JOURNAL OF VIROLOGY, Feb. 1984, p. 479–489
0022-538X/84/020479-11$02.00/0

Vol. 49, No. 2

Two Types of Deletion Within Integrated Viral Sequences Mediate Reversion of Simian Virus 40-Transformed Mouse Cells

KAZUO MARUYAMA* AND KINICHIRO ODA

Department of Tumor Virus Research, Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan

Received 11 May 1983/Accepted 23 October 1983

Simian virus 40 (SV40) DNA insertions from SV40-transformed mouse cell line W-2K-11 and its revertants M18, M31, and M42 were cloned. W-2K-11 cells contain 1.5 copies of the SV40 sequences in a partially tandem duplicated form. The endpoints of the viral sequences at the virus-host junctions are located very close to those reported by others, indicating that there are some preferred sites for integration and rearrangement in SV40 sequences. One flanking cellular sequence is a long stretch of adenine and thymine with repeated AAAT, and the other is a stretch of guanine and cytosine with repeated CCG. There are patchy homologies between the flanking cellular sequences and the corresponding parental SV40 sequences. The sequences around both junctions were retained in all the revertants, whereas most of the internal SV40 sequences coding for large T antigen were deleted. The coding sequences for small T antigen are intact, and small T antigen was expressed in all the revertants. The fragments cloned from M18 and M42 were identical and 3.9 kilobases of SV40 sequences were deleted. The parental SV40 sequences around the deletion site have sequences capable of forming a secondary structure which might reduce the effective distance between the two regions. The SV40 DNA retained in M31 is colinear with SV40 virion DNA, and a unit length of SV40 DNA was deleted within the SV40 sequences present in W-2K-11 cells. These results indicated that two types of deletion occurred during the reversion, one between homologous sequences and the other between nonhomologous sequences.

In general, the organization of a cellular genome is stable and is conserved from generation to generation. However, the genomic structure has the potential to acquire new genetic information by gene rearrangements—an induction of new covalent linkage between distant DNA sequences—such as amplification, deletion, transposition, and acquisition of exogenous sequences. The interaction of the simian virus 40 (SV40) genome with cellular genomes is a good system with which to analyze gene rearrangement in eucaryotic cells, since the covalent linkage between SV40 and cellular DNA has been found in transformed cells (24) and in the evolutionary variants generated after serial passages of SV40 at high multiplicity of infection (15).

Characterization of evolutionary variants by restriction enzyme analysis and electron microscopic heteroduplex mapping has suggested that recombination between SV40 and cellular DNA could occur at multiple sites (16, 21). Using Southern blot analysis, it has been demonstrated that there is no specific site of integration in both SV40 and cellular DNA in the transformed cells (6, 13). By means of recombinant DNA techniques, the fragments containing SV40 sequences from the transformed cells and those containing the corresponding cellular sequences from untransformed cells were cloned to determine whether the specific sequences are involved in the integrative recombination. The results have also shown that there is no evidence of homology between viral and cellular DNAs at the integration sites (5, 23). However, Stringer reported an exception in which there are five base pairs shared at the junction between SV40 and rat DNA in the SVRE9 cell line (27).

The genomic structure seems to be unstable at the site of insertion of viral sequences (7, 12). We previously isolated revertants from SV40-transformed mouse and rat cells which contained multiple copies of viral DNA serially arranged within the length of about 30 kilobases (kb) with at least two intervening cellular sequences between the viral sequences (19). The reversion occurred at a fairly high frequency and all the revertants isolated had suffered deletions within the integrated viral sequences, even if the revertants retained the intact early region and synthesized large T antigen (18).

In the present paper, the molecular mechanism of gene rearrangements which mediates reversion of SV40-transformed cells was investigated by cloning and sequencing the DNA fragments containing the viral sequences from one transformed mouse line and its large T antigen-negative revertants. The results revealed the following features. (i) The flanking cellular sequences in the transformed cells consist of either a long stretch of A+T or G+C sequences, and there are preferred viral sequences for integration and rearrangement. (ii) In all the revertants, both sides of virus-host junctions in the transformed cells were retained, but internal viral sequences were deleted. Deletion had occurred either between homologous sequences or between nonhomologous sequences. (iii) Small T antigen is expressed in all the revertants. These features are discussed in relation to the mechanisms of reversion in the transformed cells.

MATERIALS AND METHODS

Cell lines. The transformed cell line W-2K-11 was established by infection of C3H/He mouse kidney cell line C3H2K-C4 with virions of SV40 strain 777. Revertants M18, M31, and M42 were isolated from W-2K-11 cells by negative selection with bromodeoxyuridine as described previously (19).

Blot hybridization. High-molecular-weight cellular DNA and recombinant plasmid DNA were cleaved to completion with various restriction enzymes and electrophoresed on agarose slab gels. After alkali denaturation, DNA was transferred to a nitrocellulose filter or nylon membrane by the

* Corresponding author.

method of Southern (25). The fragments containing SV40 DNA sequences were detected by hybridization to ^{32}P-labeled SV40 DNA prepared by nick translation (22), followed by autoradiography as described previously (19).

DNA cloning. High-molecular-weight DNA from W-2K-11 cells was cleaved to completion with *Eco*RI. DNAs from the revertants were also cleaved to completion with *Eco*RI and separated by electrophoresis on an agarose gel. The slices of agarose containing the fragment-bearing SV40 sequences were cut out, and DNA was eluted by using glass powder (31). Phage vector λgtWESλB (28) was propagated by the PDS method described by Blattner et al. (4), using the *supF* strain of *Escherichia coli* Ymel. λgtWESλB DNA was cleaved with *Eco*RI, and the cohesive ends were annealed. The large terminal fragments (λ arms) were separated by centrifugation through a 10 to 40% sucrose gradient as described by Maniatis et al. (17). The annealed arms were ligated with the *Eco*RI fragments of cellular DNA and packaged in vitro (3). Recombinant phages containing SV40 sequences were screened by plaque hybridization (2). DNA fragments containing SV40 DNA sequences in recombinant phages were subcloned into a plasmid vector (pBR322) and characterized by restriction mapping. These recombinant plasmids were designated as pW2K, pM18, pM31, and pM42.

DNA sequencing. DNA sequences were determined by the method of Maxam and Gilbert (20). SV40 nucleotide sequences were numbered according to the SV numbering system (30). The right-hand virus-host junction at nucleotide 4311 in W-2K-11 cells (Fig. 3) was sequenced with plasmid pW2K2.8, which contained the 2.8-kb *Eco*RI fragment detected by Southern blot hybridization of W-2K-11 DNA, from the SV40 *Hin*fI site at nucleotide 4376 and from a cellular *Eco*RI site which was about 120 base pairs (bp) from the junction. The junction at nucleotide 4311 in revertant M18 was sequenced with pM18 from a cellular *Eco*RI site. The left-hand junction at nucleotide 1664 in revertant M18 was sequenced from the SV40 *Hin*dIII site at nucleotide 1493. The junction at nucleotide 1664 in revertant M42 was sequenced with pM42 from a cellular *Hin*fI site located within 30 bp of the junction. The virus-virus junction at nucleotides 161 and 1521 in revertant M18 (Fig. 3) was sequenced from the SV40 *Hin*dIII site at nucleotide 1493.

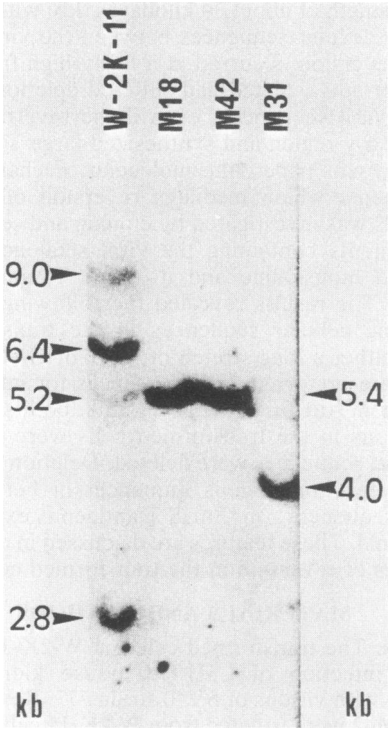

FIG. 1. Analysis of SV40 DNA sequences in W-2K-11 cells and in the revertants M18, M31, and M42. DNA (20 μg) extracted from transformed cell line W-2K-11 or from one of its revertants, M18, M31, or M42, was digested with *Eco*RI, electrophoresed on a 1.2% agarose gel, transferred to a nitrocellulose filter, and hybridized to ^{32}P-labeled SV40 DNA. The lengths of the fragments containing SV40 sequences were estimated from the positions of size marker DNA fragments which were run in parallel.

