MOLECULAR AND CELLULAR BIOLOGY, Oct. 1985, p. 2599–2607
0270-7306/85/102599-09$02.00/0

Vol. 5 No. 10

Mechanisms of Nonhomologous Recombination in Mammalian Cells

DAVID B. ROTH, THOMAS N. PORTER, AND JOHN H. WILSON*

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

Received 26 April 1985/Accepted 25 June 1985

The primary mechanism of nonhomologous recombination in transfected DNA involves breakage followed by end joining. To probe the joining step in more detail, linear simian virus 40 genomes with mismatched ends were transfected into cultured monkey cells, and individual viable recombinants were analyzed. The transfected genomes carried mismatched ends as a result of cleavage with two restriction enzymes, the recognition sites of which are located in the intron of the gene encoding the T antigen. Because the T antigen gene was split by this cleavage, the transfected genomes were inert until activated by cell-mediated end joining. Clonal descendants of the original recombinants were isolated from 122 plaques and were grouped into four classes based on the electrophoretic mobility of the junction fragment. The structures of representative junctions were determined by nucleotide sequencing. The spectrum of nonhomologous junctions analyzed here along with a large number of previously reported junctions suggest that there are two mechanisms for the linkage of DNA molecules: (i) direct ligation of ends and (ii) repair synthesis primed by terminal homologies of a few nucleotides. A paired-priming model of nonhomologous recombination is discussed.

Recombination events that show little or no dependence on nucleotide sequence homology represent important pathways for DNA rearrangement in animal cells. Examples of such nonhomologous rearrangements include chromosome translocations (14), gene amplification events (44), movements of retroviruses and transposable elements (41), developmental rearrangements of antibody and T-cell receptor genes (2, 15, 17, 28, 61), formation of processed pseudogenes (6), and programmed rearrangements in ciliates (5, 21). Although the mechanisms may differ, each process accomplishes a breakage and reunion of DNA segments resulting in a genetic rearrangement.

Nonhomologous recombination also occurs efficiently among DNA molecules transfected into animal cells in culture (22, 38, 46, 57). Although the primary pathway of nonhomologous recombination in transfected DNA apparently involves breakage followed by end joining (57), neither the mechanism of breakage nor the mechanism of end joining is known. Ample evidence indicates that both breakage (4, 8, 23, 36, 37, 51) and efficient end joining (22, 32, 38, 51, 57) occur in transfected DNA in a variety of animal cells, thereby confirming the principal components of the overall pathway. Similar breakage-joining mechanisms have been proposed for reciprocal translocations (14) and for rearrangements in the immune system (2). The similarities among these disparate processes are underscored by a peculiar type of junction, termed an insertion junction, which is common to all three systems. These junctions contain extra nucleotides of unknown origin, which appear exactly at the junction. The structure of these insertion junctions suggests that free DNA ends, to which extra nucleotides could be added by terminal transferase, are a likely intermediate (9, 57).

The individual steps in the overall mechanism of nonhomologous recombination can be conveniently analyzed in the intron of the simian virus 40 (SV40) T antigen gene, because the intron is dispensable for lytic growth. This feature allows nonhomologous recombination to be studied statistically by plaque assay of viable recombinants and individually by analysis of recombinant genomes. In this study, we examined the mechanism of end joining by analyzing the spectrum of viable recombinants that arose after transfection of linear SV40 genomes with mismatched ends: one end was blunt and the other end carried a 5′ protruding single strand. The ends of the genomes were within the intron of the gene encoding T antigen, which is required for viral replication and the subsequent expression of viral capsid genes. Because the T antigen gene is interrupted in these genomes, no viral genes could be expressed until after circularization, thereby ensuring that the enzymatic activities responsible for circularization were cellular rather than viral. The nonhomologous recombination junctions present in 122 viable genomes were classified by restriction digestion, and the structures of 28 representative junctions were determined by nucleotide sequencing. Examination of these junctions and a comprehensive list of junctions from the literature supports two mechanisms for the linkage of duplexes: (i) direct ligation of ends and (ii) repair synthesis primed by terminal homology of a few nucleotides.

MATERIALS AND METHODS

Cells, viruses, and DNAs. Procedures for the growth of the established monkey kidney CV1 cell line have been described previously (57). Construction of the DNA substrate used in this study has been described previously (38). DNA transfections were carried out as described previously (55) with DEAE-dextran and 0.04 to 0.08 ng of SV40 DNA per 60-mm dish. This quantity of linear SV40 DNA (specific infectivity, 100 to 200 PFU/ng) allows recovery of well-isolated plaques. Plaque assays were performed according to published procedures (56). All enzymes were purchased from Boehringer Mannheim Biochemicals and used according to the manufacturer's recommendations.

Miniwell preparation of DNA. Confluent monolayers of CV1 cells in 96-well microtiter plates were infected with picked plaque suspensions, and viral DNA was labeled in vivo by the addition of $^{32}P_i$ as described previously (57).

DNA sequencing. DNA samples were sequenced by the method of Maxam and Gilbert (30). DNAs from recombinant viruses were cleaved with *Taq*I or *Eco*RI and labeled at the 3′ end with the Klenow fragment of DNA polymerase I and

* Corresponding author.

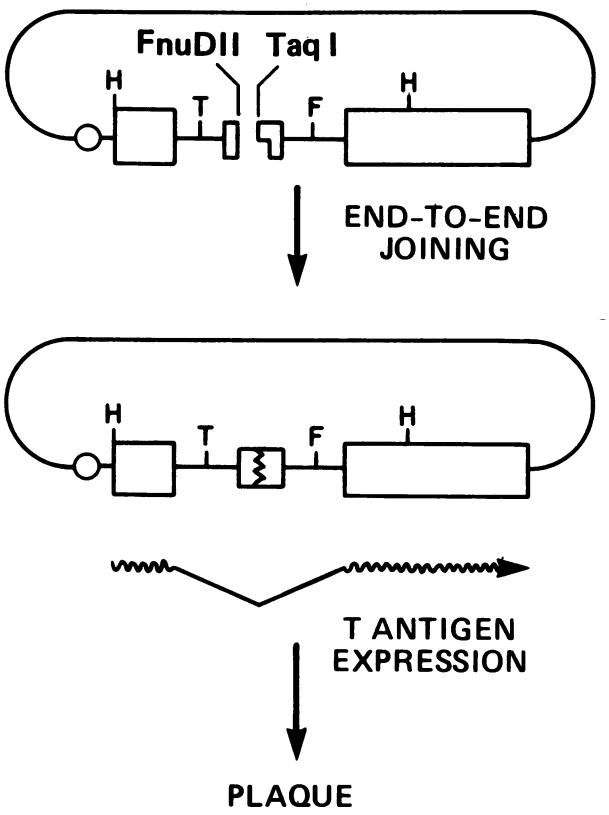

FIG. 1. DNA substrate for nonhomologous recombination. A linear derivative of the SV40 genome is shown. Large open boxes denote exons that encode large T antigen, and the open circle represents the origin of replication. The termini, which lie within the intron of the gene for T antigen, were produced by digestion with *Fnu*DII and *Taq*I. The details of construction have been described previously (38). Joining of the ends after transfection into CV1 cells reconstructs a functional T antigen gene, leading to viral replication and the eventual formation of a plaque. Clonal descendants of an original recombinant can be recovered from an isolated plaque. Internal *Taq*I (T) and *Fnu*DII (F) restriction sites, as well as the *Hin*dIII (H) sites that define the SV40 *Hin*dIII B fragment, are indicated; these sites were used to classify the junctions as described in the text and in the legend to Fig. 2.

