Copyright © 2001 by the Genetics Society of America

Capture of DNA Sequences at Double-Strand Breaks in Mammalian Chromosomes

Yunfu Lin and Alan S. Waldman

Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208

Manuscript received November 30, 2000
Accepted for publication May 11, 2001

ABSTRACT

To study double-strand break (DSB)-induced mutations in mammalian chromosomes, we transfected thymidine kinase (*tk*)-deficient mouse fibroblasts with a DNA substrate containing a recognition site for yeast endonuclease I-*Sce*I embedded within a functional *tk* gene. To introduce a genomic DSB, cells were electroporated with a plasmid expressing endonuclease I-*Sce*I, and clones that had lost *tk* function were selected. Among 253 clones analyzed, 78% displayed small deletions or insertions of several nucleotides at the DSB site. Surprisingly, ~8% of recovered mutations involved the capture of one or more DNA fragments. Among 21 clones that had captured DNA, 10 harbored a specific segment of the I-*Sce*I expression plasmid mapping between two replication origins on the plasmid. Four clones had captured a long terminal repeat sequence from an intracisternal A particle (an endogenous retrovirus-like sequence) and one had captured what appears to be a cDNA copy of a moderately repetitive B2 sequence. Additional clones displayed segments of the *tk* gene and/or microsatellite sequences copied into the DSB. This first systematic study of DNA capture at DSBs in a mammalian genome suggests that DSB repair may play a considerable role in the evolution of eukaryotic genomes.

MAMMALIAN genomes suffer a variety of types of damage each day. One type of damage is a DNA double-strand break (DSB), which may arise either spontaneously or may be induced by many agents (reviewed in Bernstein and Bernstein 1991; Hoffman 1994; Camerini-Otero and Hsieh 1995; Shinohara and Ogawa 1995; Wood 1996). It has been estimated that cells of higher eukaryotes suffer more than eight genomic DSBs daily (Bernstein and Bernstein 1991). Efficient repair of DSBs is critical for survival, since failure to do so can result in deleterious genomic rearrangements, cell cycle arrest, or cell death (Huang *et al.* 1996; Vamvakas *et al.* 1997). Recent studies have revealed that DSBs in higher eukaryotic genomes can be repaired by several different pathways, including homologous recombination, nonhomologous end-joining (NHEJ), or precise religation (Roth and Wilson 1986, 1988; Derbyshire *et al.* 1994; Choulika *et al.* 1995; Jackson and Jeggo 1995; Sargent *et al.* 1997; Taghian and Nickoloff 1997; Liang *et al.* 1998; Lin *et al.* 1999; Haber 2000). The various factors that may influence the relative contribution that each pathway makes to DSB repair have yet to be elucidated.

DSB repair in mammalian cells is frequently accomplished with little genetic change. For example, DSB-induced intrachromosomal homologous recombination is often in the form of gene conversions rather than crossovers (Taghian and Nickoloff 1997; Richardson *et al.* 1998; Lin *et al.* 1999; Johnson and Jasin 2000). The preferential induction of gene conversions may be viewed as a cellular mechanism to prevent potentially deleterious rearrangements such as translocations. DSB-induced conversion tracts also tend to be short (Taghian and Nickoloff 1997), which may also serve to minimize genetic change.

In mammalian cells, DSB repair is frequently accomplished by NHEJ of the DNA termini with no requirement for a genetic donor sequence (Roth and Wilson 1986, 1988; Derbyshire *et al.* 1994; Jackson and Jeggo 1995; Sargent *et al.* 1997; Liang *et al.* 1998; Lin *et al.* 1999). Recently, we demonstrated that precise religation of DNA termini with no accompanying loss or alteration of sequence is a common mode for DSB repair in mammalian genomes (Lin *et al.* 1999).

In contrast with the notion that genetic change accompanying DSB repair should be minimized, it has been reported that NHEJ is sometimes associated with the insertion of novel fragments of DNA into the DSB in yeast (Moore and Haber 1996a; Teng *et al.* 1996; Ricchetti *et al.* 1999; Yu and Gabriel 1999), plant cells (Salomon and Puchta 1998; Kirik *et al.* 2000), and mammalian cells (Sargent *et al.* 1997; Liang *et al.* 1998; Van De Water *et al.* 1998). The sequences captured at DSBs had various sources, including microsatellites, retrotransposable elements, and exogenous DNA sequences. However, other than a few examples of DNA capture in mammalian chromosomes men-

Corresponding author: Alan S. Waldman, Department of Biological Sciences, University of South Carolina, 700 Sumter St., Columbia, SC 29208. E-mail: awaldman@sc.edu

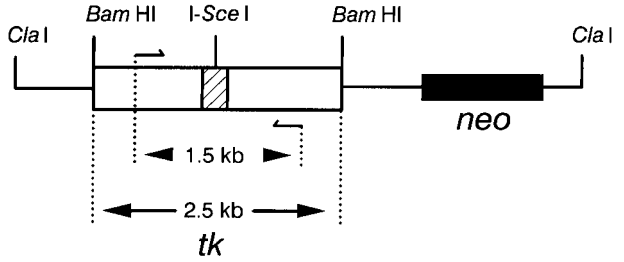

FIGURE 1.—DNA substrate for studying DSB repair. Shown is a schematic of plasmid pYL1, depicted in linear form as if linearized at the unique *Cla*I site in the vector. Plasmid pYL1 contains a *tk* gene on a 2.5-kb *Bam*HI fragment. (The plasmid also contains a *neo* gene, which is not relevant to these studies.) Inserted into the *tk* gene is a 24-bp oligonucleotide (hatched segment) containing the 18-bp recognition site for yeast endonuclease I-*Sce*I. PCR primers that flank the I-*Sce*I site and are positioned 1.5 kb apart are indicated. These primers were used to amplify *tk* sequences from cells that had undergone DSB repair for further analyses.

tioned in the literature, there has been no systematic investigation of this phenomenon in mammalian cells.

Previously, we studied DSB repair events that resulted in a gain-of-function of a herpes thymidine kinase (*tk*) gene integrated into the genome of mouse fibroblasts (LUKACSOVICH *et al.* 1994; LIN *et al.* 1999). We observed that most DSB repair occurred with little genetic change, and we did not observe any instances of DNA capture. Since the set of recoverable repair events in our previous work was restricted to gain-of-function events, we did not have the opportunity to recover mutagenic events (including DNA capture) even if they had actually occurred. To undertake a more complete examination of the types of mutations that may be associated with DSB repair, we developed a loss-of-*tk*-function assay to monitor DSB repair. We now report that while the majority of recovered DSB repair events in the genome of mouse fibroblasts results in minimal genetic change, nearly 1 in 10 recovered DSB repair events is associated with the capture of endogenous or exogenous DNA sequences of up to several kilobases in length. Our analysis of 21 cases of DNA capture in mammalian chromosomes suggests that DSB repair may play an important and ongoing role in the evolution of mammalian genomes.

