

Somatic Cells Efficiently Join Unrelated DNA Segments End-to-End

JOHN H. WILSON,^{1*} PETER B. BERGET,² AND JAMES M. PIPAS³

Marrs McLean Department of Biochemistry, Baylor College of Medicine,¹ and Department of Biochemistry and Molecular Biology, University of Texas Medical School and Graduate School of Biomedical Science,² Houston, Texas 77030, and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260³

Received 11 May 1982/Accepted 28 June 1982

Molecular substrates for probing nonhomologous recombination in somatic cells were constructed by inserting pBR322 sequences at selected sites on the simian virus 40 (SV40) genome. The chimeric products are too large to be packaged into an SV40 capsid. Therefore, production of viable progeny requires that most of the pBR322 sequences be deleted without altering any SV40 sequences that are essential for lytic infection. As judged by plaque assay, these recombination events occur at readily detectable frequencies after transfection into CV1 monkey kidney cells. Depending on the site of pBR322 insertion, the infectivities of the full-length circular or linear chimeras ranged from 0.02 to 2% of the infectivity of linear wild-type SV40 DNA. Nucleotide sequence analysis of several recombinant progeny revealed three distinct classes of recombination junction and indicated that the causative recombination events were minimally dependent on sequence homology. Potential mechanisms involving recombination at internal sites or at ends were distinguished by measuring the infectivity of chimeric molecules from which various lengths of pBR322 had been removed. These data support end-to-end joining as the primary mechanism by which DNA segments recombine nonhomologously in somatic cells. This end joining appears to be very efficient, since SV40 genomes with complementary single-stranded tails or with short non-complementary pBR322 tails were comparably infectious. Overall, this study indicates that mammalian somatic cells are quite efficient at the willy-nilly end-to-end joining of unrelated DNA segments.

Eucaryotic genomes are exceptionally plastic. The rich variety of recombination events that have been characterized can be grouped conveniently into homologous events, which depend on extensive sequence homology, and nonhomologous events, which depend on limited sequence homology. Homologous recombination during meiosis in eucaryotes and during sexual exchanges in procaryotes shuffles the alleles present in the population into new combinations in individual members but does not alter the order of genes present on the parental chromosomes. By contrast, nonhomologous recombination can create new sequence arrangements. Recombination events in this class can be site specific as are the integration of lambdoid phages, the rearrangement of variable and constant genes for both light and heavy chains in B-cell development, and the class switching of heavy chains (4, 8, 20, 26); they can occur between sites with short and imperfect sequence homology as is characteristic of many deletion and duplication events in procaryotes (10, 11,

28); or they can be nearly independent of sequence homology as is characteristic of the recombination events mediated by mobile genetic elements such as transposons (6).

DNA that is transfected into somatic mammalian cells is actively recombined, apparently by homologous and nonhomologous mechanisms (2, 14, 27, 30, 42). These results imply an assortment of enzymatic activities and a variety of potential recombination pathways. The genomes of animal viruses such as simian virus 40 (SV40) and adenovirus have proven quite useful for characterizing cellular recombinational capabilities (7, 9, 12, 13, 16, 35, 36, 39, 40, 43, 45). One advantage of these viruses is that their biological activity allows one to follow selectively the very small fraction of DNA molecules that retain function. For example, only about 1 in 10⁶ molecules of linear SV40 genomes transfected by DEAE-dextran initiates a successful infection and produces progeny virus. Results from studies with this small, but relatively well-defined, population of molecules allow one to

distinguish among potential cellular pathways of recombination.

We are using the animal virus SV40 to probe and characterize the recombination capabilities of mammalian somatic cells. Our previous experiments using *in vitro*-constructed tandem oligomers of SV40 have demonstrated that intramolecular homologous recombination occurs efficiently in transfected SV40 DNA and with approximate uniformity in repeated regions of the genome (39, 40). These results and others argue strongly against either gene conversion or site-specific recombination as the primary mechanism of recombination in SV40 DNA. Homologous recombination also occurs between different molecules, but at a much lower frequency (35, 36, 39; C. T. Wake, F. Vernaleone, and J. H. Wilson, manuscript in preparation). Among other things, these studies have demonstrated that intramolecular recombination events can be assayed selectively in transfected DNA by using suitably low DNA concentrations.

In this paper we report our genetic analysis of the intramolecular somatic recombination events that do not depend on extensive sequence homology. We have constructed chimeric molecules that are too large to be packaged into an SV40 capsid by inserting pBR322 sequences at selected sites on the SV40 genome. Production of viable progeny from such chimeras requires that most of the pBR322 sequences be deleted without altering any SV40 sequences that are essential for lytic infection. The biological constraints of packageability and genetic integrity define a target in which recombination events can be detected and analyzed. By locating these targets in nonessential regions of the SV40 genome, we have been able to study recombination between grossly nonhomologous DNA segments. We have analyzed the population of such events by plaque assay of viable progeny and have examined individual progeny by nucleotide sequencing. The results indicate that somatic monkey cells join unrelated DNA segments end-to-end with a high relative efficiency.

MATERIALS AND METHODS

Cells and viruses. The origins and procedures for growth of the established monkey kidney CV1 cell line have been described previously (41). The titers of the viruses were determined by plaque assay on CV1 cells as previously described (41). Wild-type SV40 is strain Rh911. The host range (hr) mutant hrBC1906 was derived from the wild-type strain (41). The host range mutation does not alter plaquing efficiency on CV1 cells, and its presence is immaterial to these experiments.

DNA preparation. SV40 DNA was extracted accord-

ing to Hirt (15) from CV1 cells which had been infected at a multiplicity of about 0.005 SV40 PFU per cell, when about 75% of the cells showed signs of infection. Closed-circular SV40 DNA was purified directly from the Hirt supernatant by adding CsCl to 1.56 g/cm³ and ethidium bromide to 200 µg/ml and centrifuging to equilibrium. The closed-circular DNA band was collected, and the ethidium bromide was removed by extraction with CsCl-saturated isopropanol. DNAs were phenol extracted, ethanol precipitated, and then resuspended and stored in DNA buffer (10 mM Tris [pH 7.6], 1 mM EDTA, 10 mM NaCl).

Plasmid DNA was prepared from RR1 transformants essentially as described by Bolivar and Backman (1). Cells in M9 medium were grown at 37°C to a cell density of 10⁸ cells per ml; chloramphenicol was added to 200 µg/ml, and the culture was amplified overnight. Plasmid DNA was extracted and then centrifuged to equilibrium in CsCl gradients containing ethidium bromide at 200 µg/ml. The closed-circular plasmid band was collected, and the ethidium bromide was removed by extraction with CsCl-saturated isopropanol. Plasmid DNAs were phenol extracted, ethanol precipitated, and then resuspended and stored in DNA buffer.

All SV40 and plasmid DNA preparations were labeled *in vivo* with [³H]thymidine (New England Nuclear Corp.). Specific activities ranged from 5 × 10³ to 75 × 10³ cpm/µg.

