MOLECULAR AND CELLULAR BIOLOGY, Aug. 1985, p. 2080–2089
0270-7306/85/082080-10$02.00/0

Vol. 5, No. 8

Topological Requirements for Homologous Recombination Among DNA Molecules Transfected into Mammalian Cells

CLAIRE T. WAKE,† FRANCESCA VERNALEONE, AND JOHN H. WILSON*

Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030

Received 4 March 1985/Accepted 23 May 1985

Cultured animal cells rearrange foreign DNA very efficiently by homologous recombination. The individual steps that constitute the mechanism(s) of homologous recombination in transfected DNA are as yet undefined. In this study, we examined the topological requirements by using the genome of simian virus 40 (SV40) as a probe. By assaying homologous recombination between defective SV40 genomes after transfection into CV1 monkey cells, we showed that (i) linear molecules are preferred substrates for homologous exchanges, (ii) exchanges are distributed around the SV40 genome, and (iii) the frequency of exchange is not diminished significantly by the presence of short stretches of non-SV40 DNA at the ends. These observations are considered in relation to current models of homologous recombination in mammalian cells, and a new model is proposed. The function of somatic cell recombination is discussed.

Homology-dependent recombination events are commonly observed between DNA molecules introduced into animal cells in culture. These events have been detected in primate (8, 13, 24a, 25, 27, 30, 33–37), rodent (3a, 4a, 5, 8, 11, 14, 24, 28, 28), and avian cells (3, 17) by assaying the cell-mediated reconstruction of selectable genes or viral genomes from defective input molecules. Both methods apparently detect the same events, since transfections of viral genomes and selectable genes have yielded similar results. Furthermore, the general characteristics of the recombination events appear to be the same regardless of the way the foreign DNA is introduced, be it by microinjection (4a, 5), protoplast fusion (24, 27), or transfection mediated by DEAE-dextran (24a, 25, 30, 33, 34, 36, 37) or calcium phosphate (3, 3a, 8, 11, 13, 14, 17, 28, 29, 35). Taken together, these observations suggest that the enzymes responsible for homologous recombination are normal components of animal cells.

The mechanism(s) by which cellular enzymes accomplish homologous exchange is not yet clear. However, homologous recombination in animal cells is stimulated when defined double-strand breaks are introduced into the transfected DNA in vitro (3a, 4a, 8, 11). Similar double-strand breaks are introduced naturally into transfected DNA in vivo. In genetic studies designed to assess physical (38) and mutagenic (2, 5, 10, 22, 23, 26) damage to transfected DNA molecules that reach the nucleus, apparent double-strand breaks and their presumed consequence, deletion mutations, were detected at high frequencies. The cellular location at which transfected DNA is broken is not yet defined; however, several observations indicate that breaks can occur in the nucleus (2, 7, 10; X. B. Chang and J. H. Wilson, manuscript in preparation). The possibility that double-strand breaks are a required step in the mechanism led us to examine the topological requirements for homologous recombination in transfected DNA. Here, we demonstrate that linear molecules are preferred substrates for homologous recombination between transfected simian virus 40 (SV40) genomes.

* Corresponding author.
† Present address: Damon Biotech, Needham Heights, MA 02194.

MATERIALS AND METHODS

Cells and viruses. Procedures for growth of the established monkey kidney CV1 cell line have been described previously (39, 42). The temperature-sensitive mutants of SV40 were obtained from P. Tegtmeyer and R. Martin. The map positions of the *ts* mutations and the restriction sites used in this study are shown in Fig. 1.

DNA preparation. SV40 DNA and plasmid DNA were prepared as described previously (6, 38). All SV40 and plasmid DNA preparations were labeled in vivo with [^{3}H]thymidine (New England Nuclear Corp.). Specific activities were determined and used to adjust DNA concentrations as appropriate for each experiment.

DNA transfections. DNA transfections were performed essentially by the method of McCutchen and Pagano (16) with 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 as described previously (40). After transfection, the monolayers were overlaid with agar and incubated at the appropriate temperature as in a standard plaque assay.

Recombination assays. Wild-type recombinants were detected by transfecting mixtures of *ts* mutants into CV1 cells and assaying directly at the nonpermissive temperature (41°C). In previous studies, we have demonstrated that this procedure is equivalent to the more standard method of scoring wild-type recombinants among the progeny virus generated at the permissive temperature (46, 47). The agreement between these two methods supports the argument that recombinants are generated before replication begins; the results in this paper offer further support.

Molecular cloning. Restriction enzymes, T4 DNA ligase, and DNase I were purchased from Bethesda Research Laboratories, New England Biolabs, or Boehringer Mannheim Biochemicals and used as recommended by the supplier. Nicked circular DNAs were prepared by digestion of closed circular DNA with DNase I in the presence of ethidium bromide (20). The ethidium was then removed by extraction with isoamyl alcohol. All molecular clonings were performed by standard procedures (12). SV40 mutant tsA7 was cloned in both orientations through its *Bam*HI site into

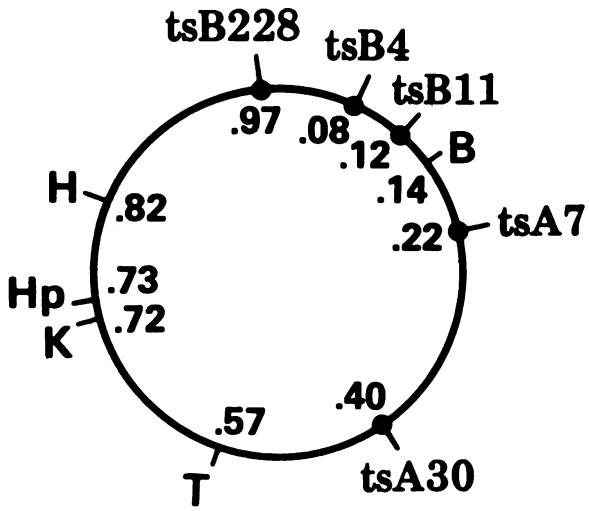

FIG. 1. Locations of mutations and restriction sites. The locations of restriction sites *Bam*HI (B), *Taq*I (T), *Kpn*I (K), and *Hpa*II (H) are given as their fractional genome length as measured from the *Eco*RI site at 0/1 (32). The location of each mutation is given as the midpoint of the restriction fragment to which it has been mapped (9); for tsA7, the midpoint of the segment defined by the *Pst*I site at nucleotide position 3204 and the C terminus of the T antigen was used (J. H. Wilson, unpublished data).

the *Bam*HI site of pBR322 to create the chimeric plasmids pFV17 and pFV22. SV40 mutant tsA30 was cloned in one orientation through its *Taq*I site into the *Cla*I site of pBR322 to create the chimeric plasmid pFV4.

RESULTS

Transfection with linear genomes. Recombination between transfected linear SV40 genomes could occur before or after circularization, since transfected linear molecules are circularized efficiently in monkey cells (38). To identify the topological form of DNA active in recombination, we compared the number of wild-type recombinants produced in crosses between aligned linear genomes (Fig. 2) and offset linear genomes (Fig. 3). In both types of cross, the parental genomes were marked with temperature-sensitive mutations so that wild-type recombinants could be selectively assayed at the nonpermissive temperature. If circularization preceded recombination, the position at which the genomes were linearized should not affect production of wild-type recombinants because any initial differences would be eliminated by circularization. On the other hand, if recombination preceded circularization, the production of wild-type recombinants should depend in a predictable fashion on the way the input genomes were linearized.