Immunoprecipitation. The W-2K-11 cell line and its revertants grown to semiconfluence were labeled with [^{35}S]methionine for 2 h. The cell extracts were prepared by lysing the cells in lysis buffer containing 25 mM Tris-hydrochloride (pH 8.0), 150 mM NaCl, 0.5% Nonidet P-40, and 150 μg of phenylmethylsulfonyl fluoride per ml. Normal hamster serum was added to the extracts, and the mixtures were incubated at 0°C for 30 min. Then, a 10% suspension of Formalin-fixed *Staphylococcus aureus* Cowan I was added, and the mixture was agitated at 4°C for 30 min. After centrifugation, the supernatants were combined with either anti-T hamster serum or normal hamster serum and incubated at 4°C overnight. The mixtures were similarly treated with bacteria, and the pellets were washed three times with Ripa buffer containing 50 mM Tris-hydrochloride (pH 7.5), 150 mM NaCl, 10% Triton X-100, 0.1% sodium dodecyl sulfate, and 150 μg of phenylmethylsulfonyl fluoride and 10 mg of deoxycholate per ml and suspended in sample buffer. After centrifugation, the supernatants were electrophoresed on a sodium dodecyl sulfate–14% polyacrylamide gel. The gel was dried and fluorographed.

RESULTS

Cloning of the viral insertions in W-2K-11 and the revertant cell lines. To analyze the detailed structures of SV40 DNA integrated into the cellular chromosomes in transformed cell line W-2K-11 and in its revertant cell lines M18, M31, and M42, we cloned the viral insertions together with flanking cellular sequences from each of these cell lines by use of λ phage vectors. As described in a previous paper (19), a possible arrangement of viral DNA sequences in these cell lines was constructed based on the results of Southern blot analysis. In the transformed cell line W-2K-11, two genome equivalents of SV40 sequences are serially arranged within a length of about 30 kb, with two intervening cellular sequences. In the revertant cell lines M18, M31, and M42, less than one full copy of the SV40 sequence is retained in a single site. A Southern blot analysis of the DNA fragments generated by digestion of cellular DNA from these lines with restriction endonuclease *Eco*RI is shown in Fig. 1.

W-2K-11 DNA yielded four bands with lengths of approximately 9.0, 6.4, 5.2, and 2.8 kb. To clarify the arrangement of these four fragments in the chromosome, W-2K-11 DNA was partially digested with *Eco*RI and cloned in λ phage Charon 4A. From 4 μg of partially digested W-2K-11 DNA, we could construct a DNA library of 1.6×10^6 recombinant phages. After amplification by the method of Maniatis et al. (17), a total of 5×10^5 recombinant phage plaques were screened for SV40 sequences and three clones were isolated. After several purification steps, restriction endonuclease digestion of these recombinant phages generated many mi-

nor fragments which hybridized to SV40 DNA, indicating an occurrence of high frequency recombination within the λ Charon 4A vector. Such recombinations were reported by Clayton and Rigby (7); recombinant phages carrying tandem duplications tended to be unstable. In our case, homologous sequences cloned in the phage vector were prone to be deleted by recombinations.

To minimize these homologous recombinations, we digested W-2K-11 DNA to completion with *Eco*RI and cloned in a λgtWESλB phage vector. Fourteen clones containing SV40 sequences were isolated by screening 1.5×10^6 recombinant phages; 10 of these clones contained a 5.2-kb fragment and 2 contained a 2.8-kb fragment, but the remaining 2 clones corresponded to none of four fragments detected by Southern blot analysis. These two clones seemed to be derived by rearrangement within phage DNA. To clone the 6.4-kb fragment detected by Southern blot analysis, we digested W-2K-11 DNA to completion with *Eco*RI, and the DNA fragments with a length of about 6.4 kb were purified and cloned in λgtWESλB. Three clones containing SV40 sequences were isolated by screening 8×10^5 recombinant phages. All of these clones contained a 6.4-kb fragment. As mentioned below, part of the 6.4- and 2.8-kb fragments are retained in the revertants, but the SV40 sequences in the 9.2- and 5.2-kb fragments are completely excised in the revertants. In this respect, the 9.0-kb fragment was not essential in the present study and was not cloned.

The revertants M18, M31, and M42 generated only one band each of about 5.4, 4.0, and 5.4 kb, respectively (Fig. 1). Cellular DNAs from each revertant were digested to completion with *Eco*RI, and the fragments corresponding to these lengths were purified. Ten clones were isolated by screening 6×10^5 recombinant phages constructed from M18. Three and seven clones each were isolated by screening 4×10^5 recombinant phages constructed from M31 and M42. The sizes of cloned *Eco*RI fragments containing SV40 sequences were identical to that of the fragment detected by the blot analysis. These *Eco*RI fragments were subcloned in the plasmid vector pBR322 and analyzed further. The recombinant plasmids containing the sequences from W-2K-11 were designated pW2K6.4, pW2K5.2, and pW2K2.8. The numbers represent the lengths of the cloned fragments in kb. The recombinant plasmids containing the sequences from the revertants were designated pM18, pM32, and pM42.

Structure of three cloned fragments from W-2K-11 cells. The physical maps of the DNA fragments cloned from transformed cell line W-2K-11 were determined by Southern blot hybridization analysis and partial restriction endonuclease digestion of end-labeled fragments. Figure 2 shows some of these analyses. *Eco*RI plus *Kpn*I digestion of pW2K2.8 (Fig. 2B) produced two fragments which hybridized to SV40 DNA. One of them comigrated with the 1,488-bp fragment corresponding to viral DNA between the *Kpn*I site at nucleotide 294 and the *Eco*RI site at nucleotide 1782 (Fig. 2A). *Eco*RI plus *Hin*cII digestion (Fig. 2C) produced the 1,263-bp fragment corresponding to viral DNA between the *Hin*cII site at nucleotide 519 and the *Eco*RI site (Fig. 2A). These results indicated that the SV40 insertion in pW2K2.8 extended counterclockwise from the viral *Eco*RI site at nucleotide 1782. Digestion with *Eco*RI and *Hin*fI (Fig. 2D) produced the viral *Hin*fI fragments A, D, G and one more fragment of about 190 bp which hybridized to SV40 DNA, indicating that the SV40 insertion in pW2K2.8 was colinear with SV40 virion DNA, and that the end of the viral fragment was located about 190 bp apart from the viral *Hin*fI site at either nucleotide 4459 or nucleotide 4376. Digestion with *Eco*RI and *Bst*NI (Fig. 2E) produced the viral *Bst*NI fragments D, E, G, and I and the 247-bp fragment corresponding to viral DNA between the *Bst*NI site at nucleotide 1535 and the *Eco*RI site (Fig. 2A) and one more fragment of about 700 bp which hybridized to SV40 DNA, indicating that the other end of the fragment was located about 700 bp apart from the viral *Bst*NI site at nucleotide 4892. These results indicated that the viral insertion extended at most to 190 bp beyond the *Hin*fI site at nucleotide 4376.

*Eco*RI plus *Hin*cII digestion of pW2K5.2 (Fig. 2C) produced the 1,263-bp fragment corresponding to the viral *Eco*RI-*Hin*cII fragment (Fig. 2A) and the 2.7-kb fragment with weak radioactivity, indicating that the SV40 DNA insertion in pW2K5.2 extended counterclockwise from the viral *Eco*RI site but not much beyond the *Hin*cII site at nucleotide 519. Digestion with *Eco*RI and *Hpa*II produced only one fragment which hybridized to SV40 DNA (Fig. 2F), indicating that the SV40 insertion did not contain the *Hpa*II site of SV40 DNA. These results indicated that the endpoint of SV40 insertion in pW2K5.2 was located between the *Hin*cII site at nucleotide 519 and the *Hpa*II site at nucleotide 346.

*Eco*RI and *Kpn*I digestion of pW2K6.4 (Fig. 2B) produced the 3,755-bp fragment corresponding to viral DNA between the *Kpn*I site at nucleotide 294 and the *Eco*RI site at nucleotide 1782 (Fig. 2A) and the 2.7-kb fragment which hybridized to SV40 DNA. *Eco*RI plus *Acc*I digestion of pW2K6.4 (Fig. 2G) produced about 3,000- and 2,000-bp fragments corresponding to viral DNA fragments between the *Eco*RI and *Acc*I sites and between two *Acc*I sites, respectively. Note that there are two *Acc*I recognition sites in SV40 strain 777 DNA, the one at nucleotide 1628 and the other at about nucleotide 4840. These results indicated that the SV40 insertion in pW2K6.4 was colinear with SV40 virion DNA and extended clockwise from the *Eco*RI site to beyond the *Acc*I site at nucleotide 1628.

