MOLECULAR AND CELLULAR BIOLOGY, Dec. 1986, p. 4295–4304
0270-7306/86/124295-10$02.00/0

Vol. 6, No. 12

Nonhomologous Recombination in Mammalian Cells: Role for Short Sequence Homologies in the Joining Reaction

DAVID B. ROTH AND JOHN H. WILSON*

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

Received 13 June 1986/Accepted 12 August 1986

Although DNA breakage and reunion in nonhomologous recombination are poorly understood, previous work suggests that short sequence homologies may play a role in the end-joining step in mammalian cells. To study the mechanism of end joining in more detail, we inserted a polylinker into the simian virus 40 T-antigen intron, cleaved the polylinker with different pairs of restriction enzymes, and transfected the resulting linear molecules into monkey cells. Analysis of 199 independent junctional sequences from seven constructs with different mismatched ends indicates that single-stranded extensions are relatively stable in monkey cells and that the terminal few nucleotides are critical for cell-mediated end joining. Furthermore, these studies define three mechanisms for end joining: single-strand, template-directed, and postrepair ligations. The latter two mechanisms depend on homologous pairing of one to six complementary bases to position the junction. All three mechanisms operate with similar overall efficiencies. The relevance of this work to targeted integration in mammalian cells is discussed.

DNA rearrangements that require little or no sequence homology are quite rare in bacteria and yeasts but occur frequently in mammalian cells. For example, targeted integration of exogenous DNA at its homologous chromosomal location in mammalian cells is masked by a 1,000-fold-higher frequency of random integration (20, 33, 34, 38). By contrast, nontargeted integration events in bacteria and yeasts are difficult to detect (11, 26, 27). Sequence analysis of junctions created by a variety of DNA rearrangements reveals minimal homology around the site where the recombining duplexes are joined (for a review, see reference 30). For that reason, these DNA rearrangements are often called nonhomologous recombination events.

Nonhomologous recombination in transfected DNA is thought to occur by a two-step process involving breakage of DNA molecules followed by end joining (43). Similar break-join mechanisms have been proposed for chromosome translocations (9), immunoglobulin and T-cell receptor gene rearrangements (2, 10), chromosome rearrangements in maize (23), the excision of integrated viral genomes from the chromosome (4), and the chromosomal integration of transfected (29) and microinjected (3, 8) DNA molecules. In these examples, the mechanisms proposed for breakage are quite different, ranging from site-specific cleavage (2, 14), to topoisomerase I cleavage (4), to damage-induced breakage (5, 28, 40), to mechanical breakage (23). In contrast, the mechanism for joining broken duplexes may be the same: the simple linking of free DNA ends.

The presence of short, 1- to 5-nucleotide homologies at roughly 60% of nonhomologous junctions has led several investigators to propose that these homologies play a role in nonhomologous recombination (1, 2, 4, 7, 13, 21, 30, 32, 36). However, because these homologies are so short, it is difficult to distinguish between mechanistic relevance and chance occurrence. Nevertheless, a statistical analysis of more than 100 junction sequences indicates that short homologies are present somewhat more often than expected by chance, suggesting that these homologies might direct the end-joining reaction (30). These studies lend circumstantial support to the idea that ends are joined by two different types of mechamism: a more common, homology-independent process and a less common, homology-dependent process.

To probe the mechanism of the end-joining reaction in more detail, we constructed a series of linear simian virus 40 (SV40) genomes carrying a variety of mismatched ends. These genomes were linearized in the intron of the SV40 T-antigen gene, so that T antigen could not be expressed until after circularization. Since T antigen is necessary for expression of the other viral genes (39), this experimental design ensured that the enzymatic activities responsible for end joining were provided by the host cell. In addition, because the intron sequences are not essential for lytic infection, all end-joining products can be recovered as viable viruses.

We transfected seven linear genomes with different pairs of mismatched ends into monkey cells and isolated individual plaques. The viral genomes from 199 plaques were analyzed by nucleotide sequencing. Comparison of the junction sequences with the structures of the input ends lends strong support to both homology-independent and homology-dependent end joining and suggests the existence of two general classes of homology-dependent joining reactions. The results of this study allow us to propose a set of rules for these joining processes. An understanding of these rules may aid future attempts at targeting DNA to its homologous chromosomal location.

MATERIALS AND METHODS

Cells, viruses, and DNAs. The CV1 monkey kidney cell line was grown according to standard procedures (43). The SV40 mutants used in this work were derived from the Rh911a wild-type strain. DNA transfections were carried out as described before (42), with DEAE-dextran and 0.005 to 0.01 ng of SV40 DNA per 60-mm dish.

Insertion of a polylinker into the intron. The starting material for construction was the SV40 mutant su1901, which has been described before (40). In this mutant, a region of the intron of the large T-antigen gene has been replaced by pBR322 sequences that contain a unique *Fnu*DII

* Corresponding author.

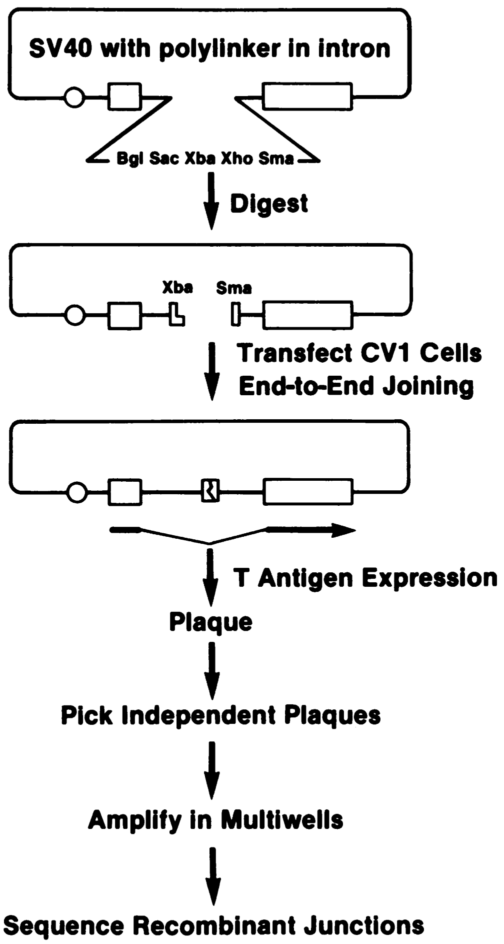


FIG. 1. Production of substrates for end joining. In the diagram of the SV40 genome, the boxes represent the exons encoding T antigen, and the circle represents the origin of replication. The SV40 strain su1910 was constructed by inserting a polylinker into the unique *Bgl*II site in the intron of the T-antigen gene, as described in the text. The 5′ end of the polylinker is separated from the 5′ exon by 90 nucleotides; the 3′ end of the polylinker is 33 nucleotides upstream of the 3′ exon. The polylinker contains the recognition sequences for the restriction enzymes indicated, all of which occur only once in the genome of su1910. The *Xma*I site used for some constructs is not shown because *Xma*I cleaves at the *Sma*I recognition sequence. Substrates for transfection were constructed by digesting circular su1910 DNA with various pairs of enzymes that cleave within the polylinker. The resulting linear DNA molecules were transfected into CV1 monkey kidney cells. (The small boxes on the ends of the linear genome represent the structures of the ends: *Xba*I leaves a 5′ overhang; the *Sma*I end is blunt.) Since the T-antigen gene was interrupted in these genomes, the expression of T antigen, and hence plaque formation, was dependent upon cell-mediated end joining. At 10 to 15 days after transfection, one plaque was picked from each plate; virus suspensions from individual plaques were used to infect monolayers of CV1 cells in 24-well plates. At 6 to 10 days after infection, viral DNA was isolated from each well, and the recombinant junctions were sequenced as described in the text.

restriction site. We removed 220 base pairs from the intron of the T-antigen gene by digesting su1901 DNA with *Fnu*DII and *Nde*I. The *Nde*I end was filled in by using the Klenow fragment of DNA polymerase I, and *Bgl*II linkers were added by using T4 DNA ligase. The resulting mutant, su1903, had wild-type infectivity (data not shown).

