned into the EcoRI site of pAT153 to form plasmid 82J. Plasmid 82JF3 was formed from plasmid 82J by cleavage with BamHI followed by circularization of the BamHI fragment containing early region A, the flanking cellular DNA and plasmid sequences. Plasmid 82JE6 was formed from plasmid 82J after cleavage with both EcoRI and BamHI and cloning of the fragment containing early region B between the EcoRI and BamHI sites of pAT153. Plasmid R82J was formed from plasmid 82JF3 by isolating the 0.85 kb PstI fragment between 50 map units (late region B) and 80 map units (early region A) (see Fig. 1) and ligating it to a 1.0 kb Py PstI–EcoRI fragment (between 80 and 100 map units). The hybrid Py molecule was then cloned between the PstI and EcoRI sites of pAT153. Plasmid R82J contains a reconstructed 5' end of the Py early region derived in part from early region A and in part from Py sequences (contains complete coding region for the Py middle and small T Ags) and a truncated regulatory region (nucleotides 5096 to 5265) (Ruley & Fried, 1983).

Table 1. Ability of plasmids to transform cells

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Transformed foci per 100 ng DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>53L</td>
<td>186</td>
</tr>
<tr>
<td>53J</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>82L</td>
<td>235</td>
</tr>
<tr>
<td>C82L</td>
<td>262</td>
</tr>
<tr>
<td>82J</td>
<td>197</td>
</tr>
<tr>
<td>82JF3</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>82JE6</td>
<td>172</td>
</tr>
<tr>
<td>R82J</td>
<td>203</td>
</tr>
<tr>
<td>Py Bam-RI</td>
<td>252</td>
</tr>
</tbody>
</table>
cloned DNA fragments possessed all the biological properties of the corresponding regions of viral DNA. Thus plasmids containing a complete middle TAg coding region (53L, 82J, 82L, C82L, 82JE6, R82J) were able to transform cells. No transformation was observed with plasmids 53J or 82JF3.

Infectious Py can be rescued from 53-Rat cells after fusion to permissive mouse cells, presumably by homologous recombination between duplicated Py sequences within the viral insert (see Fig. 1) (Lania et al., 1982). The 53L plasmid contained a complete Py linear EcoRI fragment derived from within the duplicated Py sequences (Fig. 1). When this EcoRI fragment was cleaved from 53L, circularized and transfected onto permissive mouse cells, infectious virus was produced. The circularized 53L EcoRI fragment gave a titre of $1.7 \times 10^5$ p.f.u./μg of DNA which is similar to titres ($3.1 \times 10^5$ p.f.u./μg of DNA) obtained with wild-type viral DNA which has been cleaved with EcoRI and recircularized.

Assignment of TAg species to the early regions of the viral inserts

Viral sequences required to make a complete large TAg are not present in a colinear form in 82-Rat or 53-Rat cells. Thus, although 82-Rat cells contain three (A, B and C) and 53-Rat cells contain two (A and B) early transcription units, all lack sequences required to make a complete 100K large TAg (Fig. 1). Analysis of the TAg synthesized in 82-Rat and 53-Rat cells have been previously published (Lania et al., 1980, 1981). In 82-Rat cells three novel TAg species of 63K, 40K and 32K were observed (see Fig. 3 a) whereas in 53-Rat cells no novel species were detected. Cells transfected with plasmids containing different 82-Rat early regions (see Fig. 2 and Table 1) were analysed in order to assign the 82-Rat cell novel TAg species to specific early region segments.
The TAg species expressed in cell lines established after transfection with plasmids 82J, C82L and 82JF3 are presented in Fig. 3. The 82J and C82L cells were selected by virtue of their transformed phenotype (see Table 1). Although no foci with strongly transformed phenotypes were observed after transfection with the 82JF3 plasmid, 12 areas of cells with altered morphology were detected. Analysis of cell clones derived from these regions showed that three were producing Py-related TAg species. In particular, two novel TAg species of 32K and 40K in addition to the 22K Py small TAg were produced. These novel TAg species were confirmed by three cell lines derived from these regions that produced the 32K and 40K TAg species produced by 82-Rat cells, which were also derived from early region A and that this region also codes for a complete Py 22K small TAg. Cells transformed by the 82JF3 plasmid which contains 82-Rat early region A (see Fig. 1 and 2c) were analysed and found to produce the 32K and 40K TAg species (Fig. 3b). The 56K middle TAg present in 82J cells is encoded by early region B sequence present in 82J (Fig. 1 and 2c), whereas the 22K small TAg is derived from both early regions A and B (Fig. 3b). The 56K middle TAg species present in 82J-transformed cells (Fig. 3b) but not the original 82-Rat cells results from truncation of the large TAg reading frame in early region B by joining the Py and plasmid sequences at the EcoRI site (see Fig. 2c). Finally the C82L early region is derived from joining the 5' end of early region C to the 3' end of early region B (Fig. 2b). Cells transformed by C82L produced a 63K protein which comigrated with the novel 63K TAg species found in 82-Rat, in addition to the Py middle and small TAg species (Fig. 3c).

