

Target Frequency and Integration Pattern for Insertion and Replacement Vectors in Embryonic Stem Cells

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

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

Received 31 December 1990/Accepted 24 June 1991

Gene targeting has been used to direct mutations into specific chromosomal loci in murine embryonic stem (ES) cells. The altered locus can be studied in vivo with chimeras and, if the mutated cells contribute to the germ line, in their offspring. Although homologous recombination is the basis for the widely used gene targeting techniques, to date, the mechanism of homologous recombination between a vector and the chromosomal target in mammalian cells is essentially unknown. Here we look at the nature of gene targeting in ES cells by comparing an insertion vector with replacement vectors that target *hprt*. We found that the insertion vector targeted up to ninefold more frequently than a replacement vector with the same length of homologous sequence. We also observed that the majority of clones targeted with replacement vectors did not recombine as predicted. Analysis of the recombinant structures showed that the external heterologous sequences were often incorporated into the target locus. This observation can be explained by either single reciprocal recombination (vector insertion) of a recircularized vector or double reciprocal recombination/gene conversion (gene replacement) of a vector concatemer. Thus, single reciprocal recombination of an insertion vector occurs 92-fold more frequently than double reciprocal recombination of a replacement vector with crossover junctions on both the long and short arms.

It is now possible to study the function of specific genes in the mouse by the application of gene targeting in embryonic stem (ES) cells. ES cells which have undergone a specific gene-targeting reaction may subsequently be used to establish the allele in the germ line of mice (2, 8, 12, 17, 20, 23, 27). However, the application of gene targeting is limited by random recombination, which is 30- to 40,000-fold more frequent than homologous recombination; thus, most transfected cells are not targeted (2–4, 6, 7, 9, 11, 12, 17, 20, 22–24). Understanding the parameters that influence the absolute and relative frequencies of gene targeting will have a direct impact on many experiments that require the gene to be mutated in ES cells.

Two distinct vector designs have been investigated in gene targeting experiments, replacement and insertion vectors. Replacement (omega type) vectors are intended to replace homologous gene sequences with a selectable mutation (22) and have been most commonly used to make null alleles in ES cells (2, 8, 12, 17, 20, 23, 27). Typically, replacement vectors contain sequences homologous to the target, with a *neo* expression cassette inserted within the region of homology so that it disrupts an exon and creates a mutation. The vector is linearized outside the region of homology. Following homologous recombination, sequences in the vector replace the genomic sequence by double reciprocal recombination or gene conversion, thereby inserting the *neo* mutation into the target gene (22). The nonhomologous sequences at the free ends of the construct are excised (11). Insertion vectors (O type) are the alternative vector type, which are intended to insert the entire construct into the homologous sequences (22). These vectors are cut at a site within the homology. A double-strand break in the homology has been shown to increase both the targeting frequency in yeast cells (15) and extrachromosomal recombination in mammalian cells (10, 19, 26). Insertion vectors have been

used in ES cells for gene correction (3) and gene disruption (22) at the selectable *hprt* locus.

The most widely used vectors are replacement vectors. Their general acceptance is based upon a preliminary report that recombination frequencies were similar to those with insertion vectors (22) and that positive-negative selection is applicable to this vector type (11). Recently, some reports have indicated that nonpredicted integration events were occurring when replacement vectors were used (18, 23). The types of integration events that were occurring suggested that screens such as polymerase chain reaction (PCR) and negative selection may eliminate certain classes of targeted recombinants from analysis. A description of the entire spectrum of targeted integration events, free from any bias imposed by selection and screening techniques, will facilitate an understanding of the gene-targeting pathways in ES cells. This, in turn, will enable vectors to be designed that can most efficiently utilize the host recombination machinery to target genes.

In order to study the frequency and integration events of homologous recombination in ES cells, we have chosen to target a gene involved in the purine salvage pathway, that for hypoxanthine phosphoribosyltransferase (*hprt*) (13). *hprt* is located on the X chromosome and is a single-copy gene in XY ES cells; therefore, all targeted events regardless of fidelity can be selected by the absence of *hprt* function with a toxic base analog, 6-thioguanine (TG). We show that an insertion vector targeted up to ninefold more frequently than a replacement vector with the same homologous sequences and the same *neo* expression cassette. We also report that most of the clones targeted with replacement vectors were not generated by the predicted gene replacement event; rather, the entire construct was usually integrated into the *hprt* locus. The targeting efficiency and integration pattern were unaffected even if the targeted homology was released from the plasmid. Our observations demonstrate that circularized products and/or concatemers form before replacement vectors integrate into the target locus.

* Corresponding author.