MATERIALS AND METHODS

Cell culture: Mouse L*tk*$^-$ cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum, 0.1 mM minimal nonessential amino acids (GIBCO BRL, Gaithersburg, MD), and 50 μg of gentamicin sulfate/ml. Cells were maintained at 37° in a humidified atmosphere of 5% CO_2.

DSB repair substrate: Plasmid pYL1 (Figure 1) is based on the vector pJS1 (LISKAY *et al.* 1984), which is a derivative of pSV2neo (SOUTHERN and BERG 1982). Plasmid pYL1 contains a herpes simplex virus type 1 *tk* gene on a 2.5-kb *Bam*HI fragment inserted into the unique *Bam*HI site of vector pJS1. Inserted into the *Sst*I site at position 963 of the *tk* gene (numbering according to WAGNER *et al.* 1981) is a 24-bp oligonucleotide containing the 18-bp I-*Sce*I recognition sequence. The inserted 24-bp sequence is 5′-GTAGGGATAACAGGGTAAT CAGCT-3′, with the I-*Sce*I recognition site underlined. The 24-bp sequence was inserted into the *Sst*I site as a double-stranded oligonucleotide with sticky ends compatible with *Sst*I sticky ends. Insertion of the 24-bp sequence does not disrupt *tk* function.

Establishing cell lines for studying DSB repair: Mouse L*tk*$^-$ cells (5×10^6) were resuspended in 800 μl of phosphate-buffered saline (PBS) and electroporated using a Bio-Rad Gene Pulser (set at 1000 V, 25 μF) with either 1 μg of pYL1 linearized with *Cla*I or 1 μg of the isolated 2.5-kb *Bam*HI fragment of pYL1 containing the *tk* gene. Cells were plated at a density of 1×10^5 cells/75-cm^2 flask. Two days after electroporation cells were fed with DMEM supplemented with hypoxanthine/aminopterin/thymidine (HAT; Sigma, St. Louis) to select for stable transformants with a *tk*-positive phenotype. Cell lines that, on the basis of Southern blotting analysis, appeared to contain a single integrated copy of the transfected DNA were used for further studies.

Recovery of DSB-induced *tk*-deficient mutants: Plasmid pCMV-I-*Sce*I, kindly provided by Maria Jasin (Sloan Kettering), contains a gene encoding I-*Sce*I endonuclease under the control of the cytomegalovirus promoter and expressible in mouse cells (ROUET *et al.* 1994). Prior to DSB induction, cell lines containing either pYL1 or the *Bam*HI fragment of pYL1 were grown in HAT medium for 4 days to kill cells that had spontaneously mutated to *tk*$^-$. HAT selection was removed, and cells were allowed to continue to grow in DMEM for 5 additional days before DSB induction. DSBs were induced by suspending 5×10^6 cells from the appropriate cell line in 800 μl of PBS containing 20 μg of pCMV-I-*Sce*I and by electroporating the cells. For mock experiments, cells were electroporated with PBS alone. After electroporation, cells were plated at a density of 10^5 cells/150-cm^2 flask. The cells were grown in DMEM for 6 days under no selection and then refed with medium supplemented with 5 μg of trifluorothymidine (TFT)/ml to select for *tk*-deficient clones. Cells were refed with TFT-containing medium every 3 days for ~1 week until TFT-resistant (TFT^R) colonies were picked. Individual TFT^R clones were propagated further and genomic DNA was prepared from clones as described (LIN *et al.* 1999).

Southern blotting analysis: Five micrograms of genomic DNA were used per lane in all Southern blots. DNA was transferred to nitrocellulose filters and hybridized with a ^{32}P-labeled probe specific for the herpes simplex virus type 1 *tk* gene as previously described (LUKACSOVICH *et al.* 1994).

PCR and DNA sequencing analysis: A segment of the *tk* gene spanning the position of the I-*Sce*I-induced DSB was amplified from genomic DNA isolated from TFT^R clones using the primers AL1 (5′-CCAGCGTCTTGTCATTGGCG-3′) and AW-22 (5′-CGGTGGGGTATCGACAGAGT-3′). AL1 is nucleotide (nt) 308–327 of the *tk* gene, while AW-22 is nt 1786–1767 of the noncoding strand. PCR reactions used Ready-To-Go PCR beads (Amersham Pharmacia, Piscataway, NJ) and a "touchdown" PCR protocol. The annealing temperature initially was set to 72° and was progressively decreased in steps of 2° down to 62° with two cycles at each temperature. An additional 20 cycles were run at an annealing temperature of 60°. PCR products were routinely processed with a Presequencing kit (Amersham, Buckinghamshire, UK) and then sequenced directly using sequenase kit version 2.0 (Amersham Pharmacia) using primer AW-28 (5′-TCTACACCACACAA CACCGCC-3′), nt 812–831 of the *tk* coding strand, and primer AW43 (5′-GGCAAGGTCGGCGGGATGAG-3′), nt 1111–1092 of the noncoding strand. For several longer PCR products, additional sequencing primers were used as needed.

TABLE 1

Recovery of TFT^R clones following DSB induction

Cell line[a]	TFT^R colony frequencies ($\times 10^5$)[b]			
	2 days	4 days	6 days	Mock
1	<0.3	1.0	42	1.3
2	2.3	3.3	138	1.3
3			51	1.3
4			181	1.7
5			19.3	0.4
6			13	0.2

[a] Cell lines 1 through 4 each contain a single integrated copy of pYL1 (Figure 1) while cell lines 5 and 6 contain an integrated copy of only the 2.5-kb *Bam*HI fragment of pYL1. For each cell line, 5×10^6 cells were electroporated with pCMB-I-*Sce*I or, in the case of mock experiments, with PBS alone.

[b] Calculated by dividing the number of TFT^R colonies by the number of cells plated following electroporation. Cells from lines 1 and 2 were grown for 2, 4, or 6 days after electroporation before being exposed to TFT. Data presented for all other cell lines and for all mock electroporations are from experiments in which cells were grown for 6 days after electroporation before being exposed to TFT.

