MOLECULAR AND CELLULAR BIOLOGY, Mar. 1984, p. 387–398
0270-7306/84/030387-12$02.00/0

Vol. 4, No. 3

How Damaged Is the Biologically Active Subpopulation of Transfected DNA?

CLAIRE T. WAKE,[1]† TOM GUDEWICZ,[1] TOM PORTER,[2] ALLEN WHITE,[2] AND JOHN H. WILSON[1]*

Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine,[1] and Department of Biomathematics, The University of Texas System Cancer Center, M. D. Anderson Hospital and Tumor Institute,[2] Houston, Texas 77030

Received 12 September 1983/Accepted 28 November 1983

Relatively little is known about the damage suffered by transfected DNA molecules during their journey from outside the cell into the nucleus. To follow selectively the minor subpopulation that completes this journey, we devised a genetic approach using simian virus 40 DNA transfected with DEAE-dextran. We investigated this active subpopulation in three ways: (i) by assaying reciprocal pairs of mutant linear dimers which differed only in the arrangement of two mutant genomes; (ii) by assaying a series of wild-type oligomers which ranged from 1.1 to 2.0 simian virus 40 genomes in length; and (iii) by assaying linear monomers of simian virus 40 which were cleaved within a nonessential region to leave either sticky, blunt, or mismatched ends. We conclude from these studies that transfected DNA molecules in the active subpopulation are moderately damaged by fragmentation and modification of ends. As a whole, the active subpopulation suffers about one break per 5 to 15 kilobases, and about 15 to 20% of the molecules have one or both ends modified. Our analysis of fragmentation is consistent with the random introduction of double-strand breaks, whose cause and exact nature are unknown. Our analysis of end modification indicated that the most prevalent form of damage involved deletion or addition of less than 25 base pairs. In addition we demonstrated directly that the efficiencies of joining sticky, blunt, or mismatched ends are identical, verifying the apparent ability of cells to join nearly any two DNA ends and suggesting that the efficiency of joining approaches 100%. The design of these experiments ensured that the detected damage preceded viral replication and thus should be common to all DNAs transfected with DEAE-dextran and not specific for viral DNA. These measurements of damage within transfected DNA have important consequences for studies of homologous and nonhomologous recombination in somatic cells as is discussed.

Transfection of foreign DNA into somatic cells in culture has become an extremely important tool in modern molecular genetics, permitting studies of gene function (7, 11, 12, 31, 33), gene regulation (6, 8, 23, 24, 30, 40, 55–57), oncogenesis (3, 34, 36, 47, 49), and somatic recombination (2, 4, 15, 17, 20, 21, 26, 29, 32, 37, 38, 41, 42, 45, 50–54, 60, 62) that were unimaginable only a few years ago. Despite the usefulness of DNA transfection, very little is known about the pathway by which the biologically active molecules travel from outside the cell into the nucleus or about the status of the DNA once it arrives (19, 27). Because the biologically active molecules constitute such a small portion of the total (commonly less than 1 part in 10^5), these questions are beyond the limits of routine biochemical investigation. Stated another way, experiments designed to follow the mass of input DNA can give no certain information about the fate of the biologically active molecules, since the active molecules may retain their biological integrity precisely because they travel a different path than the majority of input molecules.

For many applications of DNA transfection, these concerns are peripheral; however, for studies of somatic recombination, which is our special interest, they are central. The goal of DNA transfection experiments that investigate somatic recombination is to derive meaningful statements about cellular recombinational capabilities by comparing the structures of parental and recombinant molecules. The very small fraction of transfected molecules that retain biological activity imposes a limitation for interpreting these experiments because it is unclear whether the molecules that participated in a detected rearrangement were identical to the input molecules or were first modified in some way by the cell. This ambiguity complicates efforts to derive mechanistic details from such experiments.

To gain more specific information about the status of the molecules that recombine, we adopted a genetic approach, which allowed us to follow selectively the small subpopulation of transfected DNA molecules that includes the biologically active ones. For these studies we constructed reciprocal pairs of linear dimers, which differed only in the arrangement of two temperature-sensitive genomes of simian virus 40 (SV40). Reciprocal dimers should be equally subject to damage inflicted by the cell because they are structurally identical. However, the biological consequences of that damage differ for the two members of a pair because they are informationally distinct. By measuring the infectivities of several pairs of reciprocal dimers after transfection (by using DEAE-dextran as a carrier), we estimated the combined damage due to fragmentation of the input molecules and modification of their ends. We used independent approaches to confirm that fragmentation and end modification do occur and to estimate their individual magnitudes. Collectively, these studies indicate that the transfected DNA molecules that travel the biologically relevant pathway are moderately damaged, with about 15 to 20% of the molecules sustaining modification of at least one end and with the population as a whole suffering about one double-strand break per 5 to 15 kilobases.

MATERIALS AND METHODS

Cells, viruses, DNA assays. Procedures for growth of the established monkey kidney CV1 cell line have been de-

* Corresponding author.
† Present address: Biogen, Inc., Cambridge, MA 02142.

scribed (58, 61). The temperature-sensitive mutants of SV40 were obtained from P. Tegtmeyer and R. Martin. SV40 strain *su*1901 was isolated as a viable virus after transfection of CV1 cells with DNA from plasmid pJJ4, which contained complete SV40 and pBR322 sequences joined together at their unique *Taq*I and *Cla*I sites, respectively (60). In *su*1901, a segment of the intron in the gene for T antigen extending for 152 base pairs from the *Taq*I site toward the large exon has been replaced by a continuous segment of pBR322 DNA extending for 145 base pairs from the *Cla*I site toward the gene for ampicillin resistance. The sequence of *su*1901 will be described elsewhere. Strain *in*1902, which carries a unique *Fnu*DII site in place of the normal SV40 *Taq*I site, arose during infection with SV40 DNA that had been cleaved at the *Taq*I site (see Table 5). Presumably this strain arose by cell-mediated filling in of the *Taq*I ends followed by blunt-end ligation. The wild-type strain is Rh911a. DNA plaque assays were performed as previously described (59).

DNA preparation. SV40 DNA was extracted by the method of Hirt (18) 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/cm^3 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). All SV40 DNA preparations were labeled in vivo with [^{3}H]thymidine (New England Nuclear Corp.). Specific activities ranged from 10×10^3 to 75×10^3 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 ^{32}P (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 (18), treated with RNase (20 μg/ml), phenol extracted, ethanol precipitated, and resuspended in 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate).

Enzymatic reactions. Restriction enzymes were used according to the recommendations of the supplier. Enzymes were removed by extracting twice with redistilled phenol equilibrated with 10 mM Tris-hydrochloride (pH 7.6)–1 mM EDTA–400 mM NaCl. The DNAs were precipitated with ethanol and stored in 10 mM Tris (pH 7.6)–10 mM NaCl–1 mM EDTA. The ends of the linear molecules were verified to be undamaged by ligating a sample with T4 DNA ligase as described previously (53) and analyzing the ligated material on agarose gels. DNAs were treated with *Eco*RI methylase (kindly supplied by Paul Modrich) in reactions containing 100 mM Tris-hydrochloride (pH 8.0), 5 mM β-mercaptoethanol, 0.4 mg of bovine serum albumin per ml, 1.1 μM *S*-adenosyl-L-methionine, and 150 μg of SV40 DNA per ml. Incubation was at 37°C for 30 min. S1 nuclease digestions were done at 50 U/ml in reactions containing 30 mM sodium acetate (pH 4.6), 1 mM zinc acetate, 50 μg of tRNA per ml, and 225 ng of DNA per ml at 37°C for 30 min.

Construction of oligomeric DNA. Full-length linear dimers were constructed from pair-wise combinations of SV40 genomes that were wild type or were marked with temperature-sensitive mutations (Fig. 1). For each pair of genomes

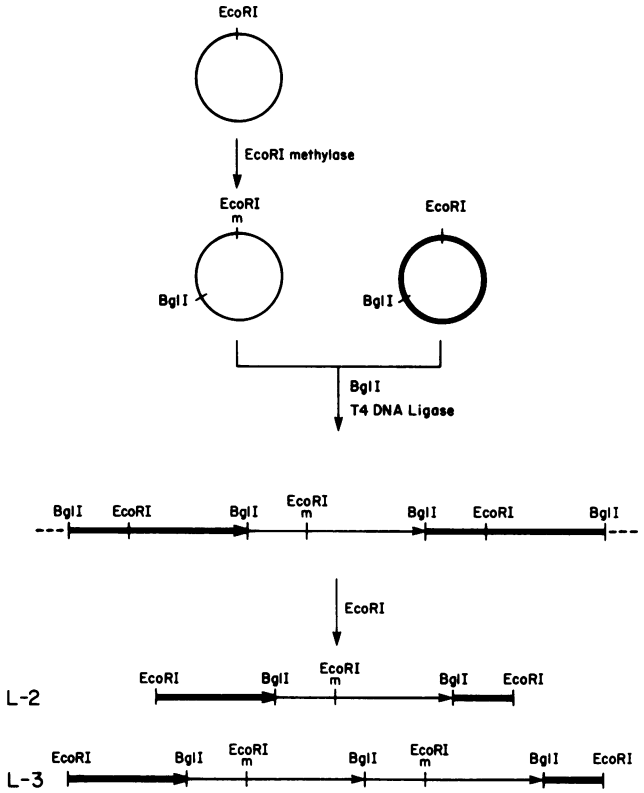

FIG. 1. Scheme for constructing dimeric DNA molecules. m denotes the site of methylation by *Eco*RI methylase. See the text for details.

to be joined, one member first was treated with *Eco*RI methylase to render its unique *Eco*RI site resistant to subsequent cleavage. Complete methylation was verified on agarose gels by the absence of linears after incubation with *Eco*RI. After phenol extraction and ethanol precipitation, the methylated DNA was mixed with a three- to fivefold molar excess of unmethylated DNA and digested with *Bgl*I. After phenol extraction and ethanol precipitation, the *Bgl*I-cut DNA was incubated with T4 DNA ligase at a DNA concentration of 400 to 800 μg/ml to favor formation of very large oligomers.