The predicted outcomes for recombination between aligned linear genomes, which were cleaved with the same restriction enzyme, are illustrated schematically in Fig. 2. Because the SV40 genome is circular, the size of the target (*T*) in which a crossover could produce a wild-type recombinant depends on whether the genomes are linearized outside or between the markers. If recombination occurred between linear molecules, the difference in target size should be reflected in the number of wild-type recombinants. The results from these transfections are shown in Table 1. Although the numbers of recombinants were small, they correlated with the size of the target on the linear genomes, suggesting that recombination occurred preferentially between linear molecules.

To extend this observation, we compared the number of wild-type recombinants produced by recombination between offset genomes, which had been linearized with different restriction enzymes. If offset genomes recombined as linear genomes, there should be substantially more wild-type recombinants produced in one cross (designated a type I cross) than in the other (designated a type II cross) (Fig. 3). The results from these transfections are shown in Table 2. Crosses 1 and 2 involved aligned linear genomes like those shown in Table 1 and yielded similarly low numbers of wild-type recombinants. Crosses 3 and 4 involved offset linear genomes and, as expected, when recombination occurred between linear genomes, the type I cross (cross 3) produced many more wild-type recombinants than did the type II cross (cross 4). These results, like those in Table 1, indicate that input linear genomes recombine as linear genomes, not as circular ones.

Transfections with circular and linear genomes. Another way to assess the topological requirements for recombination is to compare the numbers of wild-type recombinants

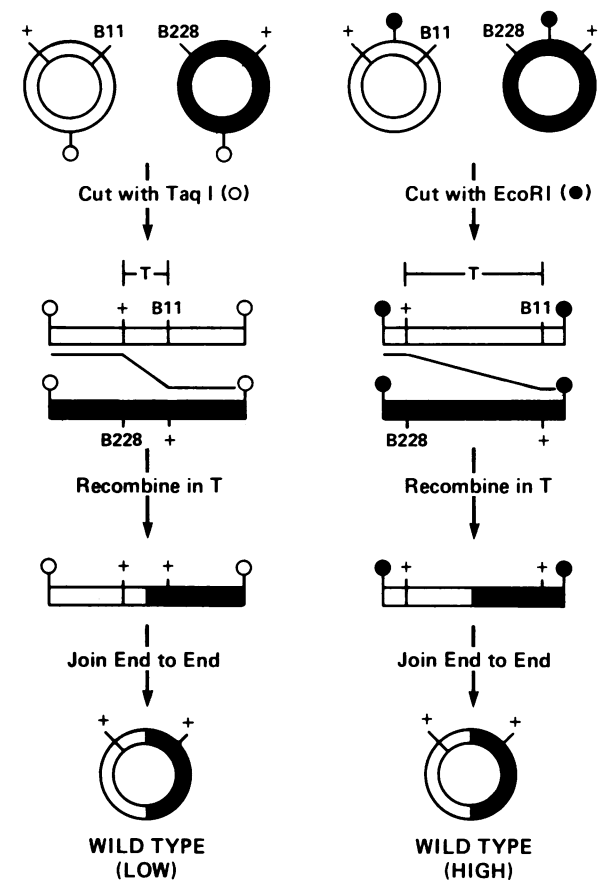

FIG. 2. Recombination between aligned linear genomes. This figure illustrates the predicted outcomes of homologous recombination between tsB228 and tsB11 linear genomes. Cleavage of both genomes outside the markers, for example, at the *Taq*I site, leaves a small target (*T*) in which an exchange could link the wild-type markers, whereas cleavage between the markers at the *Eco*RI site leaves a large target. Since recombination events are distributed throughout the genome (28, 37) (Fig. 4), the cross with the larger target should produce the greater number of wild-type recombinants. For clarity, only the wild-type products of recombination are shown; some experiments indicate that homologous recombination in transfected DNA is nonconservative, as described in the text.

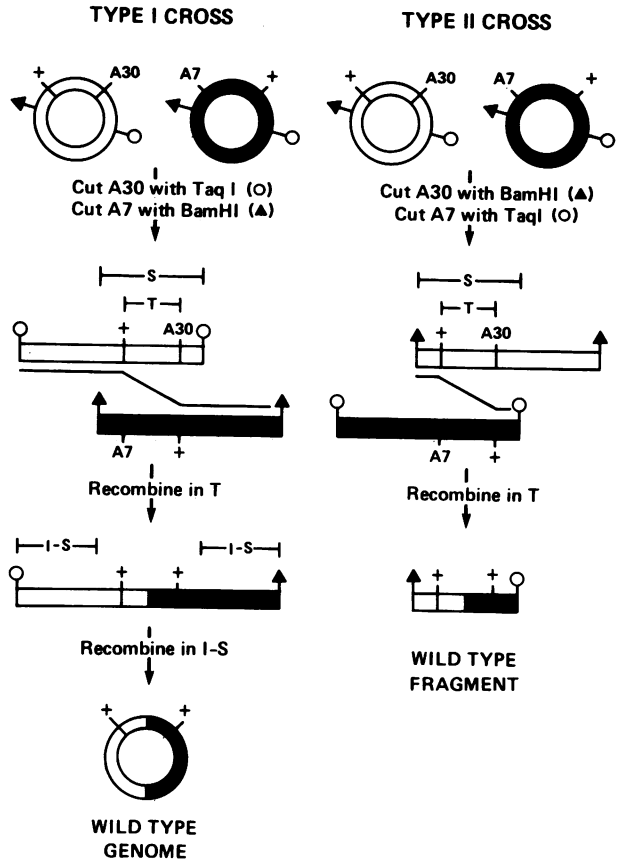

FIG. 3. Recombination between offset linear genomes. This figure illustrates the predicted outcomes of homologous recombination between tsA7 and tsA30 linear genomes. Cleavage of tsA7 with *Bam*HI and of tsA30 with *Taq*I, as shown on the left, generates a type I offset cross. An intermolecular recombination within *T* to link the wild-type markers would produce a partially dimeric genome containing terminal repeats of 1 − *S*. A subsequent intramolecular recombination in 1 − *S* would generate a circular wild-type recombinant. The efficiency of the intramolecular recombination step is expected to be independent of the length of 1 − *S*, since partially dimeric molecules with repeats from 500 to 5,000 base pairs, the range of repeat lengths generated in these experiments, have been shown to circularize by homologous recombination with approximately the same efficiency (see Fig. 5 in reference 38). The order of the recombination events in *S* and 1 − *S* should make no difference to the outcome. If the intermolecular recombination were in 1 − *S*, a partial dimer with terminal repeats of *S* would be generated; a subsequent intramolecular recombination within *T* would generate the wild-type recombinant. For clarity, only the wild-type products of recombination are shown. Cleavage of tsA7 with *Taq*I and of tsA30 with *Bam*HI, as shown on the right, generates a type II offset cross. An intermolecular recombination within *T* to link the wild-type markers, as illustrated, would produce a genomic fragment of length *S*. Such a fragment could not produce a wild-type genome by any intramolecular process. Once again, the order of events in *S* and 1 − *S* should make no difference to the outcome. If the intermolecular recombination event occurred in 1 − *S*, a partial dimer with terminal repeats of *S* would be generated; a subsequent intramolecular recombination within *T* would have the potential to generate a circular double mutant and the same wild-type fragment shown. For clarity, only the wild-type products are shown.