Miniwell preparation of DNA. Confluent CV1 cells in 96-well microtiter plates were infected with picked plaque suspensions. DNA was labeled 24 h after infection by the addition of 20 µCi of ³²P (New England Nuclear Corp.) in phosphate-free medium. Cells were harvested 5 to 10 days after infection when 50 to 75% of the cells exhibited cytopathic changes. Viral DNA was extracted as described by Hirt (15), treated with RNase (20 µg/ml), phenol extracted, ethanol precipitated, and resuspended in 0.1 × SSC (where SSC is 0.15 M NaCl plus 0.015 M sodium citrate).

Isolation of DNA from agarose gels. Agarose gel electrophoresis was performed as described previously (31) using 0.6 or 0.9% agarose in Tris-acetate buffer (40 mM Tris [pH 7.9], 5 mM sodium acetate, 1 mM EDTA). Bands of interest were cut out, dissolved in saturated KI, and passed over a 1-ml hydroxyapatite column equilibrated with saturated KI (44). The column was washed with saturated KI and then with 100 mM potassium phosphate buffer (pH 7.4), and the DNA was eluted with 400 mM phosphate buffer. Carrier RNA was added, and after dialysis against DNA buffer, the DNA was concentrated by ethanol precipitation.

DNA infections. DNA infections were essentially by the method of McCutchen and Pagano (23) using DEAE-dextran (molecular weight, 500,000, Pharmacia Fine Chemicals, Inc.) at 500 µg/ml in a volume of 0.3 ml. All DNA plaque assays were carried out on 60-mm plastic petri plates that contained freshly confluent, or slightly subconfluent, cell monolayers (1 × 10⁶ to 2 × 10⁶ cells per plate). Before infection, cell monolayers were rinsed once with TD buffer (25 mM Tris [pH 7.4], 140 mM NaCl, 5 mM KCl, 0.7 mM K₂HPO₄). Infection was carried out in TD buffer containing DNA and DEAE-dextran for 15 min with continual agitation. After infection, cell monolayers were rinsed twice with TS buffer (TD buffer plus 0.5 mM MgCl₂ and 1.0

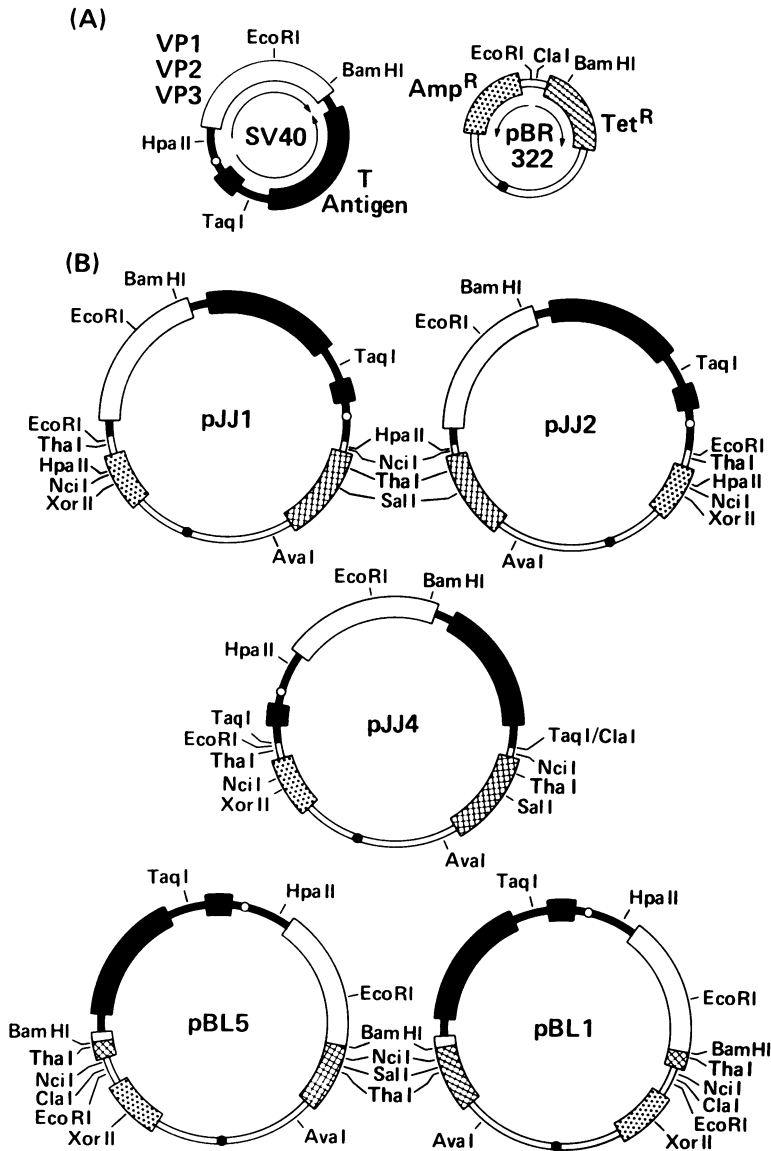


FIG. 1. Construction of the chimeric molecules. (A) SV40 and pBR322. The SV40 T antigen coding sequence is indicated by the black boxes. The small t antigen coding sequence is not shown. The late coding block for VP1, VP2, and VP3 is indicated by the open box. VP2 and VP3 are nested genes whose common C terminus overlaps slightly the N terminus of VP1 (29). The small circle represents the origin of SV40 DNA replication. On pBR322, the genes encoding ampicillin resistance and tetracycline resistance are indicated by stippled and hatched boxes, respectively. Insertion of SV40 at either the *ClaI* or *BamHI* restriction site interrupts expression of the gene for tetracycline resistance. Arrows indicate directions of transcription through these genes. (B) Chimeric molecules. Chimeras pJJ1, pJJ2, and pJJ4 were constructed by mixing and ligating *HpaII*-cut or *TaqI*-cut wild-type SV40 with *ClaI*-cut pBR322. These enzymes are isocaudamers, and their identical complementary ends can be ligated directly. The pBL1 and pBL5 chimeras were constructed by mixing and ligating *BamHI*-cut hrBC1906 SV40 with *BamHI*-cut pBR322. The second orientation of pBR322 inserted at the SV40 *TaqI* site was not found. Several restriction sites are indicated on the chimeras. *XorII*, *ClaI*, *SalI*, and *AvaI* cleave the chimeras at single sites as indicated. For the other restriction enzymes that cut pBR322, but not SV40, only the restriction sites closest to the SV40 segment are shown. These restriction enzymes were used to trim various lengths of pBR322 sequences from the chimeras (see Table 4). Note that the junctions between SV40 and pBR322 in pJJ1 and pJJ2, which were formed by ligating *HpaII*-cut SV40 with *ClaI*-cut pBR322, are not cleaved by either restriction enzyme. The junctions in pJJ4, which were formed by ligating *TaqI*-cut SV40 with *ClaI*-cut pBR322, both are cleaved by *TaqI*, and one junction also is cleaved by *ClaI*.

mM CaCl₂). Monolayers then were overlaid with agar and incubated at 37°C as in a standard plaque assay.