Ligation at *Bgl*I-generated ends assured that only head-to-tail junctions would be produced (54), and excess unmethylated DNA favored the ligation of unmethylated DNA to both ends of methylated molecules. T4 DNA ligase was inactivated by heating at 68°C for 10 min, and the reaction was diluted fourfold for *Eco*RI digestion. Because of the artificial restriction enzyme difference created by methylating the *Eco*RI site, digestion of the ligated mixture with *Eco*RI endonuclease generated a variety of linear oligomeric molecules containing one or more internal methylated genomes. Some mixtures were phenol extracted and ethanol precipitated again before being loaded onto preparative agarose gels. DNA was extracted from agarose as described previously (53).

The structure of purified linear dimers was verified by restriction enzyme analysis (data not shown). Greater than 95% of dimer-sized material was resistant to cleavage by *Eco*RI, indicating that very little of the material was the result of incomplete *Eco*RI digestion of adjacent unmethylated genomes. Digestions with *Taq*I, *Hpa*II, and *Bgl*I generat-

ed only those fragments consistent with the expected structural map.

Circular dimers were formed by ligating linear dimers at a DNA concentration of 0.3 to 0.9 μg/ml. The reaction was terminated by adding EDTA to 5 mM. Greater than 95% of the linear dimers formed circles and were not further purified. *Bgl*I digestion of a purified circular dimer generated only one band comigrating with full-length linear monomers on agarose gels, verifying that the product of the ligation reaction was circular.

Linear partial dimers were constructed as described previously (54). Portions of wild-type SV40 DNA were cleaved with one of several restriction enzymes (*Eco*RI, *Bam*HI, *Taq*I, *Hpa*II, or *Hae*II) that recognize a unique site in the genome and then treated with bacterial alkaline phosphatase to prevent subsequent in vitro ligation at those sites. These portions were digested with restriction enzyme *Bgl*I, mixed in pairs, and incubated with T4 DNA ligase. After electrophoresis on agarose gels, each mixture displayed three distinct bands: unit-length linears, a fragment of a monomer, and a partial dimer with terminal repeats of the segment between the two restriction sites at which the DNAs were cleaved initially. Partial dimers were extracted from the agarose gels as described previously (53).

RESULTS

Reciprocal linear dimers produce different frequencies of wild-type progeny. To estimate the sensitivity of transfected DNA to damaging events by the cell, we constructed several linear dimers of SV40, which contained pairs of mutant temperature-sensitive genomes linked in defined arrangements. Two distinct arrangements are possible for each pair of mutants, and one such pair of linear dimers (reciprocal dimers) is diagrammed in Fig. 2A. Either dimer can generate a monomer-length wild-type genome by intramolecular homologous recombination (Fig. 2B). However, the locations of the product wild-type sequences differ in the two dimers; in type I dimers these sequences are internal and contiguous, whereas in type II dimers they are terminal and split (see Fig. 2). As a consequence, a type II dimer can yield a wild-type genome by this pathway only if the ends of the molecule join precisely to restore the gene for the major capsid protein. By contrast, production of a wild-type genome from a type I dimer does not require precise end joining. Therefore, the production of wild-type virus by type II dimers relative to type I dimers reflects the efficiency with which ends are joined correctly inside the cell. This measurement of precise end joining reflects the extents of end modification and fragmentation suffered by the active subpopulation of transfected DNA molecules.

The results from transfections with five pairs of reciprocal dimers are presented in Table 1. Production of wild-type genomes was measured directly by plaque assay at the nonpermissive temperature. The progeny virus in the plaques that arose were judged to be wild type by testing at permissive and nonpermissive temperatures as described in Table 1, footnote *b*. The production of wild-type genomes was expressed in terms of two recombination values: %R, in which the number of wild-type plaques was compared to the infectivity of the mutant dimer assayed at the permissive temperature; and %W, in which the number of wild-type plaques was compared to the infectivity of a nonmutant (wild-type) dimer assayed at the nonpermissive temperature. Type I linear dimers behaved in two important respects like the type I, 1.84 partial dimers characterized previously

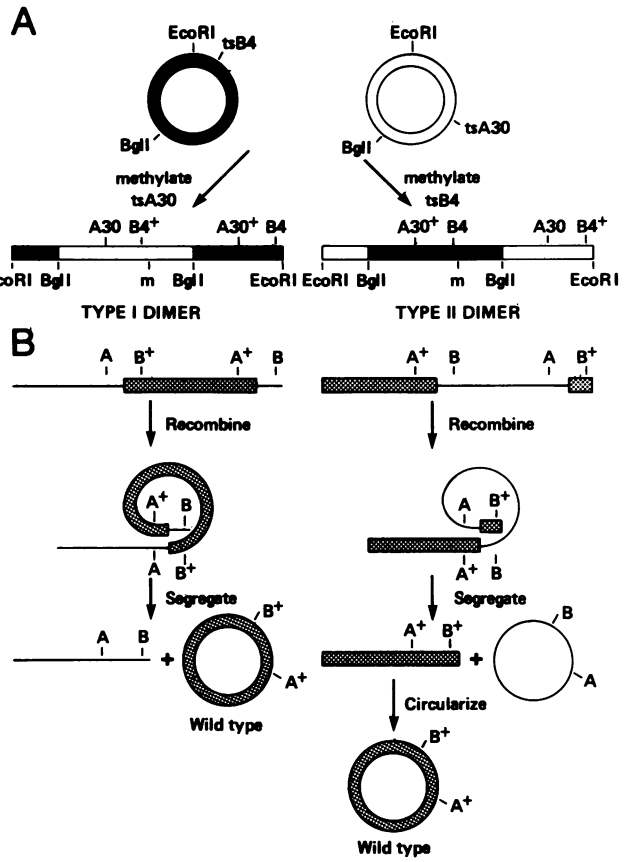

FIG. 2. Reciprocal pair of linear dimers. (A) Construction. Type I dimers were constructed by linking methylated *tsA*30 genomes to unmethylated *tsB*4 genomes as described in Fig. 1. Type II dimers were constructed by linking methylated *tsB*4 genomes to unmethylated *tsA*30 genomes. The wild-type counterparts of the mutations are indicated by a superscript +. m denotes the site of methylation by *Eco*RI methylase. (B) Production of circular wild-type genomes. Sequences that will form the product wild-type genomes are indicated as cross-hatched boxes. The possible products of homologous recombination between the sites of temperature-sensitive mutations in the reciprocal linear dimers are illustrated.

(Table 1) (54): the formation of wild type was roughly proportional to the physical distance between mutations and, as indicated by the dimers containing two *tsA* mutations, was not dependent upon normal replication of the viral DNA, which requires a functional *A* gene product (48). These two criteria indicate that homologous recombination was involved in the process of wild-type production.

In all cases, type II linear dimers produced a lower frequency of wild type than did the corresponding type I dimer, as indicated by the type II/type I ratios in Table 1. Overall, type II dimers produced wild-type genomes at about 30% the frequency of type I dimers. Although lower than those for type I dimers, type II recombination values were still significantly above the background of wild-type production due to minor recombination pathways such as gene conversion and multiple exchanges. We estimated the background from results with type II, 1.84 partial dimers, which cannot generate wild-type genomes by the recombination pathway illustrated in Fig. 2B because they are missing essential sequences at their ends (Table 1) (54). The deficiency in wild-type production by type II dimers relative to type

TABLE 1. Wild-type production by type I and type II dimers

Dimer type	Type I[a]	Recombination[b]		Type II	Recombination		Ratio of II/I	
		%W	%R		%W	%R	%W	%R
Linear	m*A*30:*B*228	50	54	m*B*228:*A*30	12	20	0.24	0.37
	m*A*30:*B*4	47	58	m*B*4:*A*30	15	10	0.32	0.17
	m*A*28:*B*4	56	52	m*B*4:*A*28	24	37	0.43	0.71
	m*A*30:*A*28	6.8	5.8	m*A*28:*A*30	1.6	0.9	0.24	0.16
	m*A*30:*A*58	3.7	4.9	m*A*58:*A*30	0.4	0.7	0.11	0.14
Circular	m*A*30:*B*228	60	73	m*B*228:*A*30	35	54	0.58	0.74
	m*A*30:*B*4	62	61	m*B*4:*A*30	71	67	1.15	1.10
	m*A*28:*B*4	29	24	m*B*4:*A*28	36	35	1.24	1.46
	m*A*30:*A*28	1.2	1.2	m*A*28:*A*30	1.1	1.1	0.92	0.92
	m*A*30:*A*58	0.5	0.4	m*A*58:*A*30	0.6	0.5	1.20	1.25
Linear partial[c]	*A*30:*B*228	45		*B*228:*A*30	0.4		0.01	
	*A*30:*B*4	31		*B*4:*A*30	0.6		0.02	
	*A*28:*B*4	33		*B*4:*A*28	0.6		0.02	
	*A*30:*A*28	6.1		*A*28:*A*30	0.4		0.07	
	*A*30:*A*58	2.4		*A*58:*A*30	0.2		0.08	

[a] Each dimer is designated type I and type II according to the definition in Fig. 2 and in the text. The physical distances between the mutations used in these experiments (expressed as percentage of the SV40 genome) are *tsA*30 to *tsB*228, 43%; *tsA*30 to *tsB*4, 32%; *tsA*28 to *tsB*4, 27%; *tsA*30 to *tsA*28, 5%; *tsA*30 to *tsA*58, 5%. These values assume that the mutations lie at the center of the restriction fragments to which they have been localized (25).