TABLE 1. Crosses between aligned linear genomes

Cross	Cleavage enzyme[b]	No. of wild-type recombinants[a]			
		B228 × B11 with a *T* value of:		B228 × B4 with a *T* value of:	
		0.15	0.85	0.11	0.89
1	*Bam*HI	8		0	
2	*Hae*II	8		2	
3	*Hpa*II	10		0	
4	*Taq*I	8		3	
5	*Eco*RI		27		8

[a] Wild-type recombinants were selected by plaque assay of the transfection mixtures under nonpermissive conditions. The plaque counts were summed from several experiments. For mixtures of tsB228 and tsB11, 173-ng portions of each genome were mixed and plaque assayed on a total of 17 plates in four experiments. For mixtures of tsB228 and tsB4, 30-ng portions of each genome were mixed and plaque assayed on a total of four plates in two experiments. *T* is expressed in fractional genome units.

[b] All linear genomes were generated by cleavage of circular genomes with the indicated restriction enzymes.

produced in crosses between the various topological forms. This analysis is complicated by two factors: input linear molecules can be circularized by ligation of ends, and input circular molecules can be linearized by breakage during transfection (38). Despite these distorting factors, the greatest number of wild-type recombinants would still be expected in transfections involving the preferred topological forms. The largest number of wild-type recombinants was produced in a cross involving two linear substrates (Table 2, cross 3). That cross yielded about eightfold more recombinants than did crosses between circular substrates (crosses 9 and 10) and about twofold more recombinants than did the best crosses involving a mixture of linear and circular substrates (crosses 5 and 7). These results suggest that

TABLE 2. Crosses with linear and circular genomes

Cross	DNA form (cleavage enzyme)[a]		Cross type[b]	No. of wild-type recombinants[c]
	tsA7	tsA30		
1	L (*Bam*HI)	L (*Bam*HI)		4
2	L (*Taq*I)	L (*Taq*I)		4
3	L (*Bam*HI)	L (*Taq*I)	I	159
4	L (*Taq*I)	L (*Bam*HI)	II	4
5	L (*Bam*HI)	C	(I)	64
6	C	L (*Bam*HI)	(II)	4
7	C	L (*Taq*I)	(I)	82
8	L (*Taq*I)	C	(II)	3
9	C	C		14
10	C	C		25

[a] Equimolar mixtures of genomes from the mutants tsA7 and tsA30 were transfected into CV1 cells (15 ng of each genome per plate) and plaque assayed at the nonpermissive temperature to select for wild-type recombinants. Three DNA forms were used in these experiments. Linear genomes (L) were linearized at the *Bam*HI site or the *Taq*I site, as indicated. Open circular genomes (C) were prepared by DNase nicking of closed circular genomes in the presence of ethidium bromide. Greater than 95% of the genomes were present in the indicated form. Nicked circular genomes were used in preference to supercoiled genomes because their specific infectivity (PFU per nanogram) is about the same as that for linear genomes (38). Less extensive but qualitatively similar results have been obtained with supercoiled genomes (C. T. Wake, Ph.D. thesis, Baylor College of Medicine, Houston, Tex., 1980).

[b] Crosses are designated type I or type II as defined in the legend to Fig. 3. The designations in parentheses were assigned as described in the text.

[c] For each mixture, the wild-type recombinants, which formed plaques under nonpermissive conditions, were summed from 30 plates assayed in three experiments.

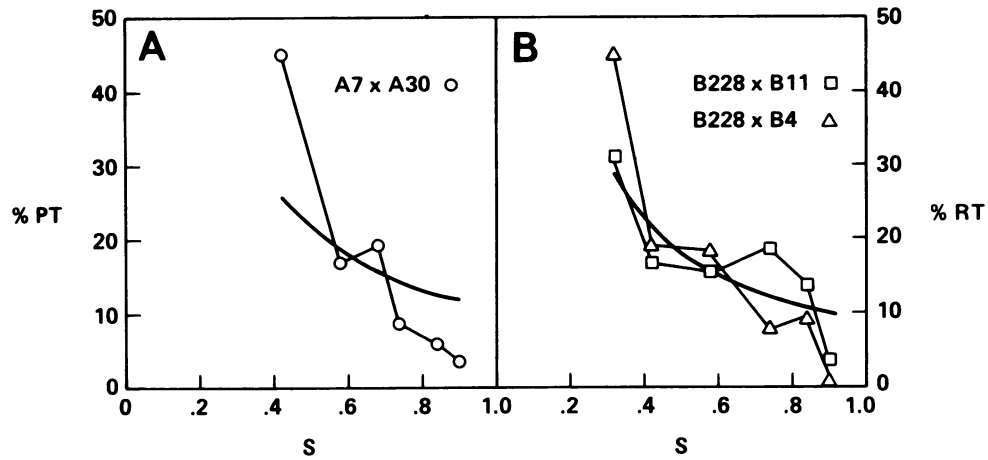

FIG. 4. Distribution of crossovers in S. The number of wild-type recombinants, expressed as a percentage of the total (%PT), is plotted as a function of the length of S. The expected number of wild-type recombinants, expressed as a percentage of the total recombinants (%RT) based on the relationship $R = k(T/S)$ as developed in the text, is also plotted as a function of the length of S. These data are the normalized values shown in Table 3. (A) Crosses with tsA7 and tsA30. (B) Crosses with tsB228 and tsB11 and with tsB228 and tsB4.

homologous recombination in transfected DNA preferentially involves two linear substrates.

That conclusion is supported by additional correlations within the data. For example, if circular molecules were the preferred substrates for recombination, then crosses 5 and 6 should produce equal numbers of recombinants, since they both would require circularization of a *Bam*HI-cut linear molecule. However, crosses 5 and 6 yielded quite different numbers of recombinants. The same reasoning suggests that crosses 7 and 8 should be equal, but they were not. These unfulfilled expectations constitute an independent argument that circular molecules are not the preferred substrates for homologous recombination.

A similar argument suggests that a linear molecule and a circular molecule are not the preferred substrates for homologous recombination. If they were the preferred topological forms, the high level of recombinants produced in cross 3 would support the argument that some of the input linear molecules circularized, thereby generating recombination events as shown in crosses 5 and 7. Even at high efficiency, such a conversion of cross 3 into cross 5 or 7 would, at best, be expected to yield a number of recombinants intermediate between crosses 5 and 7. The observed value is twofold higher than expected, implying that linear × linear crosses are converted into linear × circular crosses at greater than 100% efficiency. A high-efficiency conversion of linear × linear crosses into linear × circular crosses leads to additional expectations. If conversion were 100% efficient, then cross 1 should be intermediate between crosses 5 and 6, and cross 2 should be intermediate between crosses 7 and 8. These expectations were not met.

