The Role and Fate of DNA Ends for Homologous Recombination in Embryonic Stem Cells

PAUL HASTY, JAIME RIVERA-PÉREZ, AND ALLAN BRADLEY*

Institute for Molecular Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030

We have analyzed the gene-targeting frequencies and recombination products generated by a series of vectors which target the *hprt* locus in embryonic stem cells and found the existence of alternative pathways that depend on the location of the double-strand break within the vector. A double-strand break in the targeting homology was found to increase the targeting frequency compared with a double-strand break at the edge of or outside the target homology; this finding agrees with the double-strand break repair model proposed for *Saccharomyces cerevisiae*. Although a double-strand break in the homology is important for efficient targeting, observations reported here suggest that the terminal ends are not always directly involved in the initial recombination event. Short terminal heterologous sequences which block the homologous ends of the vector may be incorporated into the target locus. A modification of the double-strand break repair model is described to account for this observation.

Gene targeting, i.e., homologous recombination between introduced and chromosomal DNA, is widely used to mutate genes in embryonic stem (ES) cells (4, 6–8, 15, 16, 19, 32, 36, 41). In mammalian cells, integration into the homologous locus is inefficient compared with random nonhomologous integrations (4, 15, 22). There are many variables which influence the gene-targeting frequency, including but not limited to the genetic locus (4, 15), the amount of homologous sequences (7, 28, 35), the positive selection cassette (32), and the vector design (8).

Two types of vectors, insertion and replacement, have been used for gene targeting in mammalian cells (8, 35). Here, we define an insertion vector as one with a double-strand break in the homology and a positive selection marker located either inside the homology or in the bacterial plasmid. We define a replacement vector as one with a double-strand break either at the edge of or outside the target homology and a positive selection marker located within the target homology. We have shown that insertion vectors target at a 5- to 10-fold-higher frequency than do replacement vectors with the same homologous sequences (8). One explanation for the difference in the targeting frequency of our insertion and replacement vectors may be the length of uninterrupted homology. Interruption of homology with heterologous sequences has been shown to decrease some forms of homologous recombination. Increasing the length of the heterologous interruption has been shown to progressively decrease the frequency of extrachromosomal homologous recombination between cointroduced plasmids in mouse L cells (2). Inter- and intramolecular extrachromosomal and intrachromosomal homologous recombination also appear to be more efficient with uninterrupted homology (38, 39). However, increasing the size of the heterologous interruption in a replacement vector did not affect the gene targeting frequency in ES cells (16). Another explanation to account for the different gene-targeting frequencies of the two vector types is that the ends of the vector may affect the targeting frequency. Insertion vectors have a double-strand

break in the homology, while replacement vectors have a double-strand break at either or both edges of homology.

The free ends of the recombination substrate were first shown to affect the homologous recombination frequency in yeast cells (23). A double-strand break or gap in the homology of the introduced DNA increased the gene-targeting frequency, compared with no break or a break outside the vector homology. On the basis of this observation, the double-strand break repair model was proposed, which describes the complementary homologous ends invading the target locus in association with DNA repair synthesis (23, 25, 34). Resolution of the recombinant intermediate results in either integration of the entire vector (crossover event, conservative product) or the formation of heteroduplex DNA (noncrossover event, nonconservative product). With the absence of homologous ends, a less efficient pathway was utilized that required, at least in part, a different enzymatic pathway (23). A double-strand break in the *ARG4* promoter region has been shown to initiate meiotic homologous recombination most efficiently when 3' overhangs are produced at the ends (27, 33). Thus, the ends of the DNA substrates affect the frequency and pathway for gene targeting and meiotic recombination in yeast cells.

In mammalian cells, the ends also affect extrachromosomal recombination. An increase in the frequency of homologous recombination which results in either a conservative or a nonconservative product was seen between cointroduced plasmids when a double-strand break or gap was present in the homology of one substrate and no break was in the other substrate. These results support the double-strand break repair model (2, 3, 31). Other reports have found that if both introduced DNAs are linear, either inside the homology or close to the homology, a single-strand annealing pathway results in a nonconservative product that predominates over the double-strand break repair pathway. The breaks appear to be substrates for exonuclease activity which generate complementary single-strand DNA ends that anneal to each other (1, 13, 29). The location of the break does not need to be flanked by homology so long as single-strand homology is exposed after exonuclease digestion (1, 14, 29). A similar nonconservative pathway was observed for intramolecular recombination (5, 12). An alter-

---

native nonconservative pathway which is similar to the single-strand annealing pathway has been described (37). In this model, recombination is stimulated with a break in the substrates (the breaks do not need to be in the homology), and single-strand ends are generated by the unwinding of a helicase. A nonconservative pathway has also been observed for extrachromosomal recombination in *Xenopus* oocytes (17, 18). The ends of the DNA substrate thus affect the frequency and pathway for extrachromosomal recombination for DNA introduced into mammalian cells and *Xenopus* oocytes.

DNA ends may determine the pathway followed for gene targeting in mammalian cells. Insertion vectors integrate via a double-strand break in their homologous sequences through a pathway that may have some similarity to the double-strand break repair model established in yeast cells (6, 8, 24, 30, 35). This is a conservative pathway that allows the recovery of all of the substrates involved in the exchange depending upon resolution of the Holliday junctions (9, 23, 24, 34). The final conservative product is a duplication of homology that is separated by the bacterial plasmid (6, 8). Here, we refer to any conservative pathway that could recover all the homologous substrates as a vector insertion event. In contrast to insertion vectors, the mechanistic details of the integration of a replacement vector with ends that are not in the homology are essentially unknown. The integration of a single replacement vector unit is predicted to be a nonconservative event in which only one complement of the homologous sequences are recovered. Double-reciprocal recombination and gene conversion are two possible nonconservative events that may result in gene replacement (8, 15, 35). However, our analysis of gene targeting products generated with replacement vectors in ES cells has shown that the majority of targeted clones have an integration structure similar to an insertion event even though the double-strand break was not in the homology (7, 8). These insertionlike events may have arisen from extrachromosomal ligation of vector units (40) to form concatemers followed by a nonconservative recombination event or recircularization of a vector unit followed by a conservative recombination event (vector insertion). It is difficult to differentiate these two events. Thus, the final integration pattern may appear the same for insertion and replacement vectors, although the recombination pathway for the insertion vector is conservative and the recombination pathway for the replacement vector may be conservative or nonconservative. Here, we use gene replacement to describe any nonconservative pathway in which at least some of the original homologous substrates are not recovered in the final product; this includes double-reciprocal recombination of a concatemer that replaces homologous sequences in the chromosome yet appears to have integrated by a conservative pathway.

In this report, we analyze the frequency and integration pattern of gene targeting with a series of vectors designed to target the hypoxanthine phosphoribosyltransferase gene, *hprt* (20), in mouse ES cells. Since *hprt* is hemizygous in XY ES cells, targeted clones can be isolated by direct selection in 6-thioguanine (TG) for lack of *hprt* function. Since we are selecting for a loss of function, this assay will recover targeted clones even if recombination occurs with low fidelity. We have developed a series of vectors which contain the same *hprt* homologous sequences but with a variety of modifications. Analysis of the targeting frequency and the integration pattern of these vectors into the target locus has demonstrated that different homologous recombi-

nation pathways are followed when the vector has adjacent homologous ends compared with vectors which lack adjacent homologous ends. We observed a high-frequency pathway that is critically dependent upon adjacent homologous ends, although the very terminal nucleotides do not appear to be involved in the initial strand pairing. Surprisingly, short heterologous sequences that are attached to the adjacent homologous ends can be incorporated into the target locus. In the absence of adjacent homologous ends, targeted integration occurs at a lower frequency which may involve end-to-end ligation of vector units followed by gene replacement of a concatemer or vector insertion of a recircularized unit or units. A model for gene targeting in ES cells that describes the role and fate of the free ends is presented.

MATERIALS AND METHODS

Vector construction. All vectors target *hprt* and contain a *SacI-EcoRI hprt* fragment which includes exons 2 and 3. For all insertion vectors, MClneopA (Stratagene) was inserted into the pTZ (Pharmarcia) polylinker. IV6.8 contains the *hprt* fragment and MClneopA without modifications. IV6.8 was modified to make all other insertion vectors. For IV6.8X-N3, an oligonucleotide (5'-TTGCGGCCGCTAGCG GCCGCAA-3') was inserted into the *XmnI* site in the *hprt* homology to destroy this site and generate *NotI-NheI-NotI* sites. The *XmnI* site in the plasmid is now unique. For IV6.8BN, an oligonucleotide (5'-TCGATGGATCCGCGGC CGCGGATCCA-3') was inserted into the *XhoI* site to destroy this site and generate *BamHI-NotI-BamHI* sites. For IV6.8TK, a 2.3-kb MCITK cassette (19) was inserted into the *XhoI* site; only the 3' *XhoI* site was retained. The replacement vector, RV6.8, contains the same *hprt* homology as IV6.8; however, MClneopA was inserted into the *XhoI* site of exon 3.