Recombinant DNA techniques. Restriction enzymes, T4 DNA ligase, and *Escherichia coli* DNA polymerase I (Klenow fragment) were purchased from Bethesda Research Laboratories, New England Biolabs, or Boehringer Mannheim Corp. and used according to the supplier's recommendations. *E. coli* strain RR1, a *rec*⁺ derivative of strain HB101, was obtained from Kenneth Beattie. The plasmid vector, pBR322, was the gift of Michael Mann. Transformation of RR1 was essentially as described by Mandel and Higa (21) and Bolivar and Backman (1). Selection and colony screening were on L plates supplemented with 25 µg of tetracycline or 100 µg of ampicillin per ml. Minipreps of plasmid DNA were prepared by the phenol lysis and extraction procedure of Klein et al. (17); restriction enzyme digestion was used to identify appropriate transformants.

DNA sequencing. DNA samples were sequenced by the chemical method described by Maxam and Gilbert (22). DNAs from recombinant viruses with a *Cla*I site or a unique *Taq*I site were cleaved and labeled at the 3' ends using the Klenow fragment of DNA polymerase I and [³²P]dCTP. Secondary cuts were made with *Bgl*II for *Cla*I linears and with *Hind*III for *Taq*I linears.

RESULTS

Nonhomologous recombination can be probed with specially constructed molecules. Several molecules for testing nonhomologous recombination were constructed in vitro by inserting pBR322 into selected sites on the SV40 genome as indicated in Fig. 1. Each site of insertion into the SV40 genome interrupts a different functional region and, consequently, places different constraints on the permissible sites at which recombination can generate viable progeny. Insertion of pBR322 at the unique *Hpa*II site of SV40 (pJJ1 and pJJ2) interrupts the normal leader region for late SV40 mRNA but leaves both the early and late coding blocks intact; insertion at the unique SV40 *Taq*I site (pJJ4) interrupts the gene for large T antigen in an intron region that is processed from the primary transcript (it also interrupts the coding region for small t antigen, but small t antigen is nonessential for SV40 lytic growth); and insertion at the unique SV40 *Bam*HI site (pBL1 and pBL5) interrupts the late coding block near the C-terminal end of the VP1 gene. These chimeric molecules were cloned in *E. coli*, and their structures were verified by restriction mapping (data not shown).

The common feature of these chimeric molecules is that they are too large to be packaged into an SV40 capsid. An SV40 capsid can contain at least 250 extra nucleotides, which is the size of some viable constructed genomes (T. Shenk, personal communication), but that size is probably close to the packaging limit. As a result, viable progeny can be produced after

TABLE 1. Infectivity of circular and linear chimeric molecules

DNA form ^a	Infectivity of chimeras (%) ^b				
	pJJ1	pJJ2	pJJ4	pBL5	pBL1
1. Circles	0.5	0.6	1.0	0.03	0.2
2. Linears (<i>Ava</i> I)	1.2	1.7	1.1	0.05	0.3
3. Linears (<i>Sal</i> I)	0.7	0.9	1.4	0.06	0.5
4. Linears (<i>Taq</i> I)			101		
5. Linears (<i>Bam</i> HI)				97	102

^a Chimeric molecules were digested to completion with the indicated enzymes. The unfractionated digestion products in samples 4 and 5 were used for transfection.

^b Data points are the average of two to five experiments. In each experiment, 3 to 15 ng of samples 1 to 3 and 0.5 to 1 ng of samples 4 and 5 were plaque assayed on five plates each. Infectivities are expressed as a percentage of the average of samples 4 and 5. The average specific infectivity of samples 4 and 5 was 170 PFU/ng.

infection with these chimeric molecules only if most or all of the pBR322 sequences are deleted in a manner which does not remove any essential viral sequences and which is compatible with SV40 gene expression. By measuring viable progeny by plaque assay, we detect selectively those recombination events that leave essential viral sequences intact, functional, and packageable.

SV40-pBR322 chimeras produce viable progeny virus in monkey cells. To investigate the influence of DNA topology, we infected CV1 monkey kidney cells with circular and linear forms of the various chimeric molecules and followed production of viable progeny by plaque assay. The linear forms were generated by utilizing enzymes that cut once within the pBR322 DNA and thus did not interrupt the integrity of the SV40 genome. The infections yielded plaques at readily measurable frequencies between 0.02 and 2% of the number produced by linear SV40 DNA that was excised cleanly from the chimeric molecules (Table 1). No significant difference was noted between linear and circular molecules from the same chimeras.

Plaques that arose from infections with different chimeric molecules typically were delayed relative to wild-type plaques by one to several days for pJJ1, pJJ2, and pJJ4 and up to 2 weeks for pBL1. The rare plaques formed after infection with pBL5 were not distinguishable from the wild type in appearance or time of formation. To determine whether the delayed plaquing exhibited by some chimeras was a characteristic of the viable viruses themselves or resulted from interfering pBR322 sequences (19), we picked several plaques from each infection for retest-

TABLE 2. Infectivity of *SalI*-linearized chimeras before and after gel purification

Chimeric DNA	Infectivity (no. of plaques per 100 ng of DNA)	
	Before gel	After gel ^a
pJJ2	45	52
pJJ4	66	85
pBL5	2	3
pBL1	33	42
pBL1 plus SV40 ^b	>1,000	62

^a Linear chimeric DNA was subjected to electrophoresis on 0.6% agarose gels. Bands were cut out, and DNA was extracted as described in the text.

^b A 10:1 mixture of *SalI*-linearized chimeric DNA and circular SV40 DNA (measured in nanograms) was prepared and assayed before and after purification of the chimeric DNA.

ing. These progeny viruses once again plaqued with delayed kinetics, indicating that they are in some way disadvantaged relative to wild-type SV40.

To determine whether these recombinants were produced in bacteria before plasmid purification or in monkey cells after infection, we subjected *SalI*-linearized chimeric molecules to electrophoresis on agarose gels, extracted bands of appropriate size, and compared their infectivities to those molecules in the original mixture (Table 2). As indicated by the results with a mixture of linear chimeric molecules and circular SV40 DNA, isolation from agarose gels achieved a greater than 20-fold purification from SV40-size circular molecules. We would expect a similar purification from putative bacteria-formed recombinants, which should be similar in size to SV40. Because the infectivities of the chimeric molecules were unaffected by this purification, it appears unlikely that the recombinants were already present as trace components of the original mixture. Thus, we conclude that the recombination events that we detect have occurred in monkey cells.

The 10-fold difference in the infectivities of the *Bam*HI chimeras, pBL1 and pBL5, that is evident in Tables 1 and 2 was unexpected. Additional studies with these chimeras (to be reported elsewhere) have shown consistent differences of this magnitude. Because these chimeras differ only in the orientation of pBR322, we have examined the nucleotide sequences at the constructed junctions. Significantly, in pBL1, which yields the higher frequency of viable progeny, the pBR322 sequences fortuitously provide a 17-amino acid C-terminal tail for VP1 which is similar in composition to the normal 18-amino acid tail that was removed during construction. If this new tail permitted VP1 to function in capsid assem-

bly and viral infectivity, then recombination events involving pBR322 sequences beyond this tail could generate viable progeny (Fig. 2). Direct analysis of this possibility has been hampered by the poor growth characteristics of the viable progeny produced by pBL1. In pBL5, which produces a low frequency of apparently wild-type progeny, the pBR322 sequences at the junction provide an uninterrupted string of 298 inphase codons. For this reason, we think that recombination in pBL5 probably must occur more precisely to restore an acceptable, perhaps wild-type, sequence (Fig. 2). The relatively high infectivities of pJJ1, pJJ2, and pJJ4 are consistent with the ready isolation of viable deletions around both the *Hpa*II and *Taq*I sites (24, 25, 32, 34, 38) and presumably reflect a somewhat larger target for the recombination events.