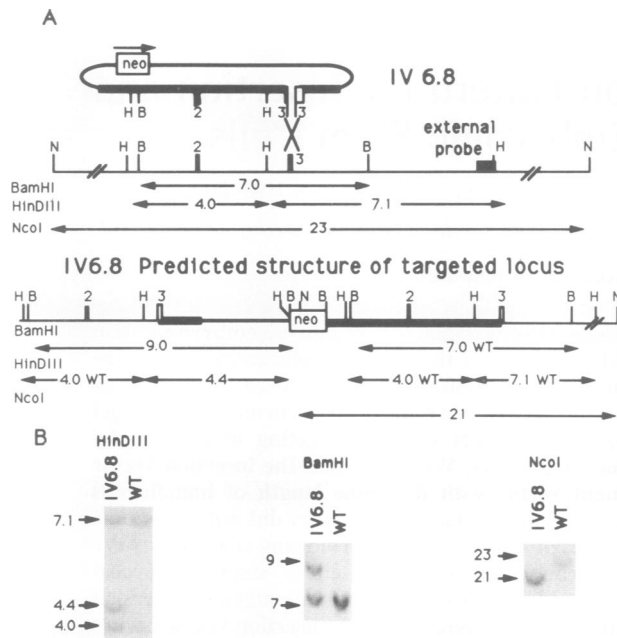


FIG. 1. Southern blot analysis of TG^r clones generated with IV6.8. (A) Wild-type (WT) fragments and predicted sizes of fragments for single reciprocal recombination seen with a *Bam*HI (B) digest, a *Hind*III (H) digest, and an *Nco*I (N) digest. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid DNA. MC1neopA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as the internal probe (taken from cDNA). The filled box is the external probe. The length of DNA is in kilobases. (B) *Bam*HI and *Hind*III digests were hybridized to exons 2 and 3. The *Nco*I digest was hybridized to the external probe.

MATERIALS AND METHODS

Construction of vectors. The insertion vector IV6.8 is a *Sac*I-*Eco*RI fragment of the murine *hprt* gene that contains exons 2 and 3 inserted into pTZ (Pharmacia) at the *Sac*I and *Eco*RI sites and contains a *neo* expression cassette, MC1neopA (Stratagene), in the polylinker. Replacement vectors RV6.8 and RV6.8PGK were made from the same homologous sequences as IV6.8; MC1neopA or PGKneopA (1, 16, 20), respectively, was inserted into the *Xho*I site of exon 3.

Electroporation and tissue culture. The DNA was prepared for electroporation by Triton lysis and banded once in CsCl. It was then linearized with the specified restriction enzyme, phenol-chloroform extracted, ethanol precipitated, and resuspended in 10 mM Tris (pH 7.5)–1 mM EDTA at a concentration between 0.5 and 1.0 μ g/ μ l. To ensure complete digestion, the cut DNA was compared with uncut DNA by a mobility shift assay with gel electrophoresis. DNA was not electroporated unless it was completely cut. Electroporations were performed on AB1 (12) cells collected from confluent 9-cm plates of trypsin-treated cells about 3 h after refeeding. The AB1 cells were centrifuged for 5 min at 1,000 rpm and resuspended in phosphate-buffered saline (Ca²⁺ and Mg²⁺ free) at a density of 10⁷/ml. DNA (25 μ g) was electroporated in 1 ml at 575 V/cm and 500 μ F 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 10⁷ cells were plated onto a 9-cm SNL76/7 feeder plate (12) with 10 ml of medium (Dulbecco's modified Eagle's medium with 15% fetal

calf serum, 2 mM glutamine, 30 μ g of penicillin per ml, 50 μ g of streptomycin per ml, and 10⁻⁴ M β -mercaptoethanol).

Three different methods were used to determine resistance to G418 and TG. Method 1: At 24 h after electroporation, 180 μ g of G418 (active ingredient) per ml was added to each 9-cm plate and maintained for 10 days in order to select and count G418-resistant 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 and TG^r colonies were counted. Method 2: Plates were also selected in G418 and TG simultaneously at the concentrations described for method 1. The G418 was added 24 h after the electroporation, and then TG was added 5 to 6 days later. Double selection was maintained for another 10 days, at which point G418^r and TG^r colonies were counted. Method 3: Cells were scored by the incorporation of [³H]hypoxanthine. Transfected cells were selected in G418 for 10 days and then labeled for 24 h in 4 ml of medium containing 0.04 μ Ci of [³H]hypoxanthine. The colonies were washed twice in phosphate-buffered saline and then fixed in methanol-acetic acid vapor (3:1) for 10 min. The colonies were extracted twice in ice-cold 10% trichloroacetic acid and washed in water overnight. The colonies were dried, and a thin layer of nuclear emulsion was applied. The emulsion was exposed in the dark for 3 days at 4°C. Silver grains were developed and fixed by standard procedures. Cells that contained a functional *hprt* gene incorporated the labeled nucleoside and were intensely black. White colonies were scored as *hprt* negative. All three methods were used for RV6.8 in order to establish the accuracy of each method. The same results were obtained for all three methods. For RV6.8PGK and IV6.8, method 1 was used on 10 to 20% of the plates to score G418^r colonies and then G418^r plus TG^r colonies, and method 2 was used on 80 to 90% of the plates to score G418^r plus TG^r colonies.