We designed a 45-nucleotide polylinker containing several restriction sites that do not occur in SV40 DNA and chemically synthesized both DNA strands. Annealing the single strands produced four nucleotide terminal extensions that were complementary to ends produced by digestion with *Bgl*II, allowing the double-stranded DNA to be ligated directly into the *Bgl*II site of su1903. The sequences present at the ends of the polylinker were arranged so that this ligation regenerated a *Bgl*II site at only one end of the insert. The nucleotide sequence of the inserted polylinker was verified by labeling at the unique *Bgl*II site and sequencing by the method of Maxam and Gilbert (22). This SV40 mutant was designated su1910.

Production of linear substrate DNAs. Linear SV40 DNA molecules were produced by digesting su1910 viral DNA with restriction enzymes that cleaved within the polylinker. Substrates containing mismatched ends were produced by digestion with a pair of enzymes. Since the restriction sites in the polylinker are separated by only a few nucleotides, it is impossible to determine whether the DNA has been cut with the second enzyme by agarose gel electrophoresis. Therefore, each enzyme was tested by digesting a separate portion of DNA; complete digestion was verified by agarose gel electrophoresis. The digested DNA samples were purified by phenol extraction and ethanol precipitation and then redigested with the second enzyme. The completeness of the double digestion was also verified by sequence analysis of the progeny genomes, since undigested or singly digested molecules should contain a complete polylinker sequence. Overall, the frequency of genomes containing the complete polylinker was less than 10%; these molecules were excluded from the analysis.

To ensure that the end modifications we observed in the recombinant genomes were generated in vivo, rather than in vitro, we tested the restricted ends in two ways. First, all enzymes were tested for in vitro end modification activity by linearizing su1910 DNA and then incubating the linear molecules with T4 DNA ligase under conditions that favor circularization. Digestion with each enzyme produced linear molecules that could be converted to at least 90% circular species as judged by agarose gel electrophoresis, and virtually all of the circular forms could be converted to linear forms by redigestion with the appropriate enzymes. Second, *Sma*I, *Sac*I, and *Xba*I were tested by digesting portions of su1910 DNA with each enzyme, transfecting the singly digested DNAs into CV1 cells, and determining the junction sequences in a number of progeny genomes. For *Sma*I linears, 13 out of 14 progeny genomes contained the complete polylinker; *Sac*I and *Xba*I gave similar results (15/17 and 8/10, respectively). These data are consistent with the degree of end modification that we have previously reported (6, 30, 40); they demonstrate that the vast majority of end modification events observed in our experiments was not generated in vitro during the production of the substrates.

*Sac*I and *Fnu*DII were from New England BioLabs, Inc.; all other restriction enzymes were from Boehringer Mannheim Biochemicals. The enzyme *Xma*I recognizes the same sequence as *Sma*I, but generates different termini; the *Xma*I site is not shown in Fig. 1.

Maxiwell preparation of DNA. To prepare viral DNA for sequencing, confluent monolayers of CV1 cells in 24-well plates (Becton Dickinson Labware) were infected with picked plaque suspensions. Cells were harvested 6 to 10 days after infection when 70 to 80% of the cells exhibited cytopathic changes. Viral DNA was extracted as described previously (12), precipitated with ethanol, resuspended, and treated for 1 h with RNase A (100 μg/ml). Lithium acetate (pH 7.5) was then added to a final concentration of 1 mol/liter; proteinase K (200 μg/ml) was added, and the samples were incubated for 24 to 72 h at room temperature. After incubation, the DNA was recovered by phenol extraction and ethanol precipitation. All of the DNA from each maxiwell was used in a single sequencing reaction. Proteinase K and RNase A were from Boehringer Mannheim.

DNA sequencing. Double-stranded viral DNAs were subjected to sequence analysis by the method of Zagursky et al. (44) with the following modifications. Each reaction contained 200 U of Maloney leukemia virus reverse transcriptase (Bethesda Research Laboratories) and 40 μCi of [^{35}S]dATP (Amersham Corp.); reactions were carried out for 15 min at 42°C. Sequencing primers were 17-mers that hybridized within the coding regions of the T-antigen gene, so that all viable mutants could be sequenced.

RESULTS

Experimental design. Our strategy for analyzing the mechanisms of end joining was to survey many end-joining junctions derived from several different substrate molecules with defined ends. Analysis of the nucleotide sequences of a large number of junctions allowed us to deduce rules for the joining process.

For these studies, we used a viable mutant of SV40, which contains a polylinker in the intron of the T-antigen gene. Seven different linear SV40 genomes with mismatched ends were generated by cleaving the polylinker with pairs of restriction enzymes (Fig. 1). The design of these molecules provides a selection for cell-mediated end-joining events, since the genome must be circularized to initiate lytic infection and plaque formation.

Several experimental procedures are relevant to our goal of studying individual, intramolecular recombination events. (i) Linear DNA molecules were introduced into CV1 monkey cells by transfection with DEAE-dextran, which delivers molecules individually rather than in aggregates as with $CaPO_4$. (ii) Extremely small quantities of DNA were used (5 pg/60-mm dish) to ensure that single DNA molecules initiated plaque formation. Even at 200 times this level (1 ng per dish), the fraction of mixedly infected cells after DEAE-dextran transfection is less than 5% of all infected cells (M. Seidman, personal communication [cited in reference 41]). (iii) A single plaque was picked from each plate of cells to ensure that each junction arose in an independent recombination event. Viral DNA was prepared from each plaque by the maxiwell method, and the recombinant junctions were sequenced directly (see above).

TABLE 1. Relative infectivities of substrates with mismatched ends

DNA end	Left[a] terminus	Right[a] terminus	Relative[b] infectivity (%)
*Xba*I-*Xba*I	5′ extension	5′ extension	95 ± 10
*Sma*I-*Sma*I	Blunt	Blunt	105 ± 10
*Xba*I-*Sma*I	5′ extension	Blunt	90 ± 30
*Sac*I-*Sma*I	3′ extension	Blunt	220 ± 90
*Bgl*II-*Sac*I	5′ extension	3′ extension	70 ± 20
*Sac*I-*Xma*I	3′ extension	5′ extension	140 ± 80
*Bgl*II-*Xho*I	5′ extension	5′ extension	100 ± 10
*Bgl*II-*Xba*I	5′ extension	5′ extension	90 ± 50
*Xba*I-*Xma*I	5′ extension	5′ extension	140 ± 60

[a] Left and right refer to the orientation shown in Fig. 1.

[b] Infectivity values represent the means ± standard deviations obtained from three separate transfections. The average infectivities of the *Sma*I and *Xba*I linears (1,370 PFU/ng) were designated 100% and used to normalize the results.