[b] CV1 monolayers were infected with 0.2 to 5 ng of purified dimers and incubated at 41°C; 10-fold lesser amounts were assayed at 33°C. The data are the average of two to four experiments, each assayed in duplicate or triplicate. Recombination frequencies, as described previously (54), were calculated as follows: %W = 100% (41°C titer of temperature-sensitive dimer DNA per nanogram)/(41°C titer of wild-type dimer DNA per nanogram); %R = 200% (41°C titer of temperature-sensitive dimer DNA)/(33°C titer of temperature-sensitive dimer DNA). Linear and circular dimers constructed from wild-type DNA were assayed at 41 and 33°C to obtain a correction factor for differences in plaquing efficiency at the two temperatures. Plaques were picked from the 41°C plates of experiments involving *tsA*28 and *tsB*4 and replated at 41 and 33°C. The number of plaques judged to contain complementing temperature-sensitive genomes out of the number tested were: type I linear, 3 of 19; type II linear, 0 of 18; type I circle, 0 of 19; type II circle, 0 of 15.

[c] The data for linear partial dimers have been reported previously (54).

I dimers could arise (i) because cell-mediated end joining, per se, occurs at low efficiency, (ii) because the ends of many input molecules in the active subpopulation are modified, and (iii) because the ends of many input molecules in the active subpopulation are separated physically from each other by fragmentation. We explored each of these possible explanations in turn. However, first we demonstrated that the difference is not a trivial consequence of the method of construction.

Reciprocal circular dimers produce the same frequency of wild-type progeny. By virtue of their mode of construction, reciprocal linear dimers differed in the position of the methylated *Eco*RI site relative to the product wild-type genomes (Fig. 2). To determine whether methylation influenced the observed recombination frequencies, we converted linear dimers to circular dimers by ligation in vitro and assayed their infectivity (Table 1). Overall, the two types of circular dimers produced wild-type progeny at essentially the same mean frequency; the average ratio of type II to type I wild-type production was 1.06. These results indicate that methylation does not differentially affect the recombinational capabilities of the two kinds of constructed dimers and thus cannot account for the different frequencies of wild type produced by type I and type II linear dimers.

Two noteworthy observations are associated with the recombination of circular dimers. First, all 34 plaques tested from infections with *A*:*B* circular dimers were wild type (see Table 1). If two monomers were the normal product of the intramolecular recombination of a circular dimer or if recombination occurred frequently within a pool of replicating dimers, one might have expected a relatively high proportion of plaques formed by complementation between *tsA* and *tsB* genomes. Second, the recombination frequencies for *A*:*A* circular dimers were consistently lower than those obtained with the corresponding type I linear dimers (Table 2). That pattern is not evident in the *A*:*B* dimers. It is not clear why *A*:*A* circular dimers, but not *A*:*B* circular dimers, are less efficient at producing wild-type recombinants. Perhaps circular molecules are poorer substrates for recombination than linear molecules and the ability of *A*:*B* dimers to increase their copy number by replication compensates for this difference. Other studies also have suggested that circular molecules are relatively poor substrates for recombination

TABLE 2. Wild-type production by linear and circular *A*:*A* dimers

Dimer[a]	Form	Recombination[b]	
		Type I	Type II
*A*30:*A*28	Linear	6.3	1.3
	Circle	1.2	1.1
*A*30:*A*58	Linear	4.3	0.6
	Circle	0.5	0.6
*A*28:*A*58	Linear	1.2	0.5
	Circle	0.6	0.4

[a] The data for *A*30:*A*28 and *A*30:*A*58 are taken from Table 1. The data for *A*28:*A*58 involves dimers with two temperature-sensitive mutations in the same restriction fragment.

[b] Recombination frequencies for the type I and type II dimers are the average of %R and %W values, which are described in Table 1, footnote *b*.

(22; C. T. Wake, F. Vernaleone, and J. H. Wilson, manuscript in preparation). We do not yet understand the molecular basis for these two observations.

Efficiency of end joining. One potential source of the difference in wild-type production by reciprocal pairs of linear dimers is a low efficiency of end joining. A type I linear dimer can yield a circular wild-type monomer directly by homologous recombination, whereas a type II linear dimer can yield a circular wild-type monomer only after a subsequent end-joining event (Fig. 2B). If the end-joining process were only 30% efficient, the differential infectivity of reciprocal pairs of linear dimers would be accounted for. (Note that linear dimers could produce circular wild-type monomers by an alternate pathway, in which the order of recombination and end-joining events is reversed; that is, by a pathway in which end joining occurs first and recombination second. We do not know which pathway predominates. However, in the alternate pathway, a reciprocal pair of linear dimers first would be converted to equivalent dimer circles, which, as demonstrated above, do not differ in their capacity to generate wild-type progeny. Consequently, for that pathway the efficiency of end joining cannot contribute to the difference in wild-type production by reciprocal pairs of linear dimers.)

Several observations indirectly indicated that intramolecular end joining occurs with high efficiency. SV40 genomes linearized at eight different sites by restriction enzymes that leave a variety of complementary ends were equally infectious even after treatment with bacterial alkaline phosphatase (data not shown). Thus, neither the composition of the complementary ends nor the absence of 5′ phosphates influences the efficiency of end joining in vivo. In other studies, in which SV40 genomes were linearized in nonessential regions and cloned into pBR322, we demonstrated that unrelated, noncomplementary DNA ends were joined together in vivo with approximately the same efficiency as complementary ends (60). In those studies, six different linear SV40 genomes, each with less than 300 base pairs of pBR322 DNA attached at their ends, averaged about 80% the infectivity of linear monomers with complementary ends.

TABLE 3. Infectivity of circular and linear forms of SV40 DNA

DNA species[a]	Event	Infectivity (%)[b]
Supercoiled circles	None	100
Nicked circles (spontaneous)	Repair of nicks	23
Nicked circles (heteroduplexes)	Repair of nicks	22
Linear monomer (*Eco*RI)	Joining of ends	13
Linear monomer (*Bgl*I)	Joining of ends	22
Linear partial dimer	Homologous recombination	16

[a] All DNA species were prepared from the same stock of wild-type DNA and were purified from agarose gels. Linear monomers were prepared by digestion with the indicated restriction enzymes, which each cleave SV40 DNA at a unique site. Linear 1.84 partial dimers contained a tandem duplication of 84% of the viral genome extending from the *Hpa*II site (0.06 map units) to the *Taq*I site (0.90 map units). Spontaneous nicked circles represent the form II DNA normally produced during isolation of viral DNA. Heteroduplex nicked circles were prepared by mixing, denaturing, and reannealing linear monomers cut with *Taq*I and *Hpa*II.

[b] Infectivity is expresssed as a percentage of that for supercoiled circles. The data are averaged from three experiments in which each sample was assayed in quadruplicate. For each DNA species, equal numbers of molecules were assayed, except for supercoiled circles, in which case only one-fifth that number of molecules was assayed. The average infectivity of supercoiled circles was 2.2×10^3 PFU/ng.

TABLE 4. Infectivity of linear forms of SV40 DNA with different kinds of ends

DNA species[a]	Ends	Infectivity (%)[b] ± SD
*su*1901 (*Taq*I)	Sticky	100
*su*1901 (*Fnu*DII)	Blunt	96 ± 18
*su*1901 (*Taq*I + *Fnu*DII)	Mismatched	94 ± 18

[a] All DNA species were prepared from the same stock of *su*1901 DNA by digestions with the restriction enzymes shown in parentheses. All digestions were judged to be complete by agarose gel electrophoresis of the products. To verify that *su*1901 (*Taq*I + *Fnu*DII) was cut by both enzymes, whose recognition sites are separated by only 130 base pairs, portions of all three samples were digested with an additional restriction enzyme, *Kpn*I, and the products were displayed by electrophoresis on agarose gels.

[b] Infectivity is expressed as a percentage of that for *Taq*I-cut *su*1901. The data were normalized to the infectivity of this sample in each of four experiments and the averages of the four experiments with standard deviations are shown. Within each experiment, 0.2 ng of each sample was assayed on each of 8 to 10 plates. The average specific infectivity of *Taq*I-cut *su*1901 was 174 PFU/ng.