[α-^{32}P]dCTP or [α-^{32}P]dATP. DNA from recombinant DR116 was cleaved with *Fnu*DII and labeled at the 5′ end with T4 polynucleotide kinase. In all cases, secondary cuts were made with *Hin*dIII.

Analysis of junctional homologies. Junctions resulting from blunt-end ligation can exhibit junctional homology due to chance identities of flanking nucleotides. For duplexes with an unbiased base composition the probability that a junction will have X nucleotides of homology is given by the equation $P(X) = (X + 1)(1/4)^X(3/4)^2$, where $(X + 1)$ is the number of different ways that chance identities could yield the specified homology, $(1/4)^X$ is the probability the X nucleotides match, and $(3/4)^2$ is the probability that nucleotides flanking the matching nucleotides do not match. By way of example, consider two parental sequences, *N*1-*N*2-*N*3-*N*4-*N*5-*N*6 and *n*1-*n*2-*n*3-*n*4-*n*5-*n*6, which have recombined by breakage and blunt-end joining to generate the recombinant sequence, *N*1-*N*2-*N*3-*n*4-*n*5-*n*6. The probability that this junction contains zero nucleotides of homology ($N3 \neq n3$ and $N4 \neq n4$) is $(3/4)^2$. The probability of finding one nucleotide of homology on either side of the junction ($N3 = n3$ or $N4 = n4$, and the outside nucleotides do not match) is given by $2(1/4)(3/4)^2$.

To test the effect of biased base composition, the probability that blunt-end junctions would exhibit homology was determined empirically. A computer program was used to generate and analyze the homology in 5,000 junctions between duplexes with a specified base composition. The fraction of junctions displaying X nucleotides of homology in such a population of junctions was taken to be equal to the

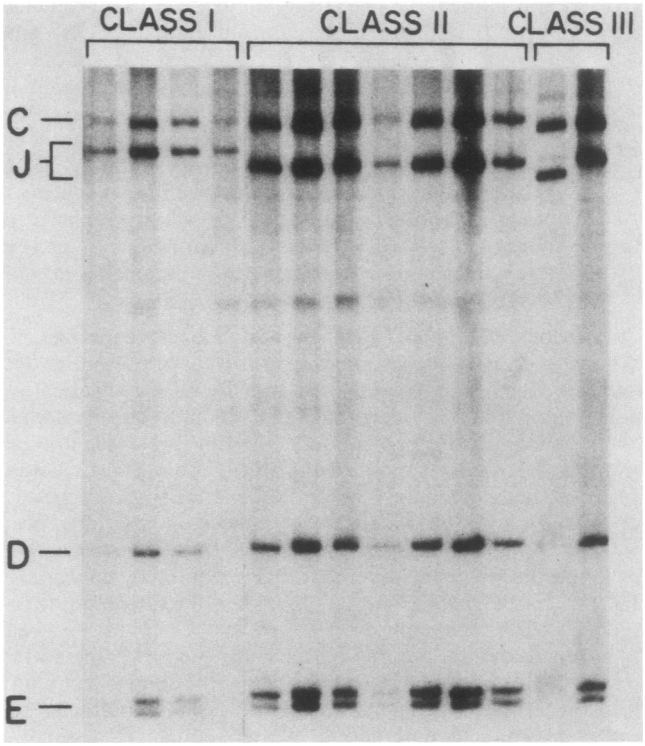

FIG. 2. Restriction enzyme analysis of recombinant genomes. The products of individual end-joining events were analyzed by digestion with the restriction enzymes *Hin*dIII and *Taq*I. ^{32}P-labeled viral DNA was prepared by the miniwell method (see the text.) Labeled DNA was restricted, and the resulting fragments were separated by electrophoresis through 5% polyacrylamide gels, which were then dried and autoradiographed. *Hin*dIII cuts SV40 into five fragments; only fragments C, D, and E are shown. The *Hin*dIII B fragment is absent because it is split into two subfragments by *Taq*I digestion (Fig. 1). The smaller of these fragments is 430 base pairs in length and migrates slightly above the E fragment. The larger fragment, which is approximately 860 base pairs in length, contains the recombinant junction, and is labeled J. The designations class I, class II, and class III are based on electrophoretic mobility differences in the junction-containing fragment. All lanes in the figure are from the same gel, but the order of the lanes in the photograph was rearranged for purposes of illustration. The *Hin*dIII C fragment in the second lane from the right migrated anomalously in this gel; in other experiments it migrated at its expected position. Electrophoretic mobility of the junction fragment is extremely sensitive to the structure of the junction and provides a useful tool for grouping junctions into like kinds. Although the mobility differences between classes are reproducible, they bear no simple relationship to the length of the junction fragment. For example, the class I junction fragments are two nucleotides shorter than the class II fragments, but the migrate more slowly in polyacrylamide gels. Similarly, one of the class III junction fragments shown in the figure contains a 12-base-pair deletion, but it migrates more slowly than class II junction fragments.