Viable progeny exhibit three distinct classes of recombination junctions. To characterize the recombination junctions in viable progeny, we picked 12 independent plaques that arose after infection with the chimera pJJ4, which had been digested with *Ava*I and *Sal*I. CV1 cells were infected with virus from the picked plaques in the presence of ³²P_i, and the resulting labeled DNA was isolated, digested with *Hind*III, subjected to electrophoresis on polyacrylamide gels, and autoradiographed (Fig. 3). Of the 10 plaques for which results were obtained, 1 (lane 4) was a mixture of two viruses, and the rest were unique. (The plaques displayed in lanes 2 and 5 have not been analyzed further.) As anticipated, each virus contained all of the wild-type *Hind*III fragments except for *Hind*III-B, which

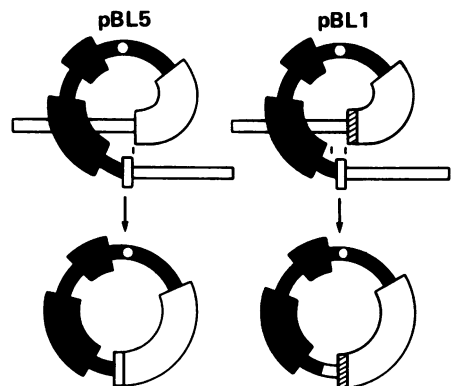


FIG. 2. Possible recombination targets in pBL1 and pBL5. For pBL5, the recombination is drawn so as to restore the normal VP1 gene. For pBL1, the recombination is drawn to eliminate excess DNA, leaving intact the junctional pBR322 sequences (hatched box) that supply the putative new C terminus for VP1 (see text). Thin vertical lines indicate the approximate target for recombination.

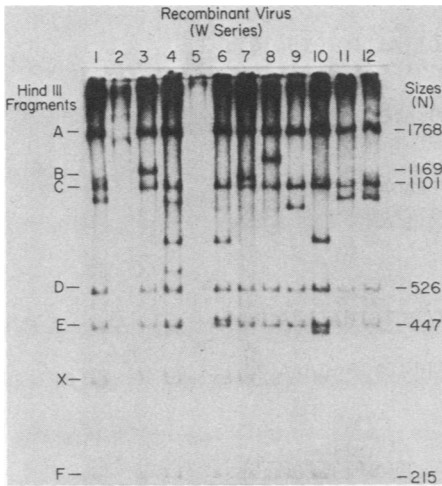


FIG. 3. Restriction analysis of the genomes from recombinant progeny. Twelve plaques were picked from assay plates of CV1 cells infected with pJJ4 that had been cleaved with *Ava*I and *Sal*I. These plaque suspensions were used to prepare ³²P-labeled DNA by the miniwell method (see the text). The labeled DNA was digested with *Hind*III and analyzed on 8% polyacrylamide gels, which were then dried and autoradiographed. The 12 plaques are numbered across the top. The positions and sizes of wild-type *Hind*III fragments are indicated on the sides. The extra band (X) is of unknown origin but shows up in all of our miniwell preparations, even of wild-type SV40. The altered B fragments in lanes 4, 6, and 10 have been cleaved into two pieces because they contain the unique pBR322 *Hind*III site, which is adjacent to the *Cla*I site in the chimera.

includes the *Taq*I site at which pBR322 was inserted. The different sizes of the new fragments and the approximate positioning of the recombination junctions by restriction mapping indicate that several different sites were involved in the observed recombination events (Fig. 4). Thus there seem to be no strong preferences for recombination sites within this region.

The recombination junctions in five recombinant viruses were analyzed in more detail by the nucleotide sequencing strategy indicated in Fig. 4B. Recombinants W4A, W6, and W10 were linearized and labeled at their unique *Cla*I sites, and recombinants W3 and W8 were linearized and labeled at their unique *Taq*I sites. Nucleotide sequence analysis of the recombination junctions in these five recombinant viruses confirmed the restriction mapping and revealed no pBR322 or SV40 sequence alterations other than those at the junctions. In Fig. 5, the nucleotide sequences around the recombination junctions in the recombinants are shown in bold type, and the parental sequences that were eliminated in the recombination event are shown in regular

type; the boxed nucleotides at some junctions are present only once in the sequences of the recombinant virus. These sequences suggest three distinct classes of junction: flush junctions, insertion junctions, and duplication junctions (Fig. 5).

At flush junctions, the recombined segments are joined to one another directly. There is some ambiguity in the precise location of the cross-overs in these junctions because of the junctional homology, indicated by the boxed nucleotides in Fig. 5. At 12 additional flush junctions that we have sequenced (Gudewicz and Wilson, unpublished data), the junctional homologies ranged from 0 to 4, yielding an overall mean of about 1.5

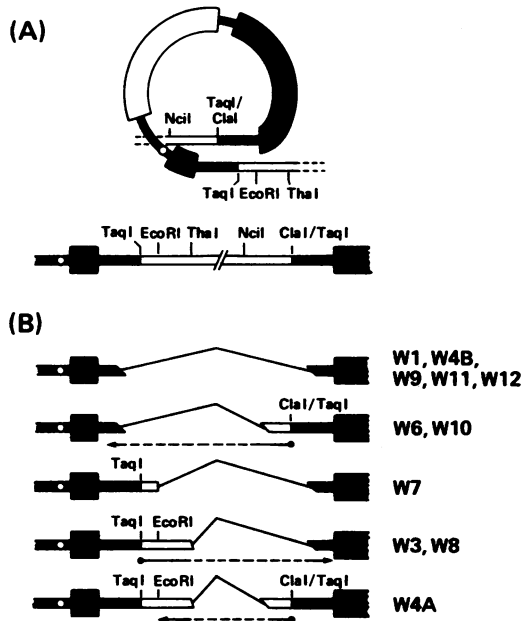


FIG. 4. Restriction mapping of the recombination junctions in viable progeny. (A) pJJ4 chimera arranged to accentuate the region in which a recombination event must occur to be detected. A linear representation is shown below. The recognition sites for several restriction enzymes that cleave in the target region are indicated. (B) Approximate positions of recombination junctions in the viable recombinants. ³²P-labeled DNA and unlabeled DNA were prepared from cells infected with the plaque suspensions, digested individually with restriction enzymes, subjected to electrophoresis through polyacrylamide gels, and then autoradiographed or stained with ethidium bromide. The 11 analyzed junctions fell into five classes defined by restriction analysis. The thin lines connecting the left and right segments indicate the junctions, which have been positioned arbitrarily in the interval defined by restriction analysis. W4A and W4B are the two species present in lane 4 of Fig. 3. Arrows below the junctions in (B) indicate the strategy for nucleotide sequence analysis.