Electroporation and tissue culture. The experiments described in Table 1 have been grouped into experiments performed on the same day (experiments A to F). This is necessary since there is some variation in the targeting frequencies from experiment to experiment. Within a group, the relative ratios from one vector to the next are consistent.

The DNA for electroporation was prepared by Triton lysis and banded once in CsCl. DNA was cut with the specified restriction enzyme to completion. Nicked circular and supercoiled DNAs were separated by three successive CsCl gradients. The upper band removed from the gradient is nicked circular DNA, and the lower band is supercoiled DNA. An equimolar aliquot from each sample was cut with a restriction enzyme and run side by side on an ethidium bromide gel and gave equimolar intensity bands, demonstrating that bacterial genomic DNA was not a significant component of the nicked circular fraction. Uncut nicked circular DNA, uncut supercoiled DNA, and linear DNA were run on a gel before and after an electroporation, which demonstrated that nicked circular DNA and the supercoiled DNA remained in their native state and were not affected by the electroporation discharge. Electroporations were performed on AB1 cells (19) cultured as previously described (26). For electroporation, the AB1 cells were resuspended in phosphate-buffered saline at a density of 10^7 cells per ml. DNA (25 μg) was electroporated in 1 ml at 575 V/cm and 500 μF with a Bio-Rad Gene Pulser. After each electroporation, the DNA and cells were incubated for 10 to 20 min at room temperature, and then 10^7 cells were plated onto a 9-cm SNL76/7 feeder plate (19).

Two methods were used to determine G418^r and TG^r. In

TABLE 1. Targeting frequency and Southern analysis for *hprt* vectors[a]

Expt	Vector	Digest[b]	No. of free homologous ends	No. of capped homologous ends	No. of electroporations[c]	Total G418[r]	Total TG[r]	TG[r]/G418[r]	TG[r] clones confirmed by Southern analysis[d]	
									Targeted	Modified
A	IV6.8X-N3	*Xho*I	2	0	10	5,500	96	1/57	15/16	1/16
	IV6.8X-N3	*Sac*I	1	0	10	2,980	11	1/271	5/5	
	IV6.8X-N3	*Xmn*I	0	0	10	7,074	10	1/707	7/8	1/8
B	RV6.8	*Nhe*I	2	0	13	23,868	1,349	1/18	9/9	
	RV6.8	*Sac*I	1	0	15	28,560	201	1/142	4/7	3/7
C	IV6.8	*Xho*I	2	0	10	10,500	108	1/97	19/20	1/20
	IV6.8	*Sac*I	1	0	13	13,602	12	1/1,133	5/5	
	IV6.8	Nicked	0	0	29	34,920	49	1/713	9/9	
	IV6.8	Supercoiled	0	0	26	9,828	3	1/3,276	3/3	
	IV6.8TK	*Xho*I	1	1	14	22,540	30	1/751	ND[e]	ND
	IV6.8TK	*Nhe*I	2	0	15	12,930	291	1/44	9/9	
D	IV6.8	*Xho*I	2	0	18	9,549	193	1/49	ND	ND
	IV6.8BN	*Not*I	0	2	19	11,115	137	1/81	21/24	3/24
	IV6.8TK	*Xho*I	1	1	17	17,314	53	1/327	40/51	11/51
E	IV6.8	*Xho*I	2	0	5	2,620	84	1/31	ND	ND
	IV6.8BN	*Not*I	0	2	5	3,680	80	1/46	ND	ND
	IV6.8TK	*Xho*I	1	1	5	3,540	11	1/322	ND	ND
F	IV6.8	*Xho*I	2	0	5	2,200	26	1/84	ND	ND
	IV6.8BN	*Not*I	0	2	5	2,640	18	1/147	ND	ND
	IV6.8TK	*Xho*I	1	1	5	4,060	12	1/338	ND	ND

[a] All vectors target *hprt*. Since there is some variation in the targeting frequencies from experiment to experiment, each letter indicates a single experiment in which all electroporations were performed at the same time and with a common batch of ES cells under the same conditions. The relative relationship between the targeting frequencies obtained with each vector is constant under these conditions.

[b] The restriction enzyme used to linearize the construct before electroporation.

[c] Number performed in a single experiment; one electroporation is 25 μg of DNA per 10^7 cells.

[d] A complete description of the Southern analysis is presented in the text and shown in Fig. 2 (IV6.8X-N3), Fig. 3 (IV6.8BN), and Fig. 4 (IV6.8TK). Figures for Southern analysis for IV6.8 and RV6.8 have been presented elsewhere (7, 8).

[e] ND, not determined.

both methods, the cells were not replated before selection such that each clone represents a single event as previously described (8). In method 1, 24 h after electroporation, 180 μg of G418 per ml (active ingredient) was added to each 9-cm plate, and the cells were maintained under selection for 10 days in order to select and count G418[r] colonies. TG was then added to a final concentration of 10^{-5} M. The cells were maintained under selection for another 21 days, after which G418[r] TG[r] colonies were counted. In method 2, plates were initially selected in 180 μg of G418 per ml (active ingredient), and after 5 to 6 days, a final concentration of 10^{-5} M TG was added; selection was maintained for another 10 days. For each experiment, method 2 was used for 80 to 90% of the plates and method 1 was used for 10 to 20% of the plates. Methods 1 and 2 always gave the same absolute targeting frequency for each construct.

Southern blot analysis of targeted colonies. For all clones targeted with *hprt* vectors, DNA (5 μg) from TG[r] clones was analyzed with a restriction enzyme digest, separated through 0.7% 0.04 M Tris-acetate–0.001 M EDTA (TAE) gels by electrophoresis, transferred to GeneScreen filters (Dupont), and hybridized to a probe internal to the vector, *hprt* exons 2 and 3 from cDNA cut with *Hinc*II and *Hpa*II (20). A *Bam*HI digest was performed on DNA from TG[r] clones generated from all of the vectors that target *hprt*. In addition to a *Bam*HI digest hybridized to exons 2 and 3, other digests and probes were used on TG[r] clones generated from the following vectors. TG[r] clones generated with IV6.8TK (linearized with *Nhe*I and *Xho*I) were digested with *Hin*dIII and *Nco*I and hybridized to *hprt* exons 2 and 3; the filters were then stripped (100°C for 10 min in 1% sodium dodecyl sulfate and 0.1× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]) and hybridized to MCITK. DNAs from TG[r] clones

generated with IV6.8BN and IV6.8TK (linearized with *Xho*I) were also analyzed with a *Bam*HI-*Xho*I digest hybridized to exons 2 and 3. TG[r] clones generated with IV6.8X-N3 (linearized with *Xho*I, *Sac*I, and *Xmn*I) were analyzed with *Xmn*I and *Nhe*I digests hybridized to exons 2 and 3. These probes were labeled with random-primed labeling, using the conditions specified by the manufacturer (Boehringer Mannheim). Southern blot analysis for RV6.8 has been described elsewhere (8). The targeted *hprt* locus from TG[r] clones can be reliably identified with an internal probe. To ensure that the internal probe was sufficient to identify targeting at the *hprt* locus, an external probe was hybridized to *Nco*I-cut genomic DNA as described previously (6–8).

Polymerase chain reaction (PCR) and sequencing. (i) PCR analysis for IV6.8BN, 5′ duplicate. The upstream sense primer, a (5′-CCTGATTTTATTTCTATAGG-3′), at the intron 2-exon 3 junction and the downstream antisense primer, b (5′-TGCTGCAAGGCGATTAAGTT-3′), in plasmid pTZ were used to amplify exon 3 in the 5′ duplicate with the following conditions in a 20-μl total volume: genomic DNA, 150 ng; AmpliTaq polymerase (Perkin-Elmer Cetus), 1 U; primers, 50 ng of each; deoxynucleoside triphosphates (dNTPs), 250 μM; 10× Cetus buffer (Perkin-Elmer Cetus), 2 μl; dimethyl sulfoxide, 1 μl; and 100 mM MgCl₂, 1.2 μl. Amplification was done in a DNA Thermal Cycler (Perkin-Elmer Cetus) for 1 cycle at 93°C for 6 min, 38 cycles at 93°C for 30 s, 37°C for 30 s, and 70°C for 1 min, and then 1 cycle at 70°C for 15 min.