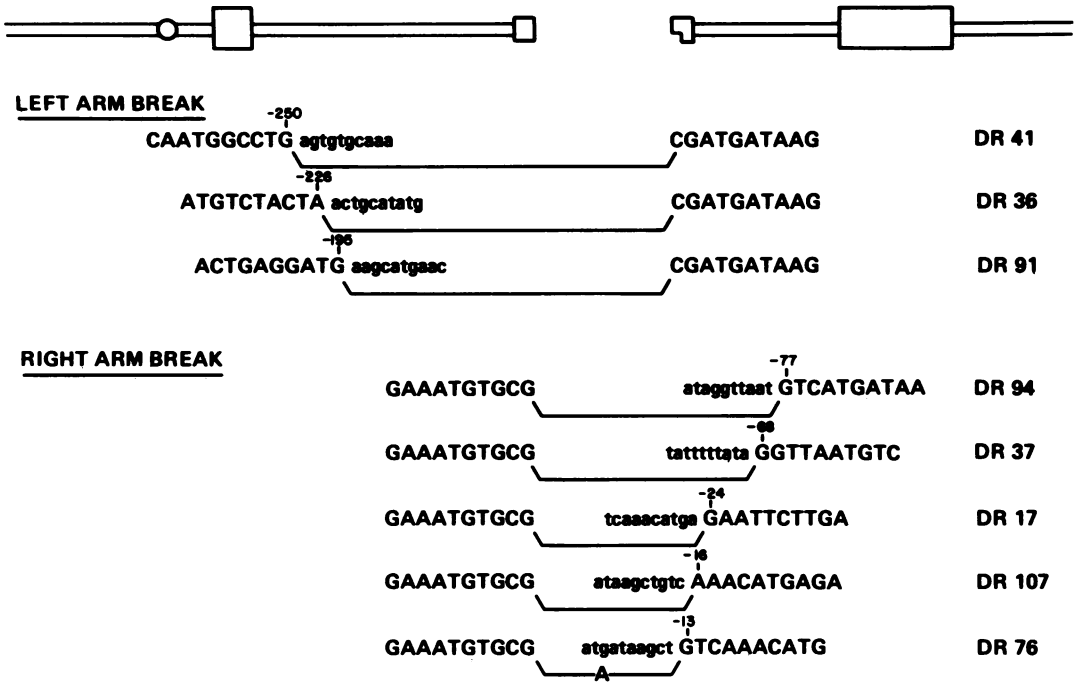

FIG. 3. Nucleotide sequences of deletion junctions. The relevant portion of the SV40 genome is diagrammed at the top of the figure. The exons encoding T antigen are shown as large open boxes; the intron is indicated by the thin lines between the exons; the ends of the input DNA are represented by the small open boxes; the open circle denotes the origin of replication. Ten base pairs on either side of the deletion junction are shown, and deleted nucleotides are represented by lowercase letters. The deletion endpoints are indicated by the numbers above each sequence; the numbering denotes the position of the junction relative to the end of the input linear DNA.

probability of finding X nucleotides of homology in an individual junction. Application of this procedure to junctions between duplexes with an unbiased base composition yielded the same probabilities as the equation described above.

RESULTS

Products of end joining. To characterize the end-joining step of nonhomologous recombination, we used a linear genome with mismatched ends, the construction of which has been described previously (38). The *Fnu*DII end was blunt, whereas the *Taq*I end carried a two-nucleotide 5′ extension (Fig. 1). The design of the construct provides a selection for recombination events, since the genome must be circularized before viral genes can be expressed. This construct has been used previously to measure the relative rates of homologous and nonhomologous recombination in transfected DNA, there being a 131-base-pair repeat at its ends (38). Here we examined the junctions in the population of genomes that arose by nonhomologous recombination. These recombinants were first grouped into classes based on the electrophoretic mobility of junction fragments (Fig. 2), and then representative examples of each class were sequenced (Table 1).

The construct was transfected into CV1 monkey kidney cells, and 122 plaques were isolated. Radiolabeled viral DNA was prepared from each plaque by the miniwell method and analyzed by restriction mapping. Viral genomes were grouped into four classes (Table 1) on the basis of the electrophoretic mobility of the restriction fragment that contained the junction (labeled J in the representative gel shown in Fig. 2). Class I (75%) and class II (12%) appeared to be homogeneous, whereas class III (7%) and class IV (7%) were heterogeneous. Genomes in class III had junction fragments with an electrophoretic mobility slightly different from that of class I or class II (Fig. 2), whereas genomes in class IV had junction fragments with quite different mobilities (38).

Sequences of junctions. To understand the basis for the electrophoretic mobility differences, we determined the nucleotide sequence of the recombination junctions in representative members from each class (Table 1). The junctions in 17 randomly chosen class I genomes were shown to be identical by nucleotide sequence analysis, all having lost a CG dinucleotide from the ends. Probability arguments suggest that most class I junctions have this structure (Table 1);

TABLE 1. Classification of nonhomologous recombination products

Classification	No. of plaques	No. of sequenced examples	Structure
Class I	91	17	Loss of CG[a]
Class II	15	2	Direct join
Class III	8	3	Small deletions[b]
Class IV	8	6	Large deletions[b]

[a] The 17 sequenced examples were identical, each having lost a CG dinucleotide from an end of the parental input DNA (Fig. 5). These results can be used to estimate the homogeneity of the class I junctions. The probability of drawing identical samples sequentially from a mixed population is $P = (N/T)(N - 1/T - 1)(N - 2/T - 2)\ldots(N - n/T - n)$, where T is the total number of junctions classified as class I, N is the number of true class I junctions, and n is the number of sequenced examples. If as few as 71 (78%) of the 91 classified junctions were true class I junctions, the probability of drawing 17 identical samples sequentially would be less than 1%. We conclude that most of the junctions in this sample are true class I junctions.

[b] Sequences of these junctions are shown in Fig. 3 and 4.

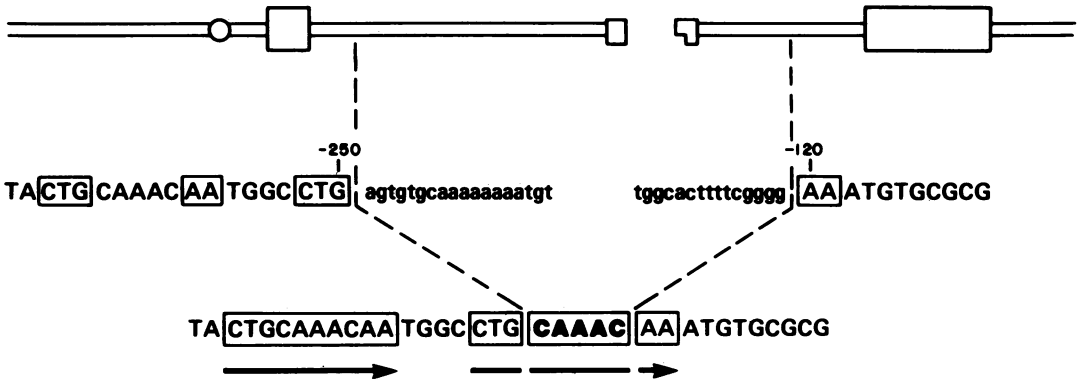

FIG. 4. Nucleotide sequence of the duplication junction in DR116. Symbols are as described in the legend to Fig. 3. Nucleotides in boldface type represent the extra bases inserted at the junction. The arrows indicate a direct repeat of 10 base pairs which is formed by insertion of the extra nucleotides. The nucleotides at the deletion endpoints are homologous to the beginning and end of the repeated unit (signified by the head and tail of the arrow).

however, it would not be unexpected to find some contamination of this class by genomes with class III junctions. The nucleotide sequences of two class II junctions revealed that the *Taq*I end, presumably after having been filled in, had been joined directly to the *Fnu*DII end. This process fortuitously regenerates a *Fnu*DII site, for which all 122 recombinants were subsequently tested by digestion with *Hin*dIII and *Fnu*DII. Only the class II genomes possessed a new *Fnu*DII site at the junction. The purity of this class emphasizes the sensitivity of the electrophoretic assay.