Structures of the cloned DNAs deduced from these experiments are schematically presented in Fig. 3 and can be summarized as follows.

SV40 sequences in pW2K2.8 extend counterclockwise from the viral *Eco*RI site to the virus-host junction located close to the SV40 *Hin*fI site at nucleotide 4376, so that the 2.8-kb fragment contains only about 100 bp of flanking cellular sequences. SV40 sequences in pW2K5.2 extend counterclockwise from the viral *Eco*RI site to the virus-host junction located between the SV40 *Hin*cII site at nucleotide 519 and the *Hpa*II site at nucleotide 346, so that the 5.2-kb fragment contains only about 1.3 kb of SV40 sequences and about 3.9 kb of flanking cellular sequences. SV40 sequences in pW2K6.4 extend clockwise from the viral *Eco*RI site at nucleotide 1782 to the virus-host junction located close to the SV40 *Acc*I site at nucleotide 1628, so that the 6.4-kb fragment contains almost a unit length of SV40 DNA sequences and 1.3 kb of flanking cellular sequences. Because of a failure in cloning them from the partially digested W-2K-11 DNA fragments, the arrangement of these three fragments in a chromosome is unclear. However, as described below, the 6.4- and 2.8-kb fragments seem to be adjacent to each other.

Structure of cloned fragments from revertants M18, M31, and M42. Figure 4 shows some of the restriction endonuclease analysis of pM18, pM31, and pM42, and their physical maps constructed from these results are schematically represented in Fig. 5. *Kpn*I digestion of pM18 and pM42 (Fig. 4A) produced 8.35- and 1.35-kb fragments hybridizing to SV40 DNA, indicating that there are two *Kpn*I sites separated by

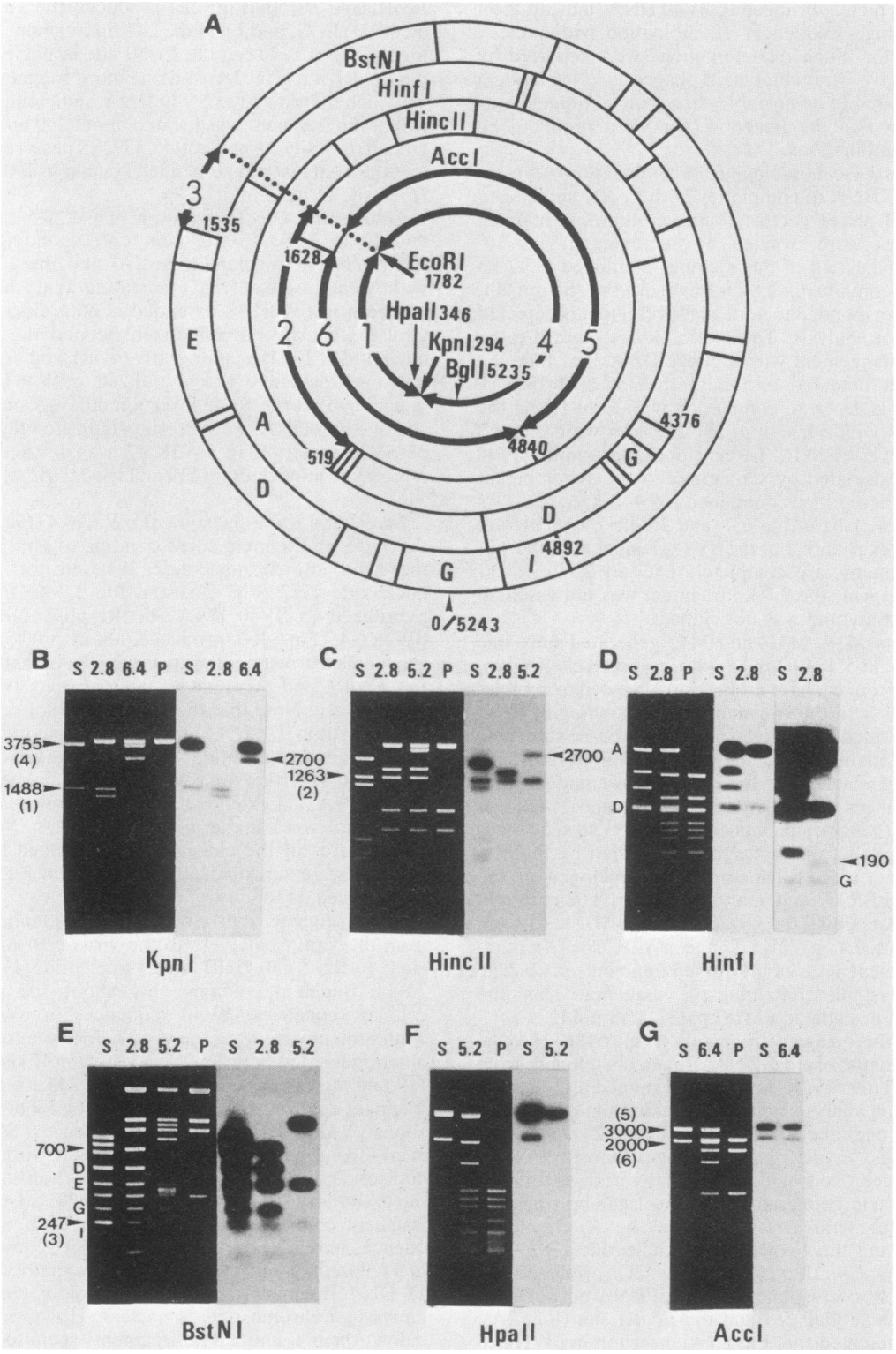

FIG. 2. Restriction enzyme mapping of cloned DNA from W-2K-11 cells. (A) Restriction map of SV40 DNA. Restriction fragments and nucleotide numbers are indicated. The fragments are: 1, *Eco*RI-*Kpn*I fragment (1,488 bp); 2, *Eco*RI-*Hin*cII fragment (1,263 bp); 3, *Eco*RI-*Bst*NI fragment (247 bp); 4, *Eco*RI-*Kpn*I fragment (3,755 bp); 5, *Eco*RI-*Acc*I fragment (3,000 bp); 6, *Acc*I-*Acc*I fragment (2,000 bp). (B) Restriction endonuclease digestion patterns and Southern blot analysis of the recombinant plasmids pW2K2.8, pW2K5.2, and pW2K6.4. DNAs were digested with *Eco*RI plus the indicated restriction endonuclease. The digests were electrophoresed, stained with ethidium bromide (white on black panels), transferred to nylon membrane, and hybridized with ^{32}P-labeled SV40 DNA (black on white panels). The DNAs used were SV40 (S), pW2K2.8 (2.8), pW2K5.2 (5.2), pW2K6.4 (6.4), and pBR322 (P). The lengths of the fragments in bp are indicated.