To assess the efficiency of end joining more directly, we measured the infectivities of a variety of circular and linear forms of SV40 DNA. The infectivities of nicked circles, linear monomers, and a 1.84 partial dimer were approximately the same (Table 3), suggesting that repair of nicks, joining of ends, and intramolecular homologous recombination all occur with comparable efficiencies. The four- to eightfold higher infectivity of supercoiled SV40 DNA may reflect some special advantage conferred by the supercoiled state. To measure the efficiency with which different kinds of ends are joined, we made use of a substitution mutant of SV40, *su*1901, which contains a unique *Taq*I site and a unique *Fnu*DII site within the nonessential intron region of the gene for T antigen. Cleavage of *su*1901 with *Taq*I leaves complementary sticky ends, cleavage with *Fnu*DII leaves blunt ends, and cleavage with *Taq*I and *Fnu*DII leaves mismatched ends. These different forms are equally infectious (Table 4), indicating that complementary ends, blunt ends, and mismatched ends are joined with the same efficiency inside the cell. The studies in Tables 3 and 4 are most easily understood if the efficiency of end joining approaches 100% for intact molecules that reach the nucleus. Although these experiments do not speak directly to the efficiency of end joining subsequent to recombination, we consider it unlikely that a low efficiency of end joining is a principal contributor to the difference in wild-type production by reciprocal pairs of linear dimers.

End modification of transfected DNA molecules. A second potential source of the difference in wild-type production by reciprocal pairs of linear dimers is modification of the ends of the input molecules during transfection. The two types of linear dimers should be differentially sensitive to end modification, where end modification is defined as a change at either terminus that alters the normal sequence at the point of end closure. Type I dimers would be expected to be insensitive to end modification, because the sequences around the termini do not contribute to the product wild-type genome. By contrast type II dimers should be exquisitely sensitive to end modification, because the sequences at their ends form an essential segment of the gene encoding the

major capsid protein in the product wild-type genome. To test these general expectations, we treated reciprocal pairs of linear dimers with nuclease S1, which is specific for single-stranded DNA and therefore should remove nucleotides from the termini. Under digestion conditions in which the infectivity of *Eco*RI-linearized wild-type monomers was reduced 10-fold, the frequency of wild type produced by type I dimers did not change significantly, whereas the frequency produced by type II dimers was reduced to 0.3 to 2% (data not shown). These results verify the differential sensitivity of reciprocal dimers to end modification.

For wild-type production from type II dimers to be reduced to 30% that of type I dimers, 70% of the input molecules would have to have had at least one end modified. To test that possibility, we analyzed viable progeny arising from transfections with monomeric SV40 genomes that had been linearized by cleavage within the nonessential intron in the gene for T antigen. For these studies we used both a wild-type SV40 and an insertion mutant, *in*1902, in which the unique *Taq*I site within the intron had been converted to a unique *Fnu*DII site. Genomes cleaved at either of these unique restriction sites have about 150 base pairs of nonessential DNA at their termini. Thus, end-modification events that do not remove more than about 150 base pairs from an end or add more than about 200 total base pairs (the approximate packaging limit) should yield viable progeny virus. Furthermore, modified genomes within these limits should be found among the viable progeny virus at the same frequency they are present inside the cell, because DNA ends of all kinds are joined with the same efficiency (Table 4).

As indicated in Table 5, 86% of *Taq*I-linearized wild-type SV40 DNA and 77% of *Fnu*DII-linearized *in*1902 DNA regenerated the respective restriction sites during transfection. The restriction digestion patterns for a selection of *Taq*I$^-$ viable genomes are shown in Fig. 3. Among those viable genomes that did not regenerate the restriction sites, the majority involved small changes, and a minority involved somewhat larger deletions or insertions (compare the bands at position B in Fig. 3). By restriction analysis with other

TABLE 5. Modification of sticky- or blunt-ended linear genomes of SV40

Transfected species[a]	Progeny type[b]	No. of plaques (%)
Wild type (*Taq*I)	*Taq*I$^+$	127 (86)
	*Taq*I$^-$ (small change)	18 (12)
	*Taq*I$^-$ (deletion)	2 (1.4)
	*Taq*I$^-$ (insertion)	1 (0.7)
*in*1902 (*Fnu*DII)	*Fnu*DII$^+$	62 (77)
	*Fnu*DII$^-$ (small change)	16 (20)
	*Fnu*DII$^-$ (deletion)	2 (2.4)
	*Fnu*DII$^-$ (insertion)	1 (1.2)

[a] Linear genomes of wild-type SV40 were prepared by digestion of viral DNA or DNA from plasmid pJJ4 (60) with the restriction enzyme *Taq*I. Results from transfection with these two sources of *Taq*I-cut linear SV40 were indistinguishable and were added together. Linear genomes of *in*1902 were prepared by digestion of viral DNA with restriction enzyme *Fnu*DII. All digestions were judged to be complete by electrophoresis on agarose gels.

[b] Progeny viruses were isolated from well-separated plaques that arose after transfection with 0.02 ng or less DNA per plate (generally less than 10 plaques per plate). Progeny virus were categorized after digestion with restriction enzymes *Hin*dIII and *Taq*I or *Tha*I (an isoschizomer of *Fnu*DII) as described in the legend to Fig. 3. Greater than 97% of all characterized plaques contained a single species of DNA. Results from mixed plaques were excluded from the table.

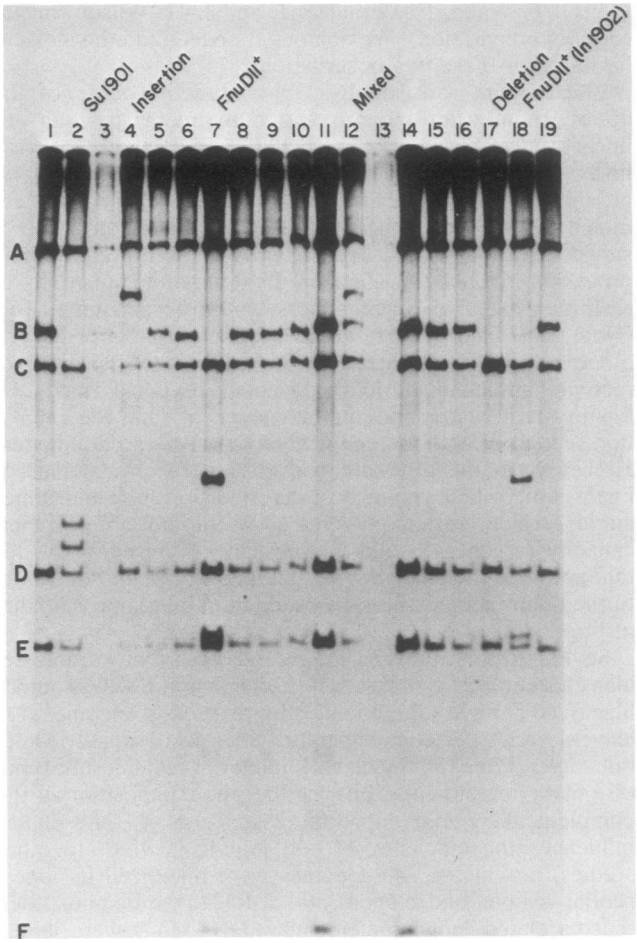

FIG. 3. Restriction analysis of *Taq*I$^-$ viable genomes that arose after transfection with *Taq*I-cleaved wild-type SV40 DNA. Seventeen viable genomes, which were lacking a *Taq*I site, were tested for the presence of a *Fnu*DII site, which could have been generated in vivo by filling in the *Taq*I ends and ligating the blunt ends. ^{32}P-labeled DNA was prepared from plaque suspensions by the miniwell method (see the text). The labeled DNA was digested with *Hin*dIII and *Tha*I (an inexpensive isoschizomer of *Fnu*DII) and electrophoresed on polyacrylamide gels, which were then dried and autoradiographed. The positions of the *Hin*dIII fragments from wild type (lane 1) are indicated at the side. *su*1901 (lane 2) was included as a positive control for *Tha*I cleavage, because wild-type SV40 contains no *Fnu*DII sites. Lanes 3 to 19 contain different viable genomes that lack the normal *Taq*I site. As expected, only band B shows any variation from wild type in these genomes, because it contains the normal *Taq*I site. The genomes displayed in lanes 7 and 18 (*in*1902) possess a new *Fnu*DII site as indicated by the cleavage of band B into two smaller fragments. The pattern generated by *Hin*dIII plus *Tha*I cleavage of these two genomes is indistinguishable from the pattern generated by *Hin*dIII plus *Taq*I cleavage of wild-type DNA, indicating that the new *Fnu*DII site is at the position of the original *Taq*I site. The majority of viable genomes show very little change in band B. However, there are clear examples of an insertion (lane 4) and a deletion (lane 17). The pattern in lane 12 indicates a mixed plaque containing an insertion and a wild type (the band at position B was cleaved by *Taq*I digestion). This plaque was not included in the results in Table 5.

multi-cut enzymes, the small changes were estimated to involve less than 25 base pairs (data not shown). Overall, these studies indicate that about 15 to 20% of the input linear monomers in the active subpopulation had one or both ends modified. If the same fraction of the ends of linear dimers was modified, wild-type production by type II dimers would have been 80 to 85% that of type I dimers. Thus, end modification does contribute to the difference in wild-type production by reciprocal pairs of linear dimers, but it is not the major source of the difference.