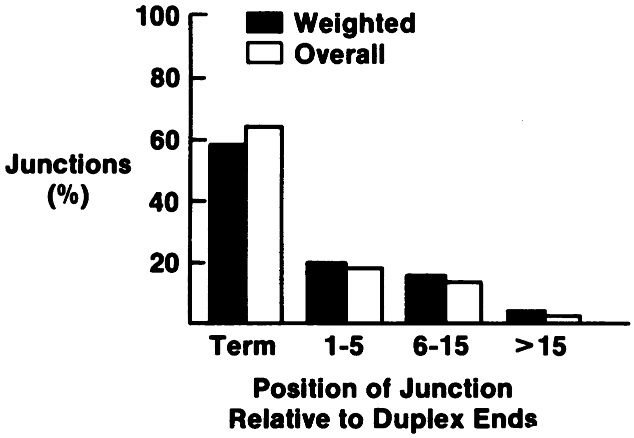


FIG. 2. Distribution of junctions relative to the ends of the input DNA. Data in Fig. 3 and 4 were used to calculate the average distribution of the junctions relative to the ends of the substrate molecules. "Term" indicates that the junctions involved both termini, where terminus includes all nucleotides of a single-stranded extension and the first nucleotide of the duplex portion of the ends. Numbers on the abscissa represent distance of the junction from the duplex portion of the termini (in nucleotides). Open bars represent the average distribution calculated as a percentage of the total number of junctions. This distribution is biased in favor of substrates with a greater number of sequenced junctions. To remove this bias, the junctions from each substrate were treated separately and expressed as percentages of the total number of junction sequences obtained for that substrate. Individual percentages were then averaged to give the weighted distribution (filled bars). In this treatment, the data from each substrate were weighted equally, regardless of the number of junction sequences obtained.

Several mismatched ends join with the same efficiency. Previous experiments demonstrated that sticky, blunt, and mismatched ends were joined by monkey cells with an efficiency approaching 100% (40). However, in that study, only one pair of mismatched ends was tested. To extend those measurements, we determined the infectivity of each of the seven different linear substrate molecules by plaque assay (Table 1). The infectivities of molecules with mismatched ends did not differ significantly from each other or from the average infectivity of sticky (*Xba*I) and blunt (*Sma*I) ends, which was defined as 100% (Table 1). These results indicate that mammalian cells can join a variety of mismatched ends with the same high efficiency as sticky or blunt ends.

Cells use the terminal few nucleotides in the joining reaction. We determined the nucleotide sequences of 199 end-joining junctions derived from the seven linear genomes with mismatched ends. Twelve of these junctions (6%) contained extra nucleotides inserted at the junction; these junctions will not be considered further. Here we focus on the 187 junctions that did not contain insertions.

Examination of the nucleotide sequences of the junctions revealed that the vast majority occurred very close to the original ends of the transfected molecules. Of the 187 junctions, 97% occurred within 15 nucleotides of the original termini, and 83% of the junctions occurred within 5 nucleotides of the duplex portion of the input termini. The overall distribution of the position of the 187 junctions relative to the ends of the input molecules is summarized in Fig. 2. These