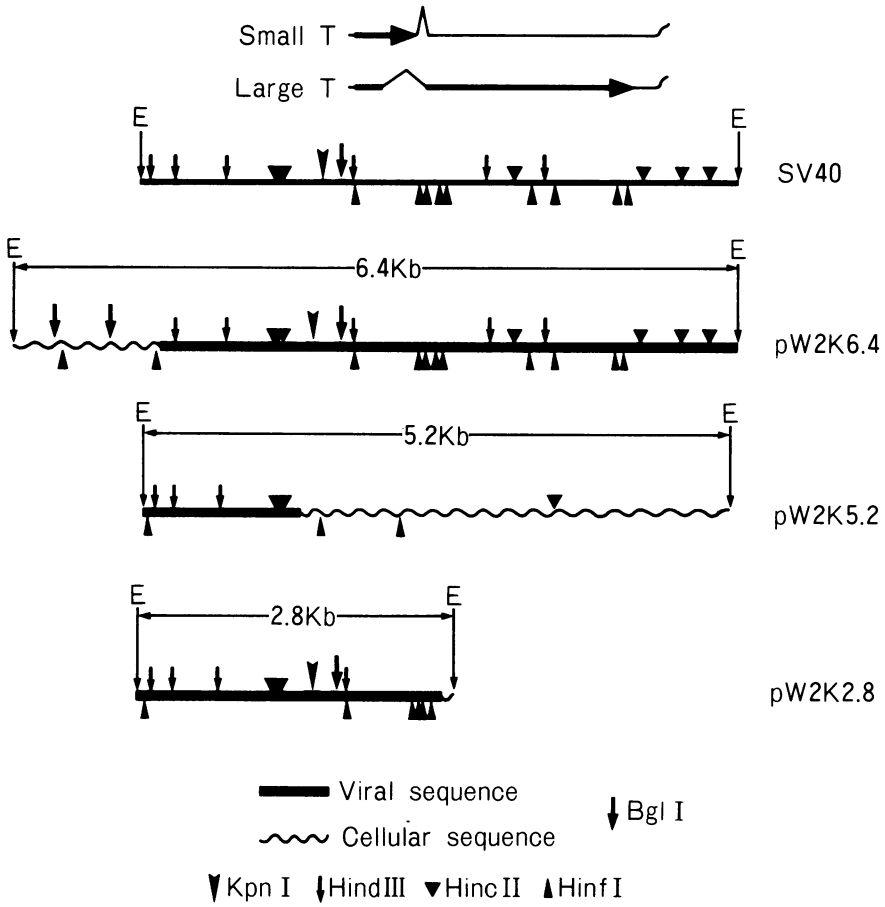

FIG. 3. Structures of three cloned fragments from W-2K-11 cells. The restriction map of SV40 DNA linearized at the single *Eco*RI site and the coding regions for large and small T antigens as well as the direction of the viral early transcripts are indicated at the top of the figure. Caret symbols represent the sites of splicing and the thin, wavy lines at the 3′ end of the transcripts represent polyadenylate. Physical maps of the cloned fragments were deduced from some of the data shown in Fig. 2. All maps are drawn such that the direction of early SV40 transcription is from left to right. On the basis of this conversion, we refer in the text to the left and right ends of the cloned fragments. E indicates the *Eco*RI site.

1.35 kb. *Kpn*I digestion of pM31 produced only one fragment of 8.35 kb, indicating that there is only one *Kpn*I site in pM31. *Eco*RI plus *Kpn*I digestion of these three recombinant plasmids (Fig. 4B) produced three fragments of 4.3, 2.7, and 1.35 kb. The 2.7- and 1.35-kb fragments hybridized to SV40 DNA sequences, and the 4.3-kb fragment corresponded to the unit length of pBR322. In pM18 and pM42, the total lengths of 2.7 and 1.35 kb are smaller than that of the inserted *Eco*RI fragment (5.4 kb) by 1.35 kb, suggesting that the 1.35-kb fragment was a doublet. These results indicated that the order of these fragments in the cloned DNA in pM18 and pM42 is 2.7-kb *Eco*RI/*Kpn*I fragment–1.35-kb *Kpn*I/*Kpn*I fragment–1.35-kb *Kpn*I/*Eco*RI fragment. If the 1.35-kb *Kpn*I/*Eco*RI fragment has no SV40 sequences, the relative radioactive intensity of the two bands in Fig. 4A must be equal to that of the two bands in Fig. 4B. The results obtained showed that this was not the case, indicating that the outside 1.35-kb fragment contained SV40 sequences. If SV40 DNA insertion is not interrupted by cellular sequences, two *Kpn*I sites separated by 1.35 kb are in SV40 sequences, and the SV40 insertions in pM18 and pM42 are not colinear with SV40 virion DNA, which has only one *Kpn*I site. As suggested previously (19), SV40 DNA insertions in M18 and M42 appeared to be identical. We confirmed this assumption by digestion of pM18 and pM42 with *Eco*RI and either *Hin*fI or *Hin*dIII. The electrophoretic patterns of DNA fragments generated were identical (Fig. 4C).

To analyze the SV40 sequences present in pM18, pM31, and pM42 in detail, DNAs were digested with *Bst*NI, which cuts SV40 and pBR322 DNA at 16 and 6 sites, respectively. The cloned DNAs in pM18 and pM42 were inserted in the reverse orientation to identify the fragments containing the end of the cloned DNA and the adjacent pBR322 sequences. All the plasmids produced the viral *Bst*NI fragments D, E, G, and I (Fig. 4D). The intensity of these fragments stained with ethidium bromide suggested that fragments D and E in pM18 and pM42 were doublets or triplets, whereas these fragments in pM31 appeared to be singlets. The result indicated that the SV40 insertion in pM31 is colinear with SV40 virion DNA and that the 1.35-kb sequence containing the *Bst*NI fragments D and E is duplicated in pM18 and pM42. Since, pM18 differs from pM42 only in the orientation of the insertion, two fragments not common to both of them must be the fragments containing the end of the cloned DNA and the adjacent pBR322 sequences. Such fragments are the 2.0- and 0.83-kb fragments in pM18 and the 2.4-kb fragment in pM42 (Fig. 4D). Because the total lengths of these

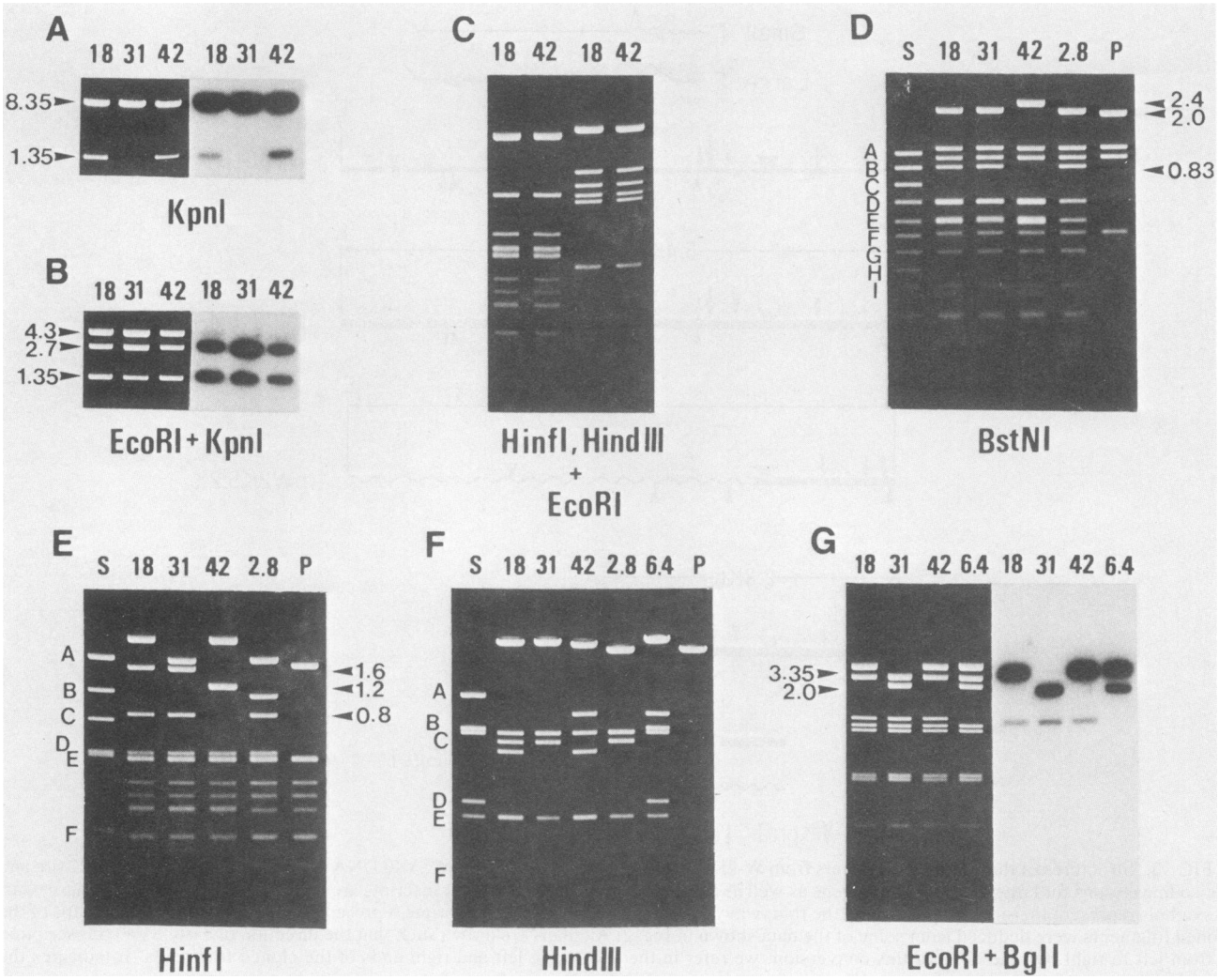

FIG. 4. Restriction enzyme mapping of cloned DNA from revertants M18, M31, and M42. DNAs were digested with the restriction endonuclease indicated below the panel. The digests were electrophoresed, stained with ethidium bromide (white on black panels), transferred to nylon membrane, and hybridized with ^{32}P-labeled SV40 DNA (black on white panels in A, B, and G). The DNAs in each panel are: pM18 (lane 18), pM31 (lane 31), pM42 (lane 42), pW2K2.8 (lane 2.8), pW2K6.4 (lane 6.4), SV40 (lane S), and pBR322 (lane P). The lengths of the fragments are indicated in kb.

fragments must be equal, pM42 must have a fragment of about 0.4 kb which may comigrate with the viral *Bst*NI fragment E.