Fragmentation of transfected DNA molecules. The third potential source of the difference in wild-type production by reciprocal pairs of linear dimers is fragmentation of the input molecules by internal breakage. Reciprocal dimers are differentially sensitive to internal breaks because of the positions of the product wild-type genomes within the dimers. Any unrepaired breakage of type II dimers will separate segments of the product wild-type genome onto different molecules, thereby lowering the frequency with which wild type can be generated. By contrast, wild-type production by type I dimers should be affected only by breaks that directly interrupt the product wild-type genome; that is, type I dimers should not be sensitive to breaks between the ends of the dimer and the sites of the mutations (see Fig. 2B). Thus type II dimers have an effective target size equivalent to 2 SV40 genome lengths, whereas type I dimers have a target size of only about 1 genome length. Actually, type I molecules would be expected to survive random breakage better than simple monomers, because they have a redundancy of information in the region between the mutations within which a recombination event can generate a wild-type genome. The theoretical survival curves for type I and type II dimers are shown in Fig. 4; the mathematical basis for the type I survival curves is described in the Appendix.

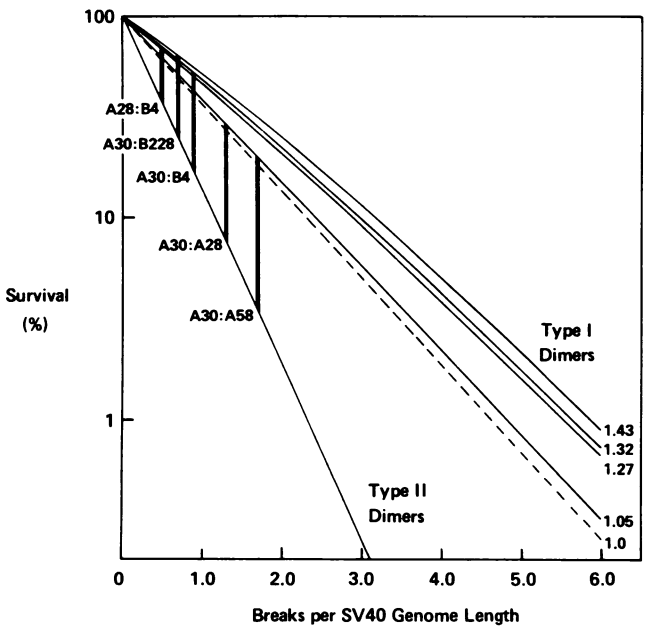

FIG. 4. Theoretical survival curves for type I and type II dimers as a function of breakage. Heavy vertical lines indicate the number of breaks at which the ratio of the survival curves matches the experimentally determined ratio of wild-type production. Type I dimers are plotted as survival = $(100\%)(1 + rH)e^{-H}$, where H is the number of breaks per SV40 genome length and r is the fractional length of the genome between the mutations (see the Appendix). The target size for type I dimers is indicated on the right (see Table 6). Type II dimers are plotted as survival = $(100\%)e^{-2H}$.

TABLE 6. Calculated number of breaks per SV40 genome length in the input dimers

Dimer	Ratio of II/I[a]	Adjusted ratio of II/I[b]	Target for Type I dimer[c]	Breaks per genome[d]
*A*30:*B*228	0.31	0.38	1.43	0.7
*A*30:*B*4	0.25	0.31	1.32	0.9
*A*28:*B*4	0.43	0.53	1.27	0.5
*A*30:*A*28	0.20	0.25	1.05	1.3
*A*30:*A*58	0.14	0.17	1.05	1.7

[a] The ratio of the wild-type production by type II and type I linear dimers was averaged from Table 1 after the highest and lowest values were discarded.

[b] The ratio of wild-type production by type II and type I dimers was divided by 0.815 to compensate for end modification.

[c] The target for type I dimers varies depending on the separation of the mutations, which define the region in which recombination can generate wild type. For this analysis, the mutations were assumed to lie at the center of the restriction fragment in which they were localized (25).

[d] Breaks per genome length were calculated from the theoretical curves according to the equation: ratio of type II/type I = $e^{-2H}/(1 + rH)e^{-H}$, where H is the number of breaks per SV40 genome length and r is the repeated region of the genome between the mutations (target − 1). See the appendix.

These curves can be used to estimate the extent of breakage by finding the number of breaks at which the ratio of the survival of type II to type I molecules matches the experimental ratio of wild-type production. These ratios are marked by vertical lines in Fig. 4. The parameters used in calculation and the estimated number of breaks per SV40 genome length are shown in Table 6. These estimates of breakage range from 0.5 to 1.7 breaks per SV40 genome length with a mean of about 1.0.

These estimates of fragmentation involve two measured parameters: the extent of end modification (15 to 20%) and the efficiency of end joining (100%). In addition they assume that the ratio of wild-type production by reciprocal dimers, which requires homologous recombination in mutationally defined regions, can be equated directly with the ratio of survival. To measure the extent of breakage in a way that is independent of these parameters, we took advantage of the predicted dependence of infectivity on the length of the repeated region in an oligomer (see the Appendix). Different length oligomers yield a family of survival curves like those for type I molecules in Fig. 4.

For these studies we constructed several nonmutant linear oligomers of SV40 ranging in size from 1.1 to 2.0 genomes in length. Production of wild-type monomers from such oligomers is independent of end modification and the efficiency of end joining; furthermore, homologous recombination anywhere within the repeated region of a molecule will generate a circular wild-type genome. The infectivities of these oligomers were measured relative to those of linear SV40 monomers, which were cleaved in a nonessential region to eliminate the effects of end modification on their infectivity. These data are displayed in Fig. 5 against a set of theoretical curves that show the predicted dependence of relative infectivity on length for several numbers of breaks per SV40 genome. These measurements are consistent with about 0.3 to 1.0 breaks per SV40 genome and agree reasonably well

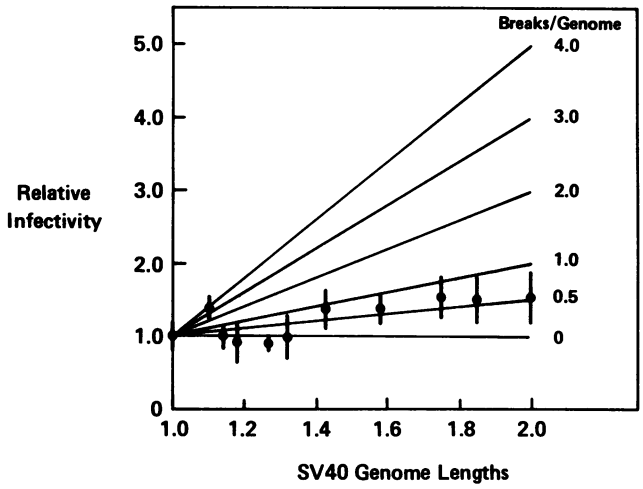

FIG. 5. Relative infectivity of different-length species of wild-type SV40 DNA. All DNA species were prepared from the same stock of wild-type DNA, and all species were isolated from agarose gels before transfection. Partial dimers and full-length dimers were constructed as described in the text. Linear monomers were isolated from the preparative gel of the pair-wise combination of molecules cut with *Hpa*II and *Taq*I, which was used to prepare a 1.85 partial dimer. Thus, the linear monomers were an equal mixture of *Hpa*II-cut and *Taq*I-cut molecules that had been cleaved and religated at the *Bgl*I site. This procedure ensured that the linear monomers had been through the same series of enzymatic steps as the partial dimers. An equivalent number of molecules (3×10^7) of each molecular species, as determined from the specific activity of the wild-type stock of DNA, were assayed in triplicate in two separate experiments. The data points with standard deviations represent the average of the two experiments. The average specific infectivity of the linear species in these experiments was 234 PFU/ng. The straight lines represent the ratio of the theoretical survival curves for partial dimers to monomers at several numbers of breaks per genome length. Ratio = $e^{-H}(1 + rH)/e^{-H} = 1 + rH$, where H is the number of breaks per SV40 genome length and r is the length of the repeated region expressed as a fraction of the genome.

with the estimates derived from reciprocal linear dimers. Below we describe another independent method for estimating breakage and show that it also yields similar estimates of the degree of fragmentation of transfected DNA. Thus, we conclude that transfected DNA is moderately fragmented and that fragmentation is primarily responsible for the difference in wild-type production by reciprocal pairs of linear dimers.

DISCUSSION

In this study we applied a genetic analysis to obtain information about the status of transfected DNA molecules that reach the nucleus. We used the carrier DEAE-dextran to mediate the transfections in this study and others because it yields the highest specific infectivity (PFU per nanogram) for SV40 DNA (14, 28). Our analysis allows us to examine selectively that subpopulation of transfected molecules that gives rise to viable viral genomes. This active subpopulation includes not only the intact molecules but also the cotravelers that have been damaged along the way or inside the nucleus. We investigated the active subpopulation by assaying the viable virus produced (i) by reciprocal pairs of mutant linear dimers, (ii) by a series of wild-type oligomers of increasing length, and (iii) by monomers linearized at nonessential sites. These measurements reflect events within the active subpopulation as a whole, thereby allowing certain inferences to be drawn about the extent of damage it has suffered. We conclude that the active subpopulation is moderately damaged; overall it received about 1 break per 5 to 15 kilobases, and about 15 to 20% of the molecules had one or both ends modified. We stress that these estimates of damage are specific for DEAE-dextran-mediated transfection of DNA into CV1 monkey kidney cells. How accurately these measurements describe transfection in other vertebrate cells with DEAE-dextran or other common methods for introducing foreign DNA remains to be determined; however, they do form a sound basis for evaluating a large body of experimental data in monkey cells, and they suggest experimental approaches for determining the extent of damage in other systems.