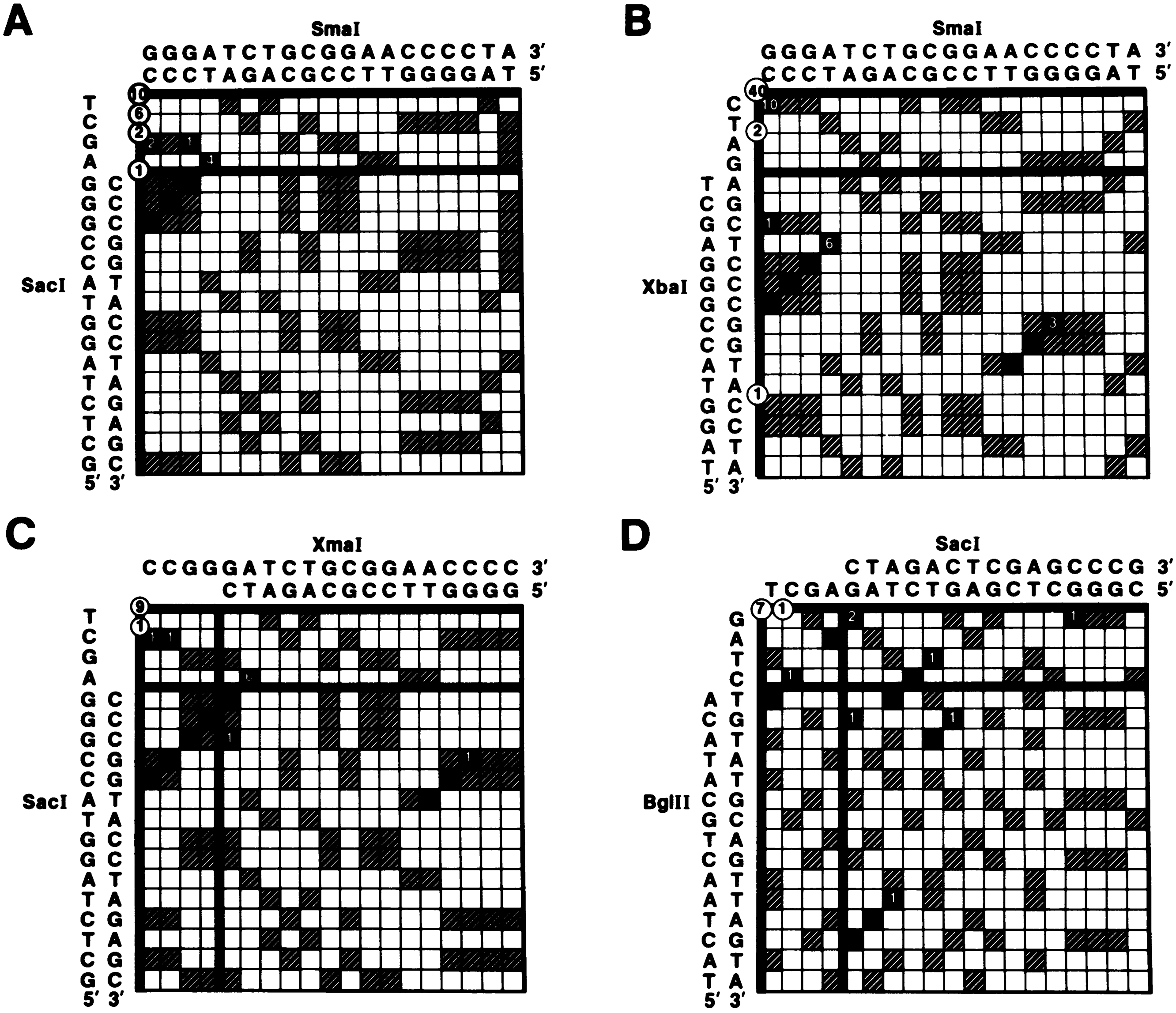


FIG. 3. Sequences of recombinant junctions from substrates that can join by single-strand ligation. (A) *Sac*I-*Sma*I; (B) *Xba*I-*Sma*I; (C) *Sac*I-*Xma*I; (D) *Bgl*II-*Sac*I. Nucleotide sequences of the ends of the indicated substrates are shown along the axes. Strands drawn at the outside edges are written 5′ to 3′ (from bottom to top and from left to right); the inner strands are 3′ to 5′. Grid lines represent the phosphodiester bonds between the bases; junctions are indicated by circles or squares containing numbers. Junctions that do not exhibit homology are represented by circles drawn at intersections of the grid lines. Nucleotide sequences of these junctions can be determined from the diagram by reading the vertical strand from bottom to top until the horizontal line indicated by the circle is reached, then by following the vertical line from the circle to the corresponding horizontal strand, and finally by continuing to the right with the nucleotides of the horizontal strand. For example, consider the 10 junctions denoted by the circle at the upper left corner of panel A. The nucleotide sequence of these junctions reads (starting with the first nucleotide of the single-stranded *Sac*I extension and reading the outside strands) 5′-. . .AGCT′GGGATCT. . .-3′, with the apostrophe indicating the position of the junction. Junctions with homology are denoted by squares drawn between the lines, since the homology makes it impossible to determine precisely which bonds were joined. Junctions with more than 1 nucleotide of homology appear as a series of filled boxes connected along a diagonal. The nucleotide sequences of junctions with homology can be read from the diagrams as described above, except that the boxed nucleotides are present only once in the junction. As an example, consider the *Xba*I-*Sma*I substrate shown in panel B. The 10 junctions represented by a filled box in the upper left corner of the matrix exhibit a 1-nucleotide homology precisely at the site where the input ends were joined. At this position, both parental sequences contain the same nucleotide. The nucleotide sequence of this junction is (starting with the first single-stranded nucleotide of the *Xba*I extension and reading the inner strands) 3′-. . .GATCCCTAGA. . .-5′. The 1-nucleotide homology at the junction is underlined. Note that the homology leads to a 1-nucleotide ambiguity in the position of the junction with respect to the ends of the parental DNA. Such ambiguities are the hallmark of junctions that exhibit homology. To indicate this uncertainty, the boxes are drawn between the lines, which represent the phosphodiester bonds. Hatched boxes represent positions where homologies exist, but no junctions were found. These potential homologies have been included to allow a visual assessment of whether junctions form randomly or are directed by homology. Thus, a hatched box was placed at every position in the grid where the same nucleotide exists in the corresponding strands (both inner strands or both outer strands) from each end. Homologies of more than one nucleotide form a diagonal that runs upward and to the right; diagonals running in the opposite direction simply denote a series of one nucleotide homologies. One of the sequences from panel D is not shown because the junction occurred more than 15 nucleotides from one end of the substrate.

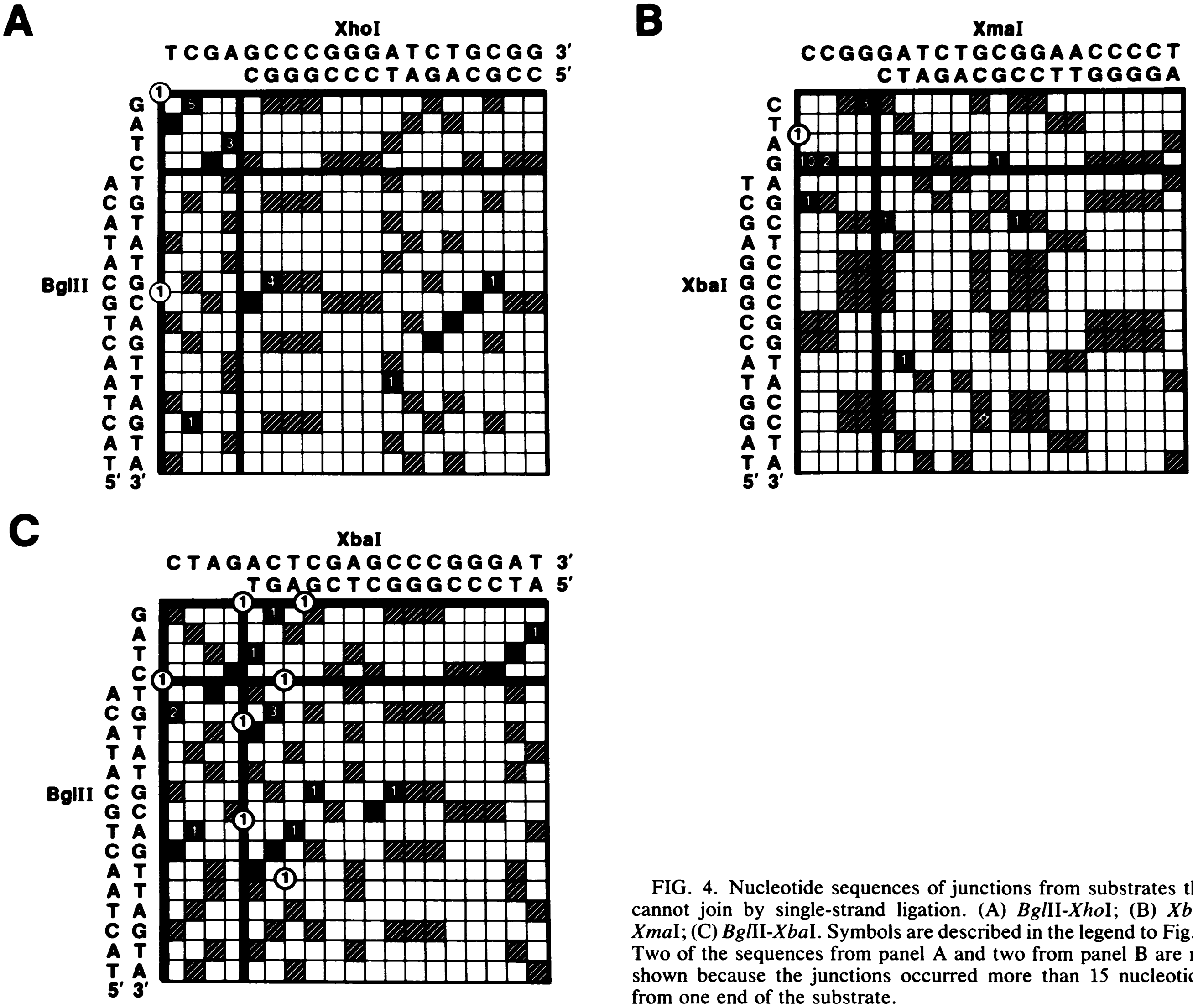


FIG. 4. Nucleotide sequences of junctions from substrates that cannot join by single-strand ligation. (A) *Bgl*II-*Xho*I; (B) *Xba*I-*Xma*I; (C) *Bgl*II-*Xba*I. Symbols are described in the legend to Fig. 3. Two of the sequences from panel A and two from panel B are not shown because the junctions occurred more than 15 nucleotides from one end of the substrate.

results emphasize that the joining reaction was confined to the immediate neighborhood of the input termini. These results also indicate that the mechanism of end joining does not require extensive modification of the input ends.

Nucleotide sequences of the junctions. The nucleotide sequences of the 187 junctions are shown in Fig. 3 and 4. In Fig. 3, the sequences of the junctions from the *Sac*I-*Sma*I, *Xba*I-*Sma*I, *Sac*I-*Xma*I, and *Bgl*II-*Sac*I linear forms are shown. These substrates were grouped together because their ends share a common feature: these pairs of mismatched ends can be joined by single-strand ligation involving protruding or blunt ends (discussed below). In contrast, the substrates represented in Fig. 4 (*Bgl*II-*Xho*I, *Xba*I-*Xma*I, and *Bgl*II-*Xba*I) contain pairs of 5′ extensions that cannot join by single-strand ligation, since the protruding single strands have the opposite chemical polarity. (For directions on how to read the sequences of the junctions from Fig. 3 and 4, see the legend to Fig. 3.) In these diagrams, the ends of the substrate molecule converge at the upper left corner. The sequences at the side and top, respectively, represent the left and right ends of the genome as it is shown in Fig. 1. The grid lines represent the phosphodiester bonds between the bases; each heavy line indicates the end of a single strand. The positions of the junctions are indicated by open circles and filled boxes; circles denote junctions with no homology, and boxes denote junctions with homology. The numbers inside the symbols indicate the number of occurrences of each junction. The hatched boxes indicate positions of potential homology in the matrix (regions where base pairing interactions could occur between the two ends).