FIG. 5. Nucleotide sequences around recombination junctions. The upper and lower sequences in each recombinant correspond to the upper and lower strands in the circular representation in Fig. 4A. The nucleotide sequence of each recombinant is shown in boldface type. Diagonal lines connecting the strands indicate potential crossover points at the junction. The homologous nucleotides at the flush junctions are boxed. Each strand is written 5' to 3' in the same sense as the early SV40 mRNA (mRNA identical strand) One nucleotide in each strand is numbered for reference (5, 33). The wavy line in the upper W8 strand marks the constructed junction between pBR322 and SV40 DNA at the *ClaI* site. The net size change in nucleotides relative to wild-type SV40 is indicated at the right.

nucleotides (Table 3). At the insertion junction in W10, the recombined segments are separated by a few nucleotides whose origin is uncertain, although they seem unlikely to have been derived from the nearby parental strands. At the duplication junction in W6 the 18 nucleotides encompassing the recombination joint are repeated directly in the adjacent pBR322 sequences (underlined in Fig. 5). These different junctional types could indicate that several mechanisms can generate recombinants, or alternatively, they could represent different fac-

ets of a common underlying mechanism. In any case, flush junctions apparently are the most common, outnumbering the other two categories nearly eight to one in our collection.

Recombinants are produced primarily by end-

TABLE 3. Junctional homology at 15 flush junctions derived from pJJ4

DNA form	No. of flush junctions with homologies of the following no. of nucleotides ^a :					
	0	1	2	3	4	5
Circles	2	1	1	0	1	0
Linears (<i>AvaI</i>)	1	0	0	0	0	0
Linears (<i>SalI</i>)	0	0	1	0	0	0
Linears (<i>SalI</i> and <i>AvaI</i>)	0	2	0	1	0	0
Linears (<i>ThaI</i>)	0	3	1	0	1	0

^a Junctional homology was determined by sequencing across the recombination joints in individual progeny derived from transfections with pJJ4 DNA cleaved with the restriction enzymes indicated in parentheses.

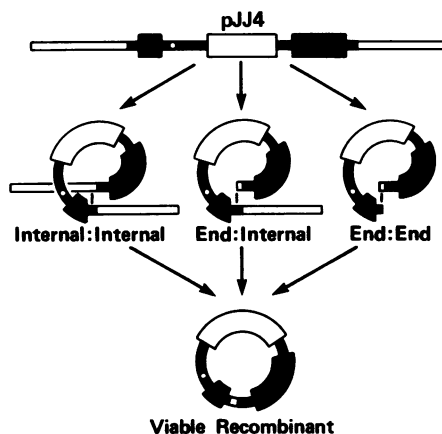


FIG. 6. Categories of nonhomologous recombination mechanisms as illustrated for the pJJ4 chimera. Internal and end are defined as broadly as necessary to include all possible intramolecular mechanisms. Thin vertical lines indicate hypothetical sites of recombination.

to-end joining. Although the structure of the input DNA was defined precisely, we do not know what modifications, if any, were introduced by the cell before the recombination

event. Thus, the structure of the input DNA and the sequence of nucleotides at the recombination junctions provide no information about whether the molecular recombination mecha-

TABLE 4. Infectivity of trimmed chimeric DNA

Restriction digestion	pBR322 tails (kb) ^a		Total length (kb)	Relative infectivity ^b (%)	Normalized infectivity ^b
	Short	Long			
pJJ1					
<i>Sall</i> ^c	0.63	3.74	9.61	0.7 ± 0.6	0.009
<i>AvaI</i> ^c	1.40	2.96	9.61	1.2 ± 0.9	0.015
<i>Sall</i> and <i>AvaI</i>	0.63	2.96	8.83	0.9 ± 0.3	0.011
<i>Sall</i> and <i>XorII</i> ^d	0.63	0.65	6.52	1.6 ± 1.4	0.020
<i>NciI</i>	0.15	0.49	5.87	11	0.14
<i>HpaII</i> ^c	0.14	0.49	5.86	12 ± 3.6	0.15
<i>ThaI</i>	0.13	0.32	5.69	57 ± 22	0.71
<i>ThaI</i> and <i>NciI</i>	0.13	0.15	5.52	80 ± 29	(1.0)
<i>ThaI</i> and <i>HpaII</i>	0.13	0.14	5.51	62 ± 1	0.78
pJJ2					
<i>Sall</i> ^c	0.63	3.74	9.61	0.9 ± 0.5	0.011
<i>AvaI</i> ^c	1.40	2.96	9.61	1.7 ± 1.6	0.020
<i>Sall</i> and <i>AvaI</i>	0.63	2.96	8.83	0.9 ± 0.4	0.011
<i>Sall</i> and <i>XorII</i> ^d	0.63	0.65	6.52	3.3 ± 3.3	0.040
<i>NciI</i>	0.15	0.49	5.87	33	0.40
<i>HpaII</i> ^c	0.14	0.49	5.86	36 ± 12	0.43
<i>ThaI</i>	0.13	0.32	5.69	33 ± 10	0.40
<i>ThaI</i> and <i>NciI</i>	0.13	0.15	5.52	83 ± 26	(1.0)
<i>ThaI</i> and <i>HpaII</i>	0.13	0.14	5.51	56 ± 5	0.67
pJJ4					
<i>Sall</i> ^c	0.63	3.74	9.61	1.4 ± 0.2	0.011
<i>AvaI</i> ^c	1.40	2.96	9.61	1.1 ± 0.5	0.014
<i>Sall</i> and <i>AvaI</i>	0.63	2.96	8.83	0.8 ± 0.4	0.008
<i>Sall</i> and <i>XorII</i> ^d	0.63	0.65	6.52	7.0 ± 5.9	0.070
<i>NciI</i>	0.15	0.49	5.87	50 ± 20	0.50
<i>ThaI</i>	0.13	0.32	5.69	68 ± 26	0.68
<i>ThaI</i> and <i>NciI</i>	0.13	0.15	5.52	100 ± 17	(1.0)
pJJ4 (Clal cut)					
<i>Clal</i> ^c	0	4.36	9.61	1.0 ± 1.0	0.012
<i>Clal</i> and <i>Sall</i>	0	3.74	8.98	0.8 ± 0.5	0.010
<i>Clal</i> and <i>AvaI</i>	0	2.96	8.20	0.9 ± 0.3	0.011
<i>Clal</i> and <i>XorII</i> ^d	0	0.65	5.89	14 ± 2.8	0.17
<i>Clal</i> and <i>NciI</i>	0	0.49	5.73	59 ± 14	0.71
<i>Clal</i> and <i>ThaI</i>	0	0.13	5.37	83 ± 13	(1.0)
Reference molecules					
(pJJ4) <i>TaqI</i> ^c	0	0	5.24	105 ± 12	
(pBL1) <i>BamHI</i> ^c	0	0	5.24	93 ± 16	
(pBL5) <i>BamHI</i> ^c	0	0	5.24	98 ± 5	

^a Tails refer to the pBR322 nucleotides that remain attached to the ends of the intact SV40 genome. See Fig. 1 for the restriction sites. kb, Kilobase.