The end-modification and fragmentation events we detected occurred early in transfection before viral replication began. This conclusion is implicit in the design of the experiments. Reciprocal pairs of linear dimers are differentially sensitive to end modification and fragmentation only before circularization, which is a prerequisite for viral replication (48). In addition, the linear genomes used to detect end-modification events and to verify fragmentation events (discussed below) were cleaved in the intron in the gene for T antigen, which is required for viral replication. As a consequence, functional mRNA for T antigen could not have been synthesized before end joining. Thus the events we detected in these studies should be common to all DNAs transfected with DEAE-dextran and not specific for SV40 DNA.

It could not be determined from our experiments whether the detected end-modification and fragmentation events occurred outside the cell, in the cytoplasm, or in the nucleus. However, it was clear that the damage occurred during transfection. The possibility that these events occurred during the construction phase was rendered unlikely by control experiments with the reciprocal dimers. Greater than 95% of linear dimers, which was the limit of detection, could be converted to circles by incubation with polynucleotide ligase, thereby ruling out a significant fraction of preexisting molecules with modified ends. In addition, transfection of these circular dimers, which were not purified further from the ligation reaction, eliminated the difference in wild-type production by reciprocal pairs, a result inconsistent with the presence of preexisting fragmented molecules.

Fragmentation. Our estimate of the extent of fragmentation of the active subpopulation is based on the production of wild-type progeny by reciprocal pairs of mutant linear dimers and on the infectivity of a series of wild-type oligomers. The agreement between these two estimates of breakage is significant because the methods involve entirely different assumptions with the exception that both methods assume breaks are delivered randomly to the active subpopulation. The analysis of reciprocal pairs of linear dimers is independent of the efficiency of uptake and transport of transfected DNA because the two types of dimers are structurally identical; however, it is dependent on events at the molecular termini. (Based on the identical infectivity of linear genomes with sticky, blunt, or mismatched ends, we used 100% as the efficiency of end joining and corrected for the frequency of end modification.) By contrast, the analysis of infectivity as a function of length of the input molecule is independent of events at the molecular termini because terminal sequences are not part of the product wild-type genomes; however, it is dependent on the efficiency of uptake and transport of different length molecules and on the

efficiency of recombination. (Our analysis assumed that the efficiency of uptake and transport was the same for molecules from 1 to 2 SV40 genome lengths and that the efficiency of recombination was 100% for all molecules.) The agreement between these two independent methods tends to validate the measurements and assumptions used in both; however, given the number of parameters involved in the two analyses, the agreement certainly could be fortuitous. Therefore, we developed a third independent argument (below); it also yields an estimate consistent with about 1 break per 5 to 15 kilobases.

If breaks were introduced randomly into the active subpopulation of transfected molecules, they should be evident among the viable progeny generated in end-joining experiments like those described in Table 5. For an SV40 genome linearized by restriction enzyme digestion within a nonessential region, an additional break in adjacent nonessential DNA would yield a viable genome with a deletion at the original restriction site. Four of the 229 progeny tested in these experiments contained substantial deletions that could be distinguished readily from the more prevalent small scale end-modification events discussed below. Given a target of 300 base pairs for detection of such deletions and given breaks distributed in a Poisson fashion among the active subpopulation, four deletions among 229 genomes corresponds to about 0.3 breaks per genome. In another study, we detected eight such deletions among 88 viable progeny that arose from genomes that were linearized within an intron about 125 base pairs longer than wild type, corresponding to 1.4 breaks per genome (D. S. Roth and J. H. Wilson, unpublished data). In that study we also analyzed four of the deletions by restriction mapping; each deletion had one endpoint close to the original end of the input linear genome, as would be expected if the deletion arose by the joining of one broken end with one input end. These studies confirm that fragmentation of input molecules does occur and also support the estimate of 1 break per 5 to 15 kilobases derived from reciprocal dimers and partial dimers. The estimates of breakage from all three methods are summarized in Table 7. The one assumption common to all three methods is that breaks are introduced randomly into the active subpopulation.

End modification. Modification of the ends of transfected DNA molecules is a less prevalent form of damage than fragmentation. Our estimate of the extent of end modification was based on an analysis of viable progeny that arose from sticky- or blunt-ended linear molecules which were generated by cleavage of circular genomes near the center of the intron in the gene for T antigen. The requirement for viability limited detection to those modifications that did not remove more than about 300 base pairs (did not remove essential DNA) or insert more than about 200 base pairs (did not exceed the packaging limit). The majority (189 of 229, 83%) of the viable genomes that were isolated from these transfections were not modified at all. Four of the 40 genomes that were altered could be readily classified as deletions and probably arose from input molecules that suffered an additional break in the intron, as discussed above. Of the remaining genomes, 34 of 36 (94%) had fewer than 25 nucleotides removed or added to the ends as estimated by analysis with restriction enzymes. The nucleotide sequence around the site of end joining in one of these genomes revealed a 9-base-pair deletion that removed the restriction site (T. Gudewicz and J. H. Wilson, unpublished data). In addition, two of the modified genomes that arose during transfection of sticky-ended molecules appeared to have been generated by filling in of the sticky ends followed by blunt-end ligation to create a characteristic new restriction site (see Fig. 4). The high efficiency with which restriction sites are regenerated by sticky- and blunt-ended molecules and the ability of somatic cells to fill in sticky ends have been noted previously (29, 46, 51). However, the degree of modification of untreated genomes that do not regenerate the relevant restriction site has not been investigated systematically. The extremely limited extent of the alterations in the majority class of modified genomes argues that end joining of modified and unmodified genomes alike is quite rapid relative to net addition or removal of base pairs at the ends of transfected DNA molecules. In particular, the studies with blunt-ended molecules argue strongly that ends of transfected DNA molecules in monkey cells are not subject to extensive degradation by single- or double-stranded exonucleases. On the contrary, the cellular enzymatic machinery seems balanced in favor of preserving sequence information by the rapid joining of DNA ends.

Addition of blocks of nucleotides (of unknown origin) to the input ends accounts for a minority (2 of 36) of end modifications in our experiments. However, the requirement for viability limits the length of added sequences due to the packaging constraint and thus probably leads to an underestimate of the true frequency of such events. Similar insertions constitute a frequent source (5 to 20%) of the detected mutational events (about 1 to 9% overall) in experiments with shuttle vectors passaged through animal cells and analyzed in bacteria (5, 35). Unfortunately, those frequencies cannot readily be related to the studies reported here. Consequently, the degree of underestimation is uncertain, although even a factor of 10 would not significantly affect the other conclusions derived from our overall estimate of end modification.

Consequences for somatic recombination. These measurements of damage to the active subpopulation of transfected DNA have a direct bearing on related investigations into somatic cell recombination. Somatic cells actively recombine foreign DNA by two general mechanisms: one that requires extensive sequence homology (homologous recombination), and a second that requires little, if any, sequence homology (nonhomologous recombination). Because homologous and nonhomologous recombination events occur at comparably high frequencies early during transfection (45, 53, 60; Roth and Wilson, unpublished data), the damaged members of the active subpopulation must be considered if mechanistic details are to be derived from such experiments.

TABLE 7. Estimates of fragmentation of transfected DNA

Basis for estimate	Breaks per SV40 genome	Kilobases per break
Reciprocal dimers	0.5–1.7	3–10
Partial dimers	0.3–1.0	5–17
Linear monomers	0.3–1.4[a]	4–17

[a] For linear monomers the number of breaks per SV40 genome length was calculated from the first two terms of the Poisson distribution (P_0 and P_1, the probabilities of receiving 0 or 1 hit). Higher order terms make no significant contribution and were ignored. Both estimates of breakage were calculated from the relationship $D = TP_1/(P_0 + TP_1) = THe^{-H}/(e^{-H} + THe^{-H}) = TH/(1 + TH)$, where D is the fraction of genomes with sizable deletions among the viable progeny, H is the number of breaks per SV40 genome length, and T is the target in which a break can yield a viable genome. For the estimate of 0.3 breaks, $T = 300/5{,}243 = 0.057$; for the estimate of 1.4 breaks, $T = 425/5{,}368 = 0.079$ (see text).

In a previous investigation of nonhomologous recombination in transfected DNA with SV40:pBR322 chimeric molecules, we concluded from a genetic analysis that the primary mechanism involved end-to-end joining (60). To account for the infectivity of full-length chimeric molecules (about 1% that of SV40 linear genomes) by an end-joining mechanism, we assumed that the input DNA was fragmented during transfection and that the broken ends joined with high efficiency. Both of these assumptions have been validated by our present studies. Thus, we reaffirm our conclusion that the principal mechanism of nonhomologous recombination in transfected DNA is end-to-end joining and raise the possibility that other nonhomologous recombination events in somatic cells, such as integration of foreign DNA into host chromosomes, also may occur by end joining, perhaps at transient chromosomal breaks (10). Certainly the limited extents of homology at the novel recombination junctions in both kinds of nonhomologous events are quite similar (15, 16, 39, 43–45, 60).