Analysis of virus–cell junctions

The sequences through the virus–host DNA junctions in clones 82J and 53J were determined. Viral early region sequences were joined to host cell DNA at nucleotide 1239 at the left end of the insert in 82-Rat and at nucleotide 2312 at the right end of the insert in 53-Rat (Fig. 4a and b). The sequences were analysed by the SEQ computer program for homologies to either strand of the Py genome and for structural features such as dyad symmetries, inverted or direct repeats or for biases in base composition. At each of the two junctions no common sequence specificity or structural feature in either the viral or cellular DNA was observed. Moreover, a comparison of the flanking cellular and displaced viral DNA sequences (i.e. viral DNA sequences adjacent to the viral inserts which have been displaced by cellular DNA) revealed that the transition from viral to cellular sequences was abrupt, whereas beyond the join there were only small patches of homology between the flanking cellular and displaced viral DNAs. The significance of these homologies was assessed by comparing sequences chosen at random for similar homologies and by the SEQ computer program. None of the homologies was statistically significant.

The sequences through the insertion/deletion in early region B of 82-Rat (Ruley & Fried, 1983) revealed the presence of a termination codon in the large TAg reading frame (Fig. 4c). This termination codon would result in the loss of 345 amino acids from the carboxy end of large TAg, with an additional 10 amino acids added by the rearranged viral sequences. These data predict that the 63K TAg is a truncated large TAg encoded by early region B of the 82-Rat viral insert and is consistent with the fact that cells transformed by plasmid C82L produce the 63K TAg species (Fig. 3c).
(a) 82-Rat join

82-Rat

Large T Ag (32K)

Thr-Pro-Ala-His-Cln-***

Middle T Ag (40K)

Pro-Cln-Leu-Ile-Asn-Asn-Ser-***

Py early A

ACC CCAGCTCATCAATAACTCA~CTTCAGGGAATTGGAGAGTACATAAC

Pro-Gln-Leu-Ile-Asn-Asn-Ser-***

Py late Z

82-Rat

Large T Ag (63K)

Pro-Leu-Leu-Leu-Leu-Ser-Pro-Trp-Pro-Ser-***

(b) 53-Rat join

53-Rat

Large T Ag

Leu-Ala-Ala-Ala-Gly-Thr-Ala-Val-Pro-Cln-Gly-Leu-Lys-Ser-Arg-Glu-Leu-Leu-***

(c) 82 Insert

82-Rat

Large T Ag (63K)