Molecules that join by single-strand ligation. The most common junctions in Fig. 3 (52%) retained all the nucleotides in the parent molecule; that is, the linear molecules were joined at their protruding ends. For the molecules cleaved with *Sac*I, which leaves a 3′ protrusion that cannot be filled in by any known DNA polymerase (16), the joining event must have involved ligation of a single strand. For molecules with a 5′ protrusion, which could be filled in by a repair polymerase, it is not clear whether end joining involved a single strand or a filled-in blunt end. However, results obtained with other substrates suggest that filling-in of 5′ protrusions is relatively rare (see below). Thus, both 5′ protrusions and 3′ protrusions may join preferentially by single-strand ligation.

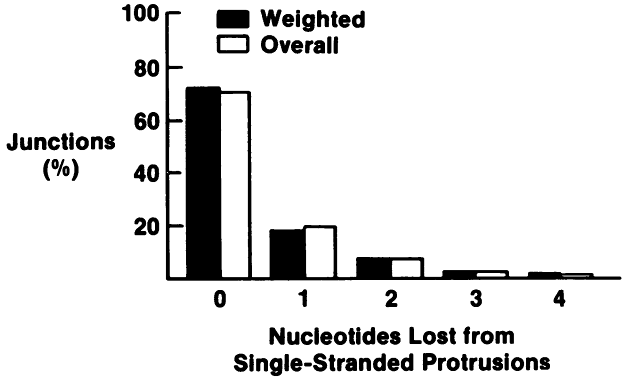


FIG. 5. Stability of short single-stranded extensions. Junction sequences represented in Fig. 3 were used to calculate the distribution of junctions that lost 0, 1, 2, 3, and all 4 nucleotides from single-stranded extensions. Open bars represent the average of all junctions shown in Fig. 3, and the filled bars give the weighted average, which was calculated as described in the legend to Fig. 2.

Although ligation of protruding or blunt strands was frequent, ligation involving a recessed strand was rare; there is only one example in all the junctions shown in Fig. 3 (Fig. 3A). The common observation that complementary, sticky ends are readily ligated in mammalian cells indicates that recessed ends can be joined if there is homology to direct the ligation (Fig. 4B). In the absence of homology, however, the cellular ligation machinery apparently prefers a pair of blunt or protruding strands to bring about end joining.

The results in Fig. 3 provide information about the stability of single-stranded protrusions. Although removal of an entire single-stranded extension to create a blunt end was uncommon, removal of some nucleotides was occasionally observed. Overall, 23% of the junctions in Fig. 3 involved loss of one or more nucleotides from a single-stranded protrusion. The distribution of junctions missing 0 to 4 nucleotides from one end is shown in Fig. 5. This distribution is consistent with progressive removal of nucleotides from the end. Removal of nucleotides from both 5′ and 3′ protrusions suggests the existence of both 5′ and 3′ exonucleases (or a single-strand-specific endonuclease). However, since more than 50% of the junctions did not lose any nucleotides from either end, the rate of removal is much less than the rate of end joining.

Molecules that join by homology-dependent processes. The junctions derived from molecules with two 5′ protrusions are shown in Fig. 4. In sharp contrast to the results shown in Fig. 3, only 1 of 61 junctions derived from molecules with mismatched 5′ protrusions retained all the nucleotides present in the transfected molecule (Fig. 4A). Since, in principle, 5′ protrusions could be filled in and then joined, the rarity of this class of junction was surprising. This result suggests that 5′ protrusions are filled in infrequently in CV1 cells and supports the conclusion that most of the junctions shown in Fig. 3B actually arose by single-strand ligation, rather than by filling in ends followed by blunt-end ligation.

These results also support the notion that recessed ends are rarely ligated in the absence of homology. In Fig. 4A and C, only 2 of 40 junctions (5%) involved ligation of a recessed strand to a protruding strand. However, in the presence of one nucleotide of homology (Fig. 4B), 13 of 21 (62%) junctions involved ligation to the recessed strand. Thus, ligation at recessed ends is highly dependent on the presence of homology.

The majority (82%) of the sequences displayed in Fig. 4 show homology at the junction. These results suggest that the principal mechanism for joining ends that cannot join by single-strand ligation involves the use of short homologies. As discussed below, several features of the data shown in Fig. 3 and 4 argue strongly that the observed homology is not due to chance but is actually used in the joining process.

DISCUSSION

In this study we used linear SV40 genomes with mismatched ends to examine the mechanisms of end joining in transfected DNA. We transfected a series of seven SV40 genomes containing different pairs of mismatched termini into monkey cells and determined the nucleotide sequences of 199 recombinant junctions derived from these genomes. In 187 of these junctions (94%), one parental sequence was joined directly to another without the addition of extra nucleotides. Analysis of these junctions defines three general mechanisms for joining DNA molecules (see below): homology-independent ligation of protruding strands (single-strand ligation), homology-dependent ligation of recessed strands (template-directed ligation), and a collection of homology-dependent joining processes that are directed by the pairing of 1 to 6 nucleotides of homology (postrepair ligation). The preferred mechanism of joining depends on the structure of the input ends. However, regardless of the nature of the input ends, the infectivities of all these genomes were essentially the same. Thus, the efficiencies of end joining by these three mechanisms are similar, even though their relative rates may be different.

Stability of single-stranded extensions. This study suggests that the ends of transfected DNA molecules remain relatively intact, since 182 of the 187 junctions (97%) occurred within 15 nucleotides of the input termini (Fig. 2). In addition, the results implicate the 5′ and 3′ ends of single strands as reactive components in the mechanism of joining, since 91% of the junctions involved the end of a strand. Short single-stranded extensions also are relatively stable, since roughly 60% of the junctions involved the single-stranded extensions or the blunt ends of both termini. In particular, the removal of a single-stranded extension to create a blunt end is quite rare. This process would show up in our analysis as homology-independent ligation of recessed ends; in the absence of homology (Fig. 3A to D and 4A and C), such ligation events were observed in only 3 of 160 junctions (Fig. 3A and 4C). As shown in Fig. 5, removal of 1, 2, or 3 nucleotides from 5′ and 3′ single-stranded extensions was observed in about 20% of the junctions in Fig. 3, suggesting the existence of 5′ and 3′ exonucleases (or a single-strand-specific endonuclease). However, the distribution of the loss of nucleotides from single-stranded extensions indicates that the rate of joining is greater than the rate of exonucleolytic removal, since more than 50% of the junctions did not lose any nucleotides from either end.

Similarly, the filling-in of recessed 3′ ends to create a blunt end is also rare. This process would show up in our analysis as retention of all nucleotides from both ends in substrates with pairs of 5′ extensions (Fig. 4); such ligation events were observed in only 1 of 61 junctions (Fig. 4A). Thus, the creation of blunt ends by filling in or chewing off short single-stranded extensions followed by blunt-end joining, while logically straightforward, is rarely used by monkey cells, even though all the necessary enzymes are present.