(ii) Sequencing for IV6.8BN, 5′ duplicate. The PCR product in the PCR buffer was incubated with Klenow enzyme at 21°C for 30 min, incubated at 70°C for 5 min, and then purified by Gene Clean (Bio 101). The purified product was cut with *Eco*RI purified on a 1% TAE gel by electrophoresis,
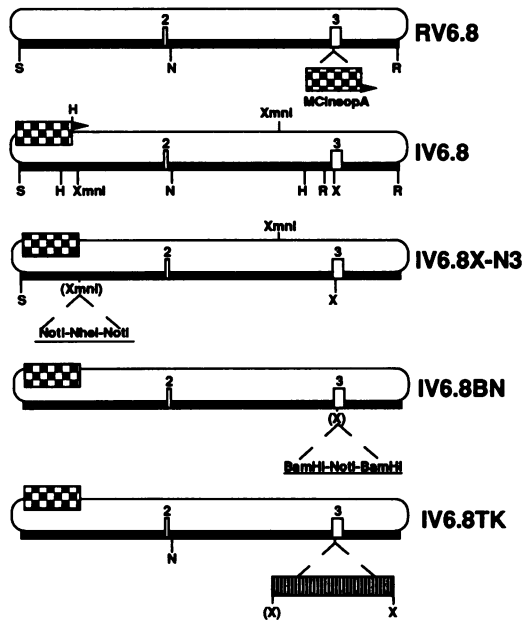
FIG. 1. Vectors that target *hprt*. The thick line is *hprt* homology, and the thin line is plasmid. All vectors were derived from the same 6.8-kb *Sac*I-*Eco*RI fragment of *hprt* with exons 2 and 3 (labeled rectangles) as described in Materials and Methods. RV, replacement vector; IV, insertion vector. The checkered box is the *neo* cassette, MC1neopA, and the arrow illustrates the direction of transcription. The striped box is MCITK. Restriction sites: H, *Hin*dIII; N, *Nhe*I; R, *Eco*RI; S, *Sac*I; X, *Xho*I. Mutated restriction enzyme sites are in parentheses.

and the 1.2-kb fragment was gel purified and cloned into the *Eco*RI and *Hin*cII sites of pKS⁻. Sequencing was performed with an automated Applied Biosystems 373A DNA sequencer, using the reverse universal primer for pKS⁻.

(iii) **PCR analysis of IV6.8BN, 3′ duplicate.** The upstream sense primer, c (5′-TTATGATTACAGTCCAGCCTGGGC-3′), in intron 2 and the downstream antisense primer, d (5′-CTCTGTGAGCCACTATGTAAGTGC-3′), in intron 3 were used to amplify exon 3 in the 3′ duplicate with the following conditions in a 100-μl total volume: genomic DNA, 750 ng; AmpliTaq polymerase, 5 U; primers, 250 ng each; dNTPs, 250 μM; Cetus buffer (Perkin-Elmer Cetus), 10 μl; and 100 mM MgCl₂, 5.5 μl. Amplification was carried out by a DNA Thermal Cycler (Perkin-Elmer Cetus) as follows: 1 cycle at 93°C for 6 min; 35 cycles at 93°C for 30 s, 60°C for 30 s, and 70°C for 1 min; and 1 cycle at 70°C for 15 min.

(iv) **Sequencing for IV6.8BN, 3′ duplicate.** The PCR product was purified with Gene Clean and then cut with *Eco*RI as previously described. The 1.3-kb fragment was purified from a gel and cloned into the *Eco*RI site of pKS⁻. Sequencing was performed as previously described.

RESULTS

The targeting frequency is determined by the position of the linearization site. We have previously shown that the targeting frequency of an insertion vector, IV6.8 (Fig. 1), was up to ninefold greater than that of a replacement vector, RV6.8 (Fig. 1), which contained the same 6.8-kb *hprt* homologous sequences and the same *neo* cassette (8). This observation indicates that the mechanism of homologous recombination was playing a rate-limiting role in the targeting frequency.

There are two potential factors that may account for these different frequencies. First, the homology in IV6.8 is uninterrupted (the *neo* cassette is located in the plasmid), while the homology in RV6.8 is divided into 5.6- and 1.2-kb fragments by the 1.1-kb *neo* cassette. The recombination frequency may be reduced when there is an interruption in the homology. Interrupted homology has been shown to decrease extrachromosomal and intrachromosomal recombination (2, 38, 39). However, increasing the lengths of heterologous sequences in replacement vectors apparently does not affect the targeting frequency (16). Alternatively, the vector ends may affect gene targeting. The insertion vector had a double-strand break in the homology that produced adjacent homologous ends, while the replacement vector had a double-strand break at either one or both edges of homology that produced only one homologous end or two nonadjacent homologous ends. A double-strand break in the homology has been shown to increase the frequency of extrachromosomal recombination in mammalian cells (2, 3, 31) and of gene targeting in yeast (23) and COS1 (10) cells. However, double-strand breaks outside the homology have also been shown to increase extrachromosomal homologous recombination in mammalian cells (1, 14, 29).

To determine whether the ends have a rate-limiting role associated with homologous recombination between introduced and chromosomal DNAs, the targeting frequency of an insertion vector, IV6.8X-N3 (Fig. 1), was measured with a double-strand break inside, at the edge of, or outside the target homology. IV6.8X-N3 was cut at the unique *Xmn*I site in the bacterial plasmid to produce ends in the heterologous sequences, at the *Sac*I site, to produce ends at the homologous-heterologous junction and at the *Xho*I site to produce ends in the homologous sequences. These linearized vectors were electroporated into ES cells and were consecutively selected in G418 for transfection and TG for *hprt* disruption. The targeting frequencies obtained demonstrated that adjacent homologous ends increased the targeting frequency by 5- and 12-fold over ends at the homologous-heterologous junction and in the heterologous sequences, respectively (Table 1, experiment A).

RV6.8, designed to be a replacement vector, was transfected into ES cells to establish whether the interruption in homology had an effect on the targeting frequency and to confirm that the position of the linearization site was a critical factor in obtaining high targeting frequencies. This vector may also be linearized in the homologous sequences to generate an insertion vector. RV6.8 was cut with either *Nhe*I to generate a double-strand break in the homology or *Sac*I to generate a double-strand break at the homologous-heterologous junction. RV6.8 cut in the homology targeted eightfold more frequently than did RV6.8 linearized at the edge of homology (Table 1, experiment B). This observation also suggests that the availability of ends in the homologous sequences is important for a high targeting frequency and that the disruption of homology with the *neo* cassette had a minimal effect on the targeting frequency.

To confirm that the TGʳ clones represented the predicted targeted events, Southern blot analysis was performed on a representative subset of clones generated with each vector type; the results are summarized in Table 1. TGʳ clones generated with IV6.8X-N3 linearized with *Xho*I, *Sac*I, and *Xmn*I showed the integration of the entire vector at the *hprt* locus for 15 of 16, 5 of 5, and 7 of 8 clones, respectively (Fig. 2). The same restriction pattern is observed for all three groups with a *Bam*HI digest hybridized to an internal probe (exons 2 and 3); the 7-kb wild-type fragment is retained, and
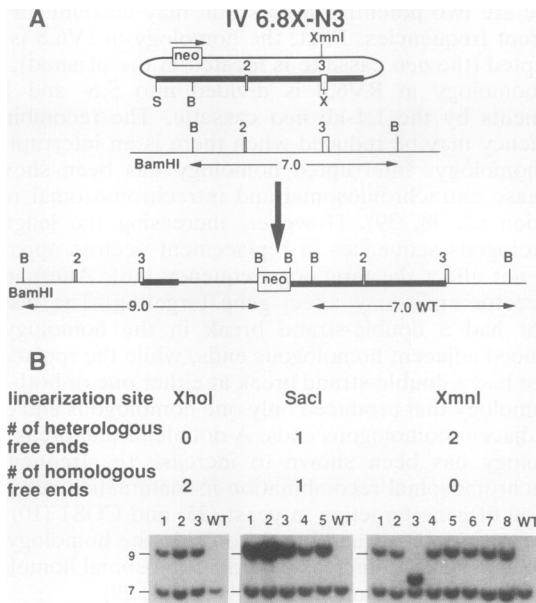
FIG. 2. Southern blot analysis of TG^r clones targeted with IV6.8X-N3. DNA was cut with *Bam*HI and hybridized to exons 2 and 3 (from cDNA). (A) Predicted size fragments for vector insertion seen with a *Bam*HI (B) digest. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid. MC1neopA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles. The length of DNA is in kilobases. B, *Bam*HI; S, *Sac*I; X, *Xho*I; WT, wild type. (B) Southern analysis. IV6.8X-N3 was linearized at the *Xho*I site, the *Sac*I site, and the *Xmn*I site; the TG^r clones derived from each construct are represented. The numbers of homologous and heterologous ends are indicated for each linearization site. Clone 3 generated by the *Xmn*I-linearized construct is a modified insertion event. WT, wild-type (AB1 DNA). The length of DNA is shown at the left in kilobases.

a 9.0-kb fragment is generated. Clones 1 to 3 and 5 from the *Sac*I-cut DNA appear to have integrated more than one unit into the target locus; the 9.0-kb mutant fragment is denser in relation to the 7.0-kb wild-type fragment compared with clone 4. Clone 3, from the *Xmn*I-linearized DNA, does not show the expected vector insertion pattern; instead of the 9.0-kb mutant fragment, a 7.4-kb mutant fragment is seen. This pattern is referred to as a modified vector integration event in which an additional modification occurred that is not understood.