RESULTS

Experimental scheme: Mouse L*tk*$^-$ fibroblast cell lines that contain a stably integrated copy of the DNA substrate pYL1 or only the 2.5-kb *Bam*HI fragment from pYL1 were isolated (Figure 1). This latter fragment of pYL1 contains a functional herpes simplex virus *tk* gene with an inserted 24-bp oligonucleotide containing the 18-bp recognition site for yeast endonuclease I-*Sce*I. To induce a genomic DSB in the integrated *tk* gene, cells were electroporated with plasmid pCMV-I-*Sce*I, which expresses endonuclease I-*Sce*I in mammalian cells. Due to the large size of the I-*Sce*I recognition site (18 bp), it is likely that only a single genomic DSB is induced by expression of I-*Sce*I. DSB repair events were recovered by selecting for cells that have lost *tk* function and thus have gained resistance to the thymidine analog TFT. Selection for loss-of-*tk*-function in principle should allow recovery of virtually any type of nonlethal DSB-induced mutation.

Recovery of TFT^R clones following DSB induction: Cell lines 1 through 4 contain the entire construct pYL1 (Figure 1) and cell lines 5 and 6 contain only the 2.5-kb *Bam*HI fragment of pYL1. Electroporation of each of the cell lines with pCMV-I-*Sce*I brought about a 39- to 106-fold increase in the number of TFT^R colonies recovered compared with mock electroporations (Table 1). Exposure of cells to TFT had to be delayed until 6 days postelectroporation to detect an increase in the number of TFT^R colonies. We surmise that several days are required after disruption of the *tk* gene for intracellular levels of TK protein to diminish sufficiently to confer resistance to TFT.

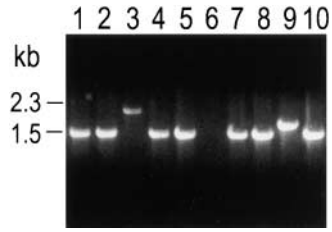

FIGURE 2.—Representative PCR analysis of DSB-induced mutations. PCR was performed on samples of genomic DNA isolated from cell line 5 and 9 representative TFT^R clones recovered from cell line 5 after electroporation with pCMV-I-*Sce*I. The pair of primers used were positioned as indicated in Figure 1. The parental cell line generated a 1.5-kb PCR product (lane 1). Most TFT^R clones (lanes 2, 4, 5, 7, 8, and 10) displayed PCR products indistinguishable in size from that of the parent. Two clones (lanes 3 and 9) displayed PCR products larger than that of the parent while 1 (lane 6) failed to produce any PCR product. A total of 253 TFT^R clones isolated from cell lines 1 through 6 were analyzed by PCR in a similar manner (see Table 2).

Small insertions and deletions associated with repair of a genomic DSB: Southern blotting analysis performed on 12 TFT^R clones recovered following mock electroporations revealed that none of the clones displayed bands that hybridize to a *tk*-specific probe (data not shown). This suggested that most clones recovered from mock electroporations arise from loss of most or all of the integrated construct and that loss of the construct occurs more frequently than does point mutation of *tk*. To study the types of mutations associated specifically with repair of a genomic DSB, *tk* sequences were PCR amplified from genomic DNA isolated from a total of 253 TFT^R colonies recovered from cell lines 1 through 6 following electroporation with pCMV-I-*Sce*I. Primers were used that were positioned ~1.5 kb apart from one another and flanked the I-*Sce*I recognition site originally present in the integrated *tk* gene (see Figure 1). As expected, a PCR product of 1.5 kb was generated when parental genomic DNA was used as a template (Figure 2, lane 1). Among the TFT^R clones, 196 (78%, Table 2) produced PCR products that appeared to be 1.5 kb, similar to the parental product (see Figure 2, lanes 2, 4, 5, 7, 8, and 10). Twenty-seven such PCR products (from cell lines 1 and 2) were sequenced. Four of these sequences displayed an insertion of a single "A" at the I-*Sce*I-induced DSB (Figure 3). The remaining 23 sequences displayed small deletions of 90 bp or less in the immediate vicinity of the I-*Sce*I site, with 16 of 23 displaying loss of 10 or fewer base pairs (Figure 3).

Capture of DNA sequences at a genomic DSB: Twenty-one (8%) of 253 clones analyzed by PCR generated PCR products that were larger than the expected 1.5 kb (see Figure 2, lane 3 and 9). The "large" PCR products were suggestive of insertions of DNA fragments at the I-*Sce*I-induced DSB. All cell lines except for cell line 1 produced clones that generated large PCR prod-

TABLE 2

DSB-induced mutations

Cell line	Small deletions or insertions[a]	DNA capture[b]	No PCR product	Multiple PCR products
1	17[c]		1	
2	38	7	8	1
3	9	2	1	
4	21	3		
5	57	6	11	1
6	54	3	11	2
Total	196 (77.5%)	21 (8.3%)	32 (12.6%)	4 (1.6%)

Classification of mutations is based on PCR products generated as described in Figure 2.

[a] Defined by a diagnostic PCR product that is indistinguishable in size from the parental 1.5-kb product (see Figure 2). Nucleotide sequences of 27 such clones are presented in Figure 3.

[b] Defined by a PCR product that is visibly larger than the parental 1.5-kb product (see Figure 2).

[c] Number of clones recovered for each category.

ucts (Table 2). Since we analyzed only 18 TFT^R clones recovered from cell line 1 (Table 2) we could not exclude the possibility that these putative insertion events occurred in cell line 1 as well.

As summarized in Table 3, sequencing of the large PCR products revealed that DNA fragments ranging from 140 to 3351 bp had been captured at the I-*Sce*I-induced DSB. In some clones, capture of DNA was accompanied by the deletion of 1–101 bp from the DSB termini, while in a few cases capture of a DNA fragment was accompanied by an additional insertion of several nucleotides. The captured DNA fragments had various sources and are described further below.

Capture of vector DNA: Ten of the 21 TFT^R clones that generated large PCR products contained a captured segment of pCMV-I-*Sce*I. The captured fragments of pCMV-I-*Sce*I ranged from 159 to 490 bp. Nine of the 10 clones contained a continuous fragment from pCMV-I-*Sce*I while 1 clone (2-2J) contained a 487-bp fragment of pCMV-I-*Sce*I with a 3-bp internal deletion. Curiously, in all clones that had captured a fragment of pCMV-I-*Sce*I, one end of the captured DNA mapped in the 27 bp a perfect palindrome (5′-CAGAGGCCGAGGCCGCCTC

