MOLECULAR AND CELLULAR BIOLOGY, Sept. 1991, p. 4389–4397
0270-7306/91/094389-09$02.00/0

Vol. 11, No. 9

# Double-Strand Gap Repair in a Mammalian Gene Targeting Reaction

VICKY VALANCIUS† AND OLIVER SMITHIES*

*Department of Pathology, School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27599-7525*

Received 23 January 1991/Accepted 10 June 1991

**To better understand the mechanism of homologous recombination in mammalian cells that facilitates gene targeting, we have analyzed the recombination reaction that inserts a plasmid into a homologous chromosomal locus in mouse embryonic stem cells. A partially deleted *HPRT* gene was targeted with various plasmids capable of correcting the mutation at this locus, and HPRT$^+$ recombinants were directly selected in HAT medium. The structures of the recombinant loci were then determined by genomic Southern blot hybridizations. We demonstrate that plasmid gaps of 200, 600, and 2,500 bp are efficiently repaired during the integrative recombination reaction. Targeting plasmids that carry a double-strand break or gap in the region of DNA homologous to the target locus produce 33- to 140-fold more hypoxanthine-aminopterin-thymidine-resistant recombinants than did these same plasmids introduced in their uncut (supercoiled) forms. Our data suggest that double-strand gaps and breaks may be enlarged prior to the repair reaction since sequence heterologies carried by the incoming plasmids located close to them are often lost. These results extend the known similarities between mammalian and yeast recombination mechanisms and suggest several features of the insertional (O-type) gene targeting reaction that should be considered when one is designing mammalian gene targeting experiments.**

The phenomenon of homologous recombination has been studied extensively in bacterial and fungal systems (see references 7 and 20 for reviews) but has only recently become amenable to study in cultured mammalian cells. With the advent of mammalian cell transformation and gene targeting techniques, the study of homologous recombination in mammalian cells has become not only possible but also practically useful. Gene targeting has been used successfully many times (see reference 6 for a review) since the initial demonstration that it was possible (24), but specifically modifying the mammalian genome is still an arduous task. The more defined the molecular processes which facilitate gene targeting are, the more accurate and efficient will the technique become.

The molecular details of homologous recombination have been studied extensively in lower eukaryotic systems, and many models have been proposed to describe these events (e.g. references 10 and 18). The double-strand break repair model, proposed by Szostak et al. (28), is one of the most widely accepted. It depicts recombination initiating from a double-strand break that is modified by an exonuclease to create single-stranded 3′ ends. These 3′ ends invade an homologous duplex and prime repair synthesis. This synthesis reaction generates a region of double-strand gene conversion that is eventually bounded on both sides by a Holliday junction (10). These junctions can branch migrate to generate regions of heteroduplex DNA that are repaired in a separate reaction (19). The Holliday junctions can then be resolved independently in either plane, resulting in a crossover event 50% of the time.

Many of the parameters of the recombination mechanisms in higher eukaryotic systems have been determined by studying extrachromosomal interactions between plasmid sequences cointroduced into mammalian cells and by studying intrachromosomal interactions between plasmid sequences stably integrated into a single chromosomal locus (see reference 3 for a review). Plasmids carrying nonoverlapping mutations can efficiently recombine in mammalian cells to generate a functional gene product, and the frequency at which this interaction occurs can be enhanced up to 300-fold by introducing either a double-strand break or gap into one of the plasmids prior to cellular transformation (4, 15, 25). Heterologies at or near the break point introduced into one of the extrachromosomal recombination substrates can be removed from the linearized or gapped plasmid during the repair/recombination reaction (5). Intrachromosomal recombination between plasmid sequences stably integrated into a single chromosomal site occurs by a mechanism that results in nonreciprocal (conversion) products 85% of the time (17). Of the reciprocal exchange (crossover) products that do arise, one-third have associated regions of gene conversion (2).

Recombination mechanisms have also been investigated by studying the interactions between stably incorporated (chromosomal) plasmid sequences and a second plasmid introduced subsequently. These gene targeting prototype studies demonstrated that both reciprocal exchange and conversion products can be obtained (16, 26). Double-strand gaps of 200 bp (12) to 1,000 bp (27) in the second plasmid were found to be efficiently repaired using homologous chromosomal sequence as template when this repair was essential for the recombinants to survive selection. The gap-repaired products were found episomally as well as integrated into the genome (13).

The yeast cell transformation experiments that were instrumental in the development of the double-strand break repair model examined the insertion of plasmid sequences

* Corresponding author.
† Present address: Department of Biochemistry, University of Oxford, Oxford OX1 3QU, England.

into a homologous chromosomal locus (21). The first experiments demonstrating gene targeting (24), and the procedures now commonly used for insertional (O-type) gene targeting in mammalian cells (8, 30, 31), have the same design as these yeast experiments. We have exploited this identity to determine the common aspects of the yeast and mammalian recombination reactions, with an emphasis on testing the predictions of the break repair model in a mammalian system.

We have targeted the directly selectable *HPRT* gene in a mouse embryonic stem (ES) cell line with various plasmids capable of correcting a deletion mutation at the target *hprt*$^-$ (i.e., partially deleted *HPRT*) locus. The recombinant loci generated were then analyzed by genomic Southern blot hybridizations to test several predictions of the double-strand break repair model in a mammalian system. Specifically, we have determined (i) the effects of double-strand breaks and gaps on the frequency of the homologous integration reaction, (ii) whether double-strand gaps are repaired using chromosomal sequences as template during the integration reaction, and (iii) the fate of various nonselected heterologies carried by the plasmids relative to the target locus. The selection system used requires that the correcting plasmids integrate into the target locus, so that all recombinants analyzed are necessarily the result of a crossover event. In contrast, however, the various gaps and heterologies that we have tested do not affect the selection scheme, and therefore their fates could be investigated in an unbiased manner.

Our results are in agreement with the predictions made from the break repair model proposed for yeast cells. We have also found evidence suggesting that plasmid-carried heterologies are frequently removed exolytically prior to or during the integration reaction. The results of these studies should prove helpful in the design of future gene targeting experiments.

## MATERIALS AND METHODS

**Plasmid construction.** Plasmid pNMR133 was described previously (8). It carries approximately 5 kb of DNA identical in sequence to the exon 3 target region of the mouse *HPRT* gene. It also carries the human promoter and exon 1 regions and the mouse exon 2 region. These elements are able to complement the deletion of the promoter and first two exons of the mutant *hprt*$^-$ gene carried by the cell line used in our studies (see below). Plasmid pNMR133 also carries a 4-bp insertion within the 5-kb region of homology to the target locus which destroys the *Hin*dIII site in intron 2. For this reason, all of the plasmids used in this study were constructed from plasmid pNMR122, the parent to pNMR133. This parent plasmid carries the intact *Hin*dIII site so that the 5 kb of homology to the target locus is uninterrupted by any modifications.