The nucleotide sequences of three class III junctions are shown at the bottom of Fig. 3 (DR17, DR107, and DR76). In each case, a few (12 to 23) nucleotides were lost from the *Taq*I end; the *Fnu*DII terminus remained intact. In one case (DR76), an A residue of unknown origin was present at the junction. Nucleotide sequences of the recombinant junctions from six class IV genomes are presented in Fig. 3 and 4. In five of the six genomes (DR116 being the exception; Fig. 4), the deletions occurred completely within one arm of the input DNA, leaving the other arm intact (Fig. 3).

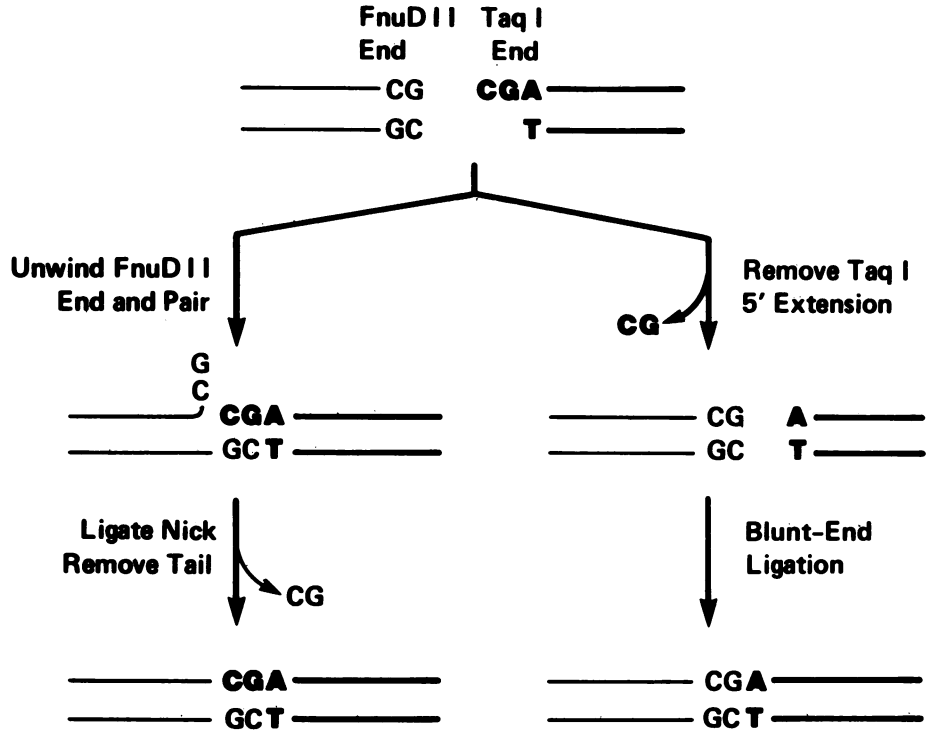

FIG. 5. Two alternative pathways for the formation of class I junctions. In the pathway on the left, the 5′ terminal extension of the *Taq*I end (in boldface type) pairs with the homologous region from the *Fnu*DII end, which is exposed by unwinding of the DNA strands. Removal of the single-stranded tail and ligation of nicks completes the joining process. Alternatively, the homologous region of the *Fnu*DII end could be exposed by the action of a 3′ to 5′ exonuclease. In the right-hand pathway, the *Taq*I end is modified by removal of the protruding 5′ extension and then joined by blunt-end ligation to the *Fnu*DII terminus.

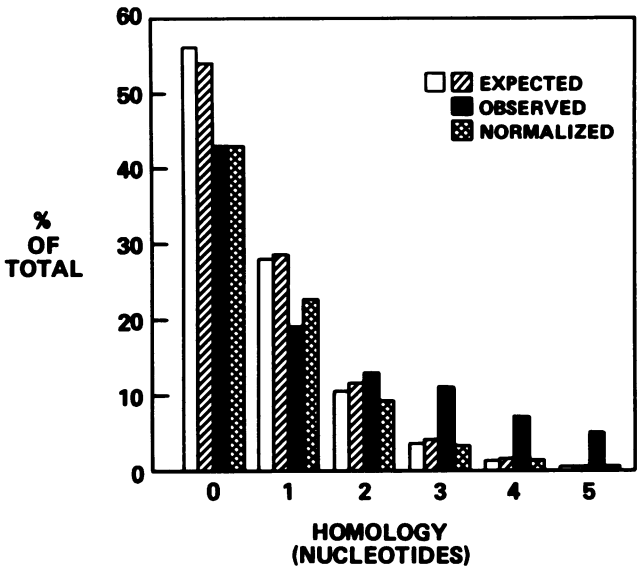

FIG. 6. Analysis of homologies at nonhomologous recombination junctions. A total of 110 junctions (1, 3, 7, 10, 11, 12, 14, 16, 18, 20, 26, 29, 31, 33, 34, 35, 39, 40, 42, 45, 46, 48, 50, 54, 57, 58, 59, 60, 62) were examined for the presence of junctional homology. The junctions were grouped into those generated during DNA transfection (46 junctions) and those generated in other ways (64 junctions), including evolutionary variants of SV40, SV40-adenovirus hybrids, integrations of viral genomes following viral infection, excisions of viral genomes from chromosomes, and chromosome translocation breakpoints. The distributions of junctional homology in these two samples were not different in a consistent way and have been added together in the figure (solid bars). The open bars represent the expected distribution, assuming an unbiased base composition. The hatched bars represent the expected distribution after adjustment for the actual base compositions of the junctions. Base composition was calculated by averaging the GC content of the 20 nucleotides surrounding each breakpoint in the parental DNAs. The distribution of junctional homologies expected for each GC content was calculated (see the text), and the overall probability distribution was determined by averaging the 110 individual distributions. The cross-hatched bars represent the expected distribution (adjusted for base composition) after normalizing to the number of junctions with zero nucleotides of homology.

The sequence of the duplication junction from mutant DR116 is shown in Fig. 4. The structure of this junction is unusual in that 5 extra nucleotides are present at the junction, and the 10 nucleotides encompassing the recombination joint are repeated directly in the sequence of the left-hand arm adjacent to the junction.