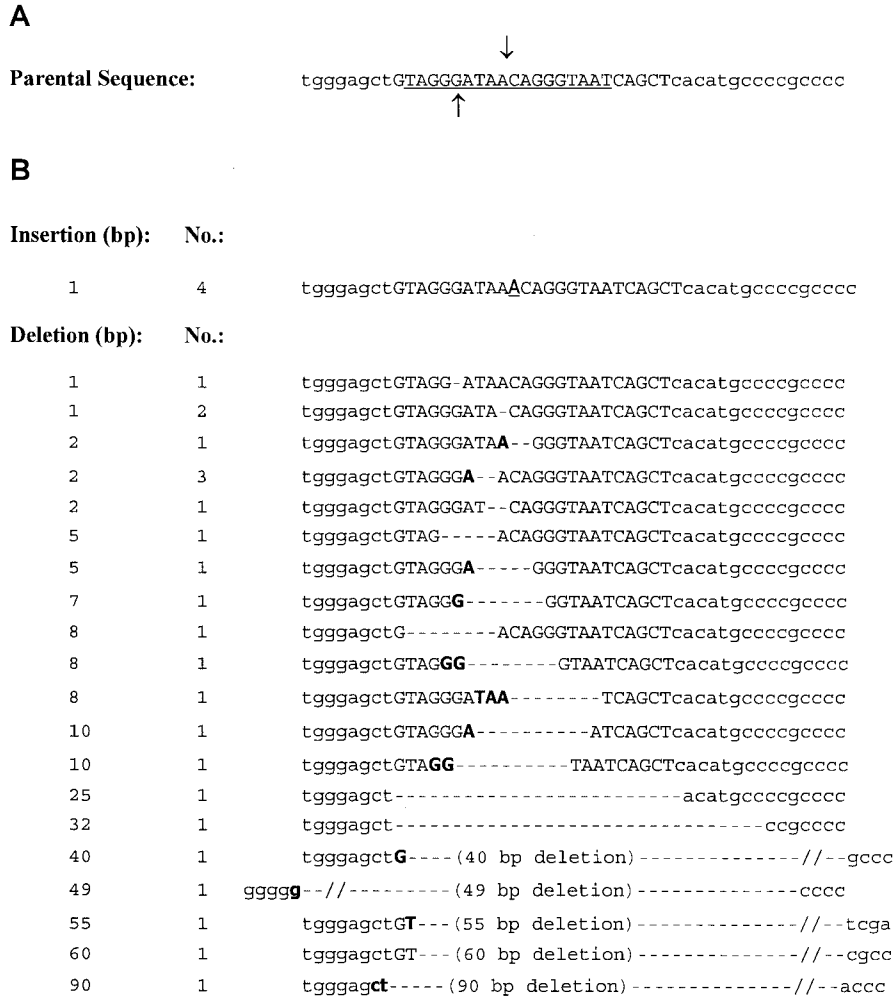

FIGURE 3.—Small insertions and deletions at a chromosomal DSB. (A) The sequence in the immediate vicinity of the I-*Sce*I recognition site within the *tk* gene in pYL1. For simplicity, only the "top" strand sequence is shown. Uppercase characters represent the sequence of the 24-bp oligonucleotide while lowercase characters represent the adjacent *tk* sequence. The actual 18-bp I-*Sce*I recognition site is underlined. Endonuclease I-*Sce*I makes staggered cuts on the top and bottom strands at the positions indicated by the arrows, producing a 3′ overhang of the 4-nt sequence ATAA on the top strand. (B) Out of 253 TFT^R clones analyzed following electroporation of cell lines 1 through 6 (Table 1) with pCMV-I-*Sce*I, 196 clones appeared to have sustained a small insertion or deletion at the I-*Sce*I site on the basis of PCR analysis (see Figure 2). The nucleotide sequence in the vicinity of the I-*Sce*I site was determined for 27 of these latter clones and these sequences are presented. Also indicated is the number of clones having each particular sequence. Four clones displayed an insertion of a single "A" (underlined), while the remaining 23 clones displayed small deletions ranging from 1 to 90 bp. The nucleotides presented in boldface type represent terminal microhomologies at the sites of DNA joining.

TABLE 3

DNA sequences captured at chromosomal breaks

Clone[a]	Upstream change (nt)[b]	Downstream change (nt)[c]	Length of captured DNA sequence (nt)	Source of captured DNA
2-1C	−29	−72	342	UI[d]/TK
2-1D	−3	−64	318	pCMV-I-*Sce*I (CW)[e]
2-2A	−3	0	353	IAP
2-2I	−3	−42	3351	*E. coli* genome
2-2J	−1	0	487	pCMV-I-*Sec*I (CW)
2-3G	−6	+2 (AA)	350	IAP
2-5L	−5	0	487	pCMV-I-*Sec*I (CW)
3-1C	+2 (AA)	0	482	pCMV-I-*Sce*I (CW)
3-3D	−4	0	468	pCMV-I-*Sce*I (CW)
4-1E	+6 (ATATAA)	0	489	pCMV-I-*Sce*I (CCW)
4-2C	+3 (TAA)	0	490	pCMV-I-*Sce*I (CCW)
4-3F	−3	−19	159	pCMV-I-*Sce*I (CW)
5-1D	−5	0	487	pCMV-I-*Sce*I (CCW)
5-1E	−19	0	309	UI
5-2A	−17	−22	1644	UI/TK
5-2F	−5	0	488	pCMV-I-*Sce*I (CW)
5-5H	−6	0	144	IAP
5-8E	−7	0	140	TK
6-1J	−8	0	321	UI/B2 repeat
6-2D	−1	−1	167	UI
6-5H	0	0	422	UI/IAP

[a] The first character of each clone name refers to the cell line (2 through 6) from which the TFT^R clone was recovered.

[b] Number of nucleotides deleted (−) or gained (+) upstream from position of the I-*Sce*I cut site on the top strand of the integrated pYL1 substrate (see Figure 3A). These changes are in addition to the captured DNA sequence. In the three cases involving the gain of several nucleotides (clones 3-1C, 4-1E, and 4-2C), the actual sequences added are indicated in parentheses. The added sequences may be derivatives of the 3′ overhang sequence ATAA produced upon staggered cleavage of the integrated substrate with I-*Sce*I.

[c] Number of nucleotides deleted or gained downstream from the I-*Sec*I cut site on the top strand of the integrated pYL1 substrate (Figure 3A). The changes are in addition to the captured DNA sequence.

[d] UI: sequences of unidentified origin. These sequences are not present in GenBank or EMBL.

[e] Refers to orientation of captured fragment of pCMV-I-*Sce*I. CW: clockwise, meaning that the SV40 origin sequence is positioned to the left of the ColE1 origin as drawn in Figure 1. CCW: counterclockwise, opposite orientation.

GGCCTCTG-3′) of the SV40 origin of replication on pCMV-I-*Sce*I (Figure 4). In 8 of the 10 clones, the other end of the captured fragment mapped in or near the ColEI replication origin on pCMV-I-*Sce*I. Neither origin of replication should have been functional in the mouse L*tk*$^-$ cells.