Plasmid pD200 was constructed from pNMR122 by removing a 200-bp *Bgl*II fragment from intron 2, within the region of homology, and destroying a *Bgl*II site in intron 1, outside the region of homology. The resulting plasmid has a unique *Bgl*II site which can be used to linearize the plasmid and create a 200-bp gap within the region of homology.

Plasmid pD200(H−) was made from plasmid pD200 by destroying the *Hin*dIII site in intron 2. pD200 was digested with *Hin*dIII, and then the ends were filled in with T4 DNA polymerase and blunt-end ligated. This generated a 4-bp insertion within the region of homology.

Plasmid pD600 was constructed by digesting the parental plasmid pNMR122 at the unique *Hin*dIII site in intron 2 and at the unique *Xho*I site in exon 3, filling in the recessed 3′ ends with T4 DNA polymerase, and then blunt-end ligating the plasmid into a circle. This recreates a unique *Hin*dIII recognition site which can be used to linearize plasmid pD600, generating a 600-bp gap within the region of homology.

Plasmid pD200D600 was made from plasmid pD200 by removing the 600-bp fragment which lies between the unique *Hin*dIII and *Xho*I sites in the same manner as pD600. Plasmid pD200D600 can be linearized either at the unique *Bgl*II site to generate a 200-bp gap within the region of homology or at the *Hin*dIII site to generate a 600-bp gap, also within the region of homology.

Plasmid pD2500 was constructed from the parent plasmid pNMR122 by removing a 2.5-kb *Xba*I fragment from within the 5-kb region of homology. This new plasmid carries a

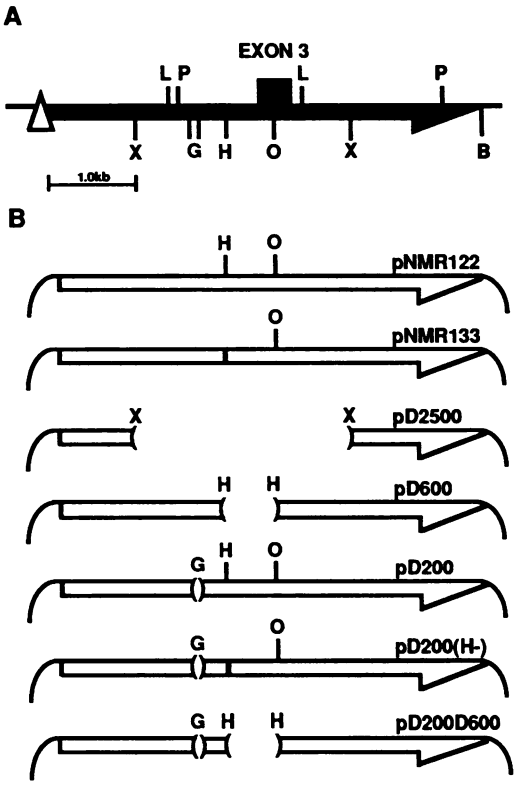


FIG. 1. Targeting plasmids. (A) Restriction enzyme map of the 5-kb *hprt*$^-$ target locus (an expansion of the target region depicted in Fig. 2A). Restriction sites indicated above the line are those used to characterize the resulting recombinants (see Fig. 3A). Sites indicated below the line correspond to critical sites in the various targeting plasmids. The arrow indicates the 5-kb target region. Abbreviations: B, *Bam*HI; G, *Bgl*II; H, *Hin*dIII; L, *Bal*I; O, *Xho*I; P, *Pvu*II; X, *Xba*I. (B) The seven targeting plasmids used. The arrows indicate the regions of homology to the target locus carried by the plasmids. Not shown are the promoter and exon 1 and 2 sequences (depicted in Fig. 2B). The vertical bar represents a 4-bp insertion that destroys the plasmid-carried *Hin*dIII site. ( ) represents regions of DNA deleted from the plasmids. Linearization of these deletion-containing plasmids at the corresponding unique restriction sites, indicated above the ( ), generates double-strand gaps in the region of homology. Linearization at other indicated sites creates plasmids carrying a deletion within the region of homology. Restriction enzyme sites are abbreviated as in panel A.

single *Xba*I site which can be used to generate a 2,500-bp gap within the region of homology.

All plasmids are depicted in Fig. 1. Prior to electroporations, they were digested with the appropriate restriction enzymes according to manufacturers' recommendations, ethanol precipitated, and resuspended in sterile TE (0.05 M Tris, 0.001 M EDTA) buffer. Plasmids introduced in their uncut (supercoiled) forms were also precipitated and resuspended prior to electroporations but were not pretreated in any other way. As judged by electrophoretic mobilities, these uncut plasmids were supercoiled at the time of their use.

**Cell culture.** The ES cell line E-14TG2a was isolated and described previously (11, 31). It carries a nonreverting deletion of the promoter and first two exons of the nine-exon, 33-kb *HPRT* gene. Because the gene is X linked and these cells were isolated from a male mouse blastocyst, the cells carry only the one deletion-mutant copy of the *HPRT* gene, rendering them phenotypically HPRT$^-$. The cells were grown in Dulbecco's modified Eagle medium (GIBCO) supplemented with 15% heat-inactivated fetal calf serum (Flow) and 10 μM 2-mercaptoethanol (Sigma). The growth medium was also supplemented with $10^6$ U of recombinant human leukemia inhibitory factor (LIF) per liter to maintain the cells in their nondifferentiated, pluripotential state (23, 33). The LIF was either purchased from Amrad Corp. or produced in vivo in Cos7 cells transfected with the LIF-expressing plasmid pC106R (Genetics Institute). Hypoxanthine-aminopterine-thymidine (HAT) selection was in the described growth medium supplemented with 120 μM hypoxanthine, 0.4 μM aminopterin, and 20 μM thymidine. All culture dishes were coated with 0.1% sterile gelatin to ensure cell adhesion. Cultures were grown at 37°C in an atmosphere of 5% $CO_2$.