The two fragments of 2.0 and 0.83 kb present in pM18 were also present in pM31 (Fig. 4D). The same relations were also seen by digestion of pM18 and pM31 with either *Hin*fI (Fig. 4E) or *Hin*dIII (Fig. 4F). These results indicated that both ends of the cloned *Eco*RI fragments in pM18 and pM31 were identical. Figure 4D, E, and F also shows that one of the fragments containing the end of the cloned fragments was also present in pW2K2.8, indicating that one end of the cloned DNA in pW2K2.8 was identical to those of pM18 and pM31. The 2.7-kb *Eco*RI-*Kpn*I fragment containing the other end of the cloned DNAs was present in pW2K6.4, pM18, pM31, and pM42 (Fig. 2B and 4B). The same relations were seen by digestion of pW2K6.4, pM18, pM31, and pM42 (Fig. 4G) with *Eco*RI and *Bgl*I, in which the *Bgl*I fragment containing the viral-host junction is 2.0 kb in pW2K6.4 and pM31 and 3.35 kb in pM18 and pM42, owing to the 1.35-kb duplication. These data suggested that the other end of the cloned fragment was also identical in pW2K6.4, pM18, pM31, and pM42.

The structures of the cloned DNAs in pM18, pM31, and pM42 are presented schematically in Fig. 5 together with the structure of the joint between the SV40 sequences within the 6.4- and 2.8-kb fragments. These structures can be summarized as follows. SV40 sequences in the 5.4-kb fragments of M18 and M42 extend clockwise from a virus-host junction located close to the SV40 *Hin*fI site at nucleotide 4376 to another virus-host junction located close to the SV40 *Acc*I site at nucleotide 1628 through a virus-virus junction. There is about 1.35 kb of duplication approximately between nucleotides 150 and 1500 of the SV40 DNA in the fragments. The SV40 sequence in the cloned 4.0-kb fragment in pM31 is colinear with SV40 virion DNA; however, both ends of the cloned fragment and the virus-host junctions are identical to those of the 5.4-kb fragments in pM18 and pM42. Furthermore, the left ends of the fragments in these plasmids are identical to the left end of the 6.4-kb fragment in pW2K6.4, and the right ends of the fragments in these plasmids are

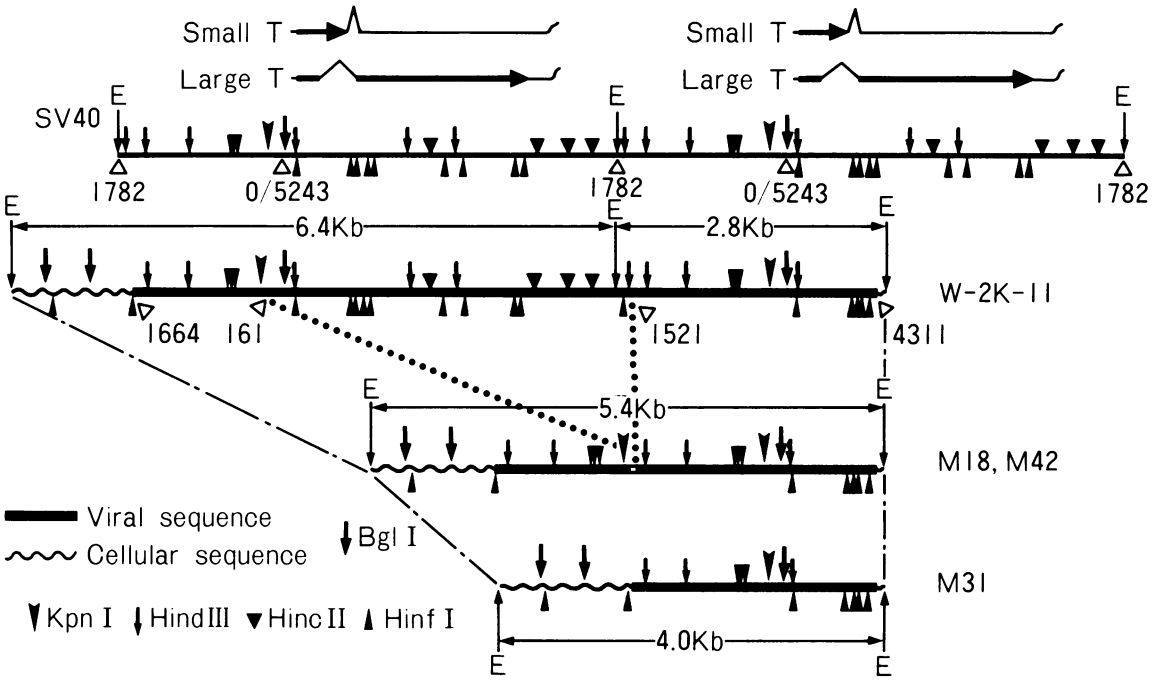

FIG. 5. Structures of cloned fragments from revertants M18, M31, and M42 and that of corresponding fragments from W-2K-11 cells. The restriction map of SV40 DNA linearized at the *Eco*RI site is shown at the top. Two copies are arranged in head-to-tail tandem arrays. The conversions are the same as in Fig. 3. The numbers below the line indicate the nucleotide numbers according to SV numbering system (30). Physical maps of the cloned fragments were deduced from some of the data shown in Fig. 4. Structures of the fragments cloned from M18 and M42 are identical to those described in the text. The clockwise direction is indicated as right to left in the text.

identical to the right end of the 2.8-kb fragment in pW2K2.8.

These results indicate that the 6.4- and 2.8-kb fragments cloned from W-2K-11 cells are adjacent to each other; i.e., the viral *Eco*RI sites of both fragments originated from identical *Eco*RI sites located inside the SV40 DNA. Thus, W-2K-11 cells contain about 1.5 copies of a partially tandem array of SV40 DNA and contain an intact early transcription unit. This structure in W-2K-11 DNA explains the precise excision of infectious SV40 after fusion with simian cells as reported previously (19). All the revertants do not contain an intact early transcription unit and cannot code for the intact large T antigen.

DNA sequences of the junctions. The SV40 sequences in revertants M18, M31, and M42 were generated by a deletion(s) within the integrated SV40 sequences in the transformed cell line W-2K-11. To analyze the patterns of integration and deletion of viral sequences, the base sequences were determined at two virus-host junctions which are present in all of these cell lines and at one virus-virus junction which is present in the revertant cell lines M18 and M42. The top panel of Fig. 6 shows the positions and directions in which the sequences were determined. The 200-bp *Hin*dIII-*Hin*fI fragment (Fig. 6A) containing the left-hand junction was obtained from pM18 and pM42, and the 180-bp *Eco*RI-*Hin*fI fragment (Fig. 6C) containing the right-hand junction was obtained from pM18 and pW2K2.8. The 160-bp *Kpn*I-*Hin*dIII fragment (Fig. 6B) containing the virus-virus junction was obtained from pM18. These fragments were end labeled, and their sequences were determined by the method of Maxam and Gilbert (20).

At the left-hand junction, SV40 sequences stretch to nucleotide 1664 and then nonviral sequences begin. The flanking cellular sequences are composed almost completely of guanines and cytosines, and they are organized in an unusual repeating sequence of CCG. At the right-hand junction, SV40 sequences stretch to nucleotide 4311 and then nonviral sequences begin. Contrary to the sequences of the left-hand junction, the first 53 nucleotides in the flanking cellular sequences are composed of adenines and thymines except for one cytosine, and they are also organized in a strikingly unusual repeating sequence of AAAT. As shown in Fig. 6C, there is patchy homology between the cellular flanking sequences and the corresponding parental SV40 sequences which were displaced by recombination. The sequences at the left- and right-hand junctions were identical in pM18 and pM42 and in pM18 and pW2K2.8, respectively. The physical maps of the cloned DNA in pW2K6.4, pW2K2.8, pM18, pM42, and pM31 were also identical around these junctions as shown above. These results strongly suggest that the sequences at both the left- and right-hand junctions in W-2K-11 cells were retained in all the revertants to the nucleotide sequence level.

DNA sequences also indicated that the SV40 insertions in all the revertants contain the coding region for viral capsid proteins VP2 and VP3 and, unexpectedly, for small T antigen. The right-hand junction at nucleotide 4311 is located downstream of the small T antigen coding region and the 5′ and 3′ splice sites but upstream of the mRNA polyadenylation signal. Since the right-hand flanking sequences are composed of exclusively A+T sequences, there are five consensus polyadenylation signals of AATAAA. In spite of their normal morphology, these three revertant cell lines seemed to synthesize SV40 small T antigen by using these cellular polyadenylation signals. This assumption was confirmed by immunoprecipitation analysis of the cell extracts prepared from the revertants. As shown in Fig. 7, small T antigen was detected in all the revertants, but intact large T antigen was not. The results indicate that SV40 small T antigen is not involved in the maintenance of transformed state.