Cell lines 1 through 4 contain integrated vector sequences from pYL1 that share substantial homology with vector sequences on pCMV-I-*Sce*I, and so it seemed possible that capture of a fragment of pCMV-I-*Sce*I could depend on homology. Cell lines 5 and 6 contain only the 2.5-kb *Bam*HI fragment of pYL1 and thus share no homology with pCMV-I-*Sce*I. Two clones (5-1D and 5-2F) isolated from cell line 5 had captured pCMV-I-*Sce*I sequences, indicating that such capture does not require significant homology.

Capture of retrotransposable elements into a genomic DSB: Four clones (2-2A, 2-3G, 5-5H, and 6-5H) had captured all or part of a long terminal repeat (LTR) from an intracisternal A particle (IAP), a moderately

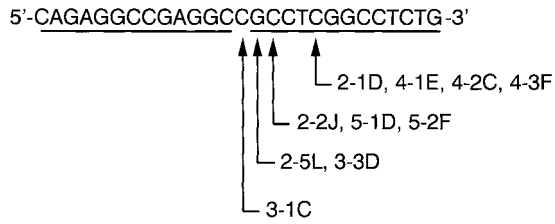

FIGURE 4.—DNA junctions within a palindromic sequence in the SV40 origin of replication. Shown at the top of the figure is one strand of a 27-bp perfect palindrome from the SV40 origin of replication, with a 13-bp perfect direct repeat underlined. As indicated in Table 3, 10 TFT^R clones had captured a fragment of pCMV-I-*Sce*I. For each of these 10 clones, one end of the captured DNA mapped within the 27-bp perfect palindrome from the SV40 origin of replication on pCMV-I-*Sce*I. Indicated is the position within the palindrome at which the captured DNA in each clone was joined to genomic DNA. Each captured DNA fragment includes the indicated nucleotide and extends to the right so that none of the final repair products retains more than half of the palindrome.

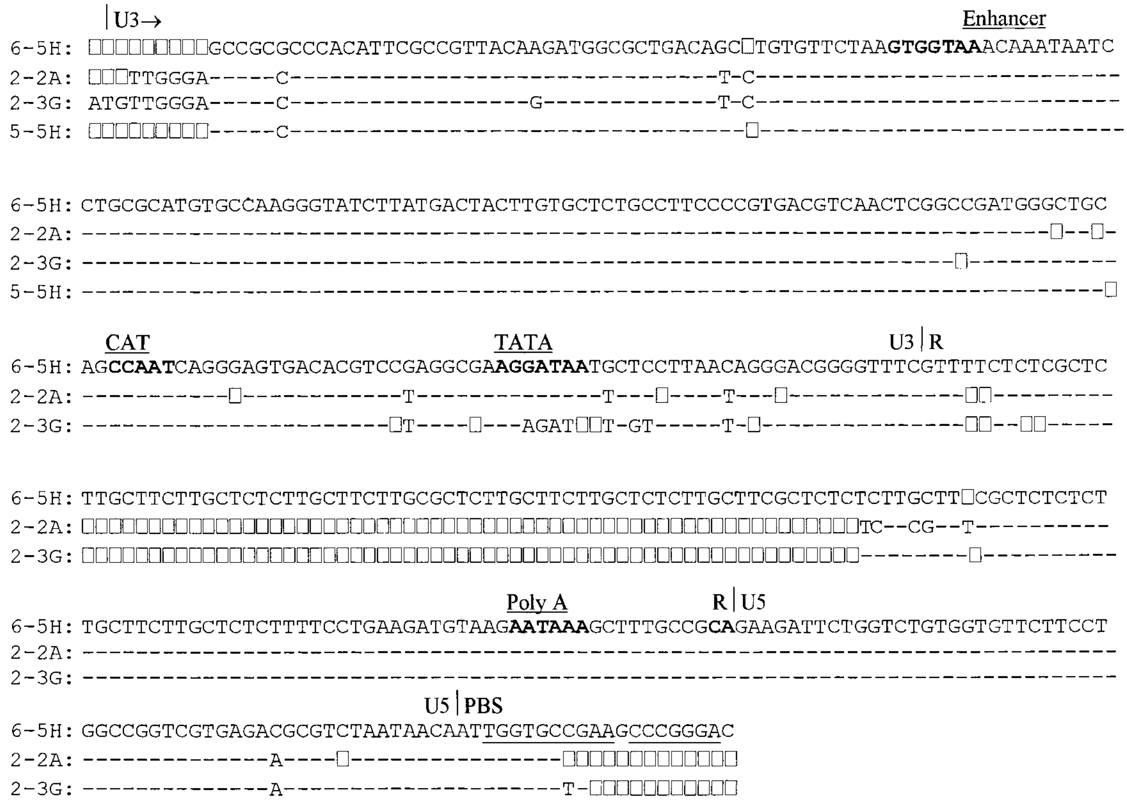

Figure 5.—IAP LTR sequences captured at a chromosomal DSB. Shown are the IAP LTR sequences captured at the induced DSB in clones 6-5H, 2-2A, 2-3G, and 5-5H. Nucleotide identity among clones is indicated with a dash, while a nucleotide gap is indicated with a box (□). Captured sequences from clones 6-5H, 2-2A, and 2-3G display the typical three-domain structure, 5′-U3-R-U5-3′, of an IAP LTR (Christy *et al.* 1985), while the sequence from clone 5-5H is truncated within the U3 region. The borders between the U3, R, and U5 regions are indicated. As shown, clone 2-3G contains a complete U3 region (which begins with the sequence TGTT) while the other clones are missing the first several nucleotides of the U3 region. Indicated in boldface type are enhancer, "CAT box," "TATA box," and polyadenylation motifs that are typically found in IAP LTRs. The primer binding sequence is denoted. Underlined bases within the primer binding site are complementary to the 3′ end of phenylalanine tRNA, which primes reverse transcription of IAP mRNA.

repetitive retrovirus-like sequence in the mouse genome (Figure 5). Clone 6-5H additionally displayed 20 bp of unknown origin immediately upstream from the IAP sequence. For each of the four clones, 2-2A, 2-3G, 5-5H, and 6-5H, there were at least 2 bp of homology at the IAP sequence/genomic DNA junction at one or both ends of the DSB (Figure 6).

For clone 6-1J, the captured DNA consisted of 106 bp of unknown origin followed by a 174-bp sequence having >98% homology to the B2 repeat sequence family from the mouse genome. Immediately downstream from the B2 sequence was a run of 41 consecutive A's and a polyadenylation signal was found 10 bp upstream from the run of A's. Thus, a portion of the DNA captured in clone 6-1J may have arisen from a B2 cDNA.