Pro-Leu-Leu-Leu-Leu-Ser-Pro-Trp-Pro-Ser-***

Fig. 4. DNA sequence and translation products at the virus-host DNA joins of 82-Rat early region A and flanking cellular DNA (a) and 53-Rat early region B and flanking cellular DNA (b) as well as the 82-Rat virus-virus join (Ruley & Fried, 1983) of early region B and late region Z (c) (see Fig. 1). The joins between the Py and flanking cell DNA (a and b) and the different Py sequences (c) are indicated by an arrow. In (a) and (b) the Py DNA sequences displaced as a result of the recombination with flanking cellular DNA are shown above the 82-Rat and 53-Rat virus-cell sequence. The small asterisks under the DNA sequence indicate nucleotide homologies between the displaced Py DNA and the flanking cellular DNA in (a) and (b). The Py nucleotide numbers are indicated beneath the Py sequences. Below the 82-Rat DNA sequence is presented the translated amino acid sequence of the 32K large (a), 40K middle (a) and 63K large (c) T Ags. Part of the amino acid sequences of both the 32K and 40K T Ags are encoded by cellular DNA sequences. Beneath the 53-Rat sequence is presented the amino acid sequence (partly Py-encoded and partly cell-encoded) of the putative large T Ag product (b). No 53-Rat large T Ag products are detected in vivo or in vitro (see text). The translational termination codons in the 82-Rat and 53-Rat flanking DNA (a and b) and in the Py late region Z (c) are underlined.

Transcription of 82-Rat early region A

Transcription from the 82-Rat early region A was analysed in greater detail by S1 mapping (Berk & Sharp, 1978; Favaloro et al., 1980) (Fig. 5). Unlabelled 5.2 kb PstI–EcoRI fragment, containing viral sequences from early region A viral and adjacent cellular sequences, was first hybridized to either cytoplasmic or nuclear RNA from 82-Rat cells. The S1-resistant RNA–DNA hybrids were analysed following alkaline gel electrophoresis by blot hybridization. Fig. 5 shows the results of the blot hybridization using the 5.2 kb EcoRI–PstI fragment and Py DNA as radioactive probes and the deduced structures of the RNAs detected. The bands present only in the nuclear fractions indicate the unspliced precursors of the spliced transcripts found primarily in the cytoplasmic fractions but also present in the nuclear fractions. All of the bands shown in Fig. 5 contained at least some viral sequences as they were positive when the blot was hybridized with a Py early region probe as well as the 5.2 kb PstI–EcoRI probe. Hybridization of the unlabelled 5.2 kb PstI–EcoRI fragment with the 82-Rat early region B and/or C RNAs resulted in S1-resistant hybrids ending at Py nucleotide 1239 (end of the viral sequences in the DNA probe fragment derived from early region A; see Fig. 4a). Thus the 450 bp band and part of the 275 bp band are derived from spliced RNAs and the 780 bp band from unspliced RNAs derived from early region B and/or C (Fig. 5). Part of the 275 bp band and all the other larger bands in Fig. 5 are entirely derived from protection of regions of the 5.2 kb PstI–EcoRI fragment by virus-host transcripts initiated in 82-Rat early region A.
Fig. 5. S1 mapping of 82-Rat early region A mRNAs. At the top is shown a map of 82-Rat early region viral (white) and adjacent cellular (black) DNA (see Fig. 1). The 5.2 kb PstI–EcoRI fragment used in the S1 mapping below is indicated. Beneath the map on the left is shown an autoradiograph of an S1 mapping gel. Either cytoplasmic (C) or nuclear (N) RNA from 82-Rat cells, or no RNA (−), was hybridized to the unlabelled 5.2 kb PstI–EcoRI fragment (above) and the S1-resistant hybrids were analysed after fractionation in a 1% alkaline agarose gel using either a Py probe (right) or the 5.2 kb fragment from early region A (left). HpaII-digested Py DNA combined with an EcoRI linear Py molecule were used as markers (M). On the right are shown diagrammatical representations of the structures of RNAs protected by the hybridization fragment deduced from the bands of the gel on the left. At the top are shown the RNAs initiated in early region A and terminated at two different points (arrowheads) in the cellular sequences separated by about 170 bp. LT, MT and ST indicate the spliced RNAs with the large, middle and small T Ag splices respectively. 275 indicates the 275 bp leader sequences between the PstI site and the middle and small T Ag splice donor sites (no large T Ag leader sequences were detected since the PstI site lies downstream from the large T Ag donor site). 1030 and 1200 indicate the 1030 and 1200 bp bodies of the viral–cellular hybrid transcripts starting at either the small, or the middle and large T Ag splice acceptor sites (not distinguished) and terminating at the two different positions in the cellular sequence. 1370 and 1550 indicate the 1370 and 1550 bp unspliced precursors of the two RNAs terminated at the two different points in the cellular sequence; these are found only in the nuclear RNA (see gel on left). At the bottom right are the RNAs derived from early region B and/or C. 275 indicates the 275 bp of the middle and small T Ag spliced RNA (see above). 450 indicates the 450 bp body of protected viral sequences from the hybridization fragment extending from the splice acceptor sites of the three T Ags to the end of the viral sequences in early region A at Py nucleotide 1239 (see Fig. 4). 780 indicates the unspliced precursor of the early region B and/or C spliced RNAs and is found only in the nuclear RNA (see gel on left).
The results are best explained by the presence of two transcripts initiated in the viral sequences in early region A and terminating at two different positions within the cellular sequences (see Fig. 5). The more abundant of these two mRNAs extends about 1030 bp from the Py early region A acceptor sites (about 580 bp into the cellular sequences). The 1370 bp band in the nuclear lane (Fig. 5) would represent the unspliced precursor of this predominant transcript. The second, less abundant transcript would extend 1200 bp from the Py early region A splice acceptor sites (about 750 bp into the cellular sequences). The faintly visible band of 1550 bp in the nuclear lane (Fig. 5) would represent the unspliced precursor of this transcript. The absence of other putative unspliced products in the nuclear lane (Fig. 5) indicates that the difference detected in the 3' ends of these two RNAs is due to differences in polyadenylation and not different splice donor sites in the cellular sequences.