DISCUSSION

In this study we used linear SV40 genomes with mismatched ends to probe the nonhomologous recombination capabilities of mammalian cells. Our previous work has demonstrated that monkey cells join mismatched ends with the same efficiency as sticky or blunt ends (51). To ensure that cellular activities were detected, the SV40 DNA molecules were linearized in the intron of the T antigen gene to render the genome inert until activated by recombination. The junctions in 122 recombinant genomes were grouped into four classes according to the electrophoretic mobility of the junction fragment, and the structures of 28 representative junctions were determined by nucleotide sequencing. Results of this analysis indicate that the majority of junctions are flush junctions, in which one parental sequence is joined directly to another. Only two examples of other types of junction were found: DR76 contained an insertion junction, in which one nucleotide of unknown origin appeared exactly at the junction; and DR116 contained a duplication junction, in which added nucleotides at the junction formed part of a direct repeat of an immediately adjacent sequence.

Flush junctions could be accounted for by direct ligation of ends subsequent to end modification by cellular enzymes. The most favored pathway, represented by the class I junctions, resulted in the loss of a CG dinucleotide from one end of the input parental DNA. A CG dinucleotide could have been removed from either the *Taq*I end or the *Fnu*DII end (Fig. 5). As shown on the right in Figure 5, exonucleolytic removal of the 5′ CG extension from the *Taq*I end followed by blunt-end ligation could generate a class I junction. Alternatively, exposure of the 5′ single strand at the *Fnu*DII end, either by a 3′ to 5′ exonuclease or a helicase (as indicated on the left in Fig. 5), followed by direct ligation to the complementary *Taq*I end, could also generate a class I junction. Which of these pathways actually produced the class I junctions cannot be deduced from these data; however, experiments with other constructs indicate that protruding 3′ and protruding 5′ single strands are relatively stable in CV1 cells, supporting the pathway on the left in Fig. 5 as the one by which class I junctions were generated (D. B. Roth and J. H. Wilson, unpublished data). A less favored pathway for end joining, which is represented by the class II junctions (12%), presumably involved the blunting of the *Taq*I end by filling in the recessed 3′ end followed by blunt-end ligation to the *Fnu*DII end. Since the number of

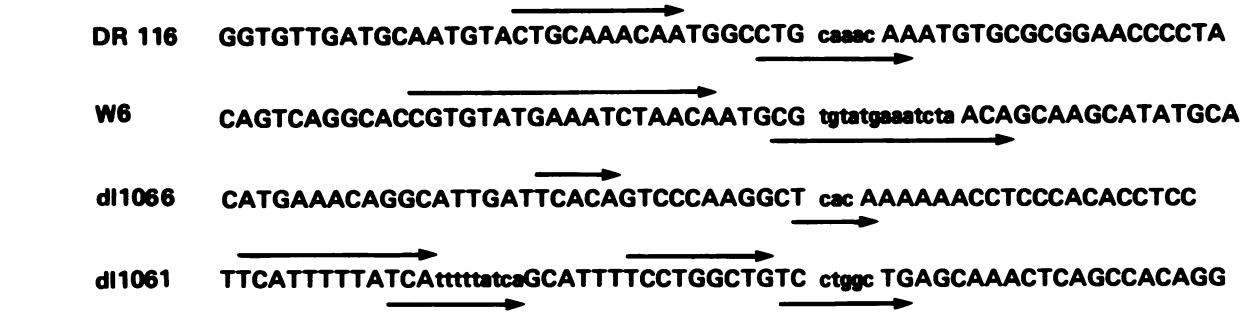

FIG. 7. Nucleotide sequences of duplication junctions. The sequence of W6 is from reference 57; dl1066 and dl1061 are from reference 35. The lowercase letters separated by a space from the uppercase letters denote the extra nucleotides added at the junction. Direct repeats are indicated by arrows. Note that dl1061 contains a second duplication to the left of the duplication junction. A potential mechanism for its formation is shown in Fig. 9.

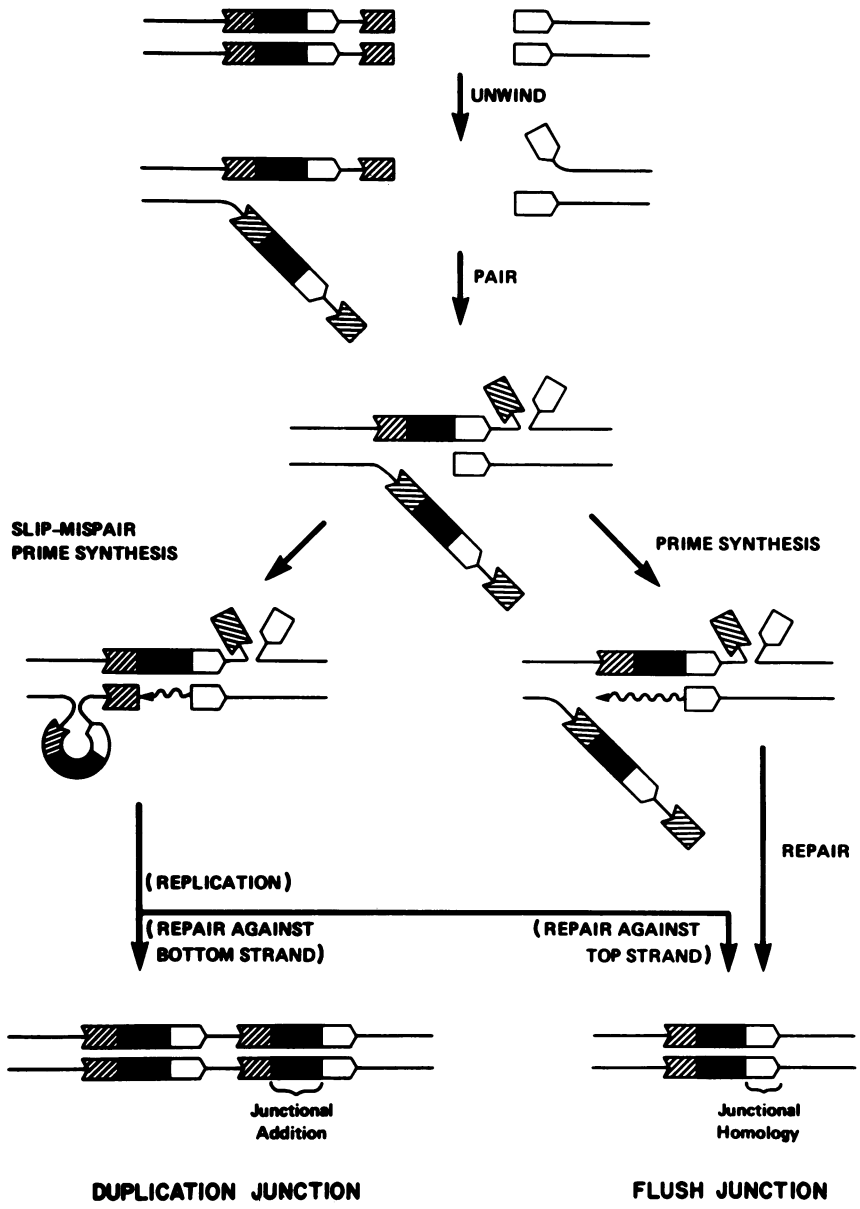

FIG. 8. A paired-priming model for nonhomologous recombination. At the top of the figure, two DNA duplexes are diagrammed. Breakage followed by unwinding uncovers short homologies (open arrowheads) through which duplexes can transiently pair. On the left, slipped mispairing occurs between direct repeats near one end (hatched arrowtails). As illustrated in this figure, pairing between the terminal homologies primes repair synthesis to fill in the gap; however, if a 3′ end slipped and mispaired, it could prime repair synthesis also. The resulting heteroduplex could be resolved by replication or repair (see the text). On the right, pairing between the terminal homologies primes repair DNA synthesis, which fills in the gap. Removal of single-stranded tails and ligation produces a flush junction with junctional homology, corresponding to the terminal homology used in pairing.