The analysis of the integration pattern for clones targeted with RV6.8 cut at the edge of homology has been previously described. Briefly, most of the clones integrated at least one unit into the target locus to give a pattern that appears to be an insertion event (7, 8). Southern analysis of clones generated from RV6.8 cut with *Nhe*I showed nine of nine to integrate by vector insertion (Table 1).

Targeting with nicked circular and supercoiled DNA. The position of the ends with respect to the homologous sequences of a targeting vector has a rate-limiting effect on the targeting efficiency. Double-strand breaks within the homology increased the number of targeted events by 5- to 12-fold. Furthermore, when a replacement vector is treated to generate a double-strand break at the edge of homology, most of the targeted events have an integration pattern like an insertion event that can only have arisen by intra- or intermolecular ligation of the vectors before integration (8).

Intramolecular ligation of a vector molecule results in a circular molecule. Do circular molecules participate in targeting?

Populations of IV6.8 (Fig. 1) supercoiled plasmid were purified from nicked circular plasmid to determine the targeting efficiency for two forms of uncut circular DNA. More than half of the transfected nicked circular molecules should contain a nick in the *hprt* homology because the nicks are random and 63% of the vector is composed of *hprt* homologous sequences. If just a single-strand break (21) is required to initiate targeting with a high frequency, then the randomly nicked circular DNA should target at half the frequency of a vector with a double-strand break in the homology. DNA was transfected into ES cells to compare the targeting efficiency of molecules with no breaks (supercoiled), single-strand breaks (nicked), and double-strand breaks in homology (*Xho*I) or at the edge of homology (*Sac*I). The results presented in Table 1, experiment C, show that IV6.8 with a double-strand break in homology targeted at a 10-fold-higher frequency than did nicked circular DNA and a 34-fold-higher frequency than did supercoiled DNA. The targeting frequency of nicked DNA was very similar to that obtained by cutting the vector at the edge of homology. Southern blot analysis on the TG^r clones generated with IV6.8 cut with *Xho*I showed 19 of 20 to be targeted as the predicted insertion event and 1 of 20 to have a modified integration structure (8). Southern blot analysis on DNA from TG^r clones generated with nicked circular IV6.8, supercoiled IV6.8, and IV6.8 linearized with *Sac*I showed an insertion-like integration pattern for nine of nine, three of three, and five of five clones analyzed, respectively (Table 1).

Terminal oligonucleotides blocking the ends of a vector are present in recombination products. Adjacent homologous ends are believed to be important initiators for targeting in yeast cells (23), for extrachromosomal recombination in mammalian cells (11, 31), and for targeting in COS1 cells (10); however, the two nonadjacent homologous ends presented at the 5' and 3' aspect of target homology in a replacement vector do not increase the targeting frequency over a replacement vector with heterologous ends (8, 15). Why then should adjacent homologous ends present in an insertion vector stimulate a more efficient gene-targeting reaction than do the nonadjacent homologous ends present in a replacement vector?

To determine the importance of adjacent homologous ends in the formation and resolution of recombination intermediates, IV6.8BN (Fig. 1) was constructed so that the adjacent homologous ends were blocked with 13 bp of heterologous sequence. This terminal heterology can be traced by an encoded *Bam*HI restriction site. We define the terminal heterology attached to the homologous end as a cap, and we define a construct with capped homologous ends as linearized in the homology. IV6.8BN contains an introduced oligonucleotide in exon 3 that may be cleaved with *Not*I to generate a double-strand break in the homology, but both ends are capped with 13 bp of heterology which include a *Bam*HI site and part of a *Not*I site. The targeting frequency of IV6.8BN linearized with *Not*I was marginally reduced compared with that of IV6.8 cut at the analogous position with *Xho*I (Table 1, experiments D to F). Although the availability of adjacent homologous ends is important in obtaining efficient integration, our observations suggest that the very terminal nucleotides are apparently not critical.

Southern blot analysis of the integration events was performed on 24 TG^r colonies generated with IV6.8BN (Table 1). The integration pattern is summarized in Fig. 3A. *Bam*HI
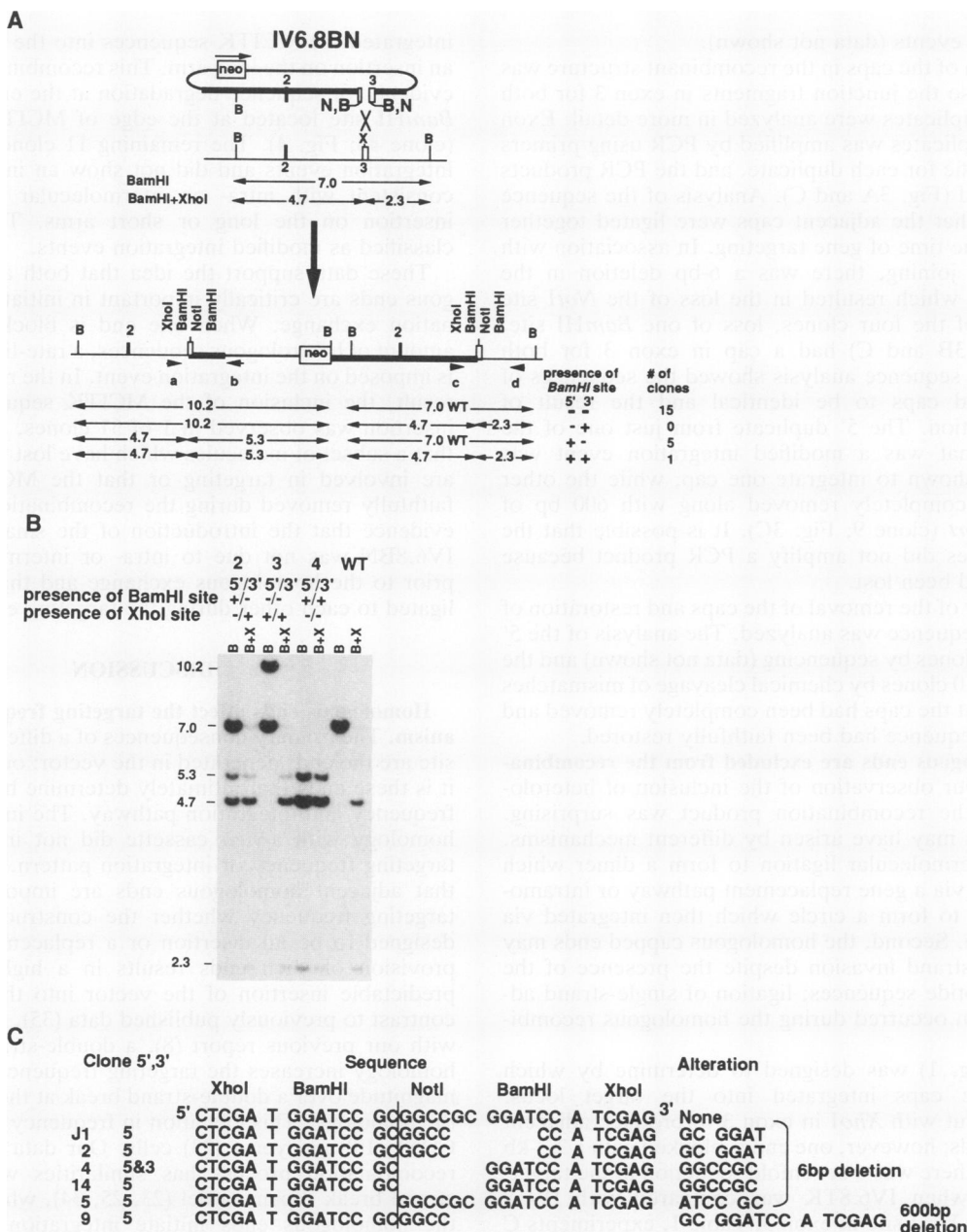
A