**Electroporations and selections.** DNA was introduced into ES cells by electroporation (1). The ES cells were grown in 100-mm culture dishes to a density of $10^7$ cells per dish, trypsinized, pelleted, and then resuspended in growth medium (see above) to a density of $0.2 \times 10^8$ to $1.2 \times 10^8$ cells per ml. Samples (0.5 ml) of the cell suspensions were added to sterile microfuge tubes containing 2.5 pmol of prepared plasmid DNA, making the final DNA concentration 5 nM. Cell-DNA mixtures were incubated on ice for 10 to 20 min and then loaded into an electroporation chamber precooled on ice (length, 5 mm; cross section, 100 mm$^2$). The mixtures were exposed to a 1-s electrical pulse from a 250-μF capacitor charged to 300 V. Treated cells were immediately removed from the chamber with a sterile 1-ml syringe, and 0.1 ml of the cell suspension was plated into each of five 100-mm culture dishes containing 7 ml of nonselective medium. These electroporation conditions were found to cause an insignificant amount of cell death. Cultures were allowed to recover overnight, and then selection was started by adding 7 ml of growth medium supplemented with HAT (described above).

Selections were maintained for 2 weeks, with replacement of medium as necessary. Targeting frequencies were then determined by counting the number of colonies surviving HAT selection. Individual HAT$^r$ colonies were picked by trypsinization, using cloning rings, into 24-well (1-ml) culture dishes and allowed to expand. Expanded colonies were trypsinized and transferred to 60-mm culture dishes. HAT selection was maintained during the entire expansion procedure. Confluent cultures were harvested by trypsinization, pelleted by centrifugation, and then frozen at −20°C until genomic DNA could be prepared from them.

**Genomic DNA preparation and characterization.** Genomic DNA was prepared from frozen cell pellets by conventional procedures. Restriction enzyme digestions were done according to manufacturers' recommendations. After electrophoresis in 0.8% agarose gels, Southern blotting onto nylon filters was accomplished by standard techniques.

**Probes.** Two probes were used: a 250-bp *Rsa*I fragment from intron 3, and a 450-bp fragment generated by the polymerase chain reaction that includes the entire mouse exon 3 element and 100 bp of the flanking intron 2 region. Both probes hybridize to sequences present in the endogenous locus as well as on the targeting plasmids. For each blot, 25 to 50 ng of purified fragment was radiolabeled with [$^{32}$P]dCTP by the random-primed oligonucleotide method, using a Boehringer Mannheim kit. Four-hour prehybridizations and overnight hybridizations were accomplished in 50% formamide solutions at 42°C. Blots were washed to a stringency of 0.1× SSC at 68°C. Washed blots were exposed to preflashed XAR-5 film at −70°C.

## RESULTS

**Effects of double-strand breaks and gaps on the frequency of the targeted integration reaction.** The first facet of the double-strand break repair model (28) that we tested was its prediction that the introduction of either double-strand breaks or gaps into the region of homology carried by targeting plasmids would stimulate recombination. To this end, we compared the targeting frequencies obtained when various plasmids were introduced in their uncut (supercoiled) forms with the frequencies obtained when these same plasmids were introduced into the recipient genome after having been predigested with restriction enzymes to generate linear molecules having either double-strand breaks or gaps in the region of DNA homologous to the target locus.

We used the method of Doetschman et al. (8), outlined in Fig. 2, to transform the ES cell line E-14TG2a. This cell line, isolated and described previously (11, 31), carries a deletion of the promoter and first two exons of the nine-exon, X-linked *HPRT* gene (Fig. 2A). The ES cells were transformed via electroporation with various plasmids capable of correcting this deletion mutation upon homologous recombination-mediated integration into the *hprt*$^-$ target locus. All of the plasmids carry a length of DNA, ranging from 2.5 to 5 kb, homologous to the exon 3 target region in addition to the human promoter and exon 1 sequences and the mouse exon 2 region (Fig. 2B). Each of the plasmids, after integration into the nonreverting, hemizygous target locus, is capable of creating a promoter, exon 1 through 9 configuration and of generating a functional *HPRT* gene. The functional gene can be generated with or without an accompanying repair of the introduced gaps; however, in the absence of gap repair, the crossover events which produce a functional gene product are more restricted. Figure 2C illustrates the recombinant locus expected to be generated when the targeting DNA either has no gaps or has been gap repaired. To provide a measure of the targeting frequencies, functional recombinants were selected for in HAT medium, and then the number of HAT$^r$ colonies obtained after 2 weeks of selection was counted.

Figure 1 depicts the targeting plasmids used for these studies and their configurations relative to the *hprt*$^-$ locus. Table 1 summarizes the results of the experiments done to measure the relative targeting frequencies. Digesting our plasmids with enzymes that created double-strand breaks in the region of DNA homologous to the target locus had the