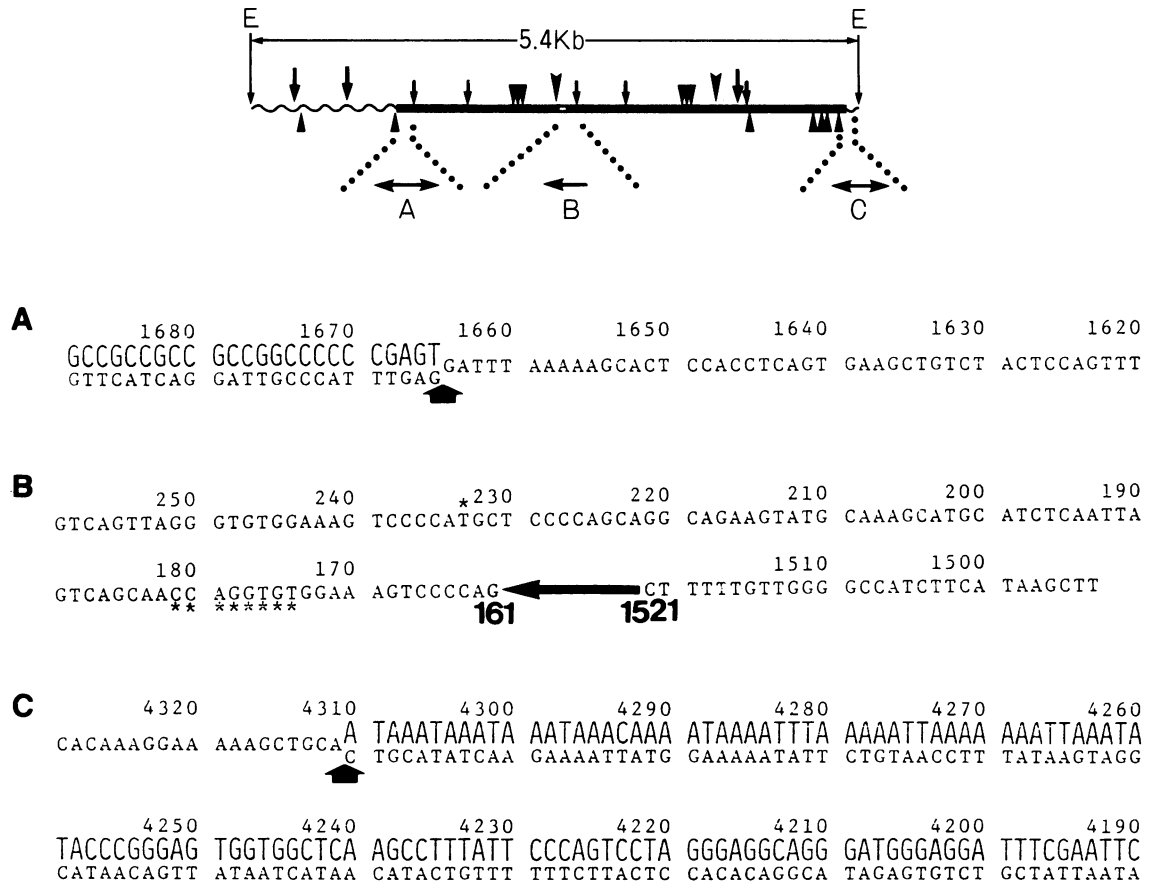

FIG. 6. Sequences at the virus-host junctions. The structure of the cloned fragment in pM18 (pM42) is shown at the top of the figure. The directions of sequencing at the left- and right-hand virus-host junctions and at the virus-virus junction are shown by the arrows A, C, and B. (A) The arrow indicates the left junction. The top line of doublet region, shown in large letters, is the flanking cellular sequences, and the bottom line shown in small letters represents the displaced viral sequences. (B) The sequences around the virus-virus junction in pM18 are shown. The arrow indicates the sequences deleted. The nucleotides underscored with asterisks are deleted, and the nucleotide overscored with an asterisk at 233 is substituted as compared with the published sequence of strain 776 (30) as described in the text. (C) The arrow indicates the right junction. The top line of doublet region is the flanking cellular sequences, and the bottom line represents the displaced viral sequences.

The sequences around the virus-virus junctions in pM18 indicate that the nucleotide at 161 joins to the nucleotide at 1521, deleting an internal 3.9 kb of viral sequences in which the intact early transcription unit is located. The parental SV40 sequences around the deletion site 161/1521 indicate that there is almost no homology between the two regions taking part in the joining, but there is a possible secondary structure which might reduce the effective distance between nucleotides 161 and 1521 (see below). There is a second deletion (eight bases between nucleotides 174 and 181) and one substitution (G → T at nucleotide 233) near the deletion site as compared with published sequences of SV40 DNA strain 776. As described above, W-2K-11 cells were established by infection of mouse kidney cell line with SV40 virions strain 777. To distinguish whether these differences had been generated concomitant with the large deletion or originated from the difference in sequences between strains 776 and 777, both SV40 DNAs were electrophoresed after digestion with *Bst*NI, whose recognition sequences are located only at the deletion and substitution sites. The result indicated that there is an eight-base deletion in strain 777 as compared with the published DNA sequence of strain 776, but there is no base substitution at nucleotide 233.

DISCUSSION

In the present paper, the mechanism of gene rearrangement with respect to viral sequences in SV40-transformed mouse cells was studied by cloning the DNA fragments containing viral sequences from the transformed cell line W-2K-11 and its revertants M18, M31, and M42. In W-2K-11 cells, about two genome equivalents per cell were serially arranged within about 30 kb, with at least two intervening cellular sequences between the viral sequences (19). To determine the relative positions of these viral sequences, large DNA fragments generated by partial *Eco*RI digestion were cloned into phage vector Charon 4A. However, the cloning of the fragments corresponding to those detected by blot analysis was unsuccessful due to recombination between the homologous sequences during propagation of the recombinant phages. The unstable feature of recombinant phages carrying tandem duplication was also reported by Clayton and Rigby (7).

The cellular DNA fragments were therefore cloned after complete digestion with *Eco*RI to minimize the homologous recombinations within λ phages. The physical maps of the fragments cloned from W-2K-11 cells and the revertants, however, revealed that two of the *Eco*RI fragments contain-

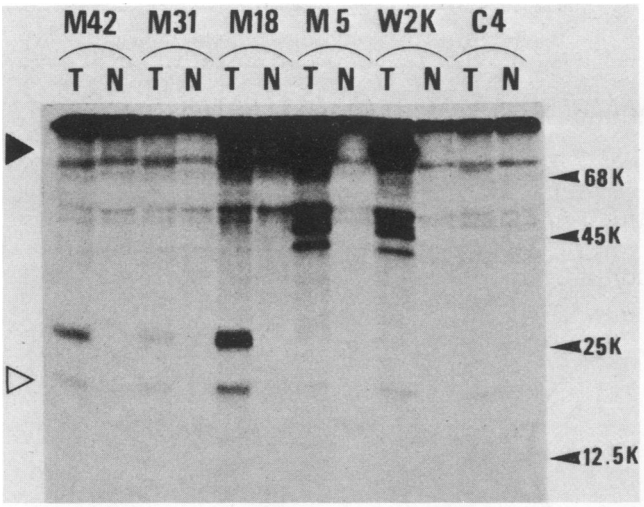

FIG. 7. Immunoprecipitation analysis of SV40 antigens synthesized in revertants. The cell extracts labeled with [^{35}S]methionine were precipitated with either anti-T hamster serum (T) or normal hamster serum (N) as described in the text. The antigens eluted from the immunoprecipitates were analyzed by electrophoresis on a sodium dodecyl sulfate–14% polyacrylamide gel, followed by fluorography. The black and white arrows indicate the positions of large and small T antigens, respectively. Molecular weights were estimated from the positions of the following marker proteins: bovine serum albumin, 68,000 (68K); ovalbumin, 45K; chymotrypsinogen A, 25K; cytochrome *c*, 12.5K. Abbreviations: C4, normal parental cell line C3H2K-C4; W2K, W-2K-11 cells; M5, large T antigen-positive revertant; M18, M31, M42, large T antigen-negative revertants.

ing the virus sequences (6.4 and 2.8 kb) are adjacent, comprising 1.5 copies of the SV40 sequences in a partially tandem duplicated form (Fig. 5). The remaining two fragments (9.0 and 5.2 kb) containing much shorter SV40 sequences may be located close to the 1.5-copy fragment.