Capture of other sequences at a genomic DSB: Three clones (2-1C, 5-2A, and 5-8E) had captured a segment of the *tk* gene. The 5′-most 141 bp of the clone 2-1C insert was of unknown origin while the 3′-most 201 bp was a *tk* segment mapping 825 bp downstream from the DSB site. Clone 5-2A had captured 1644 bp. The 5′-most 1208 bp was of unknown origin while the 3′-most 436 bp was a *tk* segment mapping almost immediately upstream of the DSB. Clone 5-8E captured 140 bp of *tk* mapping 798 bp upstream from the DSB. The *tk* sequence captured in all three clones shared several base pairs of homology with DSB termini (see Figure 6).

Clone 2-2I contained 3351 bp of *Escherichia coli* DNA, presumably a contaminant that was inadvertently electroporated into cells.

The 309-bp insert of clone 5-1E was of unknown origin and contained the microsatellite-like sequence $(GAGAA)_9$ $(AAAGG)_3(AAGGG)_3$ at one end. In addition, the abundant microsatellite sequence $(CCTTTCTT)_8(CCTT)_{13}$ (Beckman and Weber 1992) was present at the unidentified segment/*tk* gene junction in clone 5-2A, with the two sequences joined at 4 bp of homology (CCTT). The insert in clone 6-2D was of unknown origin and displayed a 53-bp imperfect palindrome 6 bp from one junction with genomic DNA.

Other DSB-induced rearrangements: Thirty-two TFT^R clones among 253 (12.6%) did not generate any PCR products (Table 2). Four of these latter clones were analyzed by Southern blotting using a *tk* probe following

Clone:	Upstream Junction:	Downstream Junction:
2-1C	N/A	TCGCCGCCC**TCC**TGTGCTAC
		GGGGTCAGG**TCC**ACGACTGG
5-2A	N/A	CATGCCCCGC**CC**CCGGCCCT
		CGTTCTGGCT**CC**TCATATCG
5-8E	GGCTGGGAG**C**TGTAGGGATA	TGTAGGGATA**ACA**GGGTAAT
	CTATCTTCT**C**ACCCGGAGGC	CTATGATGAC**ACA**AACCCGC
2-2A	GCTGTAGGGATAACAGGGTA	GTAGGGAT**AACA**GGGTAATC
	CCGGAATTCGGCACCAATTG	CGGCTCCC**AACA**ATCCCCAG
2-3G	GGAGCTGT**AG**GGATAACAGG	CTGTAGGGAT**AA**CAGGGTAA
	AAAGGGGG**AG**ATGTTGGGAG	ATTGGTGCTG**AA**GCCCGGGA
5-5H	CTGGGA**GCTG**TAGGGATAAC	CTGGGAGCTG**T**AACAGGGTA
	TGATTG**GCTG**CAGCCCATCG	TGGGGGCGGC**T**CCCAACATC
6-5H	GCTGTA**GGGA**TAACAGGGTA	N/A
	GATGTT**GGGA**GCCGCGCCCA	
2-2I	GCTGTAGGG**A**TAACAGGGTA	CCCTCACCCT**C**ATCTTCGAC
	GCATCATTT**A**AGCGTTCCAG	GCTCTGCCAC**C**GTCGCGGCC

FIGURE 6.—Microhomology at junctions between captured sequences and genomic DNA. Shown are the sequences at the sites of joining between captured DNA sequences and genomic DNA in several clones. For each junction depicted, only the top strand from genomic DNA (top line of sequence) and from the source of captured DNA (bottom line of sequence) are shown. The underlined nucleotides represent the actual sequence determined at each junction. Microhomologies at junctions are depicted in boldface type and occur where the underlined sequences for each junction overlap. As indicated in Table 3, captured DNA includes *tk* gene sequences (clones 2-1C, 5-2A, and 5-8E), IAP sequences (clones 2-2A, 2-3G, 5-5H, and 6-5H), and *E. coli* genomic sequences (clone 2-2I). We could not assess microhomologies at the junctions marked "N/A" because portions of the captured sequences had unidentified sources.

digestion of genomic DNA with *Bam*HI (data not shown). One clone failed to display any bands, consistent with the loss of most or all of the integrated construct. Of the remaining three clones, one clone displayed two *Bam*HI fragments (8 and 1.5 kb). This clone may have arisen from a translocation or an insertion of one or more fragments of DNA containing at least one *Bam*HI site. The remaining two clones each displayed a single band (2.5 or 3 kb). Each of these clones may have arisen from a deletion of a portion of the *tk* sequence or possibly from other more complex rearrangements. These clones were not studied further.

Another four clones (1.6% of the total) generated multiple PCR products (Table 2), which possibly resulted from the partial amplification of the *tk* gene. These clones were not studied further. Taking into account all clones generating either no PCR product or multiple PCR products, we inferred that at most ~14% of the recovered clones had undergone gross chromosomal rearrangements.

DISCUSSION

Most DSB repair in mammalian chromosomes results in little genetic change: From our earlier work (LUKACSOVICH *et al.* 1994; LIN *et al.* 1999) as well as the work of others (ROTH and WILSON 1986, 1988; DERBYSHIRE *et al.* 1994; JACKSON and JEGGO 1995; SARGENT *et al.* 1997; LIANG *et al.* 1998), it had been demonstrated that NHEJ serves as a common mode for DSB repair in mammalian genomes. In our previous studies (LUKACSOVICH *et al.* 1994; LIN *et al.* 1999), NHEJ in the genome of mouse L*tk*$^-$ fibroblasts was often accompanied by the loss or gain of just a few nucleotides. Our previous work involved selection for a functional *tk* gene, and so genetic selection may have influenced the types of repair events recovered. In our current work we used a loss-of-*tk*-function selection to study DSB repair and again found that NHEJ with concomitant gain or loss of a small number of nucleotides is a commonly recovered repair event. (We are cognizant that since we do not know the number of cells in our experiments that actually suffered a DSB, we do not know precisely what fraction of these cells was recovered.) Greater than 75% of the *tk*-negative clones recovered following DSB induction were products of NHEJ with deletion of <90 bp from the DSB site, and more than half of all recovered events involved loss of no more than 10 bp (Table 2 and Figure 3). Our results are similar to those reported by others for Chinese hamster ovary cells (LIANG *et al.*

1998) as well as yeast (Kramer *et al.* 1994; Moore and Haber 1996b) in which it was shown that NHEJ is often accompanied by small deletions or insertions.