junctions in each class reflects the balance of enzymatic activities within the cell, we conclude that the possibilities represented in Fig. 5 are at least five times more likely than the filling in of an end followed by blunt-end ligation.

A third pathway, represented by the class III junctions (7%), involved the removal of short stretches of DNA (less than about 25 nucleotides) from the ends of the linear construct. The existence of this class has been noted in previous studies (51). A fourth pathway, represented by the class IV junctions (7%), involved more sizeable deletions, which presumably arose from breakage of the input DNA. The frequency of breakage, which can be calculated from the frequency of deletions as described previously (51), was one break per 7 kilobase pairs in these studies, supporting our previous estimate of one break per 5 to 15 kilobase pairs (51). These deletions and the class III deletions must have occurred before circularization, because in eight of nine sequenced examples, nucleotides were lost from only one end of the input DNA.

The flush junctions in this study could be explained by a mechanism involving blunt-end ligation following modification of the ends. Since mammalian cells join blunt ends very efficiently (51), it is possible that all nonhomologous junctions might arise by blunt-end ligation. Indeed, more than

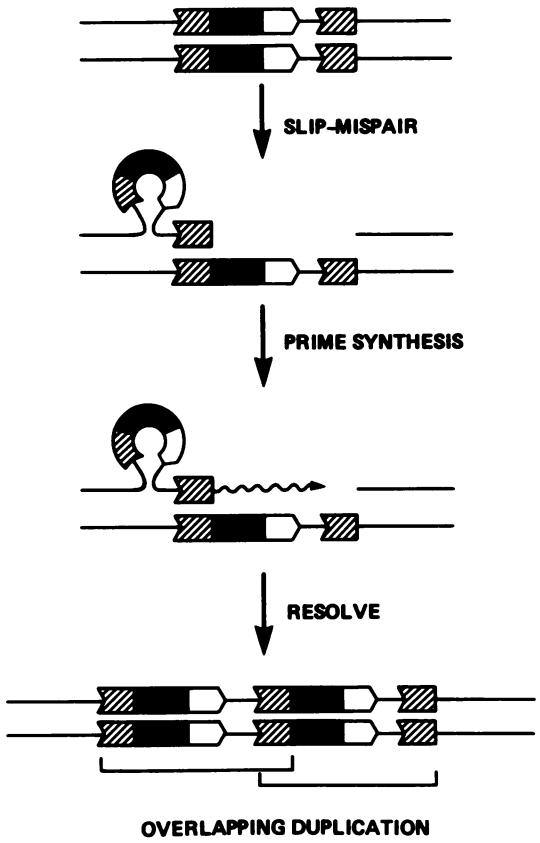

FIG. 9. A model for formation of overlapping duplications by intrastrand slipped mispairing. Symbols are as described in the legend to Fig. 8. Slipped mispairing within a duplex generates a structure that could prime DNA synthesis, if it is the 3′ end as shown, or that could be trapped by DNA synthesis, if it is the 5′ end. Appropriate resolution of the resulting structure by replication or repair would generate an overlapping duplication.

40% of nonhomologous junctions reported in the literature (see Fig. 6) display no homology at the junction, which is consistent with a blunt-end joining mechanism. However, the remaining junctions show one to a few nucleotides of homology. Do junctions with homology indicate that another mechanism in addition to blunt-end joining is used? This question is difficult to resolve, because even if all junctions arose by blunt-end joining, some would show homology due to chance identities in the flanking nucleotides.

To address this question, we examined 110 flush junctions, including 103 from the literature and the 7 sequenced class III and class IV flush junctions described in this report. (We excluded the class I and II junctions because they may be specific for the particular ends present in the substrate DNA.) These junctions were grouped according to the number of nucleotides of junctional homology present in each recombinant. In Fig. 6 the distribution of homologies at the junctions is compared with two distributions expected for blunt-end ligation: one assumes an unbiased base composition; the other assumes the base composition of the actual junctions. The actual distribution appears to be shifted slightly toward more nucleotides of junctional homology. Chi-square analysis verifies this impression, indicating that the actual distribution is significantly different from either expected distribution ($P < 0.01$).

By normalizing the expected distributions to the number of observed junctions with zero nucleotides of homology (Fig. 6), we estimate that about 20% of the observed junctions (in the range of two to five nucleotides of homology) contain more homology than expected from blunt-end ligation. Results of this statistical analysis suggest that short sequence homologies are used, in some cases, to link duplexes together. Homology could be used to facilitate direct ligations, as illustrated by the mechanism outlined on the left in Fig. 5 for the formation of class I junctions. However, a different role for homology is suggested by the unusual structures of duplication junctions, which are illustrated in Fig. 4 and 7. Duplication junctions share two common features. First, the inserted nucleotides are always identical to a sequence found close to the junction in one of the parental DNAs. Second, one to three nucleotides at the deletion endpoints in each recombinant are homologous to the nucleotides that define the beginning and end of the repeat unit. Since in most cases the duplicated segments consist of fairly long stretches of nucleotides, the probability that they are unrelated to the parental DNA is quite low.

These structural features of duplication junctions can be explained by the scheme outlined on the left in Fig. 8. Breakage of both parental duplexes provides ends which can unwind to expose single-stranded regions (Fig. 8). Pairing of the ends through terminal nucleotides defines one end of the duplicated region; slipped mispairing between homologous sequences within one duplex defines the other end of the duplicated region. This paired structure serves as a template-primer for repair DNA synthesis, leading to the formation of a heteroduplex structure containing a single-stranded loop. Replication of such a heteroduplex would generate one duplication junction and one flush junction. Alternatively, repair of the heteroduplex would generate either a duplication junction, if the loop strand were used as the template for repair, or a flush junction, if the nonloop strand were used as the template. Heteroduplexes containing single-stranded loops constructed in vitro and transfected into mammalian cells are repaired with very high efficiency; moreover, the loop strand serves as template for repair about 40% of the time (U. Weiss and J. H. Wilson, in preparation).