Short homologies can direct end joining. The majority of the junctions (80%) generated by substrates that could not join by single-strand ligation (Fig. 4) have 1 to 6 nucleotides of homology at the junction. In addition, if the junctions shown in Fig. 3 that arose by single-strand ligation are excluded, 45 of the remaining 49 junctions (92%) exhibit short homologies. These results suggest that the homologies were used in some way to direct the joining process. Several lines of evidence support this conclusion. (i) There are more junctions with homology than expected if the junctions were created in a homology-independent fashion. Random joining predicts that only about 40% of the junctions would show homology (the positions of potential homologies are indicated in Fig. 3 and 4). (ii) Junctions with 2 to 6 nucleotides of homology, especially those near the ends, often were detected multiple times. For example, 12 of the 19 sequences for the *Bgl*II-*Xho*I substrate (Fig. 4A) were distributed among just three junctions, each with a 2-nucleotide homology. These repeated patterns suggest that stability of pairing is important to the joining process. (iii) Overall, 21 junctions have 1 nucleotide of homology; however, of those homologies only 2 (10%) involved an A · T base pair. Random joining predicts that 35% (the average percentage of single AT homologies) should involve an A · T base pair. This bias toward G · C base pairs is also consistent with the idea that the stability of pairing is important for the joining process.

Three mechanisms for end joining. The majority of the 187 junctions (92%) examined in this study can be explained by one of the three models diagrammed in Fig. 6. We have termed these mechanisms single-strand, template-directed, and postrepair ligations.

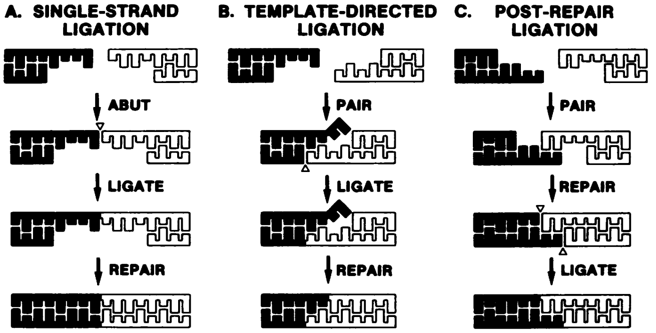


FIG. 6. Three mechanisms for end-joining. Pairs of mismatched ends containing four nucleotide single-stranded extensions are represented by the pairs of filled and open figures. (A) A possible mechanism for joining ends by ligation of single strands. In the first step, the ends are brought together (perhaps through the action of DNA binding proteins). The nick formed by the abutted single strands is indicated by the open triangle. This nick can be ligated directly, followed by repair of the gap and ligation of the other strand. (B) The two ends share 2 nucleotides of homology (represented by the complementary surfaces of the single strands). These 2 nucleotides allow the protruding single strands to pair, so that the bottom strand contains a ligatable nick, as indicated by the open triangle. Ligation of this nick joins the two molecules through their bottom strands. Subsequent removal of the unpaired nucleotides (from the upper strand) and filling in the gap allows ligation of the upper strands. The final junction contains two nucleotides of homology, corresponding to the base pairs formed during the joining reaction. (C) Pairing of the terminal two nucleotides from each single-stranded extension forms a base-paired intermediate that contains two gaps. This structure cannot be ligated directly (in contrast to panels A and B); in this case, repair synthesis must fill the gaps in order to produce ligatable nicks. The nicks can then be ligated to complete the joining process. The junction contains two nucleotides of homology.

We define single-strand ligation as a ligation reaction that joins protruding or blunt ends and does not involve a recessed strand. This definition includes ligation of two single-stranded protrusions (Fig. 6) or ligation of a protruding strand to a blunt end. One possible mechanism is shown in Fig. 6. The first step is direct ligation of single strands, producing intermediates in which the two ends are joined by a single covalent bond. These intermediates contain gaps or unpaired regions in the unligated strand which are repaired in the second step. Repair results in a molecule that contains a single nick, which can then be ligated (this step is not shown in Fig. 6). Since the formation of base pairs is not involved in the mechanism, single-strand ligation proceeds in the absence of homology.

Single-strand ligation was the predominant mode of end joining for the substrates shown in Fig. 3. Linear genomes with 3′ extensions (Fig. 3A, C, and D), which cannot be filled in by any known DNA polymerase, must have joined by ligation involving a single strand, followed by repair of the gapped region and ligation of the other strand. Similarly, results from linear molecules with 5′ extensions (Fig. 3B), which could be filled in but apparently are not, are also consistent with single-strand ligation. The efficiency with which single-stranded extensions are ligated is comparable to the efficiency of blunt-end ligation (40). Single-strand ligation is not a property of bacterial ligases (18), and such an activity has not been reported for mammalian ligases (45). Thus, the efficient ligation of single-strand extensions in mammalian cells implies a novel enzymatic activity. A precedent exists for this type of reaction; bacteriophage T4 RNA ligase ligates single strands of DNA (24).

Template-directed ligation involves ligation of a protruding strand to a recessed strand. Since ligation of recessed ends is extremely rare in the absence of homology (see above), it appears that this reaction is dependent upon the ability of the terminal nucleotide of a protruding strand from one end to pair with the first single-stranded nucleotide of the other end (Fig. 6). This base pairing allows the two strands to be closely apposed, forming a nick that can be sealed by ligase. We term this reaction template directed, because one protruding strand (top strand in Fig. 6B) serves as a template to align the other strands for the ligation reaction. We propose that after this ligation, gaps or unpaired regions on the other strand can be repaired, producing a nick (not shown in Fig. 6B), which can be ligated to complete the joining process. Since template-directed ligation requires the formation of base pairs, it is homology dependent.

Both single-strand and template-directed ligations share a distinctive feature: one strand from each end can be ligated directly, producing a covalently joined intermediate that subsequently undergoes repair, leading to ligation of the other strand. The ability of one strand to be ligated directly distinguishes these reactions from postrepair ligation, in which ligation cannot occur until after a repair event produces a nick in one or both strands (Fig. 6C). Postrepair ligation includes a heterogeneous collection of reaction pathways; however, they all involve the formation of a base-paired intermediate. Thus, postrepair ligation is homology dependent

A variety of possible paired structures for both template-directed and postrepair ligation are illustrated schematically in Fig. 7 for some of the more common junctions. In each case, the paired intermediate was drawn as if there were no modification to the input ends before pairing. The stability of single-stranded extensions provides some support for this