In contrast, similar reasoning supports the notion that linear molecules are the preferred substrates for recombination. If linear molecules were the preferred topological form, recombination in crosses involving circles would be accounted for most naturally by broken input molecules. Since transfected DNA that reaches the nuclei of monkey cells is broken roughly once per 10 kilobases (38), approximately half of the input circular genomes should be present in the nucleus as linear molecules. (A similar fraction of input linear molecules would be reduced to fragments, so that in linear × circular crosses, the concentration of input linear and broken circular molecules would be roughly equivalent.) Random breakage of input circles would generate a permuted set of linear molecules, which would recombine with the input linear molecules in crosses 5 and 7 primarily in a type I arrangement, such as in cross 3; they would recombine with the input linears in crosses 6 and 8 primarily in a type II arrangement, such as in cross 4. These expectations match the asymmetrical results obtained in those crosses. In addition, the results are quantitatively as expected. If the circular molecules in crosses 5 and 7 were randomly permuted, the number of recombinants should be near the mean of the distribution shown in Fig. 4A, as they were. These quantitative aspects extend to circular × circular crosses as well, which would be expected to yield a number of recombinants equal to about one-fourth of the mean of the same distribution (Fig. 4A), which they did.

Taken together, these arguments indicate strongly that linear molecules are the preferred topological form for homologous recombination in monkey cells.

Recombination as a function of replicative potential. Genomes in the above offset crosses carried temperature-sensitive mutations in the gene for T antigen and, therefore, were defective for replication under the restrictive conditions used in those experiments. The inability to replicate limits the concentration of circular genomes to the fraction of input genomes that circularize. Therefore, the lack of recombinants could have resulted from a low concentration of circles and not from any inherent inability to recombine. One way to assess this possibility is to compare recombination in the absence of functional T antigen with recombination in its presence, under which conditions the concentration of circles increases by 10^5 in 48 h. If circular genomes could recombine, then the disparity between type I and type II crosses might be expected to disappear as replication increased the concentration of circular genomes.

To measure disparity in the two types of cross as a function of replicative potential, we used genomes with temperature-sensitive lesions either in the A gene, which encodes T antigen, or in the B gene, which encodes a capsid protein. At the nonpermissive temperature, tsA7 and tsA30 were impaired for replication, whereas tsB228, tsB11, and tsB4 could replicate normally. The results of offset crosses

with pairs of these genomes are shown in Table 3. A consistent disparity between type I and type II crosses was detected regardless of whether replication was blocked, reinforcing the above-stated conclusion that linear molecules are the preferred substrates for recombination and suggesting, as well, that replicated circular molecules are poor recombination substrates.

Distribution of crossovers. For the three sets of type I offset crosses in Table 3, the number of wild-type recombinants tended to decrease with increasing length of the offset segment (S), reflecting in some way the distribution of crossovers within S. In previous experiments in which we examined intramolecular homologous recombination in partially dimeric genomes of SV40, we determined that crossovers were rather evenly distributed throughout the repeated regions (37). Are crossovers in intermolecular homologous recombination also evenly distributed? The generation of a circular genome by recombination between offset linear molecules requires a minimum of two crossovers: one in S and one in $1 - S$ (Fig. 3). The crossover in S must occur within T, the target defined by the mutations, to generate a wild-type recombinant. If crossovers were evenly distributed within S, the number of wild-type recombinants should be proportional to the fraction of S that is occupied by T, which is T/S.

The experimental results were compared with expectations for evenly distributed crossovers by normalizing the number of wild-type recombinants for each type I cross to the total number of recombinants observed in all type I crosses with each pair of mutations (Table 3). In a similar way, a recombination value, R, was calculated for each cross ($R = kT/S$) and then normalized to the total R summed for all type I crosses for each set of mutations (Table 3). This normalization allows the experimental data to be compared directly with the theoretical expectations for randomly distributed crossovers. This comparison is illustrated graphically in Fig. 4. The data seem consistent with randomly distributed crossovers at intermediate values of S; however, there is an excess of recombinants at low S values and a paucity of recombinants at high S values.

Deviation at both ends of the curve might be expected for several reasons. One possibility is that crossovers are not evenly distributed. For example, they might be clustered near the ends; however, a consistent gradient of recombination away from the ends has proven difficult to fit to the data (A. White and J. H. Wilson, unpublished data). Another possibility is that the recombining linear molecules were shorter than the input linear molecules. Elimination of sequences from the ends would decrease both S and $1 - S$. If the true S were less than the input S, the fraction of wild-type recombinants (T/S) would be higher than expected, leading to an apparent excess of recombinants that would be most notable at low values of input S. A second recombination event in $1 - S$ is required to complete the circle. At low values of $1 - S$ (high values of S), removal of sequences from the ends could eliminate the homology required for a recombination event, leading to an apparent paucity of recombinants at high values of S. This decrease in recombinants at high S values would also extend to aligned linear molecules ($S = 1$), which, if undamaged at their ends, would be expected to circularize efficiently by ligation after recombination, thereby producing a greater number of wild-type recombinants than was observed.

Efficiency of intermolecular recombination. Intramolecular end joining and intramolecular recombination occur with the same overall efficiency in monkey cells (38). To assess the

TABLE 3. Recombination as a function of replicative potential

Cross	Cleavage enzymes[a]	S[b]	Cross type[c]	No. of plaques[d]	%PT[e]	%RT[f]
tsA7 × tsA30						
1a	*Bam*HI, *Taq*I	0.42	I	151	45	25.8
1b	*Taq*I, *Bam*HI		II	5		
2a	*Bam*HI, *Hae*II	0.58	I	56	17	18.7
2b	*Kpn*I, *Bam*HI		II	0		
3a	*Bam*HI, *Hae*II	0.68	I	64	19	15.9
3b	*Hae*II, *Bam*HI		II	4		
4a	*Hae*II, *Taq*I	0.74	I	29	8.7	14.6
4b	*Taq*I, *Hae*II		II	2		
5a	*Taq*I, *Hpa*II	0.84	I	20	6.0	12.9
5b	*Kpn*I, *Taq*I		II	5		
6a	*Hae*II, *Hpa*II	0.90	I	12	3.6	12.0
6b	*Kpn*I, *Hpa*II		II	1		
tsB228 × tsB11						
1a	*Bam*HI, *Hae*II	0.32	I	82	31	28.8
1b	*Hae*II, *Bam*HI		II	4		
2a	*Bam*HI, *Hpa*II	0.42	I	44	17	21.9
2b	*Hpa*II, *Bam*HI		II	4		
3a	*Bam*HI, *Taq*I	0.58	I	40	15	15.8
3b	*Taq*I, *Bam*HI		II	9		
4a	*Taq*I, *Hae*II	0.74	I	49	19	12.4
4b	*Hae*II, *Taq*I		II	6		
5a	*Taq*I, *Hpa*II	0.84	I	36	14	10.9
5b	*Hpa*II, *Taq*I		II	11		
6a	*Hae*II, *Hpa*II	0.90	I	10	3.8	10.2
6b	*Hpa*II, *Hae*II		II	6		
tsB228 × tsB4						
1a	*Bam*HI, *Hae*II	0.32	I	144	45	28.8
1b	*Hae*II, *Bam*HI		II	8		
2a	*Bam*HI, *Hpa*II	0.42	I	61	19	21.9
2b	*Hpa*II, *Bam*HI		II	13		
3a	*Bam*HI, *Taq*I	0.58	I	59	18	15.8
3b	*Taq*I, *Bam*HI		II	10		
4a	*Taq*I, *Hae*II	0.74	I	25	7.8	12.4
4b	*Hae*II, *Taq*I		II	10		
5a	*Taq*I, *Hpa*II	0.84	I	30	9.3	10.9
5b	*Hpa*II, *Taq*I		II	4		
6a	*Hae*II, *Hpa*II	0.90	I	2	0.6	10.2
6b	*Hpa*II, *Taq*I		II	6		