This analysis of duplication junctions suggests a more general model to account for junctional homology (on the right in Fig. 8); it differs from the one on the left by the absence of slipped mispairing. The proposal that terminal homologies of one to a few nucleotides can prime repair synthesis is a critical element of both models, and it is being tested experimentally. Although the model is not yet verified, several disparate observations are consistent with the proposed unwinding and paired priming events. (i) The upstream duplication in d11061 (Fig. 7), tandem duplications found in minisatellite regions of chromosomes (19), duplications that restore enhancer function (47, 53) and duplications encoding SV40 super T antigens (27) generally have overlapping duplications, which are consistent with intrastrand slipped mispairing followed by repair synthesis (Fig. 9). (ii) The structure of one pair of junctions generated by a reciprocal translocation suggests that the original break occurred by unwinding from staggered nicks separated by 106 base pairs (14). (iii) The linking of duplexes during rearrangements in the immune system in some cases may involve repair synthesis primed from the 3′ extensions added by terminal transferase (2). (iv) Finally, in another context we have proposed that homologous recombination in transfected DNA involves unwinding of DNA ends, exposing single strands which then initiate homologous pairing (52).

Nonhomologous recombination of foreign DNA into chromosomes in animal cells is 10^2- to 10^5-fold more frequent than homologous integration (13, 25, 43). For that reason it has proved difficult to target exogenous DNA into homologous chromosomal loci. To achieve efficient targeted integration, it will be necessary to inhibit nonhomologous recombination, to stimulate homologous integration, or both. At present we know that DNA molecules with ends are the preferred substrates in both kinds of recombination processes. Ends stimulate homologous recombination presumably by providing sites for exposure of single strands (24, 49, 52); ends stimulate nonhomologous recombination presumably by providing termini for direct ligation or pairing-primed repair synthesis. Once the molecular steps responsible for both processes are better defined, it may prove possible to design DNA molecules that significantly enhance the relative frequency of homologous integration.

ACKNOWLEDGMENTS

We thank Kathleen Marburger and De Dieu for expert technical assistance. We thank Ursula Weiss and Cho Yeung for valuable suggestions on the manuscript, and the Molecular Biology Group in the Department of Biochemistry at Baylor for helpful criticism. We are particularly grateful to Allen White for help with the statistical analysis.

This work was supported 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. D.B.R. is in the Medical Scientist Training Program and is supported by the Edward J. and Josephine Hudson Scholars Fund.

LITERATURE CITED

1. **Adams, J. M., S. Gerondakis, E. Webb, J. Mitchell, O. Bernard, and S. Cory.** 1982. Transcriptionally active DNA region that rearranges frequently in murine lymphoid tumors. Proc. Natl. Acad. Sci. USA **79**:6966–6970.
2. **Alt, F. W., and D. Baltimore.** 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. Proc. Natl. Acad. Sci. USA **79**:4118–4122.
3. **Anderson, R. A., S. Kato, and R. D. Camerini-Otero.** 1984. A pattern of partially homologous recombination in mouse L cells. Proc. Natl. Acad. Sci. USA **81**:206–210.
4. **Ashram, 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.
5. **Austerberry, C. F., C. D. Allis, and M.-C. Yao.** 1984. Specific DNA rearrangements in synchronously developing nuclei of Tetrahymena. Proc. Natl. Acad. Sci. USA **81**:7383–7387.
6. **Bernstein, L. B., S. M. Mount, and A. M. Weaver.** 1983. Pseudogenes for human small nuclear RNA U3 appear to arise by integration of self-primed reverse transcripts of the RNA into new chromosomal sites. Cell. **32**:461–472.
7. **Bullock, P., W. Forrester, and M. Botchan.** 1984. DNA sequence studies of simian virus 40 chromosomal excision and integration in rat cells. J. Mol. Biol. **174**:55–84.
8. **Calos, M. P., J. S. Lebkowski, and M. R. Botchan.** 1983. High mutation frequency in DNA transfected into mammalian cells. Proc. Natl. Acad. Sci. USA **80**:3015–3019.
9. **Desiderio, S. V., G. D. Yancopoulos, M. Paskind, E. Thomas, M. A. Boss, N. Landau, F. W. Alt, and D. Baltimore.** 1984. Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. Nature (London) **311**:752–755.
10. **DiMaio, D., and D. Nathans.** 1980. Cold-sensitive regulatory mutants of simian virus 40. J. Mol. Biol. **140**:129–142.
11. **Doerfler, W., W. Gahlmann, S. Stabel, R. Deuring, U. Lichtenberg, M. Schulz, D. Eick, and R. Leisten.** 1983. On the mechanism of recombination between adenoviral and cellular DNAs: the structure of junction sites. Curr. Top. Microbiol. Immunol. **109**:193–228.
12. **Dunnick, W., T. H. Rabbitts, and C. Milstein.** 1980. An immunoglobulin deletion mutant with implications for the heavy-chain switch and RNA splicing. Nature (London) **286**:669–675.
13. **Folger, K., K. Thomas, and M. R. Capecchi.** 1985. Analysis of homologous recombination in cultured mammalian cells. Cold Spring Harbor Symp. Quant. Biol. **49**:123–138.
14. **Gerondakis, S., S. Cory, and J. M. Adams.** 1984. Translocation of the *myc* cellular oncogene to the immunoglobulin heavy chain locus in murine plasmacytomas is an imprecise reciprocal exchange. Cell **36**:973–982.
15. **Hedrick, S. M., D. I. Cohen, E. A. Nielsen, and M. M. Davis.** 1984. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature (London) **308**:149–153.
16. **Hogan, A., and E. A. Faust.** 1984. Short direct repeats mediate spontaneous high-frequency deletions in DNA of minute virus of mice. Mol. Cell. Biol. **4**:2239–2242.
17. **Honjo, T.** 1983. Immunoglobulin genes. Annu. Rev. Immunol. **1**:499–528.
18. **Jeang, K.-T., and S. D. Hayward.** 1983. Organization of the Epstein-Barr virus DNA molecule III. Location of the P3HR deletion junction and characterization of the *NotI* repeat units that form part of the template for an abundant 12-O-tetradacanoylphorbol-13-acetate-induced mRNA transcript. J. Virol. **48**:135–148.
19. **Jeffreys, A. J., V. Wilson and S. L. Thein.** 1985. Hypervariable 'minisatellite' regions in human DNA. Nature (London) **314**:67–73.
20. **Johnson, A. D., A. Barkan, and J. E. Mertz.** 1982. Nucleotide sequence analysis of the recombinant joints in 16 naturally arising deletion mutants of simian virus 40. Virology **12**: 464–469.
21. **Klobutcher, L. A., C. L. Jahn, and D. M. Prescott.** 1984. Internal sequences are eliminated from genes during macronuclear development in the ciliated protozoan Oxytricha nova. Cell **36**:1045–1055.
22. **Kopchick, J. J., and D. W. Stacey.** 1984. Differences in intracellular DNA ligation after microinjection and transfection. Mol. Cell. Biol. **4**:240–246.
23. **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.
24. **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.
25. **Lin, F.-L., K. Sperle, and N. Sternberg.** 1985. Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. Proc. Natl. Acad. Sci. USA **82**:1391–1395.
26. **Ling, L. E., M. M. Manos, and Y. Gluzman.** 1982. Sequence of the junction in adenovirus 2-SV40 hybrids: example of illegitimate recombination. Nucleic. Acids Res. **10**:8099–8112.
27. **Lovett, M., C. E. Clayton, D. Murphy, P. W. J. Rigby, A. E. Smith, and F. Chaudry.** 1982. Structure and synthesis of a simian virus 40 super T-antigen. J. Virol. **44**:963–973.
28. **Malissen, M., K. Minard, S. Mjolsness, M. Kronenberg, J. Goverman, T. Hunkapiller, M. B. Prystowsky, Y. Yoshikai, F. Fitch, T. W. Mak, and L. Hood.** 1984. Mouse T cell antigen receptor: structure and organization of constant and joining gene segments encoding the beta polypeptide. Cell **37**: 1101–1110.
29. **Maruyama, K., and K. Oda.** 1984. Two types of deletion within integrated viral sequences mediate reversion of simian virus 40-transformed mouse cells. J. Virol. **49**:479–489.
30. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavage. Methods Enzymol. **65**:499–560.
31. **McCutchan, T., M. Singer, and M. Rosenberg.** 1979. Structure of simian virus 40 recombinants that contain both host and viral DNA sequences. J. Biol. Chem. **254**:3592–3597.
32. **Miller, C. K., and H. M. Temin.** 1983. High efficiency ligation and recombination of DNA fragments by vertebrate cells.