^b In individual experiments, each data point was the average of five plates. Each plate was infected with 0.5 to 15 ng of the unfractionated digestion products and plaque assayed at 37°C. Relative infectivities are expressed as a percentage of the average of the reference molecules. Each value in the table represents the average of from two to seven separate experiments. Normalized infectivities were obtained by dividing each relative infectivity by the highest value (in parentheses) within each set. The average specific infectivity of the reference molecules was 140 PFU/ng.

^c Digestion with these restriction enzymes leaves complementary sticky ends on the fragment that contains SV40.

^d *XorII* cleaved only 70 to 90% of the chimeric DNA, presumably because of methylation. The data have not been manipulated to reflect this and thus they may be only 70 to 90% of their true values.

nism involved an interaction between two internal DNA segments or an interaction requiring one or both ends. However, as described below, these three categories of potential mechanism, which are illustrated for pJJ4 in Fig. 6, can be distinguished by measuring the infectivity of chimeric molecules from which different lengths of pBR322 sequences have been removed. This genetic analysis rests on two assumptions. (i) A large number of potential recombination sites are distributed throughout the chimera. (ii) Only the subset of recombination events that occur in the target zone, which is defined at its extremes by essential SV40 DNA and by the packaging limit, can be detected. Both assumptions are amply supported by the 17 recombination junctions that we have sequenced; all are located within the target, and no two are identical.

Restriction enzymes that cleave only pBR322 sequences were used to trim various lengths of pBR322 from the intact SV40 genome in the chimeras. The infectivities of the trimmed chimeras relative to a set of reference molecules were determined by plaque assay, and the results are summarized in Table 4. For comparison with theoretical expectations, these infectivities were normalized to the trimmed chimera with the maximum infectivity within each set (Table 4). If one or both ends were required for recombination, the normalized infectivity curve should rise to its maximum value only when one or both ends are positioned within the target. By contrast, if recombination occurred at random between internal DNA segments, the infectivity should be proportional to $(T/L)^2$, where T is the size of the target and L is the size of the trimmed chimera. As illustrated in Fig. 7, the actual data match most closely the curve expected for a recombination mechanism involving one or both ends. The normalized infectivity curves for the chimeras pBL1 and pBL5 also follow this same curve (data not shown).

To determine whether one or both ends were involved in the recombination event, we measured the infectivity of chimera pJJ4, which had been cut with *ClaI* in combination with other restriction enzymes (Table 4). Because *ClaI* cuts exactly at one junction of pBR322 and SV40 sequences, one end of each trimmed chimera is positioned within the target. If both ends were required for recombination, the normalized infectivity curve for *ClaI*-cut pJJ4 should rise to its maximum value only when the second end is placed within the target. Alternatively, if recombination occurred at random between an end and an internal DNA segment, the infectivity curve for the *ClaI*-cut chimera should be proportional to T/L . As shown in Fig. 8, the data most closely match the curve expected for a recombination mechanism involving both ends.

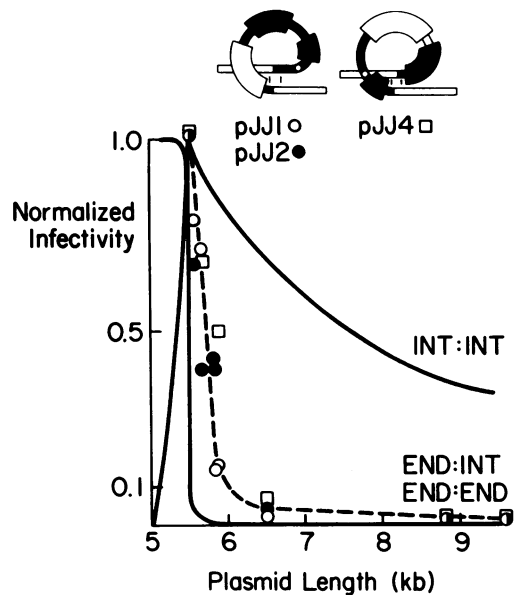


FIG. 7. Infectivity of the trimmed pJJ1, pJJ2, and pJJ4 chimeras. This figure is plotted from the data in Table 4. The solid lines represent theoretical curves for end-mediated (END) or internal (INT) recombination as indicated. These curves are calculated for a chimera which initially was linearized at a point in pBR322 equidistant from the ends of the SV40 segment and from which equal lengths of DNA were removed sequentially from both ends. For the theoretical curves, a target size of 500 base pairs was chosen arbitrarily; the general shapes of the theoretical curves are relatively independent of target size. The dotted line represents the experimental curve. The chimeras are shown above as a loop to accentuate the target for recombination (see Fig. 6). Thin vertical lines indicate approximate targets for recombination. kb, Kilobase.

One potential complication in interpreting these experiments is the presence on pBR322 of a *cis*-acting poison sequence that interferes with normal SV40 replication (19). In the experiments described in Table 4 and graphed in Fig. 7 and 8, the poison sequence remained attached only to chimeras greater than 7 kilobases in length. Because the shape of the infectivity curve is established by genomes that do not contain the poison sequence, we do not think the poison sequence affects our interpretations.

DISCUSSION

Our genetic analysis of nonhomologous recombination in transfected DNA indicates that unrelated DNA ends are joined together willy-nilly with high efficiency. Although our measurements have been made only in particular target areas of specially constructed chimeras,

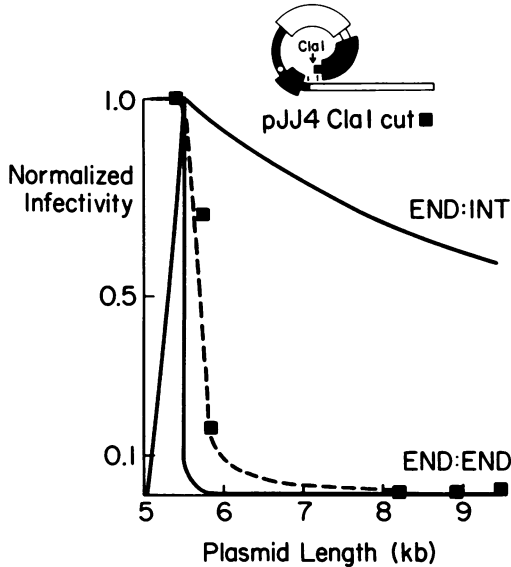


FIG. 8. Infectivity of the pJ4 chimera cut with *ClaI*. This figure is plotted from the data in Table 4. The solid lines represent the theoretical curves for recombination requiring one or both ends as indicated. These curves are calculated for a chimera which had one end positioned within the target and from which DNA was removed sequentially from the other end. For the theoretical curves, a target size of 500 base pairs was chosen arbitrarily; the general shapes of the theoretical curves are relatively independent of target size. The dotted line represents the experimental curve. The *ClaI*-cut chimera is shown above as a loop to accentuate the target for recombination (see Fig. 6). Thin vertical lines indicate the approximate target for recombination. kb, Kilobase.

they describe basic somatic cell capabilities and suggest that DNA ends will behave generally in the manner described. In this study of intramolecular recombination events, we have demonstrated that linear SV40 genomes with short, nonhomologous pBR322 tails are about as infectious as linear SV40 genomes with complementary single-stranded tails. (These results, it should be emphasized, are based on endpoint assays and thus show that the same fraction of input molecules is infectious but give no information on the relative rates of the two end-joining processes.) The viable genomes that were produced at low frequencies after infection with full-length linear and circular molecules may have arisen by a different mechanism entirely or, perhaps more likely, by an end-joining event subsequent to cell-mediated breakage of the input DNA (Wake and Wilson, manuscript in preparation).