[a] All linear genomes were generated by cleaving circular genomes with the indicated restriction enzymes. The enzyme listed first in each pair was used to cleave the genome listed first in each cross. *Kpn*I was used instead of *Hpa*II to cleave tsA7 because this mutant contains a second *Hpa*II site near the *Bam*HI site (Wilson, unpublished data).

[b] S is the fraction of the genome between the restriction sites that define the offset segment (Fig. 3).

[c] Crosses are designated type I or type II as defined in the legend to Fig. 3.

[d] Wild-type recombinants were selected by plaque assay of the transfection mixtures under nonpermissive conditions. The plaque counts were summed from several experiments. For mixtures of tsA7 and tsA30, 75-ng portions of each genome were mixed and plaque assayed on a total of 10 plates in two experiments. For mixtures of tsB228 and tsB4, 30-ng portions of each genome were mixed and plaque assayed on a total of four plates in two experiments. For mixtures of tsB228 and tsB11, 173-ng portions of each genome were mixed and plaque assayed on a total of 17 plates in four experiments.

[e] Plaque counts for the type I crosses for each pair of mutants were divided by the total number of plaques and expressed as percentage of total plaques (%PT).

[f] A recombination value (R) for each type I cross was calculated as $R = k(T/S)$, where k is an unknown efficiency constant and T is the target size in fractional genome units ($T = 0.18$ for tsA7 × tsA30, $T = 0.15$ for tsB228 × tsB11, and $T = 0.11$ for tsB228 × tsB4). Individual recombination values were divided by the total recombination value for each pair of mutants and expressed as percentage of the total recombination value (%RT). In this normalization, k and T cancel out so that %RT is related to $1/S$.

TABLE 4. Recombination in crosses with linear viral and plasmid DNAs

Cross	Cleavage enzyme for the following DNA form[a]:				Plaques (%)[b]
	tsA7		tsA30		
	Viral	Plasmid	Viral	Plasmid	
1	*Bam*HI		*Taq*I		100
2	*Bam*HI			*Taq*I	106
3		*Bam*HI	*Taq*I		69
4		*Bam*HI		*Taq*I	110
5	*Bam*HI			*Tha*I	44
6		*Tha*I	*Taq*I		99
7		*Tha*I		*Tha*I	86

[a] Linear viral and plasmid DNAs were prepared by cleavage with the indicated restriction enzyme. Cleavage with *Taq*I cleanly excises the tsA30 genome from the plasmid DNA. Cleavage with *Bam*HI cleanly excises the tsA7 genome from the plasmid DNA. Cleavage with *Tha*I leaves short pBR322 tails of 130 and 322 base pairs attached to the tsA30 genome and tails of 30 and 325 base pairs attached to the tsA7 genome. The results from the two orientations of tsA7 were not significantly different and have been averaged.

[b] Equimolar mixtures of the indicated DNA forms were transfected into CV1 cells and assayed under nonpermissive conditions to select for wild-type recombinants. For viral DNAs, 7.5 ng was assayed per plate; for plasmid DNAs, 13.8 ng was assayed per plate. The plaque counts in each of two experiments were normalized to that for cross 1, and their average is expressed as a percentage. The data for crosses 1, 2, and 5 were derived from 10 plates. The data for crosses 3, 4, 5, and 7, which include results from both tsA7 plasmids, were derived from 20 plates.

efficiency of intermolecular recombination, we compared recombination between offset linear genomes to the joining of pairs of restriction fragments. Intermolecular recombination (Fig. 5, filled triangles) and intermolecular end joining (open symbols) occupied the same part of the concentration curve, suggesting that they occur with similar efficiencies.

In addition, both kinds of intermolecular event were shifted along the concentration curve by more than two orders of magnitude from the corresponding intramolecular recombination and end-joining events, whose positions are indicated by the one-hit dilution curve (37–39). These results are consistent with independent measurements, by which it has been estimated that at 1 ng per plate, the fraction of mixedly infected cells is less than 5% of the fraction of singly infected cells (M. Seidman, personal communication). Thus, mixedly infected cells are relatively rare under our conditions of DEAE-dextran-mediated transfection. These results contrast with calcium phosphate-mediated transfections which show less of a difference between intramolecular and intermolecular events (11).

Effects of non-SV40 terminal extensions. To test whether homology at the ends of linear genomes is required for recombination, we assayed recombination between SV40 genomes that had non-SV40 sequences added to their termini. Non-SV40 sequences were added to the *Bam*HI ends of tsA7 and to the *Taq*I ends of tsA30 by cloning into pBR322 through those restriction sites and then excising the SV40 genomes with the restriction enzyme *Tha*I, which does not cleave SV40 DNA. This procedure generates non-SV40 tails of 130 and 322 base pairs for tsA7 and 30 and 325 base pairs for tsA30. The pBR322 tails on tsA7 and tsA30 are not homologous to one another.

The results of transfections with several mixtures of linear genomes with and without terminal extensions are summarized in Table 4. Crosses involving SV40 genomes with non-SV40 tails on one or both genomes (crosses 5, 6, and 7) yielded the same percentage of wild-type recombinants as did crosses involving genomes with no attached pBR322 sequences (crosses 1, 2, 3, and 4). We have obtained similar results with linear SV40 genomes that lack terminal phosphates, have blunt ends, or have dideoxynucleotides at their 3′ ends (data not shown). These results indicate that homologous recombination between linear molecules is relatively insensitive to the exact nature of the molecular termini on the input genomes.