One initially surprising result of the experiments with reciprocal pairs of dimers was that transfection so rarely produced plaques that contained complementing genomes (only 3 of 71 tested plaques). If the molecules involved in a homologous recombination event were identical to the input molecules, such a result would suggest that the mechanism of recombination is nonreciprocal, that is, only one of the two possible reciprocal products is preserved in any individual recombination event. However, our estimate of damage during transfection indicates that only about one-third to one-eighth of the dimeric molecules in the active subpopulation would be intact, and thus the likelihood of producing complementing genomes would be reduced correspondingly. Although the detected level of complementation plaques is still below the expected level for reciprocal recombination, the very existence of fragmented input molecules makes the reciprocity of homologous recombination in somatic cells a difficult aspect of the mechanism to address.

In summary, these measurements of cell-mediated damage to transfected DNA help to define an important parameter for studies of somatic recombination in monkey cells. On the one hand, the modest extent of damage renders unlikely the possibility that any common mechanism of somatic recombination depends on extensive endonucleolytic or exonucleolytic processes. On the other hand, the existence of damaged molecules in the active subpopulation means that the population of recombining molecules cannot be viewed simply as unadulterated input molecules. This fact calls for caution in deriving certain mechanistic details of recombination in somatic cells from experiments with transfected DNA.

APPENDIX

Monomers, partial dimers, and dimers of SV40 genomes are expected to survive random breakage differently because they contain different numbers of overlapping genomes (Fig. 6). For the analysis below we set as the condition for survival that a segment of DNA containing one complete genome remain intact. To illustrate the expected difference in survival, we compare monomers with one break with dimers with two breaks. This comparison is appropriate because a dimer, being twice as long as a monomer, should suffer breakage twice as often. No monomer with one break would meet the set condition for survival; however, many of the dimers with two breaks would still contain an intact genome length of SV40 and therefore would survive. Put another way, only appropriately placed pairs of breaks can interrupt all the overlapping genomes in a dimer, whereas any break in a monomer will suffice. Below we quantitate this expectation.

The derivation of the equation that describes the survival curves

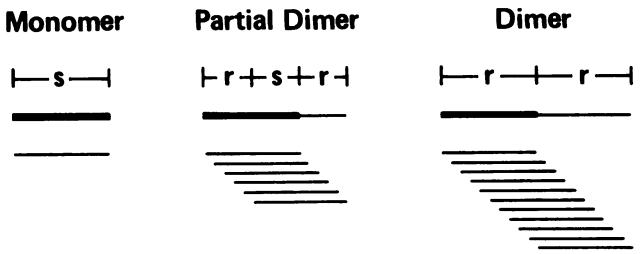

FIG. 6. Target structure of monomers, partial dimers, and dimers of SV40 genomes. For each structure, the thick line represents one genome length and the attached thin line represents the portion of the genome that is tandemly repeated. Each structure contains a set of overlapping genomes as illustrated by the thin lines below the structures. The portion of each structure that is unique and therefore sensitive to a single break is indicated by s; the portion that is repeated and therefore resistant to a single break is represented by r.

for monomers, partial dimers, and dimers is facilitated by considering first a monomer composed entirely of unique DNA (s), which is sensitive to a single break; second, a dimer composed entirely of repeated DNA (r), which is sensitive only to appropriately placed breaks; and third, a partial dimer composed of both unique and repeated segments of DNA (Fig. 6). In each case the approach we used was to derive the equation for nonsurvival by counting directly all the different breaks that would result in nonsurvival for a length of DNA consisting of equivalent sites for breakage. We first developed expressions for the probability of such effective breaks and then took the limit as the number of sites becomes large. This approach leads to a general expression for the probability of nonsurvival of monomers, partial dimers, and dimers that are subjected to random breakage. The corresponding general expression for the probability of survival is simply 1 minus the probability of nonsurvival. Our final equation for probability of survival has been confirmed by an independent derivation and checked by computer simulation.

The expression for a break in a single-hit, multitarget model, such as the monomer, is well known (9), but the derivation serves to introduce some useful notation and to demonstrate the outlines of the proof. Let us assume that there are H hits in the N sites that make up the monomer so that the probability of any given site being hit, p, is H/N, and the probability that a given site is not hit, q, is then $(1 - p)$. It is useful to treat the sites as if they were numbered from 1 to N and to consider the probability that a break does not occur until some number of sites have been passed. The probability that a break occurs at the first site is, of course, just p; the probability that a break occurs at the second site but not at the first is pq; at the third site but not at the first two is pq^2; and so on to the last site, where the probability is $pq^{(N-1)}$. Thus, the total probability of a break occurring in the molecule is:

$$f(p) = p(1 + q + q^2 + \cdots + q^{N-1}) \quad (1)$$

The summation in equation 1 can be done in closed form (equation 0.112 in reference 13) to yield:

$$f(p) = p(1 - q^N)/(1 - q) \quad (2)$$

or, substituting for p and q:

$$f(p) = 1 - (1 - H/N)^N \quad (3)$$

The limit as N becomes large gives the required final expression (equation 4.2.21 in reference 1) for breakage of the monomer:

$$f(p) = 1 - e^{-H} \quad (4)$$

The probability that breakage will interrupt all the overlapping genomes in a dimer can be developed along the same lines. However, in this case two appropriately placed breaks are required; one break must occur in each repeated segment in such a way that there are fewer than N sites (a monomer) between them. For convenience we designate the two repeated segments of the dimer as

A and B. To count all possible effective breaks, we begin with a break in segment A at position N, which is immediately adjacent to position 1 of segment B. An effective break can occur in segment B at position 1 with probability p, at position 2 with probability pq, at position 3 with probability pq^2, and so on to position N with probability $pq^{(N-1)}$. Thus, given a break in segment A at position N (which occurs with probability p), the probability of an effective break in segment B is:

$$g(p:N) = p^2(1 + q + q^2 + \cdots + q^{N-1}) \quad (5)$$

If we begin with a break in segment A at position $N-1$ (which occurs with probability pq), the same counting procedure gives the probability of an effective break in segment B as:

$$g(p:N-1) = p^2q(1 + q + q^2 + \cdots + q^{N-2}) \quad (6)$$

Repeating this process N times leads to the probability term for a starting break at position 1 of segment A:

$$g(p:1) = p^2q^{N-1} \quad (7)$$

The total probability of an effective pair of breaks is found by summing over the probabilities for each site in A to give:

$$g(p) = g(p:1) + g(p:2) + \cdots + g(p:N) \quad (8)$$

or:

$$g(p) = p^2(1 + 2q + 3q^2 + \cdots + Nq^{N-1}) \quad (9)$$

The final expression can be placed in a closed form (equations 0.112 and 0.113 in reference 13) yielding:

$$g(p) = p^2\{1 - q^N[N(1 - q) + 1]\}/(1 - q)^2 \quad (10)$$

Substituting for p and q:

$$g(p) = 1 - (1 - H/N)^N(1 + H) \quad (11)$$

In the limit as N becomes large, the probability of an effective pair of breaks occurring in a dimer is:

$$g(p) = 1 - e^{-H}(1 + H) \quad (12)$$

The equations for effective breakage of monomers and dimers can be combined to describe the effective breakage of partial dimers, which contain both unique and repeated segments. Let s, expressed as a fraction of 1 genome length, represent the unique portion of a partial dimer; let r, also expressed as a fraction of genome length, represent the repeated portions of a partial dimer. With these definitions, $r + s = 1$ genome length, and the length of a partial dimer is $1 + r$ (see Fig. 6). The probability of an effective break due to a single hit in the unique region is given by the fraction of hits in the unique region or:

$$f(p) = 1 - e^{-sH} \quad (13)$$

Similarly, the probability of an effective pair of breaks in the repeated region is given by:

$$g(p) = 1 - e^{-rH}(1 + rH) \quad (14)$$

The combined probability of effective breakage is given by the probability of effective breaks of either type less the probability of effective breaks of both types, or:

$$F(p) = f(p) + g(p) - f(p)g(p) \quad (15)$$

Substituting for $f(p)$ and $g(p)$ gives:

$$F(p) = 1 - e^{-H}(1 + rH) \quad (16)$$

Equation 16 is a general expression for the probability of nonsurvival of monomers, partial dimers, and dimers; when appropriate values of r are substituted it reduces to equation 4 for monomers ($r = 0$) and equation 12 for dimers ($r = 1$). Because the probability of survival is simply 1 minus the probability of nonsurvival, the general equation for survival of monomers, partial dimers, and dimers is:

$$S(p) = e^{-H}(1 + rH) \quad (17)$$

ACKNOWLEDGMENTS

We thank Kathleen Marburger for expert technical assistance and Kirk Aune, Peter Berget, David Roth, and Michael Seidman for valuable discussions. This work was supported by Public Health Service grants CA 15743 (to J.H.W.) and CA 11430 (to A.W.) from the National Institutes of Health.