Southern blot analysis of targeted colonies. A *Bam*HI digest and a *Hind*III digest of genomic DNA were hybridized to two probes internal to the vector, *hprt* exons 2 and 3 (cDNA cut with *Hinc*II and *Hpa*I [12]), and plasmid sequences; *Nco*I, *Nco*I-*Sal*I, and *Hind*III digests of genomic DNA were hybridized to a 3' external probe not present in the vector (a 0.5-kb fragment cut with *Pst*I and *Hind*III just 5' to exon 4). Isolated TG^r colonies were passaged once onto gelatinized plates without feeder cells to reduce feeder cell DNA. *Bam*HI-*Hind*III-digested DNA (5 μ g) was separated by electrophoresis through a 0.7% agarose gel for 18 h at 60 V in TAE buffer. The *Nco*I and *Nco*I-*Sal*I-digested DNA was separated by electrophoresis through a 0.7% gel in TAE buffer for 24 h by field inversion gel electrophoresis at 4°C. The buffer was circulated. DNA was transferred onto a GeneScreen Plus filter (Du Pont) and hybridized to the specified probe under standard conditions. The probe was random primer labeled under the manufacturer's specified conditions (Boehringer Mannheim).

RESULTS

Insertion vectors target the *hprt* locus more efficiently than replacement vectors. To test the targeting efficiency of vectors with the same homologous sequences, we constructed insertion and replacement vectors which contain 6.8 kb of homologous sequence. The insertion vector IV6.8 contains the MC1neopA expression cassette in the plasmid backbone (Fig. 1A), while the replacement vectors RV6.8 (see Fig. 3A) and RV6.8PGK (see Fig. 2A) contain MC1neopA and PGKneopA, respectively, in the *Xho*I site of exon 3. IV6.8 was

TABLE 1. Insertion and replacement vectors which target the *hprt* locus

Vector	Restriction digest	No. of electroporations	Total no. of clones		Ratio, <i>hprt</i> ⁻ /G418 ^r	No. of TG ^r clones analyzed	Integration pattern ^a (no. of TG ^r clones)					
			G418 ^r	<i>hprt</i> ⁻			L1&Lx or VILA	S1&Sx or VISA	VIM	L1&S1	L1&Sx	UD
IV6.8	<i>Xho</i> I	46	23,884	610	1:39	20	ND	ND	1	ND	ND	—
RV6.8	<i>Sac</i> I	8	7,952	36	1:221	7	4	0	3	0	0	0
RV6.8	<i>Sal</i> I	10	9,894	28	1:353	10	8	0	0	0	1	1
RV6.8	<i>Sac</i> I + <i>Sal</i> I	29	15,530	59	1:263	14	8	0	3	0	0	3
RV6.8PGK	<i>Sal</i> I	21	28,702	58	1:495	9	5	0	1	2	0	1

^a VILA, vector insertion, long arm. VISA, vector insertion, short arm; VIM, vector integration, modified; L1, long arm of the first unit in a multimer; S1, short arm of the first unit in a multimer; Lx, long arm of unit x in a multimer; Sx, short arm of unit x in a multimer; UD, undetermined; ND, not determined.

linearized at the *Xho*I site in exon 3. RV6.8 was cut at the 5' edge of homology with *Sac*I. Both RV6.8 and RV6.8PGK were cut at the 3' edge of homology with *Sal*I, which leaves 30 bp of heterologous sequence on the 3' end. The DNA was electroporated into ES cells, and the transfection events were determined by selection in G418. A second selection with TG was used to isolate clones without a functional *hprt* gene. IV6.8 targeted at a relative frequency of 1 TG^r to 39 G418^r colonies (Table 1). RV6.8 cut with *Sac*I, RV6.8 cut with *Sal*I, and RV6.8PGK cut with *Sal*I yielded TG^r/G418^r colony ratios of 1:221, 1:353, and 1:495, respectively (Table 1). PGKneopA yields more G418^r colonies per electroporation than MC1neopA (20), and this accounts for the slightly lower targeting frequency seen with RV6.8PGK. The absolute targeting frequencies were the same. IV6.8 proved to be about five- to ninefold more efficient at gene targeting than RV6.8 cut with *Sac*I and *Sal*I, respectively.

IV6.8 integrates the entire vector into the *hprt* locus during homologous recombination. Targeted integration of IV6.8 will form a duplication of the 6.8 kb of *hprt* homology separated by the plasmid and *neo* sequences (Fig. 1). To confirm that the TG^r clones had the predicted structure, *Bam*HI and *Hind*III digests were hybridized to an internal probe, *hprt* exons 2 and 3 from cDNA. A *Bam*HI digest on wild-type DNA gives a 7.0-kb fragment. The targeted insertion events should result in a duplication; the 7.0-kb wild-type fragments were unchanged, and the predicted additional 9.0-kb fragments were formed. A *Hind*III digest was hybridized to the same internal probe; as expected, the 7.1- and 4.0-kb wild-type fragments were retained and additional 4.4-kb fragments were generated. The novel fragments on the *Bam*HI and *Hind*III digests verified the presence of a 6.8-kb *hprt* duplication which contains exons 2 and 3. To ensure that the 6.8-kb duplication was separated by a plasmid sequence and a *neo* gene, the *Bam*HI and *Hind*III blots were stripped and hybridized to plasmid sequences. The 9.0-kb fragment on the *Bam*HI blot and the 4.4-kb fragment on the *Hind*III blot hybridized to the plasmid sequences, as predicted (data not shown). This analysis showed that 19 of 20 TG^r clones had the predicted restriction pattern. Concatemers appeared to have integrated into the target locus in 3 of these 19 clones. One clone gave a restriction pattern that was not expected and could not be interpreted.