DISCUSSION

The topological requirements for homologous recombination between DNA molecules transfected into monkey cells

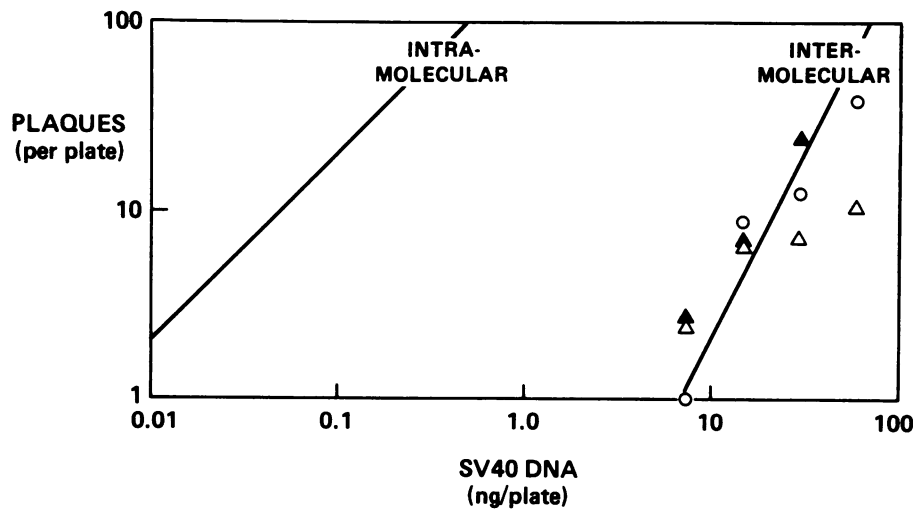

FIG. 5. Dilution curves for intermolecular events. In these experiments, the numbers of plaques per 60-mm plate are plotted as a function of the total nanograms of SV40 DNA added per plate. Open symbols show results of experiments in which complementary restriction fragments needed to join to create a wild-type genome. Open triangles show the results obtained with a mixture of *Pst*I-cleaved tsA7 and tsA30 DNAs; the data are averaged from five plates. Open circles show the results obtained with a mixture of *Bam*HI- and *Eco*RI-cleaved tsB228 and tsB4 DNAs; the data are averaged from five plates. Filled triangles show the results obtained with a mixture of *Taq*I-cleaved tsA30 and *Bam*HI-cleaved tsA7 DNAs, which have to recombine to generate a wild-type genome; the data are averaged from eight plates.

were defined in this study by assaying recombination between genomes of the animal virus SV40. Crosses between various topological forms (Table 2) as well as crosses between aligned linear genomes (Table 1) and offset linear genomes (Tables 2 and 3) all indicate that homologous recombination occurs preferentially between linear DNA molecules. Because SV40 genomes that were nonfunctional until activated by recombination were used in these experiments, they presumably detected normal cellular processes.

An apparent preference for linear molecules could arise as a result of a low intracellular concentration of circular genomes rather than an inherent inability to recombine. However, that possibility seems unlikely because crosses between replication-competent genomes gave the same results as crosses between replication-incompetent genomes (Table 3). Those crosses differ enormously in the concentration of circular molecules, due to their replicative potential, but show the same diagnostic disparity between type I and type II crosses, thereby suggesting that recombination is relatively rare among replicated circular molecules.

A relatively constant level of wild-type plaques was observed in type II crosses of offset linear genomes (Table 3). Thus far, our experiments give little information about the nature or timing of these background events, which are above the level of reversion, independent of the position of the breaks, and independent of replicative potential. One possibility consistent with recombination as linear molecules is that the background events represent repair of heteroduplex intermediates in the recombination process, that is, by conversion-like events. Although heteroduplexes introduced into monkey cells by transfection are repaired very efficiently (1, 9, 18, 37, 39), nothing is known about the fate of mismatches in the presumptive heteroduplex intermediates in homologous recombination.

A requirement for linear recombination substrates is consistent with other studies in animal cells, which show that double-strand breaks stimulate homologous recombination between transfected DNA molecules (3a, 4a, 8, 11). In those studies, the level of stimulation relative to background ranged from 5- to 100-fold, which is quite similar to the range observed in our experiments (Tables 2 and 3). Stimulation presumably occurs as a consequence of providing proper substrates to the cellular recombination machinery. In addition, since viral genomes, which recombine extrachromosomally, were stimulated to the same extent as selectable genes, which were assayed after chromosomal integration, both kinds of experiment probably are detecting the same extrachromosomal process.

The finding that linear substrates are required for recombination focuses attention on the ends of the molecules as playing a critical role in the mechanism. One possibility is that cellular enzymes, such as exonucleases or helicases, act at DNA ends to expose single-stranded regions, which can initiate homologous pairing to begin the recombination process. Whatever process exposes single strands must be able to uncover sequences that are a few hundred nucleotides in from the ends, since short terminal nonhomologies do not decrease the efficiency of recombination (Table 4). Furthermore, regardless of how pairing is initiated, crossovers must be distributed around the genome (28, 36, 37) (Fig. 5). Finally, the mechanism may need to allow for removal of terminal sequences, as suggested by the data in Fig. 5.

An example of a potential mechanism to account for these data is illustrated in Fig. 6 for an offset linear cross involving genomes with terminal nonhomology. The steps in this model are as follows: (i) single strands are exposed at ends

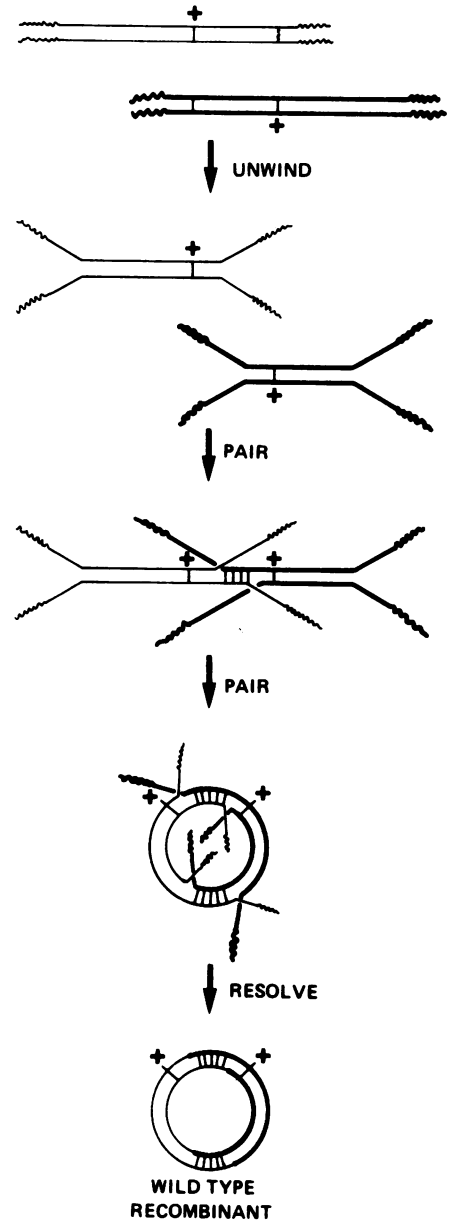

FIG. 6. An unwinding model for homologous recombination in transfected DNA. The steps in the proposed model are illustrated for recombination between offset linear genomes with terminal nonhomology. Individual steps include (i) unwinding from the ends, (ii) pairing between complementary single strands, (iii) branch migration of the junction (not illustrated), and (iv) repair of protruding single-strand tails. The mechanism is nonconservative; it produces only one of the two potential products expected from a conservative exchange. Here we use the terms conservative and nonconservative, because the more common terms, reciprocal and nonreciprocal, are used with two distinct meanings in the literature. Reciprocal and nonreciprocal are used in a global sense, in the same way as conservative and nonconservative, to describe the number of products. However, they are also used in a local sense to describe the exchange of information in the region of the crossover: an unequal representation of parental information in the region of the crossover signifies a nonreciprocal event or, more commonly, a gene conversion event. The potential for confusion can be illustrated by reference to the double-strand-break repair model (31). This model is reciprocal in the global sense, since two products are produced, but nonreciprocal in the local sense, since the parental contributions are unequal in the region of the crossover.