IV6.8BN

neo

BamHI ←————— 7.0 —————→
BamHI+XhoI ←—— 4.7 ——→←— 2.3 —→

	presence of BamHI site 5' 3'	# of clones
	− −	15
	− +	0
	+ −	5
	+ +	1

←——————— 10.2 ———————→   ←——————— 7.0 WT ———————→
←——————— 10.2 ———————→   ←———— 4.7 ————→←— 2.3 —→
←— 4.7 —→←—— 5.3 ——→   ←——————— 7.0 WT ———————→
←— 4.7 —→←—— 5.3 ——→   ←———— 4.7 ————→←— 2.3 —→

B

	2	3	4	WT
	5'/3'	5'/3'	5'/3'	
presence of BamHI site	+/−	−/−	+/+	
presence of XhoI site	−/+	+/+	−/−	
	B B+X	B B+X	B B+X	B B+X

10.2 −
7.0 −
5.3 −
4.7 −

2.3 −

C

Clone	5',3'	Sequence					Alteration	
		XhoI	BamHI	NotI	BamHI	XhoI		
		5' CTCGA T	GGATCC	GCGGCCGC	GGATCC	A TCGAG 3'	None	
J1	5	CTCGA T	GGATCC	GCGGCC	CC	A TCGAG	GC GGAT	
2	5	CTCGA T	GGATCC	GCGGCC	CC	A TCGAG	GC GGAT	
4	5&3	CTCGA T	GGATCC	GC	GGATCC	A TCGAG	GGCCGC	6bp deletion
14	5	CTCGA T	GGATCC	GC	GGATCC	A TCGAG	GGCCGC	
5	5	CTCGA T	GG	GGCCGC	GGATCC	A TCGAG	ATCC GC	600bp
9	5	CTCGA T	GGATCC	GCGGCC		GC GGATCC	A TCGAG	deletion

FIG. 3. Southern analysis for TGʳ clones targeted with IV6.8BN. DNA was cut with BamHI and with BamHI plus XhoI and hybridized to exons 2 and 3. (A) Alternative size fragments for vector insertion events were seen with a BamHI digest based upon the incorporation of a BamHI site in the 13-bp cap. At the right of the predicted fragments are indicated the presence or absence of the BamHI site in exon 3 of both 5' and 3' duplicates and the number of clones seen for each (three clones do not match any of these patterns). The thick line is hprt DNA of vector origin, the thin line is hprt DNA of genomic origin, and the line of intermediate thickness is plasmid. MC1neopA is a box labeled neo. Exons 2 and 3 are numbered rectangles. The lines bound by exon 3 and the large B,N represent the 13-bp cap that contains the BamHI site and half a NotI site. The primers used for PCR are represented by small arrowheads; a and b were used to amplify the 5' duplicate, and c and d were used to amplify the 3' duplicate. The length of DNA is in kilobases. B, BamHI; N, NotI; X, XhoI; WT, wild type. (B) Southern analysis. BamHI (B) and BamHI-XhoI (B+X) digests were done on TGʳ clones (clones 2, 3, and 4 are represented). The presence or absence of the BamHI and XhoI sites in exon 3 for both 5' and 3' duplicates is indicated above the lanes. WT, wild type (AB1 DNA). The length of DNA is shown at the left in kilobases. (C) Sequence analysis for all clones that contained a BamHI site in exon 3. Exon 3 in the 5' duplicate was amplified for clones J1, 2, 4, 5, 9, and 14 and in the 3' duplicate for clone 4 in order to evaluate the BamHI mutation. The complete sequence is listed for the oligonucleotide insertion into the XhoI site (top line). A vertical line separates the sequence where NotI cuts. The sequence is the same in both duplicates for clone 4. A deletion of CCGCGG is also possible for clones 4 and 14.

digests and BamHI-XhoI digests were performed to detect the presence or absence of the caps (Fig. 3). Fifteen clones lost the caps and regenerated the XhoI site (representative clone 3; Fig. 3B); however, five clones included the cap in the 5' duplicate and retained the XhoI site in the 3' duplicate (representative clone 2; Fig. 3B). One clone retained the cap in both duplicates (representative clone 4; Fig. 3B). Three clones gave restriction patterns that are described as modi-

fied integration events (data not shown).

The inclusion of the caps in the recombinant structure was not predicted, so the junction fragments in exon 3 for both the 5' and 3' duplicates were analyzed in more detail. Exon 3 from both duplicates was amplified by PCR using primers that were specific for each duplicate, and the PCR products were sequenced (Fig. 3A and C). Analysis of the sequence demonstrated that the adjacent caps were ligated together prior to or at the time of gene targeting. In association with the end-to-end joining, there was a 6-bp deletion in the oligonucleotide which resulted in the loss of the NotI site and, in three of the four clones, loss of one BamHI site. Clone 4 (Fig. 3B and C) had a cap in exon 3 for both duplicates, and sequence analysis showed the sequences of the two altered caps to be identical and the result of end-to-end ligation. The 5' duplicate from just one of the three clones that was a modified integration event was amplified and shown to integrate one cap, while the other cap had been completely removed along with 600 bp of downstream hprt (clone 9; Fig. 3C). It is possible that the other two clones did not amplify a PCR product because primer sites had been lost.

The accuracy of the removal of the caps and restoration of the wild-type sequence was analyzed. The analysis of the 5' duplicate of 2 clones by sequencing (data not shown) and the 3' duplicate of 10 clones by chemical cleavage of mismatches (41) showed that the caps had been completely removed and the wild-type sequence had been faithfully restored.

Long heterologous ends are excluded from the recombination product. Our observation of the inclusion of heterologous ends in the recombination product was surprising. These products may have arisen by different mechanisms. The first is intermolecular ligation to form a dimer which then integrated via a gene replacement pathway or intramolecular ligation to form a circle which then integrated via vector insertion. Second, the homologous capped ends may be capable of strand invasion despite the presence of the terminal nucleotide sequences; ligation of single-strand adjacent ends then occurred during the homologous recombination event.

IV6.8TK (Fig. 1) was designed to determine by which mechanism the caps integrated into the target locus. IV6.8TK was cut with XhoI in exon 3 to produce adjacent homologous ends; however, one end is blocked with a 2.3-kb cap, MCITK. There was a sevenfold reduction in the targeting frequency when IV6.8TK was compared with IV6.8 (both cut at the XhoI site in exon 3; Table 1, experiments C to F). To confirm that the 2.3 kb of heterologous sequence did not have a general effect on the targeting frequency, IV6.8TK was cut on the long arm of homology with NheI to give two uncapped adjacent homologous ends. The targeting frequency of IV6.8TK was 17-fold higher when the plasmid was cut at the NheI site compared with the XhoI site (Table 1, experiment C).

Targeted clones generated with IV6.8TK were analyzed for the integration structure (Table 1). IV6.8TK cut at the NheI site showed the predicted vector insertion on the long arm in nine of nine clones tested so that MCITK was in the 5' duplicate (representative clone 1; Fig. 4). The integration structure in targeted clones generated with IV6.8TK cut at the XhoI site was determined for 51 TG^r clones. Thirty-nine clones had a vector insertion pattern that removed the cap so that MCITK sequences were not present (representative clones 50 and 51; Fig. 4); the mutated XhoI site in the vector was regenerated and present in both duplicates, as determined by a BamHI-XhoI digest (data not shown). One clone

integrated the MCITK sequences into the target locus with an insertion on the long arm. This recombinant allele showed evidence of sequence degradation at the ends such that the BamHI site located at the edge of MCITK was removed (clone 46; Fig. 4). The remaining 11 clones were modified integration events and did not show an integration pattern consistent with intra- or intermolecular ligation prior to insertion on the long or short arms. These clones are classified as modified integration events.

These data support the idea that both adjacent homologous ends are critically important in initiating the recombination exchange. When one end is blocked with a large amount of heterologous sequences, a rate-limiting constraint is imposed on the integration event. In the recombinants that result, the inclusion of the MCITK sequences by vector insertion was observed in 1 of 51 clones, suggesting either that a subset of molecules which have lost the MCITK gene are involved in targeting or that the MCITK element is faithfully removed during the recombination event. This is evidence that the introduction of the small cap seen with IV6.8BN was not due to intra- or intermolecular ligation prior to the homologous exchange and that the caps were ligated to each other during the targeting event.