The restriction patterns seen for the *Bam*HI and *Hind*III digests would also be seen if a concatemer of the vector integrated randomly in the genome adjacent to *Bam*HI and *Hind*III sites at the appropriate distance in association with a spontaneous mutation of *hprt* in a location other than the targeting homology. It is unlikely for this to occur; however, in order to definitively determine that targeting occurred at

hprt, a probe outside the vector was used for Southern analysis on a subset of clones. An *Nco*I digest hybridized to a 3' external probe was used to ensure that the restriction pattern seen on the *Bam*HI and *Hind*III blots was due to targeting at the *hprt* locus (Fig. 1). An *Nco*I wild-type fragment is 23 kb long, and the insertion of IV6.8 will introduce an *Nco*I site (in the *neo*) to form a 21-kb fragment. Every clone analyzed with an *Nco*I digest had the 21-kb fragment. The combination of *Bam*HI and *Hind*III digests hybridized to the internal probes and the *Nco*I digest hybridized to the external probe verified the predicted insertion event at the *hprt* locus.

Two of nine TG^r clones targeted with linearized RV6.8PGK had gene replacement events. A gene replacement event with RV6.8PGK should introduce *Bam*HI, *Hind*III, and *Nco*I sites into exon 3. To test the structure of the targeted locus, DNA from nine clones was digested with *Bam*HI and hybridized to exons 2 and 3 (Fig. 2). The wild-type 7.0-kb fragment should be cut into two fragments of 5.9 and 2.3 kb. Surprisingly, only two of nine TG^r clones analyzed showed the predicted structure (Table 1), clones 3 and 4 (Fig. 2B). A *Hind*III digest and an *Nco*I digest hybridized to the external probe confirmed that only these clones were targeted by gene replacement. The *Hind*III digest shows that the wild-type 7.1-kb fragment is reduced to 6.6 kb, and the *Nco*I digest shows that the wild-type 23-kb fragment is reduced to 13 kb (Fig. 2B). As expected, these clones did not hybridize to plasmid sequences (data not shown).

Entire construct integrated in majority of targeted events generated with linearized RV6.8PGK. The *Bam*HI, *Hind*III, and *Nco*I digests used to identify the two replacement events also identified changes in seven targeted clones that could not be explained as gene replacement events (Fig. 2). Analysis of the *Bam*HI digest hybridized to exons 2 and 3 showed that clones 2 and 6 through 9 had targeted events that included the plasmid and made a 6.8-kb duplication of the homologous sequences, with the *neo* located in the 5' duplicate, so that the 7.0-kb wild-type fragment was retained and new 4.8- and 5.9-kb fragments were formed. The 4.8-kb fragment hybridized to plasmid sequences (data not shown). The *Hind*III digest hybridized to the 3' external probe showed only the 7.1-kb wild-type fragment. This indicates that the 3' duplicate was unaltered. An *Nco*I-*Sal*I digest hybridized to the external probe identified the introduction of the *Sal*I site in the plasmid polylinker, which reduces the wild-type 23-kb fragment to 22 kb. The 22-kb fragment hybridized to plasmid sequences (data not shown). These restriction and hybridization patterns are consistent with integration of the entire vector rather than a replacement (Fig. 3A). Integration of the entire vector can occur via