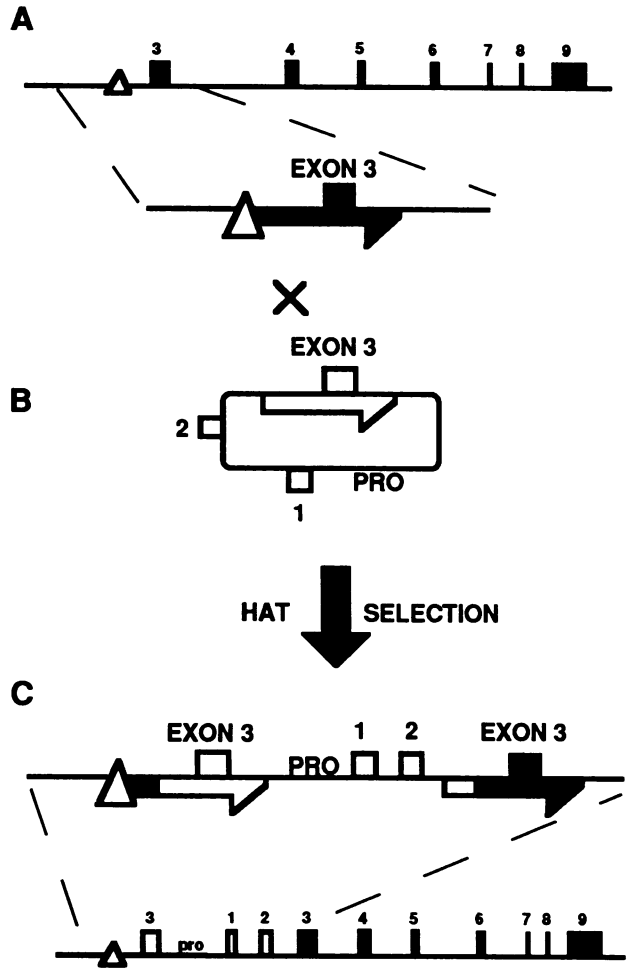


FIG. 2. Gene targeting reaction. (A) The E-14TG2a *hprt*$^-$ gene. Cell line E-14TG2a carries a deletion (△) of the promoter and first two exons of the nine-exon *HPRT* gene. In the enlargement, the exon 3 target region of 5 kb in length is indicated by the arrow and exon 3 is indicated by the box. (B) Depiction of the targeting plasmid pNMR122. The 5-kb region of DNA homologous to the target locus is depicted by the arrow. Boxes 1 and 2 represent exons 1 and 2; exon 3 is indicated. PRO, promoter region. (C) Expected recombinant locus. The insertion event creates a duplication of the 5-kb target region (arrows) separated by the remainder of the plasmid sequences. This creates the promoter, exon 1 through 9 configuration necessary to generate the HAT$^r$ phenotype. A crossover in intron 2 is depicted for illustrative purposes.

expected stimulatory effect on targeting frequency. We obtained HAT$^r$ recombinants at an average frequency of $0.06 \times 10^{-6}$ (Table 1) when the full-length plasmids pNMR122 and pNMR133 were introduced in their uncut (supercoiled) forms, but this average jumped over 30-fold, to $2.0 \times 10^{-6}$ (Table 1), when these same plasmids were linearized within the region of homology prior to electroporation. This finding confirms that double-strand breaks in the region of homology are potent stimulators of the (targeting) integrative recombination reaction, presumably because the generated free ends are more efficient at initiating the recombination event that results in the HPRT$^+$ phenotype.

Less expected was our finding that double-strand gaps, as well as extra heterologies compared with the target locus, had essentially no negative effect on the targeting frequencies compared with the frequency ($2.0 \times 10^{-6}$) obtained with the plasmids carrying a simple double-strand break. The frequencies obtained with the gapped plasmids, predigested to generate free DNA ends in the region of homology, averaged $1.4 \times 10^{-6}$ (Table 1). Thus, double-strand breaks and gaps are essentially equivalent as initiators of the integrative recombination reaction. Sequence heterologies located near the double-strand break points, ranging from a 4-bp insertion to deletions of 200 and 600 bp, did not reduce the targeting frequency and therefore did not hinder the recombination reaction.

Interestingly, the full-length plasmids in their uncut forms produced several times more HAT$^r$ colonies (averaging $0.06 \times 10^{-6}$) than did the uncut forms of the deletion-containing plasmids (averaging $0.01 \times 10^{-6}$), yet their respective cut and gapped forms gave almost identical frequencies ($2.0 \times 10^{-6}$ versus $1.4 \times 10^{-6}$). This may reflect the fact that uncut deletion-containing plasmids lacking part or all of exon 3 can produce HAT$^r$ colonies only when a crossover event occurs 5′ to the plasmid-carried deletions. This restriction does not apply to gap-repaired or full-length plasmids.

Although, in general, the gapped plasmids generated recombinants at essentially the same frequency as the full-length cut plasmids, the targeting frequency obtained with plasmid pD2500, when linearized at the unique *Xba*I site to create a 2,500-bp gap in the region of homology, was lower than any other plasmid. This plasmid carries only half the length of homology to the target locus carried by the full-length plasmids pNMR122 and pNMR133.

**Double-strand gaps are repaired using chromosomal target sequences as template during the recombination reaction.** The second facet of the break repair model that we tested was its prediction that double-strand gaps are repaired, thereby producing regions of gene conversion, during the homologous integration reaction. To test this aspect of the model, genomic DNAs isolated from several of the HAT$^r$ recombinants generated in the experiments described above were analyzed by Southern blot hybridizations. If the predictions of the double-strand break repair model are met, we would expect the same recombinant locus to be generated by all plasmid constructs tested. The expected locus is depicted in Fig. 3A: a duplication of the entire 5-kb target region (large arrows) separated by the remainder of the plasmid-derived sequences.

The HAT$^r$ recombinant loci generated in the experiments described in the previous section were characterized by digesting the genomic DNAs with the four enzymes indicated in Fig. 3A. Southern blots of the digested DNAs were hybridized to either the intron 3 probe, a 250-bp *Rsa*I fragment from intron 3, or the exon 3 probe, a 450-bp fragment generated by the polymerase chain reaction that corresponds to the entire exon 3 element and 100 bp from 5′ to it (Fig. 3A). *Hin*dIII digestion of the recombinants that are repaired as predicted by the break repair model will generate two fragments that hybridize to both probes: a 7-kb fragment from the target locus and a 12-kb fragment that is introduced by the targeting vectors. The larger fragment is diagnostic in that it is equivalent to the length of the plasmid integrated into the target locus. If the insertions, deletions, and gaps carried by the various plasmids are repaired, this fragment will be 12 kb. If any of the deletions or gaps are not repaired, this diagnostic band will be smaller by the size of the deletion or gap. If the 4-bp insertion is not repaired, there should be only one hybridizing fragment 19 kb in length. Likewise, *Bam*HI digestion should produce two hybridizing bands: one of 9.4 kb, corresponding to the target locus, and one of 6.9

TABLE 1. Targeting frequencies

| Plasmid | Targeting frequency ($10^6$)[a] | | | | | No. of expts[b] |
|---|---|---|---|---|---|---|
| | Uncut[c] | Cut[d] | | | | |
| | | *Hin*dIII | *Xho*I | *Xba*I | *Bgl*II | |
| Full | | | | | | |
| pNMR122 | 0.05 ± 0.01 | 1.5 ± 0.9 | 1.5 ± 1.1 | | | 2, 5, 3 |
| pNMR133 | 0.08 | | 4.0 ± 0.7 | | | 1, 2 |
| Avg | 0.06 ± 0.01 | | 2.0 ± 0.9 | | | 3, 10 |
| Gapped | | | | | | |
| pD200 | 0.01 ± 0.01 | 0.8 ± 0.6 | 0.7 ± 0.5 | | 0.6 ± 0.3 | 2, 2, 2, 2 |
| pD200(H-) | | | 1.5 | | 1.2 | 1, 1 |
| pD600 | ≤0.02 | 0.9 ± 0.8 | | | | 1, 2 |
| pD200D600 | | 2.1 | | | 2.6 | 1, 1 |
| pD2500 | 0.01 ± 0.01 | | | 0.3 ± 0.3 | | 4, 4 |
| Avg | 0.01 ± 0.01 | | 1.4 ± 0.8[e] | | | 7, 14 |

[a] Average number of HAT$^r$ colonies obtained per $10^6$ electroporated cells ± standard deviation, calculated by averaging results from the number of experiments indicated.
[b] Number of experiments performed with each plasmid, uncut or cut at the various restriction sites, as indicated in the left-to-right order.
[c] Plasmids were introduced as circular molecules.
[d] The restriction enzymes used to linearize or gap the plasmids within the region of homology prior to electroporation are indicated (see Fig. 1).
[e] Average value for all four restriction enzymes.

kb, corresponding to the fragment introduced by the targeting plasmids.