The results presented in Fig. 5 also indicate an unusual distribution between the nucleus and cytoplasm of the spliced transcripts initiated in early region A compared to those initiated in early region B and/or C. This is indicated by the differences in the ratio of the amount of spliced RNA in the cytoplasm to that in the nucleus of the different early region-derived RNAs. The 1030 bp and 1200 bp bands which represent the spliced mRNAs initiated in early region A are predominantly in the cytoplasm (Fig. 5). On the other hand there was a large amount of the 450 bp band, which represents the spliced mRNAs initiated in early region B and/or C, in the nucleus relative to the cytoplasm (Fig. 5). As expected, the 780 bp band derived from unspliced transcripts initiated in early region B and/or C as well as the 1370 bp and 1550 bp bands derived from unspliced early region A-initiated transcripts were only detected in the nucleus. The different distribution between the nucleus and cytoplasm of the spliced transcripts must be a consequence of the different 3' ends of the RNAs of the three early region transcripts.

**DISCUSSION**

Integrated viral and adjacent host sequences from two Py-transformed cell lines have been cloned in prokaryotic vectors. The cloned DNAs were analysed to correlate the structure of viral sequences with their expression and to understand the nature of the recombination events in Py-transformed cells. Regions in which viral and cellular DNAs have recombined have been sequenced. Alterations in protein-coding sequences produced by these events account for the three novel TAg species found in 82-Rat cells. The truncated TAg species synthesized in 82-Rat cells were further mapped to different regions of the viral insert by analysis of the T Ags synthesized in cells transfected by cloned DNA fragments.

The viral inserts in 82-Rat and 53-Rat cells together contain four early transcription units, however none encodes a complete 100K large TAg. The inability to synthesize the 100K large TAg reflects the fact that the 82-Rat and 53-Rat cell lines were initially chosen for further analysis because they contained single inserts and lacked free forms of viral DNA (Lania et al., 1980). Utilizing DNA transfection the biological activity and integrity of the cloned viral sequences from 82-Rat and 53-Rat have been assessed, and the viral proteins synthesized in these cell lines have been assigned to regions of the integrated viral DNA.