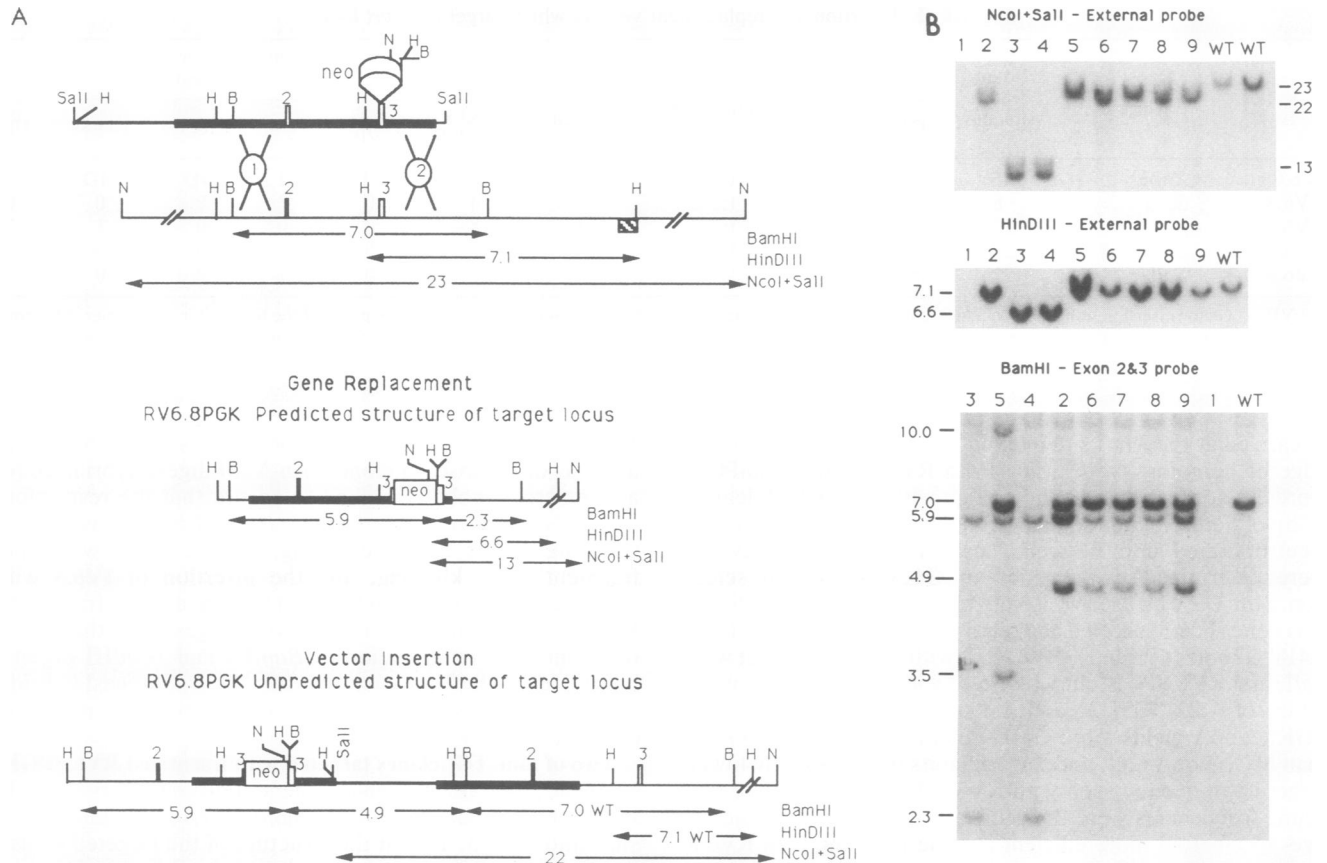


FIG. 2. Southern blot analysis of TG^f clones generated with RV6.8PGK. (A) Wild-type (WT) fragments and predicted sizes of fragments for double reciprocal recombination (gene replacement) and single reciprocal recombination (vector insertion) seen with a *Bam*HI (B) digest, a *Hind*III (H) digest, and an *Nco*I (N) 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 DNA. PGKneobpA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as the internal probe (taken from cDNA). The striped box is the external probe. The length of DNA is in kilobases. (B) The *Bam*HI digest was hybridized to exons 2 and 3. The *Hind*III and *Nco*I digests were hybridized to the external probe.

double reciprocal recombination of a concatemer on the long arm (the *neo* divides the homology into 5.6- and 1.2-kb fragments, the long and short arm) of the first unit, L1, and the long arm of a downstream unit designated unit x, Lx. This we refer to as an L1&Lx integration. The insertion may also occur via single reciprocal recombination on the long arm of a recircularized construct or circularized concatemer. This is referred to as vector insertion on the long arm (VILA).

Two TG^f clones targeted with RV6.8PGK gave restriction patterns that were not consistent with simple replacement or insertion events. In Fig. 2B, clone 5 appears to have been targeted by a modified insertion event. The *Hind*III and *Nco*I digests have the fragment sizes expected for an insertion event on the long arm. However, the *Bam*HI digest showed an aberrant pattern. Clone 1 has a large deletion that does not hybridize to exons 2 and 3 or to the external probe; this may be due to a spontaneous *hprt* deletion.

Majority of targeted clones generated with linearized RV6.8 occurred by an L1&Lx or VILA integration pattern. Ten TG^f clones were analyzed to determine the vector integration pattern of RV6.8 linearized with *Sal*I. A *Bam*HI digest hybridized to exons 2 and 3 and a *Hind*III digest hybridized to the external probe showed that the vector had inserted in eight clones (Table 1). The restriction pattern is similar to that seen for RV6.8PGK. Of the eight clones targeted by the

integration of the entire construct, seven appeared to have integrated more than one unit. In Fig. 3C, clone 7 showed the insertion of a single unit and clone 8 shows the insertion of multiple units.

Clone 4 (Fig. 3C) was the result of a gene replacement event characterized by integration of a concatemer with one crossover on the long arm of the first unit (L1) and the other crossover on the short arm of a downstream unit designated unit x (Sx). We call this an L1&Sx integration pattern. By comparing the fragment densities of clone 4 with those of a single-unit insert, clone 7, it appears that unit x is the third or fourth unit in the concatemer. For the L1&Sx gene replacement event, fragments of 5.6, 4.2, and 2.3 kb are seen on a *Bam*HI digest hybridized to exons 2 and 3, and a fragment shift from the 7.1-kb wild-type fragment to 8.3 kb was seen with a *Hind*III digest hybridized to the external probe. One clone had a restriction pattern that is not representative of any of the possibilities discussed.

Southern analysis was also performed on seven clones targeted with RV6.8 cut with *Sac*I. Vector insertion was observed for four clones. All four clones appeared to have integrated multiple units. A modified integration was found in three clones (Table 1).