Recombinants generated with the gapped plasmids pD200 and pD600 were further characterized by digesting these genomic DNAs with either *Bal*I or *Pvu*II then hybridizing them to the exon 3 probe. These enzymes are expected to produce double-intensity bands of 1.8 and 3.0 kb, respectively, if the predicted locus is generated. If the deletions and gaps on the transforming plasmids are not repaired, these enzymes should produce two fragments which hybridize to the probe: one of 1.8 or 3.0 kb, as already described, and one which should be smaller by an amount equivalent to the size of the deletion or gap in the introduced plasmid (200 or 600 bp).

Figure 3B presents a sampling of the results, obtained with plasmid pD2500, that are typical of the results obtained with the simple gapped plasmids. The predicted 7- and 12-kb *Hin*dIII fragments are seen in lanes 6, 8, and 10. The predicted 6.9- and 9.4-kb *Bam*HI bands are seen in lanes 5, 7, and 9. The starting locus is seen in lanes 1 and 2 for comparison.

We examined a total of 47 recombinants that had been generated with the simple gapped plasmids (Table 2). Nineteen were generated with plasmid pD2500, linearized at the unique *Xba*I site to create a 2,500-bp gap in the region of homology relative to the target locus; 19 were generated with plasmid pD600, linearized at the *Hin*dIII site to create a 600-bp gap; and 9 were generated with pD200, linearized at the unique *Bgl*II site to create a 200-bp gap. (The structures of these plasmids are shown in Fig. 1.) All 47 recombinants analyzed revealed the fragments predicted by the break repair model, the 7- and 12-kb *Hin*dIII fragments and the 6.9- and 9.4-kb *Bam*HI fragments, upon hybridization to the intron 3 probe. Digestion of the 9 pD200 and the 19 pD600 recombinants with *Bal*I likewise produced the predicted 1.8-kb double-intensity bands upon hybridization to the exon 3 probe. Since the gapped plasmids all carried fragments smaller than the predicted sizes by the amounts of their respective gaps, yet in all cases the full-length bands resulted, the gaps must have been repaired. All of the simple gapped plasmids produced the same recombinant locus, indicating that double-strand gaps are efficiently repaired using the chromosomal target sequences as template during the homologous integration reaction.

We examined one HAT$^r$ recombinant that was generated when plasmid pD2500 was introduced as an uncut (supercoiled) molecule and found the *HPRT*$^+$ locus produced to be highly unusual. The 7-kb *Hin*dIII fragment associated with the target locus is present (Fig. 3B, lane 4), but it is the only *Hin*dIII fragment that hybridized to our probes. *Bam*HI digestion revealed only a single hybridizing band of 13 kb (Fig. 3B, lane 3). Neither the 7-kb *Hin*dIII nor the 13-kb *Bam*HI band hybridized to a bacterial vector-specific probe (data not shown), indicating that the entire pD2500 plasmid is not inserted into the genome of this recombinant. Rather, it appears as if the sequences necessary to confer HAT resistance to the cells may have inserted into the target locus by a semi-illegitimate recombination reaction. A homologous crossover event between plasmid sequences located 5′ to the deletion and the identical chromosomal sequences, accompanied by a nonhomologous recombination event between plasmid sequences located 5′ to the promoter element and upstream chromosomal sequences, could have generated the *HPRT*$^+$ locus that we observed.

**Heterologies near the breaks and gaps are often coconverted.** The double-strand break repair model makes no specific predictions as to the fates of any small heterologies in otherwise homologous sequences carried by the incoming DNA. However, one likely extension of the model is that heterologies found close to the initiating double-strand break or gap may be lost by exolytic degradation prior to the gap repair reaction. Such heterologies could also be lost if they are incorporated into heteroduplex DNA and repaired after the integration reaction. If the incoming heterologies are degraded, all recombinant products will be converted to target locus sequences; if they are incorporated into heteroduplex DNA, conversion could occur in either direction, generating products that retain either the incoming or the target sequences. To examine the possibility of exolytic degradation (or heteroduplex DNA formation) during the