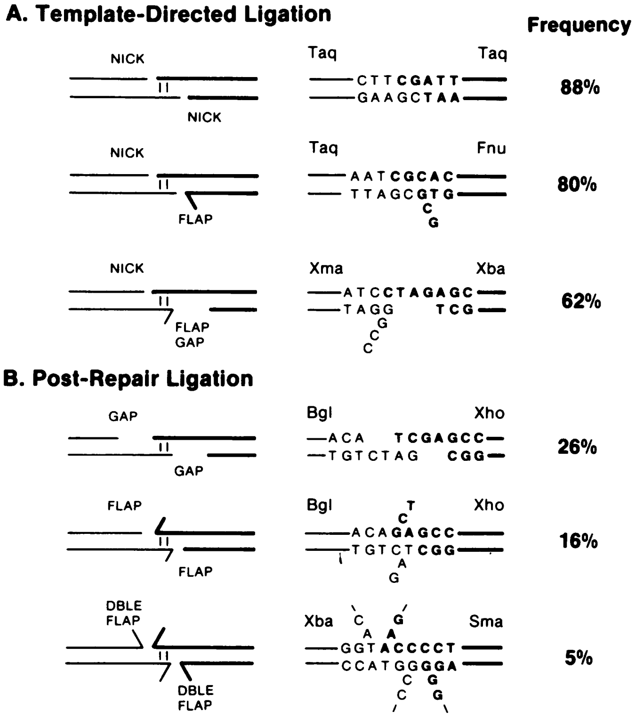


FIG. 7. Pairing mechanisms for homology-directed end joining. Six possible paired intermediates are diagrammed on the left. Corresponding examples of presumed intermediates deduced from our sequence analysis are shown on the right. The column marked "frequency" indicates the percentage of the junctions from each substrate that are consistent with the intermediate shown. (A) Three examples of template-directed ligation. Ligation of complementary termini (*Taq*I-*Taq*I substrate) is a special case of template-directed ligation in which both nicks can be ligated directly. Data from the *Taq*I-*Fnu*DII substrate indicate that mismatched ends can direct joining with almost the same frequency as complementary termini. A more complex intermediate involving only one nucleotide of homology (*Xma*I-*Xba*I substrate) is used with a lower frequency. (B) Three examples of possible postrepair mechanisms. The first example is similar to the mechanism shown in Fig. 6, in which gaps must be filled in before ligation can take place. The other two examples involve unpaired single strands (termed flaps); the more complicated intermediates, such as the one shown for the *Xba*I-*Sma*I substrate, gave rise to junctions less frequently.

assumption; however, the sequence of the junction does not define the order of the pairing and modification steps in the joining process. It is not unreasonable to imagine that cellular repair activities could join such intermediates. Note that when two base pairs can form, the joining of mismatched ends (*Taq*I-*Fnu*DII linear) occurs with almost the same specificity (in the sense that one junction species predominates) as the ligation of complementary *Taq*I ends. With 1 nucleotide of homology (*Xba*I-*Xma*I linear), the specificity is reduced, although the majority of the junctions from this substrate (62%; Fig. 4B) have structures that are consistent with this mechanism. It is somewhat surprising that such short homologies can be used so efficiently to link duplexes together. We are currently examining cell extracts for such a presumptive pairing activity.

Conclusions. The results described here define three general mechanisms for joining mismatched ends: (i) single-strand ligation, (ii) template-directed ligation, and (iii) postrepair ligation. Single-strand ligation is independent of homology, whereas template-directed and postrepair ligation are homology dependent. This study supports the results of a previous statistical analysis, which concluded that both homology-independent and -dependent processes were important for nonhomologous recombination in mammalian cells (30). Figure 8 summarizes the major features derived from our analysis of the junction sequences. It is apparent from this summary that the structure of the ends of the transfected molecule strongly influences the nature of the preferred joining reaction. For example, the majority of the junctions derived from substrates with ends that can undergo single-strand ligation (*Sac*I-*Sma*I, *Xba*I-*Sma*I, *Sac*I-*Xma*I, and *Bgl*II-*Sac*I) have structures consistent with single-strand ligation. Substrates such as *Xba*I-*Xma*I, *Bgl*II-*Xba*I, and *Bgl*II-*Xho*I cannot undergo single-strand ligation because of the configuration of strands at the termini (Fig. 8). Apparently, these ends join preferentially by homology-dependent mechanisms. This proposal is supported by the results obtained in a previous study with a *Taq*I-*Fnu*DII substrate (30). This substrate contained one blunt end and one end with a 5′ extension (Fig. 8). These ends cannot join by single-strand ligation because, during the construction of this molecule, the phosphate was removed from the 5′ extension. The ends of this substrate were joined predominantly (80% of junctions) by a homology-dependent mechanism (Fig. 8).

The structures of the nonhomologous junctions determined in this study are quite similar to the nonhomologous junctions observed in chromosome translocations (9), in immunoglobulin and T-cell receptor rearrangements (2, 10), in chromosome rearrangements in maize (23), in excision of viral DNA from chromosomes (4), and in the integration of foreign DNA in chromosomes (3, 8, 29). Although these processes differ in the mechanism of breakage of DNA, they may all use the same general end-joining processes that were detected in this study with transfected DNA. For example, immunoglobulin rearrangements, which are initiated by site-specific cleavage, are thought to proceed through a stage involving free DNA ends because inserted nucleotides are often found at the junctions. Overall, immune system rearrangements differ from the ones reported here only in the relative frequencies of junctions that contain extra nucleotides inserted at the junction (D. B. Roth and J. H. Wilson, unpublished data). The lower frequency of inserted nucleotides in fibroblasts may reflect a lower terminal transferase activity.

Previous work has shown that mammalian cells join sticky, blunt, or mismatched ends with extremely high efficiency (15, 25, 31, 43). This observation may be useful in the context of attempts to "target" the integration of exogenous DNA into the mammalian genome by homologous recombination. Recent attempts to control the site of integration have met with only limited success because nonhomologous integration is 100- to 10,000-fold more frequent than homologous integration (20, 33, 34, 38). Since free DNA ends stimulate homologous recombination (17, 19, 35, 37, 41), these investigators introduced molecules that had been linearized within the desired region of homology. However, since end joining and homologous recombination have similar relative rates (31), we have suggested that efficient intra- or intermolecular end joining might decrease the effective concentration of linear molecules. Therefore, conditions that decrease the rate or the overall efficiency of end joining might increase the frequency of homologous integration of transfected DNA molecules into the genome. Efficient end joining may also be responsible for the high background of integration at nonhomologous chromosomal sites, since the pathway of nonhomologous integration may

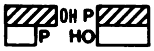
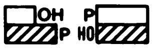
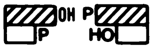
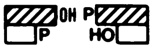
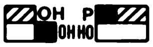
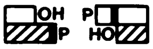
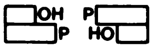
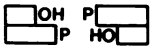
			MECHANISM OF END-JOINING				
			HOMOLOGY INDEPENDENT		**HOMOLOGY DEPENDENT**		
ENDS	**TOTAL NUMBER**	**STRUCTURE**	**Single-Strand**	**Other**	**Template-Directed**	**Post-Repair**
Sac Sma	**26**		**70%**	**3%**	—	**27%**
Xba Sma	**63**		**63%**	**5%**	—	**32%**
Sac Xma	**20**		**50%**	**0%**	—	**50%**
Sac Bgl	**17**		**47%**	**0%**	—	**53%**
Taq Fnu*	**114**		**13%**	**?**	**80%**	**?**
Xba Xma	**21**		—	**5%**	**52%**	**43%**
Bgl Xba	**21**		—	**43%**	—	**57%**
Bgl Xho	**19**		—	**10%**	—	**90%**