DISCUSSION

Homologous ends affect the targeting frequency and mechanism. The primary consequences of a different linearization site are the ends generated in the vector; our data show that it is these ends that ultimately determine both the targeting frequency and integration pathway. The interruption of the homology with a neo cassette did not interfere with the targeting frequency or integration pattern. We have shown that adjacent homologous ends are important for a high targeting frequency whether the construct was originally designed to be an insertion or a replacement vector. The provision of such ends results in a highly efficient and predictable insertion of the vector into the target site. In contrast to previously published data (35), but in agreement with our previous report (8), a double-strand break in the homology increases the targeting frequency by an order of magnitude over a double-strand break at the edge or outside the homology. This elevation in frequency was also seen in COS1 (10) and yeast (23) cells. Our data suggest that the recombination process has similarities with the double-strand break repair model (23, 25, 34), which predicts that the homologous ends initiate integration into the target locus. However, the majority of the integration events derived from a vector cut at the edge of or outside the homology follow a nonpredicted integration pathway that appears to be an insertion event (7, 8). In our experiments, only 20% of the targeted colonies generated with a replacement vector resulted in the predicted nonconservative replacement event, that is, the introduction of the neo cassette without the bacterial plasmid. This observation was unexpected and is different from the situation in yeast cells in that gene replacement events (substitution) constitute 70 to 85% of the transformants when the vector is cut at the edge of or outside the homology (23).

How do vectors insert into the target locus without a double-strand break in the homology? A double-strand break at the edge of or outside the vector homology may result in the integration of a recircularized vector by vector insertion or a concatamer by gene replacement. While these integration products are often indistinguishable, the observation that nicked circles target at a frequency similar to that of
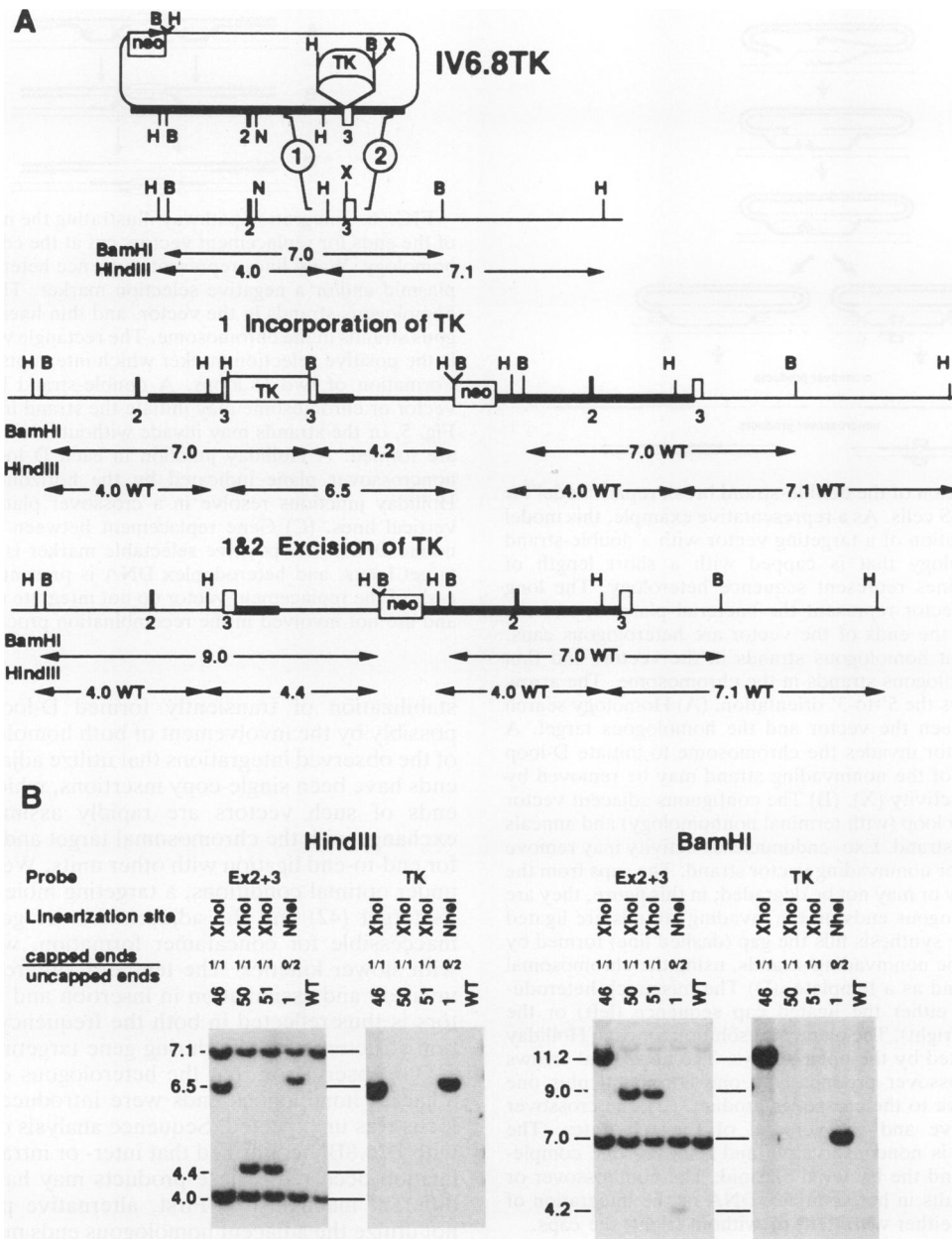
FIG. 4. Southern analysis for TG^r clones targeted with IV6.8TK. DNA was cut with BamHI and HindIII. The length of DNA is in kilobases. (A) Predicted size fragments for vector insertion events that incorporate and excise MCITK (TK). The thick line is hprt DNA of vector origin, the thin line is hprt DNA of genomic origin, and the line of intermediate thickness is plasmid. Lines represented by an encircled 1 and 2 represent points for strand invasion. MC1neopA is a box labeled neo. MCITK is the box labeled TK. Exons 2 and 3 are numbered rectangles. B, BamHI; H, HindIII; N, NheI; X, XhoI; WT, wild type. (B) Southern blot analysis. The HindIII and BamHI digests were hybridized to exons 2 and 3, and then the filter was stripped and hybridized to MCITK (TK). Linearization site, the restriction site where the vector was cut prior to the electroporation; cap, terminal heterology (MCITK) that blocks the homologous end. Clones were generated with IV6.8TK linearized at the XhoI site (clones 46, 50, and 51) and the NheI site (clone 1). Clones 1 and 46 integrated the MCITK sequence; clone 46 lost the BamHI site adjacent to the MCITK sequence; clones 50 and 51 removed the MCITK sequence. WT, wild type (AB1 DNA).

vectors cut at the edge of the homology suggests that integration of recircularized units may be the predominant event. A recircularized vector may require only one D loop and the generation of one double- or single-strand break for a vector insertion event, which should occur more frequently than the formation of two D loops and two breaks for a gene replacement event.