A high-efficiency end-to-end joining provides a natural explanation for the indiscriminate recombination of SV40 DNA with procaryotic,

polyoma, or cloned mouse DNA observed by Winocour and Keshet in DNA cotransfection experiments using DEAE-dextran as a carrier (42). It also supports the mechanism suggested for formation of the extensive arrays of input DNA commonly observed after transformation with CaPO_4 coprecipitates of exogenous DNA (27, 30). The particles in such precipitates contain many molecules of DNA and thus may provide an ideal substrate for end-to-end joining. In addition, one proposed model for excision of chromosomal DNA involves "onion skin" replication to release free linear molecules, which from our results, would be expected to circularize with high efficiency (3, 37). In line with these expectations, the low-frequency excision of single-copy SV40 DNA from transformed cells does lead to a heterogeneous set of circular DNAs often containing both SV40 sequences and adjacent cellular sequences (2, 14).

The recombination junctions sequenced in this study and in others where nuclease-treated linear SV40 DNA was transfected into cells (13) are consistent with end joining as the primary mechanism for linking unrelated DNA segments. In all of these studies, flush junctions were the most common kind of novel joint. Such junctions might be expected if DNA ends were trimmed and compared continually until a sufficient zero- to four-nucleotide joining homology was encountered. The actual joining reaction then might be no different than the better-characterized joining of restriction fragments with complementary termini. If there are any homology requirements outside of the junction itself, they must be quite subtle because the parental sequences surrounding the flush junctions seem unremarkable in base composition, and computer analysis suggests that the occasional matched nucleotides and similar sequence blocks are not greater than random. The homology at the crossovers in flush junctions is somewhat less than in similar procaryotic deletion and duplication junctions, which typically occur at 5- to 15-nucleotide homologies (10, 11, 28).

The rarer insertion junctions and duplication junctions observed by us and by others (13) are most curious. Inserted nucleotides could have been generated in a number of ways, such as the following: (i) from multiple deletions or rearrangements that have obscured their relationship to the parental molecule, (ii) from the repair of mismatched heteroduplexes between the parental strands, and (iii) from the random addition of nucleotides or short DNA segments to free ends. Although multiple events are a potential explanation for our insertion junctions, they are an unlikely explanation for the junctions observed in the straightforward circularization experiments of Gutai (13). Mismatch repair of

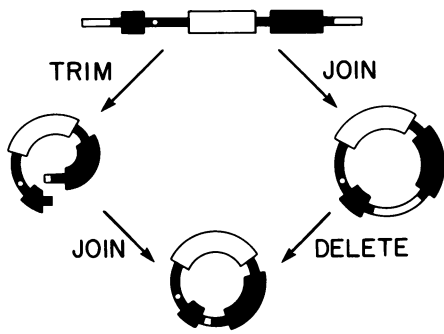


FIG. 9. Alternative pathways for the removal of short stretches of DNA from the input linear DNA.

heteroduplexes also seems improbable because the sequences of inserted nucleotides generally are an unlikely derivative of the surrounding parental sequences. By elimination, random addition to free ends, e.g., by a terminal transferase-like activity, becomes an attractive possibility. It is notable in this regard that similar inserted nucleotides have been observed at the junctions between the D and J gene segments in rearranged immunoglobulin heavy chain genes (18).

Our analysis of the infectivity of linear chimeras with different lengths of attached pBR322 sequences indicates that end-to-end joining is a primary mechanism for nonhomologous recombination in somatic cells. However, the experimental curves in Fig. 7 and 8 are displaced slightly to the right of the theoretical curves for end joining, suggesting that small amounts of DNA are removed efficiently from the infecting genomes. We have verified these expectations in some cases. For example, the SV40 genomes cut from pJJ4 with *NotI* carry 452 extra nucleotides of pBR322 DNA at their ends and are 68% as infectious as the linear reference genomes. The sequenced junctions in five viable progeny that arose from such infections (Table 3) indicate that 485 nucleotides on average had been removed from the infecting genomes. In principle, the efficient removal of these extra DNA sequences could have occurred before or after the end-joining event as illustrated in Fig. 9.

In an independent study, we have measured the extent of exonucleolytic trimming of nucleotides from the ends of transfected DNA by examining individual progeny from an infection with SV40 genomes that had been linearized with *TaqI*. These linear molecules have about 150 nucleotides of nonessential DNA at each end and thus are ideal substrates for analyzing end trimming. Of the 22 viable progeny that had lost the *TaqI* site (15% of the total), only 1 contained a deletion larger than 25 nucleotides

(Gudewicz and Wilson, unpublished data). These measurements would seem to indicate that exonucleolytic trimming of ends before end joining is much too limited to account for the efficient removal of hundreds of nucleotides suggested by the data in Fig. 7 and 8. Although these measurements do not eliminate trimming as a possibility, since trimming processes that removed long segments of DNA in a single operation would not have been detected, they tend to favor the alternative possibility that small amounts of DNA are eliminated subsequent to end joining. If such deletion events occurred even rarely within a pool of replicating genomes, they might be selected efficiently if the resulting genomes were packageable. Further studies will be required to distinguish these possibilities.

ACKNOWLEDGMENTS

We thank Kathleen Marburger and Thomas Gudewicz for expert technical assistance and the participants at the Cedar Bay Conference on Molecular Biology for lively and critical discussions.

This work was supported by research grants from the National Science Foundation (PCM 81-04523) and from the United Way (X-3) and by Public Health Service grants (CA-15743 and GM-28952) from the National Institutes of Health.