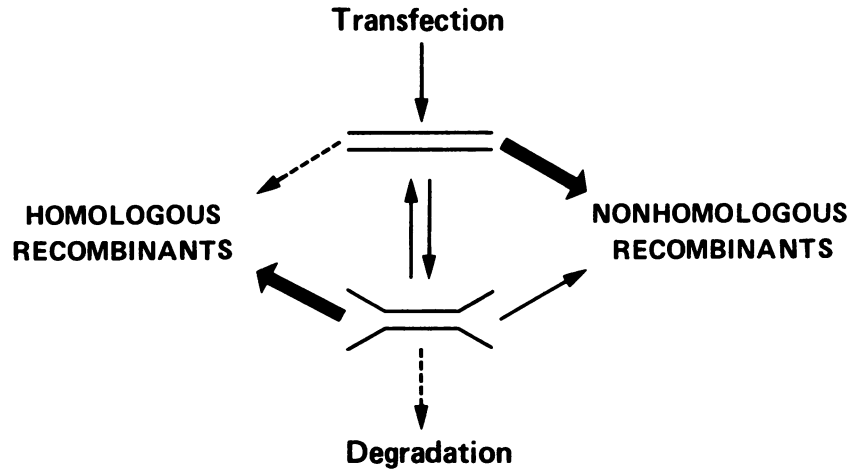

FIG. 7. Possible relationship between substrates for homologous and nonhomologous recombination. Transfected DNA in the nucleus may be unwound, as shown here, or stripped, perhaps as a first step to degradation (vertical arrows). Intact ends are the preferred substrates for nonhomologous recombination, whereas unwound ends are the preferred substrates for homologous recombination (thick arrows). Molecules with intact ends may give rise to some homologous recombinants, but not by any of the mechanisms discussed here (dashed arrow). Molecules with unwound ends may serve as the source of some nonhomologous recombinants as suggested by the existence of duplication junctions (thin arrow).

by unwinding, (ii) pairing is initiated by heteroduplex formation between complementary single strands, (iii) branch migration allows junctions to distribute, and (iv) repair of the single-strand tails fixes the crossover points. Terminal sequences could be removed as a consequence of nicks introduced before or after unwinding. This model does not represent a unique solution to the data; indeed, two other models have been proposed (11, 31).

The double-strand-break repair model (31) was formulated in response to observations on break-stimulated recombination in yeast cells and constitutes a productive way of thinking about meiotic recombination. In accord with observations made in those experiments, it is proposed in this model that the two potential products of an exchange are preserved; that is, this is a conservative mechanism (see the legend to Fig. 6). This feature is not compatible with the apparently nonconservative nature of intramolecular homologous recombination in transfected DNA, in which only one of the potential products seems to be preserved in any individual exchange. For example, homologous recombination within transfected circular dimers of SV40 apparently produces only one monomer genome (38). More extensive studies with a shuttle-vector system have failed to detect conservative recombination in transfected DNA: intramolecular inversion through inverted repeats, which can only occur by a conservative mechanism, was not detected; intramolecular recombination in a circular plasmid containing direct repeats produced only one product (M. Seidman, personal communication). If intermolecular homologous recombination in transfected DNA is also nonconservative, then the double-strand-break repair model as originally formulated is unlikely to describe it, although nonconservative versions of the model might apply (29a).

The polar stripping model of Lin et al. (11), on the other hand, is nonconservative, as is the mechanism in Fig. 6. In this model, nucleotides in one strand are stripped from the ends of molecules by the polar action of a 5′-to-3′ exonuclease. In other respects, this mechanism is much like the one depicted in Fig. 6 and accounts well for the existing data on homologous recombination in transfected DNA. However, one class of nonhomologous junctions, duplication junctions (21, 41; D. B. Roth, T. N. Porter, and J. H. Wilson, Mol. Cell. Biol., in press), is more readily explained by unwinding at termini than by degradation. These relatively rare junctions (about 5% of the total) carry inserted nucleotides at the junction that repeat an immediately adjacent sequence; a possible mechanism for their formation will be presented elsewhere (Roth et al., in press). In any case, further experiments will be required to determine whether exposed single strands at the ends of transfected DNA are critical to the mechanism of homologous recombination and, if so, whether an exonuclease or a helicase, if either, is responsible for it.

In studies of nonhomologous end joining (38; Roth et al., in press), very little degradation was detected at the ends of transfected DNA molecules. However, removal of terminal sequences before homologous recombination was suggested as an explanation for the discrepancy between the theoretical and experimental curves in Fig. 4. Are these two views compatible? One possibility is that transfected molecules in the nucleus exist in two forms which constitute separate sources of recombination substrates. Molecules with duplex ends might preferentially recombine nonhomologously by end ligation, whereas molecules with unwound ends, for example, might preferentially recombine homologously through the exposed single strands (Fig. 7). Although the sizes of the two putative populations are undefined by existing data, they may be similar, since homologous and nonhomologous recombination events occur in transfected DNA with similar efficiencies, whether measured intramolecularly (38) or intermolecularly (Fig. 5), and with similar rates as measured in a kinetic competition assay.

Although animal cells efficiently rearrange foreign DNA by homology-dependent and homology-independent mechanisms, the functions of these recombination processes remain an enigma. Almost certainly, the normal object is not the shuffling of genetic information usually associated with meiotic recombination, and yet the widespread occurrence of homologous and nonhomologous recombination in eucaryotic cells suggests an underlying importance. At the

least, these recombination processes represent obstacles that must be understood and overcome for the controlled targeting of exogenous DNA sequences into homologous chromosomal loci. What we know at present is that molecules with ends are the preferred substrates in both kinds of recombination process. Broken ends presumably stimulate nonhomologous recombination by providing termini for efficient end joining (15, 41); they presumably stimulate homologous recombination by providing sites for exposure of single strands. The final result in both cases is a reduction in the number of ends. Perhaps, as has been suggested for yeast cells (19), the enzymes responsible for homologous and nonhomologous recombination exist in animal cells primarily for the repair of double-strand breaks. If so, recombination in transfected DNA may prove an important analytical tool for characterizing cellular repair processes.

ACKNOWLEDGMENTS

We thank Kathleen Marburger for expert technical assistance. We thank Michael Seidman, in particular, for valuable discussions. David Roth, Ursula Weiss, Allen White, the Cedar Bay Conference participants, and the Molecular Genetics Group in the Biochemistry Department at Baylor College of Medicine, Houston, Tex., also provided useful discussions and critical suggestions. We thank Michael Seidman, Nat Sternberg, Raju Kucherlapati, Daniel Camerini-Otero, and Mario Capecchi for communicating their results to us in advance of publication.