The difference in the targeting frequencies and integration

patterns between insertion and replacement vectors may be due to the formation and stability of the D loop. A vector with two adjacent homologous ends requires the formation of one D loop to insert the vector, and the involvement and strand pairing of all of the ends may be critical in stabilizing this intermediate (Fig. 5). A vector without two adjacent homologous ends may require the formation and stabilization of two D loops, which may be rate limiting (Fig. 6). One
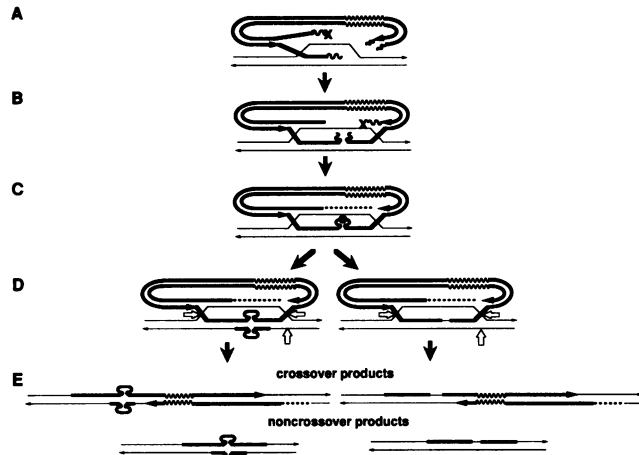
FIG. 5. Modification of the double-strand break repair model for vector insertion in ES cells. As a representative example, this model describes the integration of a targeting vector with a double-strand break in the homology that is capped with a short length of heterology. Wavy lines represent sequence heterology. The long wavy lines in the vector represent the bacterial plasmid, and the short wavy lines at the ends of the vector are heterologous caps. Thick lines represent homologous strands in the vector, and thin lines represent homologous strands in the chromosome. The arrow on each strand shows the 5'-to-3' orientation. (A) Homology search and alignment between the vector and the homologous target. A strand from the vector invades the chromosome to initiate D-loop formation. The cap of the noninvading strand may be removed by exo- endonuclease activity (X). (B) The contiguous adjacent vector strand invades the D loop (with terminal nonhomology) and anneals to the chromosomal strand. Exo- endonuclease activity may remove the cap from the other noninvading vector strand. The caps from the invading strands may or may not be degraded; in this figure, they are not. (C) The heterologous ends on the invading strands are ligated (filled circle). Repair synthesis fills the gap (dashed line) formed by the degradation of the noninvading strands, using the chromosomal complementary strand as a template. (D) The mismatch heteroduplex is repaired to either the ligated cap sequence (left) or the wild-type sequence (right). The plane of resolution for each Holliday junction is represented by the open arrows. The horizontal arrows resolve to a noncrossover product, and one horizontal plus one vertical arrow resolve to the crossover product. (E) The crossover event is conservative and recovers all of the substrates. The noncrossover event is nonconservative and removes one complement of homology and the bacterial plasmid. The noncrossover or crossover event results in heteroduplex DNA or the integration of the targeting vector either with (left) or without (right) the caps.

consequence of unstable D-loop formation might be that the heterologous ends of the replacement vector are substrates for the abundant end-joining capabilities of a mammalian cell (40). Concatemers and circles may form before the vector integrates into the target locus. Ligated molecules will become the substrates for homologous recombination in a temporal framework in which end-to-end joining is very efficient and the formation and resolution of a stable D loop is slow. Consistent with this hypothesis is the observation that concatemers were frequently found to integrate into the target locus with replacement vectors but not with insertion vectors (8).

Since a double-strand break in the homology promotes a high targeting frequency by vector insertion, it can be argued that the utilization of a single D loop by adjacent homologous ends promotes an efficient integration event. This efficiency may be at the level of strand invasion or relate to the rapid
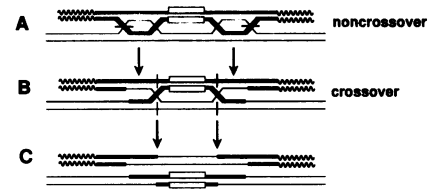
FIG. 6. Integration pathway illustrating the nonparticipating role of the ends for replacement vectors cut at the edge of or outside the homology. Wavy lines represent sequence heterology: the bacterial plasmid and/or a negative selection marker. Thick lines represent homologous strands in the vector, and thin lines represent homologous strands in the chromosome. The rectangle within the thick lines is the positive selection marker which interrupts the homology. (A) Formation of two D loops. A double-strand break in either the vector or chromosome may initiate the strand invasion as shown in Fig. 5, or the strands may invade without a break. Heteroduplexes are formed. A Holliday junction in each D loop is resolved in a noncrossover plane indicated by the horizontal lines. (B) Both Holliday junctions resolve in a crossover plane indicated by the vertical lines. (C) Gene replacement between the vector and genomic DNA. The positive selectable marker is integrated into the target locus, and heteroduplex DNA is present. The heterologous ends of the replacement vector do not integrate into the target locus and are not involved in the recombination process.

stabilization of transiently formed D-loop intermediates, possibly by the involvement of both homologous ends. Most of the observed integrations that utilize adjacent homologous ends have been single-copy insertions, which argues that the ends of such vectors are rapidly assimilated by strand exchange with the chromosomal target and are not available for end-to-end ligation with other units. We hypothesize that under optimal conditions, a targeting molecule rapidly finds its target (42) and the adjacent homologous ends become inaccessible for concatamer formation, which must occur with slower kinetics. The temporal difference between end invasion and assimilation in insertion and replacement vectors is thus reflected in both the frequency and recombination structures formed during gene targeting.

Our observation that the heterologous caps blocking the adjacent homologous ends were introduced into the target locus was unexpected. Sequence analysis of clones targeted with IV6.8BN confirmed that inter- or intramolecular vector ligation occurred. These products may have arisen by two different mechanisms. First, alternative pathways that do not utilize the adjacent homologous ends may be responsible for integration of the caps so that inter- or intramolecular ligation generated substrates that target by gene replacement of concatemers or vector insertion of circles (8). Second, the invasion of the homologous ends may be possible when the ends are blocked by short heterologous sequences so that the end-to-end ligation of strands occurs during the recombination process.

A comparison of the targeting frequencies and integration patterns between IV6.8BN (short caps) and IV6.8TK (long cap) suggests the latter pathway; that is, invasion of homologous ends does not require removal of the short caps. The data support this hypothesis in several ways. First, the targeting frequencies of IV6.8 and IV6.8BN are similar, while IV6.8 targets at a sevenfold-higher frequency than does IV6.8TK. Although the targeting frequency for IV6.8BN is higher than that of IV6.8TK, the number of clones which integrate the caps should be the same if end-to-end ligation preceded targeting because the size of the heterologous mutation should not affect the frequency of replacement

events if a concatemer formed (16) or of insertion events if the vector recircularized (Table 1, experiment C). The frequency of recovering the caps is, however, much lower for IV6.8TK than for IV6.8BN; the frequency of generating clones which included the caps was 1 of 350 G418^r colonies for IV6.8BN (21% of the targeted clones; Table 1, experiments D to F) and 1 of 23,000 G418^r colonies for IV6.8TK (2% of the targeted clones; Table 1, experiments C to F). The inclusion of the caps is thus 65-fold more common for IV6.8BN than for IV6.8TK. Although inter- or intramolecular extrachromosomal ligation of the vector cannot be excluded as a possibility for the integration of the heterologous caps, the most likely explanation for the majority of these products is that strand invasion is not inhibited by short (13-bp) terminal heterologous sequences (Fig. 5) but is blocked by long (2.3-kb) heterologous sequences.

Removal of both the short and long heterologous caps and restoration of the wild-type sequence was demonstrated in most clones generated with IV6.8BN and IV6.8TK and is evidence for gap formation and repair. The removal of the short and long caps and the restoration of the endogenous XhoI site during the integration of IV6.8BN and IV6.8TK demonstrates exo- or endonuclease activity and subsequent repair using the colinear chromosomal sequence as a template. Gap repair, a necessary requirement of the double-strand break repair model in yeast cells, has also been shown to occur in COS1 (10), CHO (24), and ES cells (36).

Modification of the double-strand break repair model for gene targeting in ES cells. The data presented here on the fate of adjacent homologous ends capped with terminal heterologies are not readily explained by the current double-strand break repair model. A modification of the double-strand break repair model that makes predictions as to the fate of the adjacent homologous ends for insertion events is required.

The model in Fig. 5 describes the insertion of a vector with adjacent homologous ends that are capped with a short heterologous sequence. The critical modification to the model is that adjacent ends of the same strand, not complementary strands, invade the target locus, and that short heterologous caps may be incorporated into the target locus. After one strand invades the target locus, the adjacent end of the same strand invades to stabilize the D loop with two Holliday junctions. The ends of the invading and noninvading strands are subject to exo- endonuclease activity to remove the caps. Removal of the cap is not essential or necessarily complete; any residual heterologous sequence may integrate into the target locus. The contiguous ends are ligated, and if any heterologous sequence contributed by the caps still exists, a mismatch heteroduplex is formed with the chromosomal pairing strand. The mismatch heteroduplex is repaired to either the mutation or the wild-type sequences. The mutation may be present in both duplicates if mismatch heteroduplex repair resolves in favor of the mutation in both duplicates.

Clone 4, generated by IV6.8BN, had the ligated cap mutation in both duplicates (Fig. 3). Repair synthesis to the mutation of two sets of mismatch heteroduplex sequences, one on each duplicate, may result in this integration pattern. Alternatively, only one mismatch heteroduplex may have formed and repaired to the mutation in one duplicate; subsequently, these strands may have dissociated by branch migration of at least one Holliday junction and reannealed to the strands of the other duplicate to form two mismatch heteroduplexes, both of which were then repaired to the mutation. The latter pathway seems inefficient; however,

recovery of the mutation in both duplicates was uncommon, and the identical junction sequence was present in both duplicates even though 6 bp was deleted from the original sequence. More clones must be analyzed to evaluate the mechanism responsible for this type of insertion.