FIG. 8. Summary of junction sequences. The seven linear substrates used in this study, as well as one substrate studied previously (30) are drawn schematically on the left. Bars represent single DNA strands. Hatched bars denote strands that can be ligated without prior modification (single-strand or template-directed ligation); open bars represent strands that require chemical modification (such as removal of noncomplementary nucleotides) before they can be joined. Filled segments indicate regions of homology that are positioned to allow template-directed ligation. Other homologies are not shown. 3′ termini are indicated by OH; phosphorylated 5′ termini are denoted by P. The 5′ hydroxyl of the *Taq*I end in the *Taq*I-*Fnu*DII linear is indicated by OH, indicating that the 5′ phosphate was removed by treatment with phosphatase. Percentages were calculated by dividing the number of junctions that were consistent with each mechanism by the total number of junction sequences obtained for each substrate (shown in the column marked total number). The presence of a horizontal line in a column indicates that that mechanism could not be used because of the structure of the ends. For molecules with 3′ extensions, all junctions that contained at least one nucleotide of the 3′ extension and all nucleotides from the other end were considered products of single-strand ligation (unless homologies were present). For example, for the *Sac*I-*Sma*I substrate (Fig. 3A), 18 out of 26 junctions (70%) must have been formed by single-strand ligation, since there is no other known mechanism to fill in the 3′ extension. For molecules with 5′ extensions (*Xba*I-*Sma*I construct, Fig. 3B), we considered the 40 junctions that contained all nucleotides from both ends products of single-strand ligation. The colum marked "other" represents all junctions with no homology that could not be explained by single-strand ligation. Junctions were classified as products of template-directed ligation when homologies were positioned so that direct template-directed ligation (as described in the legend to Fig. 6) could occur. All other junctions that exhibited homology were classified as postrepair ligation products. Some of the junctions classified as "homology-dependent" undoubtedly arose by homology-independent mechanisms; however, the evidence indicates that this number is relatively small (see the text). The asterisk indicates that data for this substrate are from Roth et al. (30). A total of 114 junctions were examined by restriction analysis and nucleotide sequencing of representative junctions. Percentages for single-strand and template-directed ligations are accurate, but the distribution of the remaining junctions could not be derived from the data.

involve joining input DNA ends to transient chromosomal breaks (3, 8, 29).

The information we have obtained about the mechanisms responsible for end joining provides a starting point for construction of recombination substrates designed to minimize end joining. If the joining step of nonhomologous recombination can be blocked, it might be possible to reduce the efficiency of nonhomologous integration, thereby increasing the relative efficiency of homologous integration.

ACKNOWLEDGMENTS

We thank De Dieu, Yue-Xian Hou, and Kathleen Marburger for expert technical assistance, and Joseph Mollick for help with some of the sequencing. We thank Peter Berget, Susan Berget, Rod Kellems, Tom Porter, Sharon Roth, and Ursula Weiss for valuable suggestions on the manuscript. We thank Susan Berget for suggesting the use of matrix diagrams to display junction sequences.

This work was supported by Public Health Service grants CA 15743 and GM 33405 from the National Institutes of Health and by 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. **Albertini, A. M., M. Hofer, M. P. Calos, and J. H. Miller.** 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions.

Cell **29**:319–328.

2. **Alt, F. W., and D. Baltimore.** 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-J_H fusions. Proc. Natl. Acad. Sci. USA **79**:4118–4122.
3. **Brinster, R. L., H. Y. Chen, M. E. Trumbauer, M. K. Yagle, and R. D. Palmiter.** 1985. Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. Proc. Natl. Acad. Sci. USA **82**:4438–4442.
4. **Bullock, P., J. J. Champoux, and M. R. Botchan.** 1985. Association of crossover points with topoisomerase I cleavage sites: a model for nonhomologous recombination. Science **230**:954–958.
5. **Calos, M. P., J. S. Lebkowski, and M. R. Botchan.** 1983. High mutation frequency in DNA transfected into mammalian cells. Proc. Natl. Acad. Sci. USA **80**:3015–3019.
6. **Chang, X.-B., and J. H. Wilson.** 1986. Formation of deletions after initiation of simian virus 40 replication: influence of packaging limit of the capsid. J. Virol. **58**:393–401.
7. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A.E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human beta-globin gene family. Cell **21**:653–668.
8. **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.
9. **Gerondakis, S., S. Cory, and J. 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.
10. **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.
11. **Hinnen, A., J. B. Hicks, and G. R. Fink.** 1978. Transformation of yeast. Proc. Natl. Acad. Sci. USA **75**:1929–1933.
12. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cultures. J. Mol. Biol. **26**:365–369.
13. **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.
14. **Hope, T. J., R. J. Aguilera, M. E. Minie, and H. Sakano.** 1986. Endonucleolytic activity that cleaves immunoglobulin recombination sequences. Science **231**:1141–1145.
15. **Kopchick, J. J., and D. W. Stacey.** 1984. Differences in intracellular DNA ligation after microinjection and transfection. Mol. Cell. Biol. **4**:240–246.
16. **Kornberg, A.** 1980. DNA replication, p. 95. W. H. Freeman & Co., San Francisco.
17. **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. USA **81**:3153–3157.
18. **Lehman, I. R.** 1974. DNA ligase: structure, mechanism, and function. Science **186**:790–797.
19. **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.
20. **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.
21. **Marvo, S. L., S. R. King, and S. R. Jaskunas.** 1983. Role of short regions of homology in intermolecular illegitimate recombination events. Proc. Natl. Acad. Sci. USA **80**:2452–2456.
22. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
23. **McClintock, B.** 1984. The significance of responses of the genome to challenge. Science **226**:792–801.
24. **McCoy, M. I. M., and R. I. Gumport.** 1980. T4 ribonucleic acid ligase joins single-strand oligo(deoxyribonucleotides). Biochemistry **19**:635–642.
25. **Miller, C. K., and H. M. Temin.** 1983. High efficiency ligation and recombination of DNA fragments by vertebrate cells. Science **220**:606–609.
26. **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. USA **80**:4417–4421.
27. **Orr-Weaver, T. L., J. W. Szostak, and R. J. Rothstein.** 1981. Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. USA **78**:6354–6358.
28. **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.
29. **Robins, D. M., S. Ripley, A. Henderson, and R. Axel.** 1981. Transforming DNA integrates into the host chromosome. Cell **23**:29–39.
30. **Roth, D. B., T. N. Porter, and J. H. Wilson.** 1985. Mechanisms of nonhomologous recombination in mammalian cells. Mol. Cell. Biol. **5**:2599–2607.
31. **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.
32. **Ruley, H. E., and M. Fried.** 1983. Clustered illegitimate recombination events in mammalian cells involving very short sequence homologies. Nature (London) **304**:181–184.
33. **Smith, A. J. H., and P. Berg.** 1984. Homologous recombination between defective *neo* genes in mouse 3T6 cells. Cold Spring Harbor Symp. Quant. Biol. **49**:171–181.
34. **Smithies, O., R. G. Gregg, S. S. Boggs, M. A. Koralewski, and R. S. Kucherlapati.** 1985. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature (London) **317**:230–234.
35. **Song, K.-Y., L. Chekuri, S. Rauth, S. Ehrlich, and R. Kucherlapati.** 1985. Effect of double-strand breaks on homologous recombination in mammalian cells and extracts. Mol. Cell. Biol. **5**:3331–3336.
36. **Stringer, J. R.** 1982. DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. Nature (London) **296**:363–366.
37. **Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl.** 1983. The double-strand-break repair model for recombination. Cell **33**:25–35.
38. **Thomas, K., K. R. Folger, and M. Capecchi.** 1986. High frequency targeting of genes to specific sites in the mammalian genome. Cell **44**:419–428.
39. **Tooze, J. (ed.).** 1980. DNA tumor viruses, p. 144. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
40. **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.
41. **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.
42. **Wilson, J. H.** 1977. Genetic analysis of mutant host range viruses suggests an uncoating defect in SV40-resistant monkey cells. Proc. Natl. Acad. Sci. USA **74**:3503–3507.
43. **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.
44. **Zagursky, R. J., K. Baumeister, N. Lomax, and M. L. Berman.** 1985. Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. Gene Anal. Tech. **2**:89–94.
45. **Zimmerman, S. B., and B. H. Pheiffer.** 1983. Macromolecular crowding allows blunt-end ligation by DNA ligases from rat liver or *E. coli*. Proc. Natl. Acad. Sci. USA **80**:5852–5856.