The sequence data at the SV40-host junctions in a number of different cell lines (5, 26, 27) and in SV40 evolutionary variants (9–12) have suggested that there is no specific integration site in either cellular or virus DNA, but there is some preference for A- and T-rich sequences in the SV40 DNA at the junction and patchy homology between the viral and cellular sequences. In W-2K-11 cells, the one flanking sequence is a long stretch of adenines and thymines and the other is a stretch of guanines and cytosines, and these flanking sequences have characteristic repeats of AAAT and CCG, respectively. Similar features were reported in SV40-transformed rat cell lines in which the flanking sequences are composed of a simple sequence of CA repeat or a stretch of adenines and guanines (26). These simple and specific sequences are not present at every flanking cellular sequence at SV40-host junctions so far sequenced; however, the high frequency of appearance seems to suggest some contribution of these sequences to integrative recombination or merely reflect the frequent occurrence of these simple sequences in mammalian cells. The cluster of A+T might facilitate the formation of the initial recombination owing to instability of hydrogen bonds between adenine and thymine as suggested by Gutai and Nathans (11).

It was suggested that the virus-host junctions appear to be clustered in the 5′ half of the late transcription unit between nucleotides 300 and 1100 (7). In W-2K-11 cells, the SV40 nucleotide numbers at the left and right virus-host junctions are 1664 and 4311, and those numbers at the virus-virus junctions in revertants M18 and M42 are 1521 and 161. These four sites at the junction all relate to the junctions in the transformed cells and the evolutionary variants previously reported by other investigators. The left virus-host junction at nucleotide 1664 is located close to the right virus-host junction at nucleotide 1668 in the SV40-transformed rat cells 14B (5). The right junction at nucleotide 4311 is located close to the virus-virus junction at nucleotide 4306 in the evolutionary variant ev1114 (9). The virus-virus junction at nucleotide 161 is located close to the virus-virus junction at nucleotide 157 in ev1117 (9), and the junction at nucleotide 1521 is located close to the virus-host junction at nucleotide 1518 in the transformed rat cells SVRE17 (26). The distance between these related junctions is only three to five bases. These relations might be an accident or reflect the selective advantage in an integration at the late region to maintain the transformed state or selective replicative advantage on the variants, but these reasons alone cannot account for the above similarities. These data suggest that there are specific sites in SV40 DNA sequences to facilitate integrative recombination and rearrangement. The viral sequences in the preferred recombination sites might be useful for the introduction of exogenous DNA into the cellular chromosomes. It should be noted that the viral sequences preferentially used for integrative recombination with cellular sequences are also used for the formation of virus-virus junctions. These results suggest that virus integration into cellular sequences and the formation of virus-virus junction might be mediated identically or that virus-host junctions do not reflect the initial integration site but reflect the secondary rearrangements after the initial integration event or both.

The parental SV40 sequences displaced around the virus-virus junction in M18 and M42 share little homology but have sequences capable of forming a secondary structure which might reduce the effective distance between nucleotides 161 and 1521. A few potential secondary structures can be constructed between two 50-nucleotide sequences around nucleotides 161 and 1521. The structure shown in Fig. 8 is the most stable structure deduced from calculations by the method of Tinoco et al. (29). The relation between the deletion at the repeated sequences and the secondary structures has been inferred in *E. coli* (1) and in the human β-globin gene family (8) in which "slipped mispairing" during DNA replication might be involved in the deletion. As shown in Fig. 8, there is a short repeat near the secondary structure. If DNA replication was started from the flanking cellular sequences or from the SV40 replication origin, the single-stranded intermediate could form this secondary structure.

Figure 9 summarizes the structures of the SV40 sequences present in the transformed and revertant cells. There are two types of deletion during the reversion of W-2K-11 cells. One deletion is seen in M31 in which the SV40 insertion is colinear with SV40 virion DNA as a result of a deletion between homologous sequences within the duplicated SV40 DNA region. This result suggests that even in nonpermissive cells a precise excision of SV40 DNA occurs without "a permissive factor(s)" supplied by fusion with permissive monkey cells. The permissive factors may be required only for the replication of the excised DNA and not for the excision of viral DNA from the cellular genomes. Lania et al. suggested similar possibilities in polyoma virus-transformed cells (14). The other deletion is seen in M18 and M42 in which the SV40 DNA insertion is not colinear with SV40 virion DNA as a result of deletion between nonhomologous regions (nucleotides 161 and 1521) within the SV40 sequences.

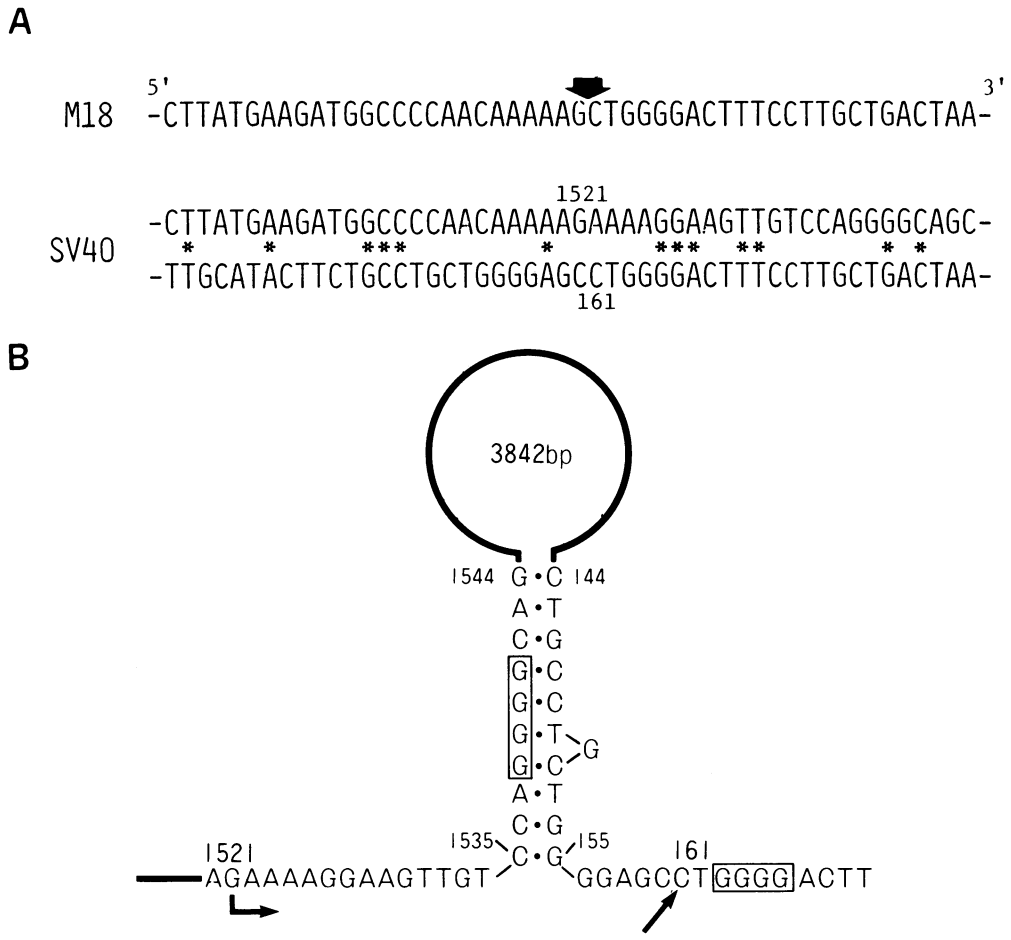

FIG. 8. Nucleotide sequences and a possible secondary structure around the virus-virus junction in revertant M18. (A) SV40 sequences around the junction (nucleotides 1521/161) are compared with those in two regions taking part in the joining. The junction is indicated by the arrow, and sequences common to the two regions are indicated by asterisks. (B) Potential secondary structures around the junction were screened in the sequences. The most stable secondary structure deduced from calculation by the method of Tinoco et al. (29) is shown. The short repeat sequences are indicated by the box.