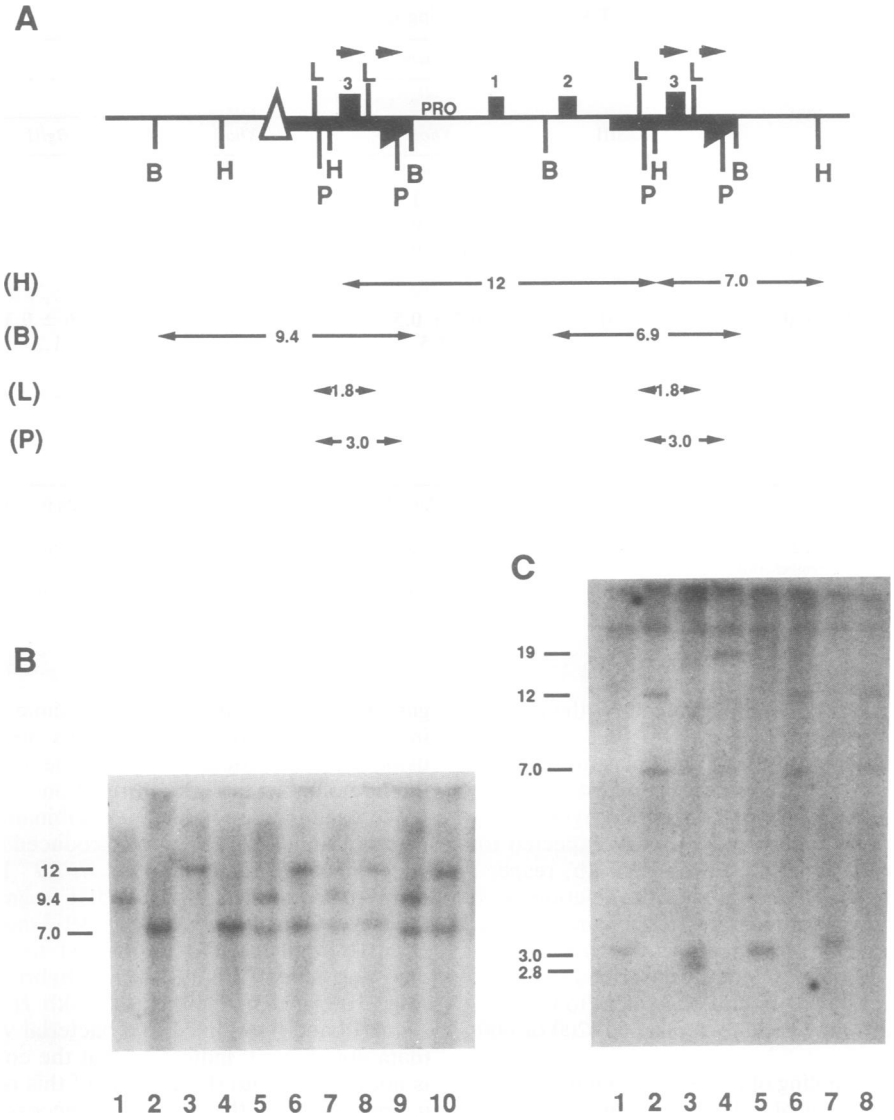


FIG. 3. Southern blot hybridization analyses of the recombinants. (A) The predicted recombinant locus: a duplication of the entire 5-kb target region (large arrows) generated by the repaired insertion of the incoming cut or gapped plasmids into the target locus, separated by the remainder of the plasmid-carried sequences. The four characterizing restriction enzymes indicated generate the depicted fragments that hybridize to the probes. Regions used as probes are shown as small arrows above the locus. Abbreviations: H, *Hin*dIII; B, *Bam*HI; L, *Bal*I; P, *Pvu*II. (B and C) Representative Southern blots of the recombinant loci generated. (B) Recombinants generated with plasmid pD2500 hybridized to the intron 3 probe. Lanes: 1, 3, 5, 7, and 9, *Bam*HI digestions; 2, 4, 6, 8, and 10, *Hin*dIII digestions; 1 and 2, *hprt*$^-$ target locus; 3 and 4, HAT$^r$ recombinant produced with uncut plasmid; 5 through 10, recombinants generated with *Xba*I-linearized plasmid. (C) Recombinants generated with plasmid pD200(H−). Lanes: 1, 3, 5, and 7, *Pvu*II digestions; 2, 4, 6, and 8, *Hin*dIII digestions; 1 through 4, recombinants generated with *Xho*I-linearized plasmid; 5 through 8, recombinants generated with *Bgl*II-gapped plasmid.

gap repair/integration reaction, we introduced heterologies into the region of homology carried by the targeting plasmids then determined the fate of these markers.

To conduct these studies, plasmids pD200, pD200(H−), and pD200D600 were digested with various restriction enzymes to produce linear molecules carrying deletions and insertions in the region of homology. These plasmids, and their configurations relative to the *hprt*$^-$ target locus, are depicted in Fig. 1; their relative targeting frequencies are listed in Table 1. Resulting recombinants were analyzed by genomic Southern blot hybridization using the characterizing enzymes and probes described in the previous section (illustrated in Fig. 3A). A summary of the Southern blot hybridization data obtained with these plasmids is presented in Table 2.

Plasmid pD200, when linearized at a site other than the *Bgl*II site, carries a 200-bp deletion in the region of homology and can be used to examine the fate of such a heterology when it is located either 200 bp from the double-strand break (after linearization with *Hin*dIII) or when it is located 800 bp from the break point (after linearization with *Xho*I). Plasmid pD200(H−) carries a 200-bp deletion and a 4-bp insertion, identified by a destroyed *Hin*dIII site, in the region of homology. Digestion with *Bgl*II generates a gapped plasmid

TABLE 2. Summary of hybridization analyses

| Plasmid[a] | No. of recombinants | | | Distance (bp)[b] |
|---|---|---|---|---|
| | Examined | Fully corrected[c] | Retain het.[d] | |
| pD2500 uncut | 1 | 0 | NA | |
| pD2500/X | 19 | 19 | NA | |
| pD600/H | 19 | 19 | NA | |
| pD200/G | 9 | 9 | NA | |
| Total | 48 | 47 | | |
| pD200/H | 5 | 4 | 1 | 200 (D200) |
| pD200/O | 5 | 2 | 3 | 800 (D200) |
| pD200(H-)/G | 5 | 5 | 0 | 200 (I4) |
| pD200(H-)/O | 5 | 4 | 1 | 600 (I4 + D200) |
| pD200D600/H | 4 | 4 | 0 | 200 (D200) |
| pD200D600/G | 4 | 4 | 0 | 200 (D600) |
| Total | 28 | 23 | 5 | |

[a] Plasmid transformed into ES cells to generate the HAT$^r$ recombinants, using the indicated enzyme to linearize the plasmid: /G, *Bgl*II; /H, *Hin*dIII; /O, *Xho*I; /X, *Xba*I.

[b] The distance between the double-strand break or gap and the nearest heterology is indicated together with the nature of the heterology (in parentheses). For example: 600 (I4 + D200) indicates that there are two heterologies, an insertion of 4 bp and a deletion of 200 bp, with the closest (4-bp insertion) being located 600 bp from the double-strand break.

[c] Recombinants that carry the predicted locus depicted in Fig. 3A with all gaps, deletions, and insertions converted to chromosomal configuration.

[d] Recombinants that retained plasmid-carried heterologies (indicated in the rightmost column). NA, not applicable. Simple gaps are not heterologies; they are all repaired. The uncut plasmid has no initiating double-strand break, and the recombinant produced is highly unusual.

carrying a 4-bp insertion located 200 bp 3′ to a 200-bp double-strand gap; digestion with *Xho*I generates a linear plasmid carrying both a 4-bp insertion, located 600 bp 5′ to a double-strand break, and a 200-bp deletion, located 800 bp 5′ to the break. Plasmid pD200D600 carries a 600-bp deletion located 200 bp 3′ to a 200-bp gap when digested with *Bgl*II; when digested with *Hin*dIII, it carries a 200-bp deletion located 200 bp 5′ to a 600-bp gap.