Science **220**:606–609.
33. **Neuberger, M. S., and F. Calabi.** 1983. Reciprocal chromosome translocation between c-*myc* and immunoglobulin gamma 2b genes. Nature (London) **305**:240–243.
34. **Pickup, D. J., B. S. Ink, B. L. Parsons, W. Hu, and W. K. Joklik.** 1984. Spontaneous deletions and duplications of sequences in the genome of cowpox virus. Proc. Natl. Acad. Sci. USA **81**:6817–6821.
35. **Pipas, J. M., K. W. 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.
36. **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.
37. **Razzaque, A., H. Mizusawa, and M. Seidman.** 1983. Rearrangement and mutagenesis of a shuttle vector plasmid after passage in mammalian cells. Proc. Natl. Acad. Sci. USA **80**:3010–3014.
38. **Roth, D. B., and J. H. Wilson.** 1985. Relative rates of homologous and nonhomologous recombination in transfected DNA. Proc. Natl. Acad. Sci. USA **82**:3355–3359.
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. **Seif, I., G. Khoury, and R. Dhar.** 1980. A rapid enzymatic DNA sequencing technique: determination of sequence alterations in early simian virus 40 temperature sensitive and deletion mutations. Nucleic. Acids Res. **8**:2225–2240.
41. **Shapiro, J. A.** 1983. Mobile genetic elements. Academic Press, Inc., New York.
42. **Sheflin, L., A. Celeste, and M. Woodworth-Gutai.** 1983. Recombination in simian virus 40-infected cells. J. Biol. Chem. **258**:14315–14321.
43. **Smithies, O., M. A. Koralewski, K.-Y. Song, and R. S. Kucherlapati.** 1985. Homologous recombination with DNA introduced into mammalian cells. Cold Spring Harbor Symp. Quant. Biol. **49**:161–170.
44. **Stark, G. R., and G. M. Wahl.** 1984. Gene amplification. Annu. Rev. Biochem. **53**:447–491.
45. **Stringer, J. R.** 1982. DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. Nature (London) **296**:363–366.
46. **Subramani, S., and P. Berg.** 1983. Homologous and nonhomologous recombination in monkey cells. Mol. Cell. Biol. **3**:1040–1052.
47. **Swimmer, C., and T. Shenk.** 1984. A viable simian virus 40 variant that carries a newly generated sequence reiteration in place of the normal duplicated enhancer element. Proc. Natl. Acad. Sci. USA **81**:6652–6656.
48. **Sylla, B. S., D. Huberdeau, D. Bourgaux-Ramoisy, and P. Bourgaux.** 1984. Site-specific excision of integrated polyoma DNA. Cell **37**:661–667.
49. **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.
50. **Wakamiya, T., T. McCutchan, M. Rosenberg, and M. Singer.** 1979. Structure of simian virus 40 recombinants that contain both host and viral DNA sequences. J. Biol. Chem. **254**: 3584–3591.
51. **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.
52. **Wake, C. T., F. Vernaleone, and J. H. Wilson.** 1985. Topological requirements for homologous recombination among DNA molecules transfected into mammalian cells, Mol. Cell. Biol. **5**:2080–2089.
53. **Weber, F., J. deVilliers, and W. Schaffner.** 1984. An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell **36**:983–992.
54. **Westphal, H.** 1981. Genomic arrangement of an adenovirus-simian virus 40 hybrid virus, $Ad2^{+}ND^{4}_{del}$. J. Virol. **40**:526–532.
55. **Wilson, J. H.** 1977. Genetic analysis of host range mutant viruses suggests an uncoating defect in simian virus 40-resistant monkey cells. Proc. Natl. Acad. Sci. USA **74**:3503–3507.
56. **Wilson, J. H.** 1978. Interference in SV40 DNA infections: a possible basis for cellular competence. Virology **91**:380–388.
57. **Wilson, J. H., P. Berget, and J. Pipas.** 1982. Somatic cells efficiently join unrelated DNA segments end-to-end. Mol. Cell. Biol. **2**:1258–1269.
58. **Woodworth-Gutai, M.** 1981. Recombination in SV40-infected cells: nucleotide sequences at viral-viral recombinant joints in naturally arising variants. Virology **109**:344–352.
59. **Woodworth-Gutai, M.** 1981. Recombination in SV40-infected cells: viral DNA sequences at sites of circularization of transfecting linear DNA. Virology **109**:353–365.
60. **Woodworth-Gutai, M. 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.
61. **Yanagi, Y., Y. Yoshikai, K. Leggett, S. P. Clark, I. Aleksander, and T. W. Mak.** 1984. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. Nature (London) **308**:145–149.
62. **Zain, B. X., and R. J. Roberts.** 1978. Characterization and sequence analysis of a recombination site in the hybrid virus ad2+ND1. J. Mol. Biol. **120**:12–31.