The preferred use of the Py alternative poly(A) signal by early region B mRNAs in 53-Rat cells may be a property of either viral or cellular sequences 3' to the early region B EcoRI site. However, circular 53L molecules containing 5' sequences and the minor poly(A) signal from early region B efficiently produce infectious virus and thus can act as templates for the synthesis of Py large TAg. In 82-Rat cells, three novel TAg species of 63K, 40K and 32K are produced in addition to the Py middle and small TAg (Fig. 3a). The 63K protein has been assigned to 82-Rat early region B
as a truncated form of Py large T Ag terminating in sequences altered by an addition/deletion in the large T Ag reading frame (Fig. 1 and 4c). The 40K and 32K proteins have been assigned to early region A which is truncated by recombination with host sequences (Fig. 1 and 4a). The 32K protein is a truncated form of large T Ag and the 40K protein a truncated form of middle T Ag. The transcripts from early region A are initiated in the viral sequences and extend into the adjacent cellular sequences where they utilize cellular polyadenylation signals probably between 450 and 750 bp past the virus–host DNA join (Fig. 5). The reading frames for both proteins terminate in cellular sequences immediately after the virus–host join (Fig. 4a). The 40K truncated middle TAg is not capable of inducing transformation (Table 1). An enhancer region (nucleotides 5021 to 5265) is associated with the Py early region (de Villiers & Schaffner, 1981).

As a result of the rearrangements in 82-Rat cells, the viral early region A contains only a portion of this sequence (nucleotides 5096 to 5265) (Ruley & Fried, 1983). This sequence is sufficient for efficient gene expression, as the plasmid R82J, in which a complete transforming region has been reconstituted from early region A (Fig. 2c), transforms efficiently (Table 1).

It is not clear whether early region C is efficiently expressed in 82-Rat cells. No T Ag species are detected either in vivo or in vitro (Kamen et al., 1980) other than those assigned to early regions A and B (see above). It is possible that an early region C-encoded aberrant T Ag species would have been missed as a consequence of its comigration with either the early region B-encoded 63K or the early region A-encoded 30K and 42K T Ags. The cloning of the 3' end and adjacent host sequences of early region C and an analysis as described above will be required to determine the quality and quantity of early region C expression. In any case the 5' portion of early region C, in the absence of the 3' end and adjacent cellular DNA sequences, is biologically active as evidenced by the transforming activity of plasmids 82L and C82L (Table 1) and the production of middle and small T Ags after transfection with plasmid C82L (Fig. 3c).

A difference was detected between the cytoplasmic and nuclear distribution of the RNA initiated in 82-Rat early region A compared to RNA initiated in early region B and/or C (Fig. 5). This difference must be a consequence of the different 3' ends of the different RNAs as each RNA is apparently initiated at the same Py cap site in each of the viral early regions. The differences may be due to stability of the RNAs in the different cellular compartments and/or differences in transport of the RNAs from the nucleus to the cytoplasm and most likely reflect the effects of the cellular sequences adjacent to early regions A and/or C.

The sequences determined through the virus–host DNA junction in 82-Rat and 53-Rat cells (Fig. 4) give no indication that a sequence-specific mechanism led to their formation. The sequences thus lack direct repeats characteristic of transposition elements (Calos & Miller, 1980; Majors & Varmus, 1981; Shimotohno et al., 1980) and lack structural features such as symmetries, specific sequences or base compositions which might facilitate recombination. However until the host DNA sequences present around the site of recombination with viral sequences are cloned and analysed, no comment on the role of homologous recombination can be made. Similar results have been observed for joins between Py and host sequences in other Py-transformed cell lines (Hayday et al., 1982; Ruley et al., 1982; Williams & Fried, 1986) and also for virus–host DNA junctions in simian virus 40-transformed cell lines (Stringer, 1981; Bullock et al., 1984).

The cloning and analysis of the integrated viral and adjacent cellular DNA sequences from the 82-Rat and 53-Rat cell lines have allowed the definition of alterations in the viral inserts, the assignment of novel virus-encoded proteins to different regions of the viral inserts, an assessment of the integrity of the viral sequences, investigation of the transcription of the various early regions and a test of their biological activity, and have provided sequence information about the virus–host DNA joins.

We are indebted to Chris Norbury and Brendan Davies for their helpful comments in the preparation of this paper and Mrs G. Briody for assistance in typing this manuscript.

REFERENCES