DSB repair in mammalian chromosomes may be accompanied by capture of DNA sequences: We found that ~8% of the recovered *tk*-deficient clones had captured DNA fragments at the DSB site. Capture of DNA fragments into genomic DSBs has been observed by other groups studying DSB repair in yeast (Moore and Haber 1996a; Teng *et al.* 1996; Ricchetti *et al.* 1999; Yu and Gabriel 1999), plants (Salomon and Puchta 1998; Kirik *et al.* 2000), and mammalian cells (Sargent *et al.* 1997; Liang *et al.* 1998; Van De Water *et al.* 1998). Our current study of 21 cases of DNA capture in a mammalian chromosome is the most complete such study to date. We discovered that nearly 50% of the clones with inserts (10 out of 21) contained a fragment of pCMV-I-*Sce*I (Table 3). For all 10 clones that captured part of pCMV-I-*Sce*I, one end of the captured DNA mapped in a 27-bp perfect palindrome the SV40 origin of replication on pCMV-I-*Sce*I (Figure 4); for 8 of these clones, the other end of the captured fragment mapped in or near the ColE1 origin. It is not presently clear what role the origins played in the capture mechanism, although there is some evidence that palindromic structures may be recombinagenic in mammalian chromosomes (see Waldman *et al.* 1999). The question of how pCMV-I-*Sce*I sequences are so efficiently recruited to the DSB remains an intriguing issue, but recent work suggests that transfected DNA in general may be efficiently recruited to genomic DSBs (Y. Lin and A. S. Waldman, unpublished results).

Clones 2-1C, 5-2A, and 5-8E each contain a captured *tk* segment. The genesis of these clones is economically explained by a one-sided invasion of a 3′ DNA terminus into the *tk* gene upstream or downstream from the DSB at a site of microhomology, followed by extension of the terminus via DNA synthesis using a portion of the *tk* gene as template. This newly synthesized strand of DNA may then be joined by NHEJ to the other side of the DSB either directly (clone 5-1E) or after first becoming joined to another sequence (clones 2-1C and 5-2A). Each of the junctions between the inserted *tk* sequences and genomic DNA displays microhomology of 1 to 3 bp (Figure 6). There have been previous reports of events that appear to have involved copying of genomic sequences into DSBs in mammalian (Sargent *et al.* 1997; Tremblay *et al.* 2000) and plant (Salomon and Puchta 1998) genomes. Repair of DSBs may provide one mechanism for sequence duplication in the evolution of eukaryotes.

The repetitive sequence $(GAGAA)_9(AAAGG)_3(AAGGG)_3$ was found at one end of the insert in clone 5-1E while the sequence $(CCTTTCTT)_8(CCTT)_{13}$ was present at the unidentified portion/*tk* sequence junction in the insert of clone 5-2A. Others have also reported capture of microsatellites at DSBs in a mammalian genome (Liang *et al.* 1998). DSB repair may thus provide a mechanism for the genesis and/or spread of microsatellite repeats in the evolution of mammalian genomes.

Clones 2-2A, 2-3G, 5-5H, and 6-5H each displayed an insert of an LTR region from an IAP sequence (Figure 5). IAPs are endogenous, retrovirus-like sequences residing in the mouse genome at ~1000 copies (Leuders and Kuff 1977). The LTR of an IAP consists of three regions referred to as U3, R, and U5 (Christy *et al.* 1985). In clones 2-2A, 2-3G, and 6-5H, the presence of continuous U3, R, and U5 sequences and a portion of the tRNA primer binding site at which cDNA synthesis is normally initiated suggests that the inserts were derived from plus strand strong-stop replication intermediates. Clone 6-1J captured what appears to be a cDNA copy of a B2 repeat, a retrotransposable element found at ~100,000 copies in the mouse genome (Krayev *et al.* 1982; Kass *et al.* 1997). There have been other reports of capture of retrotransposons into genomic DSBs in mammalian cells (Sargent *et al.* 1997; Tremblay *et al.* 2000). The apparently somewhat selective capture of retrotransposon-related sequences may be due to an abundance of such sequences in the nucleus.

Good evidence for capture of retrotransposition intermediates into genomic DSBs in yeast has been presented (Moore and Haber 1996a; Teng *et al.* 1996; Yu and Gabriel 1999). Inserts captured at an HO endonuclease-induced DSB in yeast commonly contained LTR sequences from the yeast retrotransposon Ty1. In striking concordance with our findings reported here for mammalian cells, many of the Ty1 sequences captured in yeast resembled plus strand strong-stop replication intermediates. Our results would appear to challenge a recent inference (Kirik *et al.* 2000) that the origin of sequences captured at DSBs differs strongly between lower and higher eukaryotes.

It is of interest to note that an abundance of solo LTRs from retrotransposons have been found dispersed through the genomes of higher and lower eukaryotes (Kim *et al.* 1998; Liao *et al.* 1998; Jordan and McDonald 1999; Goodwin and Poulter 2000). Our current work, and earlier work in yeast (Moore and Haber 1996a; Yu and Gabriel 1999), provides an explanation for the origin of solo LTRs, namely that they may be genomic "scars" marking sites of DSB repair. It has also been observed that solo LTRs together with microsatellite sequences may be a general feature of meiotic recombination hotspots in mammalian genomes (Shiroishi *et al.* 1995; Liao *et al.* 1998). It seems plausible to propose that such recombination hotspots may be sites that are prone to DSB formation since DSBs are known to be recombinagenic (Lin *et al.* 1999 and references therein). The presence of solo LTRs and microsatellites at the hotspots could then be explained by DNA capture since, as reported here, these are precisely two types of sequences that are commonly captured at DSBs.

In summary, although genomic stability in mamma-

lian cells is undoubtedly safeguarded by DSB repair pathways that generally minimize genetic change, the diversity of sequence alterations that may be engendered at a DSB implicates chromosome breakage as an important catalyst for evolution. The seemingly efficient recruitment of extrachromosomal DNA to the site of a genomic DSB is a puzzle that warrants further investigation.

We thank Tamas Lukacsovich for his efforts in the initial design of DNA constructs and Barbara Criscuolo Waldman for helpful comments on the manuscript. This work was supported by grants from the American Cancer Society (NP-949) and the National Institute of General Medical Sciences (GM47110) to A.S.W.

LITERATURE CITED

Beckman, J. S., and J. L. Weber, 1992 Survey of human and rat microsatellites. Genomics **12:** 627–631.

Bernstein, C., and H. Bernstein, 1991 *Aging, Sex, and DNA Repair.* Academic Press, San Diego.

Camerini-Otero, R. D., and P. Hsieh, 1995 Homologous recombination proteins in prokaryotes and eukaryotes. Annu. Rev. Genet. **29:** 509–552.

Choulika, A., A. Perrin, B. Dujon and J.-F. Nicolas, 1995 Induction of homologous recombination in mammalian chromosomes by using the I-*Sce* I system of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **15:** 1963–1973.

Christy, R. J., A. R. Brown, B. B.Gourlie and R. C. C. Huang, 1985 Nucleotide sequences of murine intracisternal A-particle gene LTRs have extensive variability within the R region. Nucleic Acids Res. **13:** 289–302.