A detailed analysis of the results obtained with one of these plasmids, pD200(H−), is presented to illustrate the data. As shown in Fig. 1, plasmid pD200(H−) carries a 200-bp deletion and a 4-bp insertion in intron 2, within the 5-kb region of homology. Five recombinants were analyzed that were obtained when this plasmid was linearized at the *Bgl*II site to create a 200-bp gap in the region of homology. The gap was repaired in all five (as described in the previous section), but none of the five retained the 4-bp insertion that was located 200 bp 3′ to the generated 200-bp gap. *Hin*dIII digestion of these genomic DNAs revealed both the 7- and 12-kb bands upon hybridization to the exon 3 probe (Fig. 3C, lanes 6 and 8), indicating that the 4-bp insertion was removed to restore the *Hin*dIII site. Had the insertion not been removed, a single 19-kb *Hin*dIII fragment would have resulted. *Pvu*II digestion of these same DNAs (Fig. 3C, lanes 5 and 7) revealed only the double-intensity 3.0-kb band upon hybridization, confirming that the 200-bp gap had also been repaired during the integration reaction. (The high-molecular-weight band seen in all the lanes of Fig. 3C was commonly seen when the exon 3 probe was used. As the nature of this extraneous hybridizing DNA does not affect our results, its origin was not investigated further.)

We also examined five recombinants that were produced when plasmid pD200(H−) was linearized at the unique *Xho*I site in exon 3, generating a plasmid carrying a double-strand break located 600 bp 3′ to a 4-bp insertion and 800 bp 3′ to a 200-bp deletion. One of the five recombinants retained both the insertion and deletion heterologies. DNA from this recombinant (Fig. 3C, lanes 3 and 4) hybridized to a single 19-kb *Hin*dIII band and two *Pvu*II bands of lengths 3.0 and 2.8 kb. The other four recombinants analyzed had lost both the 4-bp insertion and the 200-bp deletion (Fig. 3C, lanes 1 and 2) and thus contained the locus depicted in Fig. 3A; *Hin*dIII digestion of the DNAs revealed the 7- and 12-kb bands and *Pvu*II digestion revealed the double-intensity 3.0-kb band upon hybridization to the exon 3 probe.

The results of the Southern blot hybridization experiments (Table 2) can be summarized as follows. Seventy of 76 (92%) of the recombinants analyzed carried the same fully corrected targeted locus, indicating that plasmid-carried gaps and heterologies were usually lost during the integration reaction. All simple double-strand gaps (47 of 47 analyzed) were repaired using the chromosomal sequences as template during the integration reaction. Five of a total of 28 (17%) recombinants analyzed that were generated with plasmids carrying heterologies to the target locus retained the plasmid-carried heterologies. Three of them, obtained with plasmid pD200 linearized at the *Xho*I site, carried a 200-bp deletion located 800 bp from the double-strand break; one of them, obtained with plasmid pD200(H−) linearized at the *Xho*I site, carried and retained both a 4-bp insertion located 600 bp from the break and a 200-bp deletion located 800 bp from the break; and one of them, obtained with plasmid pD200 linearized at the *Hin*dIII site, carried a 200-bp deletion located only 200 bp from the double-strand break. Thus, the heterologies carried by the incoming DNA are frequently lost, but they can be retained and such retention is more likely when the heterology is distanced from the double-strand break point. The data obtained with plasmid pD200(H−) further revealed that both markers were either lost (four of five recombinants analyzed) or retained (one of five) together. None of the recombinants analyzed showed correction toward the plasmid-carried sequences.

## DISCUSSION

We have tested several facets of the yeast-derived double-strand break repair model of homologous recombination (28) in a mammalian system and have found the model able to explain the recombinant loci generated in our gene targeting experiments. The frequency at which a targeting plasmid integrated into an homologous genomic (target) locus to produce HAT$^r$ recombinants was 33- to 140-fold higher when the plasmid was digested to produce either a double-strand break or gap in the region of DNA homologous to the target locus prior to cellular introduction than when it was introduced in its uncut (supercoiled) form. This result is consistent with models that depict homologous recombination as initiating at free DNA ends located in the region of homology. This result also supports current protocols for gene targeting that recommend linearization of the plasmids within the region of homology prior to cellular introduction.

Double-strand gaps, ranging in size from 200 to 2,500 bp, were efficiently repaired using the chromosomal target sequences as template during the integration reaction. These gaps did not significantly affect the frequency of the integration event, except possibly to the extent that they reduced the amount of homology available. The targeting frequency obtained with pD2500, which carries only half the length of homology to the target locus as the other plasmids, was at least twofold lower than the frequency calculated for any

other plasmid. This result is in agreement with the previously published finding that there is a correlation between length of available homology and targeting frequency (22, 29).

We found that a 4-bp insertion and deletions of 200 and 600 bp within the region of homology carried by targeting plasmids had no negative effect on the targeting frequencies. This result suggests that the plasmid-carried heterologies do not interfere with the homologous alignment necessary for a crossover event. We also found that such heterologies were usually absent from the resulting targeted chromosome (23 of 28 recombinants analyzed), suggesting that they were often removed from the targeting plasmids prior to the repair/integration reaction. When a plasmid was linearized at a site located 800 bp from a 200-bp deletion, three of the five colonies examined retained the 200-bp deletion. When the same plasmid, or a similar plasmid (pD200D600), was linearized at a site located only 200 bp from a deletion, only 1 of 13 colonies examined retained the deletion. Thus, the further a heterologous marker is located from an introduced double-strand break point, the more likely it is to be retained. The recombinants obtained with plasmid pD200(H−) showed that either both the 200-bp deletion and the 4-bp insertion carried by this plasmid were retained (one of the five colonies examined) or neither marker was retained (four of the five colonies examined). The recombinants obtained with plasmid pD200D600 showed correction of both deletions, one of 200 bp located 5′ to one linearization site and one of 600 bp located 3′ to the other linearization site, indicating that there is no (5′ or 3′) directional bias in the repair system. None of the recombinants showed correction toward the plasmid-carried sequence heterologies. Together, these data show that the plasmid-carried heterologies, when not more than 800 bp from a break or gap, are more often removed prior to the gap repair reaction than incorporated into a recombinant. The simplest explanation of this loss, and the joint loss of double heterologies, is that gaps are often extended exolytically during the integration reaction.