Majority of targeted clones generated with RV6.8 released from the plasmid occurred by integration of ligated units. To

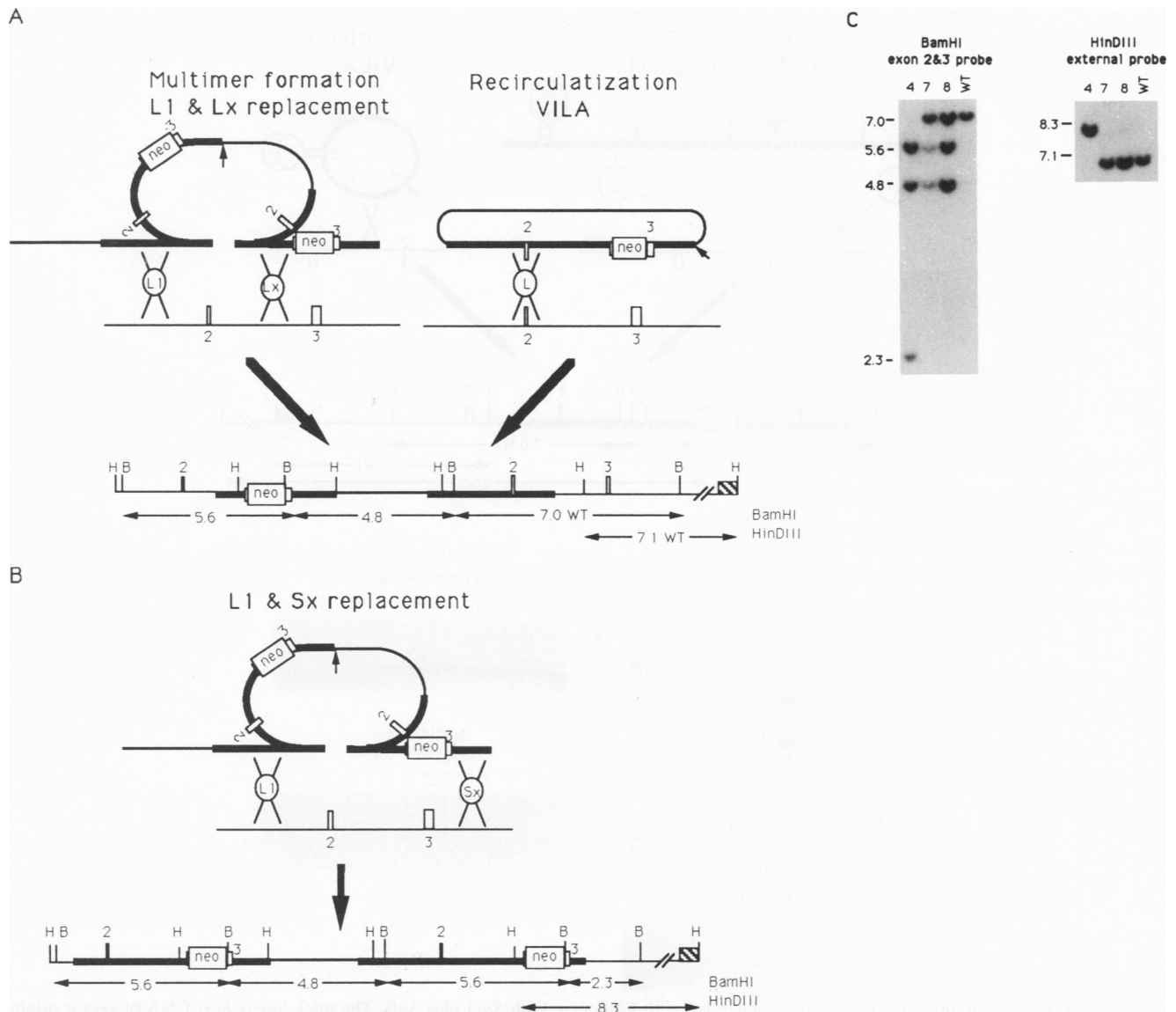


FIG. 3. Southern blot analysis of TG^f clones generated with RV6.8 linearized with *Sal*I. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid DNA. MCIneoA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as the internal probe (taken from cDNA). The striped box is the external probe. The length of DNA is in kilobases. The small arrows point to the location for end-to-end joining for multiple units of the construct. (A) Diagram of a multimer involved in gene replacement with the integration on the long arms, L1 and Lx. Diagram of a recircularized construct or circularized concatemer involved in VILA. Wild-type (WT) fragments and predicted sizes of fragments for gene replacement and vector insertion seen with a *Bam*HI (B) digest hybridized to exons 2 and 3 and a *Hind*III (H) digest hybridized to the external probe. (B) Diagram of a multimer involved in gene replacement with the crossover on the long arm of the first unit (L1) and the short arm of unit x (Sx). (C) The *Bam*HI digest was hybridized to exons 2 and 3. The *Hind*III digest was hybridized to the external probe. Sizes are in kilobases.