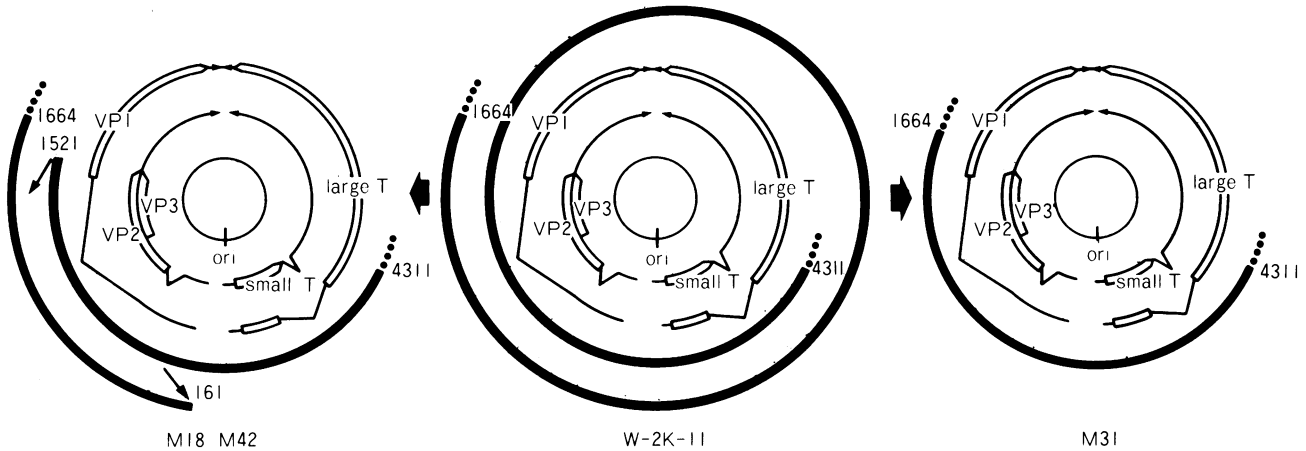

FIG. 9. Schematic representation of viral sequences integrated in W-2K-11 cells and the revertants. The viral sequences integrated into a cellular chromosome are shown by closed bars. The dotted lines represent the flanking cellular sequences. In W-2K-11 cells, about 1.5 copies of viral sequence are integrated in tandem array with partially duplicated sequences. The deletion of viral sequences had occurred in two ways; one occurred between homologous sequences, displacing a unit length of SV40 DNA (M31), and the other occurred between nonhomologous sequences, displacing an internal 3.9 kb of viral sequence (M18, M42). The virus-virus junction (nucleotide numbers 161→1521) is shown by small arrows.

The SV40 sequences retained in M18 and M42 are identical despite the fact that they were isolated independently and have different morphology. Two possible reasons for the occurrence of revertants which have an identical deletion pattern can be considered. (i) In parental W-2K-11 cells, there may be a prototype of the revertants which has some advantage of existence, e.g., a high efficiency of plating. (ii) The deletion between nucleotides 161 and 1521 occurs preferentially by forming the secondary structure as stated above.

Nucleotide sequences of SV40 DNA and flanking cellular sequences in revertants M18, M31, and M42 predicted the expression of small T antigen by using cellular polyadenylation signals (AATAAA) in the flanking sequences. In these revertants, the SV40 insertion stretches to nucleotide 4311 and contains the small T antigen coding region and the 5′ and 3′ splice sites for 19S mRNA but lacks its polyadenylation signal. However, there are five consensus polyadenylation signals (AATAAA) within the adjacent cellular sequences. The 19S mRNA for small T antigen must be polyadenylated by using one of these signals. All the revertants isolated from W-2K-11 cells expressed small T antigen, suggesting a possible role for small T antigen in the selective advantage of these revertants. These revertants might help clarify the functions of SV40 small T antigen.

ACKNOWLEDGMENTS

We thank Hideo Ideda, Yukio Mishima, Hiroyuki Kato, and Hiroko Sato for expert advice on in vitro packaging, colony hybridization, base sequencing, and small T antigen assay, respectively. We also thank Nobuo Yamaguchi and Sumio Sugano for valuable discussions.

This work was supported by grants from the Ministry of Education, Science, and Culture of Japan and by grants from the Mitsubishi Foundation.

LITERATURE CITED

1. **Albertini, A. M., M. Hofer, M. P. Calos, and J. H. Miller.** 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell **29**:319–328.
2. **Benton, W. D., and R. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques. Science **196**:180–182.
3. **Blattner, F. R., A. E. Blechl, K. D. Thomson, H. E. Faber, J. E. Richards, J. L. Slightom, P. W. Tucker, and D. Smithies.** 1978. Cloning human fetal γ globin and mouse α-type globin DNA: preparation and screening of shotgun collections. Science **202**:1279–1284.
4. **Blattner, F. R., B. G. Williams, A. E. Blechl, K. D. Thompson, H. E. Feber, L. A. Furlong, D. J. Grunwald, D. O. Kieffer, D. D. Moore, J. W. Schumm, E. L. Sheldon, and O. Smithies.** 1977. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science **196**:161–169.
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. **Botchan, M., W. Topp, and J. Sambrook.** 1976. The arrangement of simian virus 40 sequences in the DNA of transformed cells. Cell **9**:269–287.
7. **Clayton, C. E., and P. W. J. Rigby.** 1981. Cloning and characterization of the integrated viral DNA from three lines of SV40-transformed mouse cells. Cell **25**:547–559.
8. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
9. **Gutai, M. W.** 1981. Recombination in SV40-infected cells: nucleotide sequences at viral-viral recombinant joints in naturally arising variants. Virology **109**:344–352.
10. **Gutai, M. W., and D. Nathans.** 1978. Evolutionary variants of simian virus 40: nucleotide sequence of a conserved SV40 DNA segment containing the origin of viral DNA replication as an inverted repetition. J. Mol. Biol. **126**:259–274.
11. **Gutai, M. W., and D. Nathans.** 1978. Evolutionary variants of simian virus 40: cellular DNA sequence and sequences at recombinant joints of substituted variants. J. Mol. Biol. **126**:275–288.
12. **Hiscott, J. B., D. Murphy, and V. Defendi.** 1981. Instability of integrated viral DNA in mouse cells transformed by simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **78**:1736–1740.
13. **Ketner, G., and T. J. Kelly.** 1976. Integrated simian virus 40 sequences in transformed cell DNA: analysis using restriction endonucleases. Proc. Natl. Acad. Sci. U.S.A. **73**:1102–1106.
14. **Lania, L., S. Boast, and M. Fried.** 1982. Excision of polyoma virus genomes from chromosomal DNA by homologous recombination. Nature (London) **295**:349–350.
15. **Lavi, S., and E. Winocour.** 1972. Acquisition of sequences homologous to host deoxyribonucleic acid by closed circular simian virus 40 deoxyribonucleic acids. J. Virol. **9**:309–316.
16. **Lee, T. N. H., W. W. Brockman, and D. Nathans.** 1975. Evolutionary variants of simian virus 40: cloned substituted variants containing multiple initiation sites for DNA replication. Virology **66**:53–69.
17. **Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eucaryotic DNA. Cell **15**:687–701.
18. **Maruyama, K., T. Furuta, K. Ueda, and K. Oda.** 1982. Properties of T antigen-positive phenotypic revertants isolated from SV40-transformed rat and mouse cells. Microbiol. Immunol. **26**:31–40.
19. **Maruyama, K., T. Hiwasa, and K. Oda.** 1981. Characterization of flat revertant cells isolated from simian virus 40-transformed mouse and rat cells which contain multiple copies of viral genomes. J. Virol. **37**:1028–1043.
20. **Maxam, A. M., and A. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
21. **McCutchan, T., M. Singer, and M. Rosenberg.** 1979. Structure of simian virus 40 recombinants that contain both host and viral DNA sequences. J. Biol. Chem. **254**:3592–3597.
22. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–252.
23. **Sambrook, J., R. Green, J. Stringer, T. Mitchson, S. L. Hu, and M. Botchan.** 1980. Analysis of the sites of integration of viral DNA sequences in rat cells transformed by adenovirus 2 or SV40. Cold Spring Harbor Symp. Quant. Biol. **44**:569–584.
24. **Sambrook, J., H. Westphal, P. R. Srinivasan, and R. Dulbecco.** 1968. The integrated state of viral DNA in SV40-transformed cells. Proc. Natl. Acad. Sci. U.S.A. **60**:1288–1295.
25. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–515.
26. **Stringer, J. R.** 1981. Integrated simian virus 40 DNA: nucleotide sequences at cell-virus recombinant junctions. J. Virol. **38**:671–679.
27. **Stringer, J. R.** 1982. DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. Nature (London) **296**:363–366.
28. **Tiemeier, D., L. Enquist, and P. Leder.** 1976. Improved derivative of a phage λ EK2 vector for cloning recombinant DNA. Nature (London) **263**:526–527.
29. **Tinoco, I., O. C. Uhlenbeck, and M. D. Levine.** 1971. Estimation of secondary structure in ribonucleic acids. Nature (London) **230**:362–367.
30. **Tooze, J.** 1980. The molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Vogelstein, B., and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. U.S.A. **76**:615–619.