Our results strongly support the capability of the double-strand break repair model to depict the homologous recombination reactions that facilitate insertional gene targeting in cultured mammalian cells. Although the molecular events which yield targeted recombinants in mammalian cells appear to be the same as those that occur in yeast cells, there are still substantial differences in the ways that the two organisms recombine exogenous sequences into their genomes. Yeast cell transformation studies have shown that transformed DNA always integrates into a homologous chromosomal (target) locus (9), whereas studies conducted in mammalian cells have shown that transformed DNA recombines with a homologous target locus in only 1 of 1,000 transformed cells (24, 30). Since the break repair model is consistent with the molecular products of recombination found in both systems, the difference between the two systems, resulting in a 1,000-fold disparity in targeting frequencies, must lie in some other aspect of the overall transformation/gene targeting process.

A practical application of our results concerns the selection of primers for use in polymerase chain reaction amplification to find homologous (targeted) recombinants in mammalian cells (14). If a gap is included in an integrating (O-type) targeting vector, sequences from within this gap may be chosen as one primer binding site, and sequences from another region of the plasmid may be chosen as the other primer binding site. Only homologous recombinants will repair the gap to produce the full-length (starting) plasmid upon integration into the genome and thus place the two primer binding sites within an amplifiable distance. This primer choice scheme will facilitate the design of insertional plasmids for the in-out targeting procedure (32). Less encouraging are our results showing that plasmid-carried heterologies are often lost during the homologous integration reaction. However, we have found that the further the heterologies are located from the double-strand gap or break point, the more likely they are to be retained. Thus, loss of heterologies should not prevent the achievement of subtle modifications of the mammalian genome by the in-out procedure.

## ACKNOWLEDGMENTS

We thank Mike Cox, Steve Barclay, and members of our laboratory group for critical readings of the manuscript.

This work was supported by grants GM20069 and HL37001 from the National Institutes of Health.

## REFERENCES

1. **Boggs, S. S., R. G. Gregg, N. Borenstein, and O. Smithies.** 1986. Efficient transformation and frequent single-site, single-copy insertion of DNA can be obtained in mouse erythroleukemia cells transformed by electroporation. Exp. Hematol. **14**:988–994.
2. **Bollag, R. J., and R. M. Liskay.** 1988. Conservative intrachromosomal recombination between inverted repeats in mouse cells: association between reciprocal exchange and gene conversion. Genetics **119**:161–169.
3. **Bollag, R. J., A. S. Waldman, and R. M. Liskay.** 1989. Homologous recombination in mammalian cells. Annu. Rev. Genet. **23**:199–225.
4. **Brenner, D. A., A. C. Smigocki, and R. D. Camerini-Otero.** 1985. Effect of insertions, deletions, and double-strand breaks on homologous recombination in mouse L cells. Mol. Cell. Biol. **5**:684–691.
5. **Brenner, D. A., A. C. Smigocki, and R. D. Camerini-Otero.** 1986. Double-strand gap repair results in homologous recombination in mouse L cells. Proc. Natl. Acad. Sci. USA **83**:1762–1766.
6. **Capecchi, M. R.** 1989. Altering the genome by homologous recombination. Science **244**:1288–1292.
7. **Cox, M. M., and I. R. Lehman.** 1987. Enzymes of general recombination. Annu. Rev. Biochem. **56**:229–262.
8. **Doetschman, T., R. G. Gregg, N. Maeda, M. L. Hooper, D. W. Melton, S. Thompson, and O. Smithies.** 1987. Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature (London) **330**:576–578.
9. **Hinnen, A., J. B. Hicks, and G. R. Fink.** 1978. Transformation of yeast. Proc. Natl. Acad. Sci. USA **75**:1929–1933.
10. **Holliday, R.** 1964. A mechanism for gene conversion in fungi. Genet. Res. (Cambridge) **5**:282–304.
11. **Hooper, M., K. Hardy, A. Handyside, S. Hunter, and M. Monk.** 1987. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature (London) **326**:292–295.
12. **Jasin, M., J. de Villiers, F. Weber, and W. Schaffner.** 1985. High frequency of homologous recombination in mammalian cells between endogenous and introduced SV40 genomes. Cell **43**:695–703.
13. **Jasin, M., and P. Berg.** 1988. Homologous integration in mammalian cells without target gene selection. Genes Dev. **2**:1353–1363.
14. **Kim, H. S., and O. Smithies.** 1988. Recombinant fragment assay for gene targeting based on the polymerase chain reaction. Nucleic Acids Res. **16**:8887–8903.
15. **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.

16. **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.
17. **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.
18. **Meselson, M. S., and C. M. Radding.** 1975. A general model for genetic recombination. Proc. Natl. Acad. Sci. USA **72**:358–361.
19. **Orr-Weaver, T. L., A. Nicolas, and J. W. Szostak.** 1988. Gene conversion adjacent to regions of double-strand break repair. Mol. Cell. Biol. **8**:5292–5298.
20. **Orr-Weaver, T. L., and J. W. Szostak.** 1985. Fungal recombination. Microbiol. Rev. **49**:33–58.
21. **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.
22. **Shulman, M. J., L. Nissen, and C. Collins.** 1990. Homologous recombination in hybridoma cells: dependence on time and fragment length. Mol. Cell. Biol. **10**:4466–4472.
23. **Smith, A. G., J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, and D. Rogers.** 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature (London) **336**:688–690.
24. **Smithies, O., R. G. Gregg, S. S. Boggs, M. A. Koralewski, and R. S. Kucherlapati.** 1985. Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. Nature (London) **317**:230–234.
25. **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.
26. **Song, K.-Y., F. Schwartz, N. Maeda, O. Smithies, and R. Kucherlapati.** 1987. Accurate modification of a chromosomal plasmid by homologous recombination in human cells. Proc. Natl. Acad. Sci. USA **84**:6820–6824.
27. **Subramani, S.** 1986. Rescue of chromosomal T-antigen sequences onto extrachromosomally replicating, defective simian virus 40 DNA by homologous recombination. Mol. Cell. Biol. **6**:1320–1325.
28. **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.
29. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51**:503–512.
30. **Thomas, K. R., K. R. Folger, and M. R. Capecchi.** 1986. High frequency targeting of genes to specific sites in the mammalian genome. Cell **44**:419–428.
31. **Thompson, S., A. R. Clark, A. M. Pow, M. I. Hooper, and D. W. Melton.** 1989. Germline transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell **56**:313–321.
32. **Valancius, V., and O. Smithies.** 1991. Testing an "in-out" targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. Mol. Cell. Biol. **11**:1402–1408.
33. **Williams, R. L., D. J. Hilton, S. Pease, T. A. Willson, C. L. Stewart, D. P. Gearing, E. F. Wagner, D. Metcalf, N. A. Nicola, and N. M. Gough.** 1988. Myeloid leukemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature (London) **336**:684–687.