LITERATURE CITED

1. **Abramowitz, M., and I. A. Stegun. (ed.).** 1970. Handbook of mathematical functions, National Bureau of Standards, Applied Mathematics Series 55, Ninth Printing. Government Printing Office, Washington, D.C.
2. **Anderson, R. A., T. Krakauer, and R. D. Camerini-Otero.** 1982. DNA-mediated gene transfer: recombination between cotransferred DNA sequences and recovery of recombinants in a plasmid. Proc. Natl. Acad. Sci. U.S.A. **79**:2748–2752.
3. **Bishop, J. M.** 1983. Cancer genes come of age. Cell **32**:1018–1020.
4. **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.
5. **Calos, M. P., J. S. Lebkowski, and M. R. Botchan.** 1983. High mutation frequency in DNA transfected into mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **80**:3015–3019.
6. **Chandler, V. L., B. A. Maler, and K. R. Yamamoto.** 1983. DNA sequences bound specifically by glucocorticoid receptor *in vitro* render a heterologous promoter hormone sensitive *in vivo*. Cell **33**:489–499.
7. **Crowley, C. N., C.-C. Liu, and A. D. Levinson.** 1983. Plasmid-directed synthesis of hepatitis B surface antigen in monkey cells. Mol. Cell. Biol. **3**:44–55.
8. **Dierks, P., A. von Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissmann.** 1983. Three regions upstream from the Cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**:695–706.
9. **Elkind, M. M., and G. F. Whitmore.** 1967. The radiobiology of cultured mammalian cells. Gordon and Breach, New York.
10. **Folger, K. R., E. A. Wong, G. Wahl, and M. R. Capecchi.** 1982. Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. Mol. Cell. Biol. **2**:1372–1387.
11. **Gething, M. J., and Sambrook, J.** 1982. Construction of influenza haemagglutinin genes that code for intracellular and secreted forms of the protein. Nature (London) **300**:598–603.
12. **Goodenow, R. S., M. McMillan, M. Nicolson, B. T. Sher, K. Eakle, N. Davidson, and L. Hood.** 1982. Identification of the class I genes of the mouse major histocompatibility complex by DNA-mediated gene transfer. Nature (London) **300**:231–237.
13. **Gradshteyn, I., and I. M. Ryzhik.** 1965. Tables of integrals, series and products, 4th ed. Academic Press, Inc., New York.
14. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
15. **Gutai, M. W.** 1981. Recombination in SV40 infected cells: viral DNA sequences at sites of circularization of transfecting linear DNA. Virology **109**:353–365.
16. **Gutai, M. W., and D. Nathans.** 1978. Evolutionary variants of simian virus 40: cellular DNA sequences and sequences at recombinant joints of substituted variants. J. Mol. Biol. **126**:275–288.
17. **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.
18. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cultures. J. Mol. Biol. **26**:365–369.
19. **Howard, B., M. Estes, and J. Pagano.** 1971. The uptake of SV40 DNA by nonpermissive cells in the presence of DEAE-dextran. Biochim. Biophys. Acta **228**:105–116.
20. **Howell, S. H., L. L. Walker, and R. M. Walden.** 1981. Rescue of *in vitro* generated mutants of cloned cauliflower mosaic virus

genome in infected plants. Nature (London) **293**:483–486.
21. **Israel, M., J. Byrne, and M. Martin.** 1978. Biologic activity of oligomeric forms of SV40 DNA. Virology **87**:239–246.
22. **Jaenisch, R., and A. Levine.** 1971. Infection of primary African green monkey cells with SV40 monomeric and dimeric DNA. J. Mol. Biol. **61**:735–738.
23. **Khoury, G., and P. Gruss.** 1983. Enhancer elements. Cell **33**:313–314.
24. **Kratzer, P. G., V. M. Chapman, H. Lambert, R. E. Evans, and R. M. Liskay.** 1983. Differences in the DNA of the inactive X chromosomes of fetal and extraembryonic tissues of mice. Cell **33**:37–42.
25. **Lai, C., and D. Nathans.** 1975. A map of temperature-sensitive mutants of simian virus 40. Virology **66**:70–81.
26. **Lebeurier, G., L. Hirth, B. Hohn, and T. Hohn.** 1982. *In vivo* recombination of cauliflower mosaic virus DNA. Proc. Natl. Acad. Sci. U.S.A. **79**:2932–2936.
27. **Loyter, A., G. A. Scangos, and F. H. Ruddle.** 1982. Mechanisms of DNA uptake by mammalian cells: fate of exogenously added DNA monitored by the use of fluorescent dyes. Proc. Natl. Acad. Sci. U.S.A. **79**:422–426.
28. **McCutchen, J., and J. Pagano.** 1968. Enhancement of the infectivity of SV40 deoxyribonucleic acid with diethyl-amino-ethyl-dextran. J. Natl. Cancer Inst. **41**:351–357.
29. **Miller, C. K., and H. M. Temin.** 1983. High efficiency ligation and recombination of DNA fragments by vertebrate cells. Science **220**:606–609.
30. **Mulligan, R. C., and P. Berg.** 1981. Factors governing the expression of a bacterial gene in mammalian cells. Mol. Cell. Biol. **1**:449–459.
31. **Neff, N. F., J. H. Thomas, P. Grisafi, and D. Botstein.** 1983. Isolation of the β-tubulin gene from yeast and demonstration of its essential function *in vivo*. Cell **33**:211–219.
32. **Perucho, M., D. Hanahan, and M. Wigler.** 1980. Genetic and physical linkage of exogenous sequences in transformed cells. Cell **22**:309–317.
33. **Pipas, J. M., K. W. C. Peden, and D. Nathans.** 1983. Mutational analysis of simian virus 40 T antigen: isolation and characterization of mutants with deletions in the T-antigen gene. Mol. Cell. Biol. **3**:203–213.
34. **Pulciani, S., E. Santos, A. V. Lauver, L. K. Long, S. A. Aaronson, and M. Barbacid.** 1982. Oncogenes in solid human tumors. Nature (London) **300**:539–542.
35. **Razzaque, A., H. Mizusawa, and M. M. Seidman.** 1983. Rearrangement and mutagenesis of a shuttle vector plasmid after passage in mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **80**:3010–3014.
36. **Reddy, E. P., R. K. Reynolds, E. Santos, and M. Barbacid.** 1982. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. Nature (London) **300**:149–152.
37. **Robert de Saint Vincent, B., and G. Wahl.** 1983. Homologous recombination in mammalian cells mediates formation of a functional gene from two overlapping gene fragments. Proc. Natl. Acad. Sci. U.S.A. **80**:2002–2006.
38. **Robins, D., S. Ripley, A. Henderson, and R. Axel.** 1981. Transforming DNA integrates into the host chromosome. Cell **23**:29–39.
39. **Ruley, H. E., and M. Fried.** 1983. Clustered illegitimate recombination events in mammalian cells involving very short sequence homologies. Nature (London) **304**:181–184.
40. **Santangelo, G. M., and C. N. Cole.** 1983. Preparation of a functional library of African green monkey DNA fragments which substitute for the processing/polyadenylation signal in the herpes simplex virus type I thymidine kinase gene. Mol. Cell. Biol. **3**:643–653.
41. **Schaffner, W.** 1980. Direct transfer of cloned genes from bacteria to mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **77**:2163–2167.
42. **Small, J., and G. Scangos.** 1983. Recombination during gene transfer into mouse cells can restore the function of deleted genes. Science **219**:174–176.
43. **Stringer, J. R.** 1981. Integrated simian virus 40 DNA: nucleotide sequences at cell-virus recombinant junctions. J. Virol. **38**:671–679.
44. **Stringer, J. R.** 1982. DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. Nature (London) **296**:363–366.
45. **Subramani, S., and P. Berg.** 1983. Homologous and nonhomologous recombination in monkey cells. Mol. Cell. Biol. **3**:1040–1052.
46. **Subramanian, K. S.** 1979. Segments of SV40 DNA spanning most of the leader sequence of the major later viral messenger RNA are dispensible. Proc. Natl. Acad. Sci. U.S.A. **76**:2556–2560.
47. **Tabin, C. J., S. M. Bradley, C. I. Bargmann, R. A. Weinberg, A. G. Papageorge, E. M. Scolnick, R. Dhar, D. R. Lowy, and E. H. Chang.** 1982. Mechanism of activation of a human oncogene. Nature (London) **300**:143–194.
48. **Tegtmeyer, P.** 1972. Simian virus 40 deoxyribonucleic acid synthesis: the viral replicon. J. Virol. **10**:591–598.
49. **Tooze, J.** 1980. DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
50. **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.
51. **Upcroft, P., B. Carter, and C. Kidson.** 1980. Mammalian cell functions mediating recombination of genetic elements. Nucleic Acids Res. **8**:5835–5844.
52. **Volkert, F. C., and C. S. H. Young.** 1983. The genetic analysis of recombination using adenovirus overlapping terminal DNA fragments. Virology **125**:175–193.
53. **Wake, C., and J. 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.
54. **Wake, C., and J. Wilson.** 1980. Defined oligomeric SV40 DNA: a sensitive probe of general recombination in somatic cells. Cell **21**:141–148.
55. **Weintraub, H.** 1983. A dominant role for DNA secondary structure in forming hypersensitive structures in chromatin. Cell **32**:1191–1203.
56. **Wieringa, B., F. Meyer, J. Reiser, and C. Weissmann.** 1983. Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit β-globin gene. Nature (London) **301**:38–43.
57. **Wigler, M. H.** 1981. The inheritance of methylation patterns in vertebrates. Cell **24**:285–286.
58. **Wilson, J.** 1977. Genetic analysis of host range mutant viruses suggests an uncoating defect in simian virus 40-resistant monkey cells. Proc. Natl. Acad. Sci. U.S.A. **74**:3503–3507.
59. **Wilson, J.** 1978. Interference in SV40 DNA infections: a possible basis for cellular competence. Virology **91**:380–388.
60. **Wilson, J., P. Berget, and J. Pipas.** 1982. Somatic cells efficiently join unrelated DNA segments end-to-end. Mol. Cell. Biol. **2**:1258–1269.
61. **Wilson, J., 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 40 infection. J. Virol. **20**:391–399.
62. **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.