determine whether the proportion of vector replacement events would increase when the *hprt* homology was released from the plasmid, RV6.8 was cut with *Sac*I and *Sal*I and transfected into ES cells. The targeting frequency was unaltered (Table 1). Southern analysis was performed on TG^f clones with a *Bam*HI digest hybridized to exons 2 and 3 and a *Hind*III and *Nco*I digest hybridized to the external probe. We did not recover any gene replacement events in 15 clones analyzed. Instead, a variety of integration patterns that required the ligation of the *hprt* homologous sequence to itself and/or to the plasmid sequence prior to integration were seen (Fig. 4). The most common pattern appeared to

occur by an L1&Lx or VILA integration of an *hprt*-plasmid-*hprt* concatemer, as seen in 6 of 15 clones analyzed with *Bam*HI, *Hind*III, and *Nco*I digests (Fig. 4B, clones 2, 5, 7, 8, 9, and 12). Clones 1, 3, and 6 are very similar to the forementioned samples in that the fragment shift on the *Nco*I digest is the same. However, a *Bam*HI digest hybridized to exons 2 and 3 revealed that although the 7.0- and 5.6-kb fragments were the same, the expected 4.8-kb fragment was only 4.6-kb. The size difference and hybridization pattern can be explained by limited exonuclease digestion associated with the end-to-end ligation (25). The pattern seen in clone 13 is an L1&Lx integration of an *hprt*-*hprt* dimer or a VILA

ogous free ends are important in determining the frequency of gene targeting, contrary to another report (22). IV6.8 cut within the homology region to give two adjacent homologous free ends targets at a higher frequency than RV6.8 cut at either the 5' or 3' edge of homology to give one free homologous end or cut at both ends to give two nonadjacent homologous free ends. In yeast vectors, two adjacent homologous free ends are critically important in initiating the most efficient mechanism for recombination exchange, while vectors with one or no free homologous ends utilize a separate, less efficient mechanism (15). It is possible that ES cells also utilize two separate mechanisms for homologous recombination and that the most efficient mechanism requires two adjacent homologous free ends.

(iii) **Uninterrupted homology.** The replacement and insertion vectors are different in the configuration of homology in that the *neo* cassette interrupts the homology in the replacement vector but not in the insertion vector. It is possible that internal heterology might reduce the frequency of targeting for replacement vectors.

An earlier report has described the same targeting frequency for insertion and replacement vectors that target the 3' end of *hprt* (22). The insertion vectors used in those studies are different from the one used here in that the 5' and 3' homologous ends were directly joined together and the *neo* gene interrupts the homology. In our vectors, the 5' and 3' *hprt* ends were joined to the plasmid and the *neo* marker, thereby retaining a long uninterrupted stretch of homology. Thus, the entire *hprt* homology is topologically more similar to the target gene; this may improve the gene-targeting efficiency.

Insertion versus replacement vector: mechanism. We have shown that our replacement vectors commonly integrated the entire construct, including the nonhomologous free ends, into the targeted locus. This was not expected because the *neo* gene was located within the homology and the vector was linearized at the edge of homology, so that gene replacement events should predominate (11, 22). There are two possible explanations for the integration of the entire replacement vector: gene replacement (double reciprocal recombination/gene conversion of a concatemer) and vector insertion (single reciprocal recombination of a recircularized vector or a circularized concatemer) (Fig. 3A).

If replacement vectors form a head-to-tail concatemer of two units, there are potentially four different patterns that would be identified on Southern blot, depending on the location of the crossover points. The *neo* gene divides the replacement vector homology into two parts, a 5' long arm and a 3' short arm. We have designated the first and second long arms as L1 and L2, respectively, and the first and second short arms as S1 and S2, respectively. Head-to-tail dimers could recombine on the long or short arm of the first and second unit to integrate the entire vector at the following crossover points: L1&L2, S1&S2, and L1&S2. Concatemers of greater than two units may integrate *x* units into the target locus, so that the crossover points may be L1&L_x, S1&S_x, and L1&S_x. A gene replacement event utilizing only one unit may also occur, L1&S1. Of the 17 clones identified to have integrated the entire construct, 13 appear to have integrated more than one unit when replacement vectors were linearized at a single site. Others have also observed concatemers at the target locus (18, 23, 24).

There are three possible recombination products if a single vector or a concatemer recircularizes. A single crossover point on either the long or short arm may occur and integrate

the entire construct, or crossover junctions may form on both the long and short arm for a gene replacement event.

Southern analysis of targeted clones generated with replacement vectors showed that 23 of 41 targeted clones had insertion events that involved crossover points on only the long arm. This integration pattern cannot differentiate between the alternative pathways of a double crossover on L1 and L_x of a multimer or a single crossover on the long arm of a recircularized vector or circularized multimer. Targeting with a crossover point on the short arm may not be observed due simply to fivefold-lesser homology. An L1&S_x replacement was observed for only one clone. This integration pattern can only be generated by double reciprocal recombination of a multimer.