LITERATURE CITED

- Bolivar, F., and K. Backman. 1979. Plasmids of *Escherichia coli* as cloning vectors. *Methods Enzymol.* **68**:245-267.
- 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.
- Botchan, M., W. Topp, and J. Sambrook. 1978. Studies on simian virus 40 excision from cellular chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* **43**:709-719.
- Brack, C., M. Hiram, R. Lenhard-Shuller, and S. Tonegawa. 1978. A complete immunoglobulin gene is created by somatic recombination. *Cell* **15**:1-14.
- Buchman, A. R., L. Burnett, and P. Berg. 1980. The SV40 nucleotide sequence, p. 799-829. In J. Tooze (ed.), *DNA tumor viruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Calos, M. P., and J. H. Miller. 1980. Transposable elements. *Cell* **20**:579-595.
- Chia, W., and P. W. J. Rigby. 1981. Fate of viral DNA in nonpermissive cells infected with simian virus 40. *Proc. Natl. Acad. Sci. U.S.A.* **78**:6638-6642.
- Davis, M., S. Kim, and L. Hood. 1980. DNA sequences mediating class switching in α -immunoglobulins. *Science* **209**:1360-1365.
- Dubbs, D., M. Rachmeler, and S. Kit. 1974. Recombination between temperature-sensitive mutants of SV40. *Virology* **57**:161-174.
- Edlund, T., and S. Normark. 1981. Recombination between short DNA homologies causes tandem duplications. *Nature (London)* **292**:269-271.
- Farabaugh, P. J., U. Schmeissner, M. Hoffer, and J. H. Miller. 1978. Genetic studies of the lac repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J. Mol. Biol.* **126**:847-863.
- Goff, S., and P. Berg. 1977. Structure and formation of circular dimers of SV40 DNA. *J. Virol.* **24**:295-302.
- Gutai, M. W. 1981. Recombination in SV40 infected cells: viral DNA sequences at sites of circularization of transfecting linear DNA. *Virology* **109**:353-365.
- Hanahan, D., D. Lane, L. Lipsich, M. Wigler, and M.

- Botchan. 1980. Characteristics of an SV40-plasmid recombinant and its movement into and out of the genome of a murine cell. *Cell* 21:127-139.
15. Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cultures. *J. Mol. Biol.* 26:365-369.
 16. Israel, M., J. Byrne, and M. Martin. 1978. Biologic activity of oligomeric forms of SV40 DNA. *Virology* 87:239-246.
 17. Klein, R. D., E. Selsing, and R. D. Wells. 1980. A rapid microscale technique for isolation of recombinant plasmid DNA suitable for restriction enzyme analysis. *Plasmid* 3:88-91.
 18. Kurosawa, Y., and S. Tonegawa. 1982. Organization, structure, and assembly of immunoglobulin heavy chain diversity segments. *J. Exp. Med.* 155:201-218.
 19. Lusky, M., and M. Botchan. 1981. Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. *Nature (London)* 293:79-81.
 20. Maki, R., A. Trauneker, H. Sakano, W. Boeder, and S. Tonegawa. 1980. Exon shuffling generates an immunoglobulin heavy chain gene. *Proc. Natl. Acad. Sci. U.S.A.* 77:2138-2142.
 21. Mandel, M., and A. Higa. 1970. Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* 53:159-162.
 22. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
 23. McCutchen, J. H., and J. S. Pagano. 1968. Enhancement of the infectivity of SV40 deoxyribonucleic acid with diethyl-aminoethyl-dextran. *J. Natl. Cancer Inst.* 41:351-357.
 24. Mertz, J. E., and P. Berg. 1974. Viable deletion mutants of simian virus 40: selective isolation by means of a restriction endonuclease from *Hemophilus parainfluenzae*. *Proc. Natl. Acad. Sci. U.S.A.* 71:4879-4883.
 25. Mertz, J. E., J. Carbon, M. Herzberg, R. W. Davis and P. Berg. 1974. Isolation and characterization of individual clones of simian virus 40 mutants containing deletions, duplications and insertions in their DNA. *Cold Spring Harbor Symp. Quant. Biol.* 39:69-84.
 26. Nash, H. A. 1981. Integration and excision of bacteriophage λ : the mechanism of conservative site specific recombination. *Annu. Rev. Genet.* 15:143-167.
 27. Perucho, M., D. Hanahan, and M. Wigler. 1980. Genetic and physical linkage of exogenous sequences in transformed cells. *Cell* 22:309-317.
 28. Pribnow, D., D. C. Sigurdson, L. Gold, B. S. Singer, C. Napoli, J. Brosius, T. J. Dull, and H. F. Noller. 1981. rII cistrons of bacteriophage T4. DNA sequence around the intercistronic divide and positions of genetic landmarks. *J. Mol. Biol.* 149:337-376.
 29. Reddy, V. B., B. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, P. K. Ghosh, M. L. Celma, and S. M. Weissman. 1978. The genome of simian virus 40. *Science* 200:494-502.
 30. Robins, D. M., S. Ripley, A. S. Henderson, and R. Axel. 1981. Transforming DNA integrates into the host chromosome. *Cell* 23:29-39.
 31. Sharp, P. A., B. Sugden, and J. Sambrook. 1973. Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. *Biochemistry* 12:3055-3063.
 32. Shenk, T. E., J. Carbon, and P. Berg. 1976. Construction and analysis of viable deletion mutants of simian virus 40. *J. Virol.* 18:664-671.
 33. Sutcliffe, J. G. 1978. Complete nucleotide sequence of *Escherichia coli* plasmid pBR322. *Cold Spring Harbor Symp. Quant. Biol.* 43:77-90.
 34. Thimmappaya, B., and T. Shenk. 1979. Nucleotide sequence analysis of viable deletion mutants lacking segments of the simian virus 40 genome coding for small t antigen. *J. Virol.* 30:668-673.
 35. Upcroft, P., B. Carter, and C. Kidson. 1980. Analysis of recombination in mammalian cells using SV40 genome segments having homologous overlapping termini. *Nucleic Acids Res.* 8:2725-2736.
 36. Upcroft, P., B. Carter, and C. Kidson. 1980. Mammalian cell functions mediating recombination of genetic elements. *Nucleic Acids Res.* 8:5835-5844.
 37. Varshavsky, A. 1981. On the possibility of metabolic control of replicon "misfiring": relationship to emergence of malignant phenotypes in mammalian cell lineages. *Proc. Natl. Acad. Sci. U.S.A.* 78:3673-3677.
 38. Volckaert, G., J. Feunteun, L. V. Crawford, P. Berg, and W. Fiers. 1979. Nucleotide sequence deletions within the coding region for small-t antigen of simian virus 40. *J. Virol.* 30:674-682.
 39. Wake, C. T., and J. H. Wilson. 1979. Simian virus 40 recombinants are produced at high frequency during infection with genetically mixed oligomeric DNA. *Proc. Natl. Acad. Sci. U.S.A.* 76:2876-2880.
 40. Wake, C. T., and J. H. Wilson. 1980. Defined oligomeric SV40 DNA: a sensitive probe of general recombination in somatic cells. *Cell* 21:141-148.
 41. Wilson, J. H., M. DePamphilis, and P. Berg. 1976. Simian virus 40-permissive cell interactions: selection and characterization of spontaneously arising monkey cells that are resistant to simian virus infection. *J. Virol.* 20:391-399.
 42. Winocour, E., and I. Keshet. 1980. Indiscriminate recombination in simian virus 40-infected monkey cells. *Proc. Natl. Acad. Sci. U.S.A.* 77:4861-4865.
 43. Wolgemuth, D., and M.-T. Hsu. 1980. Visualization of genetic recombination intermediates of human adenovirus type 2 DNA from infected HeLa cells. *Nature (London)* 287:168-171.
 44. Wu, R., E. Jay, and R. Roychoudhury. 1976. Nucleotide sequence analysis of DNA. *Methods Cancer Res.* 12:87-176.
 45. Young, C., and S. Silverstein. 1980. The kinetics of adenovirus recombination in homotypic and heterotypic genetic crosses. *Virology* 101:503-515.