Gene replacement pathways do not utilize the ends. We have demonstrated that the heterologous nucleotides on the adjacent homologous ends of an insertion vector integrate into the target locus. Why is this not observed with the heterologous ends in a replacement vector with a break at the edge of or outside the homology? Most observations with gene replacement events have demonstrated the loss of large pieces of heterologous sequence from the construct, such as negative selection markers and bacterial plasmids (15), without a decrease in the targeting frequency (8). Although we observed the elimination of 2.3 kb of heterology from the end of an insertion vector, IV6.8TK, this was accompanied by a sevenfold decrease in the targeting frequency. These observations suggest that vector ends play different roles for insertion and replacement vectors. There are a number of pathways that could describe the integration of a replacement vector into the chromosome, including the double-strand break repair and single-strand annealing models. If the double-strand break repair model (23, 25, 34) is involved, then a break(s) may be induced in the chromosome, allowing chromosomal strands to invade the replacement vector. Our modification of the double-strand break repair model suggests that for replacement vectors, terminal heterologous nucleotides would not be incorporated in the recombinant structure because there is not a contiguous homologous strand to ligate to these ends (Fig. 6). A single-strand break(s) (21) may also be used to initiate the strand invasion (Fig. 6B and C). If the single-strand annealing model (12–14) is relevant, then a break in the chromosome homology must be induced. Exonuclease activity must degrade the ends on both the chromosome and vector to generate complementary overhangs so that the vector may anneal to the chromosome. By any of these models, the heterologous ends contributed by the replacement vector do not participate in strand invasion, nor will they be incorporated into the target locus unless they are involved in end-to-end ligations to form concatemers or circles prior to homologous recombination.

Our data confirm that a double-strand break in the homology is critical for efficient gene targeting in ES cells at the hprt locus as predicted by the double-strand break repair model. Terminal heterologies that capped the adjacent homologous ends were incorporated into the target locus. We believe that capped adjacent homologous ends generated by the break participate in the initial strand exchange and may or may not be removed by endo- or exonuclease activity. The data also suggest that the invading homologous strands are rapidly assimilated into the recombination complex and are not readily available for extrachromosomal end-to-end joining. We have shown that a double-strand break in the homology determines both the frequency and predictability of the genetic exchange at the hprt locus in ES cells.

REFERENCES

1. **Anderson, R. A., and S. L. Eliason.** 1986. Recombination of homologous DNA fragments transfected into mammalian cells occurs predominantly by terminal pairing. Mol. Cell. Biol. **6:**3246–3252.

2. **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.

3. **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.

4. **Capecchi, M. R.** 1989. Altering the genome by homologous recombination. Science **244:**1288–1292.

5. **Chakrabarti, S., and M. M. Seidman.** 1986. Intramolecular recombination between transfected repeated sequences in mammalian cells is nonconservative. Mol. Cell. Biol. **6:**2520–2526.

6. **Hasty, P., R. Ramírez-Solis, R. Krumlauf, and A. Bradley.** 1991. Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells. Nature (London) **350:**243–246.

7. **Hasty, P., J. Rivera-Pérez, and A. Bradley.** 1991. The length of homology required for gene targeting in embryonic stem cells. Mol. Cell. Biol. **11:**5586–5591.

8. **Hasty, P., J. Rivera-Pérez, C. Chang, and A. Bradley.** 1991. Targeting frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. Mol. Cell. Biol. **11:**4509–4517.

9. **Holliday, R.** 1964. A mechanism for gene conversion in fungi. Genet. Res. **5:**282–304.

10. **Jasin, M., and P. Berg.** 1988. Homologous integration in mammalian cells without target gene selection. Genes Dev. **2:**1353–1363.

11. **Kucherlapati, R. S., E. M. Eves, K. 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.

12. **Lin, F., K. Sperle, and N. Sternberg.** 1984. Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. Mol. Cell. Biol. **4:**1020–1034.

13. **Lin, F. M., K. Sperle, and N. Sternberg.** 1990. Intermolecular recombination between DNAs introduced into mouse L cells is mediated by a nonconservative pathway that leads to crossover products. Mol. Cell. Biol. **10:**103–112.

14. **Lin, F. M., K. Sperle, and N. Sternberg.** 1990. Repair of double-stranded DNA breaks by homologous DNA fragments during transfer of DNA into mouse L cells. Mol. Cell. Biol. **10:**113–119.

15. **Mansour, S. L., K. R. Thomas, and M. R. Capecchi.** 1988. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature (London) **336:**348–352.

16. **Mansour, S. L., K. R. Thomas, C. Deng, and M. R. Capecchi.** 1990. Introduction of a lacZ reporter gene into the mouse int-2 locus by homologous recombination. Proc. Natl. Acad. Sci. USA **87:**7688–7692.

17. **Maryon, E., and D. Carroll.** 1991. Involvement of single-stranded tails in homologous recombination of DNA injected into Xenopus laevis oocyte nuclei. Mol. Cell. Biol. **11:**3268–3277.

18. **Maryon, E., and D. Carroll.** 1991. Characterization of recombination intermediates from DNA injected into Xenopus laevis oocytes: evidence for a nonconservative mechanism of homologous recombination. Mol. Cell. Biol. **11:**3278–3287.

19. **McMahon, A. P., and A. Bradley.** 1990. The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. Cell **62:**1073–1085.

20. **Melton, D. W., D. S. Konecki, J. Brennand, and C. T. Caskey.** 1984. Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA **81:**2147–2151.

21. **Meselson, M. S., and C. M. Radding.** 1975. A general model for genetic recombination. Proc. Natl. Acad. Sci. USA **72:**358–361.

22. **Murnane, J. P., M. J. Yezzi, and B. R. Young.** 1990. Recombination events during integration of transfected DNA into normal human cells. Nucleic Acids Res. **18:**2733–2738.

23. **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.

24. **Pennington, S. L., and J. H. Wilson.** 1991. Gene targeting in chinese hamster ovary cells is conservative. Proc. Natl. Acad. Sci. USA **88:**9498–9502.

25. **Resnick, M. A.** 1976. The repair of double-strand breaks in DNA: a model involving recombination. J. Theor. Biol. **59:**97–106.

26. **Robertson, E. J.** 1987. Embryo-derived stem cell lines, p. 71–112. In E. J. Robertson (ed.), Teratocarcinomas and embryonic stem cells: a practical approach. IRL Press, Oxford.

27. **Schultes, N. P., and J. W. Szostak.** 1991. A poly(dA-dT) tract is a component of the recombination initiation site at the ARG4 locus in Saccharomyces cerevisiae. Mol. Cell. Biol. **11:**322–328.

28. **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.

29. **Siedman, M. M.** 1987. Intermolecular homologous recombination between transfected sequences in mammalian cells is primarily nonconservative. Mol. Cell. Biol. **7:**3561–3565.

30. **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.

31. **Song, K., 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.

32. **Soriano, P., C. Montgomery, R. Geske, and A. Bradley.** 1991. Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. Cell **64:**693–702.

33. **Sun, H., D. Treco, and J. W. Szostak.** 1991. Extensive 3′-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site. Cell **64:**1155–1161.

34. **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.

35. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51:**503–512.

36. **Valancius, V., and O. Smithies.** 1991. Double-strand break repair in a mammalian gene targeting reaction. Mol. Cell. Biol. **11:**4389–4397.

37. **Wake, C. T., F. Vernaleone, and J. H. Wilson.** 1985. Topological requirements for homologous recombination among DNA molecules transfected into mammalian cells. Mol. Cell. Biol. **5:**2080–2089.

38. **Waldman, A. S., and R. M. Liskay.** 1987. Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. Proc. Natl. Acad. Sci. USA **84:**5340–5344.

39. **Waldman, A. S., and R. M. Liskay.** 1988. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol. Cell. Biol. **8:**5350–5357.

40. **Wilson, J. H., P. B. Berget, and J. M. Pipas.** 1982. Somatic cells efficiently join unrelated DNA segments end to end. Mol. Cell. Biol. **2:**1258–1269.

41. **Zheng, H., P. Hasty, M. A. Brenneman, M. Grompe, R. A. Gibbs, J. H. Wilson, and A. Bradley.** 1991. Fidelity of targeted recombination in human fibroblasts and murine embryonic stem cells. Proc. Natl. Acad. Sci. USA **88:**8067–8071.

42. **Zheng, H., and J. Wilson.** 1990. Gene targeting in normal and amplified cell lines. Nature (London) **344:**170–173.