This work was suppported by Public Health Service grants CA 15743 and GM 33405 from the National Institutes of Health and grant Q-977 from the Robert A. Welch Foundation.

LITERATURE CITED

1. **Abastado, J.-P., B. Cami, T. H. Dinh, J. Igolen, and P. Kourilsky.** 1984. Processing of complex heteroduplexes in *E. coli* and COS-1 monkey cells. Proc. Natl. Acad. Sci. U.S.A. **81**:5792–5796.
2. **Ashman, C. R., and R. L. Davidson.** 1984. High spontaneous mutation frequency in shuttle vector sequences recovered from mammalian cellular DNA. Mol. Cell. Biol. **4**:2266–2272.
3. **Bandyopadhyay, P. K., S. Watanabe, and H. M. Temin.** 1984. Recombination in transfected DNAs in vertebrate cells in culture. Proc. Natl. Acad. Sci. U.S.A. **81**:3476–3480.

3a. **Brenner, D. A., S. Kato, R. A. Anderson, A. C. Smigocki, and R. D. Camerini-Otero.** 1984. The recombination and integration of DNAs introduced into mouse L cells. Cold Spring Harbor Symp. Quant. Biol. **49**:151–160.

4. **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.

4a. **Folger, K., K. Thomas, and M. R. Capecchi.** 1984. Analysis of homologous recombination in cultured mammalian cells. Cold Spring Harbor Symp. Quant. Biol. **49**:123–138.

5. **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.
6. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cultures. J. Mol. Biol. **26**:365–369.
7. **Kelly, T. J., and D. Nathans.** 1977. The genome of simian virus 40. Adv. Virus Res. **21**:85–173.
8. **Kucherlapati, R. S., E. M. Eves, K.-Y. Song, B. S. Morse, and O. Smithies.** 1984. Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. Proc. Natl. Acad. Sci. U.S.A. **81**:3153–3157.
9. **Lai, C., and D. Nathans.** 1975. A map of temperature-sensitive mutants of simian virus 40. Virology **66**:70–81.
10. **Lebkowski, J. S., R. B. DuBridge, E. A. Antell, K. S. Greisen, and M. P. Calos.** 1984. Transfected DNA is mutated in monkey, mouse, and human cells. Mol. Cell. Biol. **4**:1951–1960.
11. **Lin, F.-L., K. Sperle, and N. Sternberg.** 1984. Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. Mol. Cell. Biol. **4**:1020–1034.
12. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **Mautner, V., and N. MacKay.** 1984. Recombination in adenovirus: analysis of crossover sites in intertypic overlap recombinants. Virology **139**:43–52.
14. **May, P., M. Resche-Rigon, J. Borde, C. Breugnot, and E. May.** 1984. Conversion through homologous recombination of the gene encoding simian virus 40 115,000-molecular-weight super T antigen to a gene encoding a normal-size large T antigen variant. Mol. Cell. Biol. **4**:1141–1151.
15. **McClintock, B.** 1984. The significance of responses of the genome to challenge. Science **226**:792–801.
16. **McCutchen, J., and J. Pagano.** 1968. Enhancement of the infectivity of SV40 deoxyribonucleic acid with diethylaminoethyl-dextran. J. Natl. Cancer Inst. **41**:351–357.
17. **Miller, C. K., and H. M. Temin.** 1983. High efficiency ligation and recombination of DNA fragments by vertebrate cells. Science **220**:606–609.
18. **Miller, L., B. Cooke, and M. Fried.** 1976. Fate of mismatched base-pair regions in polyoma heteroduplex DNA during infection of mouse cells. Proc. Natl. Acad. Sci. U.S.A. **73**:3073–3077.
19. **Orr-Weaver, T. L., and J. W. Szostak.** 1983. Yeast recombination: the association between double-strand gap repair and crossing-over. Proc. Natl. Acad. Sci. U.S.A. **80**:4417–4421.
20. **Parker, R. C., R. M. Watson, and J. Vinograd.** 1977. Mapping of closed circular DNAs by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis. Proc. Natl. Acad. Sci. U.S.A. **74**:851–855.
21. **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.
22. **Razzaque, A., S. Chakrabarti, S. Joffee, and M. Seidman.** 1984. Mutagenesis of a shuttle vector plasmid in mammalian cells. Mol. Cell. Biol. **4**:435–441.
23. **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.
24. **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.

24a. **Roth, D. B., and J. H. Wilson.** 1985. Relative rates of homologous and nonhomologous recombination in transfected DNA. Proc. Natl. Acad. Sci. U.S.A. **82**:3355–3359.

25. **Rubnitz, J., and S. Subramani.** 1984. The minimum amount of homology required for homologous recombination in mammalian cells. Mol. Cell. Biol. **4**:2253–2258.
26. **Sarkar, S., U. B. Dasgupta, and W. C. Summers.** 1984. Error-prone mutagenesis detected in mammalian cells by a shuttle vector containing the *supF* gene of *Escherichia coli*. Mol. Cell. Biol. **4**:2227–2230.
27. **Schaffner, W.** 1980. Direct transfer of cloned genes from bacteria to mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **77**:2163–2167.
28. **Shapira, G., J. L. Stachelek, A. Letson, L. K. Soodak, and R. M. Liskay.** 1983. Novel use of synthetic oligonucleotide insertion mutants for the study of homologous recombination in mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **80**:4827–4831.
29. **Small, J., and G. Scangos.** 1983. Recombination during gene transfer into mouse cells can restore the function of deleted genes. Science **219**:174–176.

29a. **Stahl, F. W., I. Kobayashi, and M. M. Stahl.** 1985. In phage lambda, cos is a recombinator in the red pathway. J. Mol. Biol. **181**:199–209.

30. **Subramani, S., and P. Berg.** 1983. Homologous and nonhomologous recombination in monkey cells. Mol. Cell. Biol. **3**:1040–1052.
31. **Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein, and F. S. Stahl.** 1983. The double-strand-break repair model for recombination. Cell **33**:25–35.
32. **Tooze, J.** 1980. DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **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.
34. **Upcroft, P., B. Carter, and C. Kidson.** 1980. Mammalian cell functions mediating recombination of genetic elements. Nucleic Acids Res. **8**:5835–5844.
35. **Volkert, F. C., and C. S. H. Young.** 1983. The genetic analysis of recombination using adenovirus overlapping terminal DNA fragments. Virology **125**:175–193.
36. **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.
37. **Wake, C., and J. Wilson.** 1980. Defined oligomeric SV40 DNA: a sensitive probe of general recombination in somatic cells. Cell **21**:141–148.
38. **Wake, C. T., T. Gudewicz, T. Porter, A. White, and J. H. Wilson.** 1984. How damaged is the biologically active subpopulation of transfected DNA? Mol. Cell. Biol. **4**:387–398.
39. **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.
40. **Wilson, J.** 1978. Interference in SV40 DNA infections: a possible basis for cellular competence. Virology **91**:380–388.
41. **Wilson, J. H., P. B. Berget, and J. M. Pipas.** 1982. Somatic cells efficiently join unrelated DNA segments end-to-end. Mol. Cell. Biol. **2**:1258–1269.
42. **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 40 infection. J. Virol. **20**:391–399.