Derbyshire, M. K., L. H. Epstein, C. S. H. Young, P. L. Munz and R. Fishel, 1994 Nonhomologous recombination in human cells. Mol. Cell. Biol. **14:** 156–169.

Goodwin, T. J., and R. T. Poulter, 2000 Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans*. Genome Res. **10:** 174–191.

Haber, J. E., 2000 Partners and pathways: repairing a double-strand break. Trends Genet. **16:** 259–264.

Hoffman, G. R., 1994 Induction of genetic recombination: consequences and model systems. Environ. Mol. Mutagen. **23**(Suppl. 24): 59–66.

Huang, L.C., K. C. Clarkin and G. M. Wahl, 1996 Sensitivity and selectivity of the DNA damage sensor responsible for activating p53-dependent G1 arrest. Proc. Natl. Acad. Sci. USA **93:** 4827–4832.

Jackson, S. P., and P. A. Jeggo, 1995 DNA double-strand break repair and V(D)J recombination: involvement of DNA-PK. Trends Biochem. Sci. **20:** 412–415.

Johnson, R. D., and M. Jasin, 2000 Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. EMBO J. **19:** 3398–3407.

Jordan, I. K., and J. F. McDonald, 1999 Comparative genomics and evolutionary dynamics of *Saccharomyces cerevisiae* Ty elements. Genetica **107:** 3–13.

Kass, D. H., J. Kim, A. Rao and P. L. Deininger, 1997 Evolution of B2 repeats: the muroid explosion. Genetica **99:** 1–13.

Kim, J. M., S. Vanguri, J. D. Boeke, A. Gabriel and D. F. Voytas, 1998 Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res. **8:** 464–478.

Kirik, A., S. Salomon and H. Puchta, 2000 Species-specific double-strand break repair and genome evolution in plants. EMBO J. **19:** 5562–5566.

Kramer, K. M., J. A. Brock, K. Bloom, J. K. Moore and J. E. Haber, 1994 Two different types of double-strand breaks in *Saccharomyces cerevisiae* are repaired by similar *RAD*52-independent, nonhomologous recombination events. Mol. Cell. Biol. **14:** 1293–1301.

Krayev, A. S., T. V. Markusheva, D. A. Kramerov, A. P. Ryskov, K. G. Skryabin *et al.*, 1982 Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res. **10:** 7461–7475.

Leuders, K. K., and E. L. Kuff, 1977 Sequences associated with intracisternal A particles are reiterated in the mouse genome. Cell **12:** 963–972.

Liang, F., M. Han, P. J. Romanienko and M. Jasin, 1998 Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc. Natl. Acad. Sci. USA **95:** 5172–5177.

Liao, D., T. Pavelitz and A. M. Weiner, 1998 Characterization of a novel class of interspersed LTR elements in primate genomes: structure, genomic distribution, and evolution. J. Mol. Evol. **46:** 649–660.

Lin, Y., T. Lukacsovich and A. S. Waldman, 1999 Multiple pathways for repair of DNA double-strand breaks in mammalian chromosomes. Mol. Cell. Biol. **19:** 8353–8360.

Liskay, R. M., J. L. Stachelek and A. Letsou, 1984 Homologous recombination between repeated chromosomal sequences in mouse cells. Cold Spring Harbor Symp. Quant. Biol. **49:** 183–189.

Lukacsovich, T., D. Yang and A. S. Waldman, 1994 Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-*Sce* I. Nucleic Acids Res. **22:** 5649–5657.

Moore, J. K., and J. E. Haber, 1996a Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature **383:** 644–646.

Moore, J. K., and J. E. Haber, 1996b Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **16:** 2164–2173.

Ricchetti, M., C. Fairhead and B. Dujon, 1999 Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature **402:** 96–100.

Richardson, C., M. E. Moynahan and M. Jasin, 1998 Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocation. Genes Dev. **12:** 3831–3842.

Roth, D. B., and J. Wilson, 1986 Nonhomologous recombination in mammalian cells: role of short sequence homologies in the joining reaction. Mol. Cell. Biol. **6:** 4295–4304.

Roth, D., and J. Wilson, 1988 Illegitimate recombination in mammalian cells, pp. 621–653 in *Generic Recombination*, edited by R. Kucherlapati and G. R. Smith. American Society for Microbiology, Washington, DC.

Rouet, P., F. Smith and M. Jasin, 1994 Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc. Natl. Acad. Sci. USA **91:** 6064–6068.

Salomon, S., and H. Puchta, 1998 Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J. **17:** 6086–6095.

Sargent, R. G., M. A. Brenneman and J. H. Wilson, 1997 Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. Mol. Cell. Biol. **17:** 267–277.

Shinohara, A., and T. Ogawa, 1995 Homologous recombination and the roles of double-strand breaks. Trends Biol. Sci. **20:** 387–391.

Shiroishi, T., T. Koide, M. Yoshino, T. Sagai and K. Moriwaki, 1995 Hotspots of homologous recombination in mouse meiosis. Adv. Biophys. **31:** 119–132.

Southern, P. S., and P. Berg, 1982 Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early promoter. J. Mol. Appl. Genet. **1:** 327–341.

Taghian, D. G., and J. A. Nickoloff, 1997 Chromosomal double-strand breaks induce gene conversion at high frequency in mammalian cells. Mol. Cell. Biol. **17:** 6386–6393.

Teng, S. C., B. Kim and A. Gabriel, 1996 Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. Nature **383:** 641–644.

Tremblay, A., M. Jasin and P. Chartrand, 2000 A double-strand break in a chromosomal LINE element can be repaired by gene conversion with various endogenous LINE elements in mouse cells. Mol. Cell. Biol. **20:** 54–60.

Vamvakas, S., E. H. Vock and W. K. Lutz, 1997 On the role of DNA double-strand breaks in toxicity and carcinogenesis. Crit. Rev. Toxicol. **27:** 155–174.

Van De Water, N., R. Williams, P. Ockelford and P. Browe, 1998 A 20.7 kb deletion within the factor VIII gene associated with LINE element insertion. Thromb. Haemost. **79:** 938–942.

Wagner, M. J., J. A. Sharp and W. C. Summers, 1981 Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. Proc. Natl. Acad. Sci. USA **78:** 1441–1445.

Waldman, A., H. Tran, E. Goldsmith and M. A. Resnick, 1999 Long inverted repeats are an at-risk motif for recombination in mammalian cells. Genetics **153:** 1873–1883.

Wood, R. D., 1996 DNA repair in eukaryotes. Annu. Rev. Biochem. **65:** 135–167.

Yu, X., and A. Gabriel, 1999 Patching broken chromosomes with extranuclear cellular DNA. Mol. Cell **4:** 873–881.

Communicating editor: M. Lichten