There have been two reports that showed that replacement vectors target *hprt* only by gene replacement (4, 22). However, neither identified targeted clones by Southern analysis with external probes, and both analyzed only one arm for targeting either by PCR (4) or by Southern analysis (22). The true structure of these targeted clones may not be fully understood. We show here that targeted colonies in which the entire vector is integrated will give wild-type fragments (*Hind*III digest, external probe for all L1&L_x and VILA integration patterns) or gene replacement fragments (*Hind*III digest, external probe for the L1&S_x integration), depending upon the digest and probes used. In order to fully understand the integration structure of targeted clones, the DNA must be tested with multiple digests and hybridized with probes both internal and external to the targeting sequences.

We released the plasmid from the targeting homology in a replacement vector in order to evaluate the effect it had on the targeting frequency and integration pattern. The targeting frequency was unaltered, and we did not detect any gene replacement events in 15 clones analyzed. Concatemers of *hprt*-plasmid-*hprt* were found in seven clones. Another three clones showed a very similar pattern, except that the 4.8-kb fragment appeared to be about 200 bp shorter, probably the result of exonuclease activity involved in the end-to-end ligation (25). One clone was shown to be an L1&L2 integration of an *hprt*-*hprt* dimer or VILA integration of a circularized *hprt* unit.

The inclusion of heterology on the free ends of a replacement vector in the recombination product has not been reported previously, perhaps owing to the selection and screening techniques used to isolate targeted events (2, 6, 7, 9, 11, 12, 20, 23, 27). Positive and negative selection involves the integration of a positively selectable marker that interrupts the homology and the removal of a negatively selectable marker located at the edge of homology (11). Targeted clones that incorporate external heterology will include an intact negatively selectable marker and die in selection. There has been one reported clone for which positive and negative selection has recovered a multiple repeat at the target locus although the negatively selectable markers were removed (23). We have observed clones targeted by L1&L_x or VILA of a replacement vector that show an alteration possibly due to exonuclease activity associated with the joining of free ends. Such an alteration may remove the negatively selectable marker and allow tandem repeats to survive in negative selection.

Use of PCR in screening for targeted clones involves amplification of a junction fragment that is only generated following homologous recombination (2, 7, 9, 12, 20, 27) and absolutely requires crossover events on the short arm of homology. The use of PCR will screen against integration

events with crossover points restricted to the long arm of the vector. Most of the TGF clones targeted with RV6.8 and RV6.8PGK integrated the entire construct, with the crossover points restricted to the long arm; therefore, only 3 of 40 clones described here would be detected as positive by a PCR assay.

The insertion vector IV6.8 was shown to integrate the entire construct, as expected, although clones with gene replacement events would not integrate the *neo* and would die in selection. Our experiments do not determine whether the IV6.8 integration pattern was due to single reciprocal recombination at the adjacent homologous free ends or due to double reciprocal recombination of a concatemer, although the frequencies suggest that the former pathway is more likely. The differences in the targeting frequency between insertion and replacement vectors suggest that a different mechanism is occurring with insertion vectors, such as double-strand-break repair (21).

DNA transfected into a cell can clearly follow a variety of integration pathways. With replacement vectors, our data cannot discriminate whether end-to-end ligation to form concatemers or circles occurs more quickly than the targeted integration reaction or whether a concatenated or circularized molecule is a more suitable template for recombination. With insertion vectors electroporated at the same DNA concentration, we rarely observed insertions of concatemers. This suggests that the adjacent homologous free ends are rapidly assimilated in the targeting reaction and are not available for end-to-end joining. The differences in the targeting frequency between insertion and replacement vectors suggest that recircularized replacement vectors or circularized concatemers can use the more efficient pathway, although there are likely to be mechanistic differences in these events compared with those with vectors which have a double-strand break in homology.

ES cells appear to have a high affinity for targeting by the L1&Lx or VILA pathway and frequently incorporate heterologous free ends of the replacement vector into the target locus; the majority of events we analyzed were of this type. The use of vectors that utilize the most efficient pathway for homologous recombination may be important for genes that target at a low efficiency. The use of both PCR and negative selection with replacement vectors will enhance recovery of targeted clones and purify the population for replacement events; however, these strategies will also eliminate 95% of the targeted events. The observations of different classes of targeted recombinants has important implications for the design, screening, and analysis of targeted clones that will ultimately be used for genetic studies in transgenic mice.

ACKNOWLEDGMENTS

We acknowledge J. D. Wallace for technical assistance and thank our colleagues Mark Brenneman, Ann Davis, Ramiro Ramírez-Solis, John Wilson, and Hui Zheng for critical review of the manuscript.

This work was supported by grants from the NIH (A.B.), the Searle Scholars Program/Chicago Community trust (A.B.), the Granada Corporation (P.H.), and the Cystic Fibrosis Foundation (A.B., P.H., and J.R.).

REFERENCES

- Adra, C. N., P. H. Boer, and M. W. McBurney. 1987. Cloning and expression of the mouse *pgk-1* gene and the nucleotide sequence of its promoter. *Gene* **60**:65–74.
- DeChiara, T. M., A. Efstratiadis, and E. J. Robertson. 1990. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature (London)* **345**:78–80.
- Doetschman, T., R. G. Gregg, N. Maeda, M. L. Hooper, D. W. Melton, S. Thompson, and O. Smithies. 1987. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature (London)* **330**:576–578.
- Doetschman, T., N. Maeda, and O. Smithies. 1988. Targetted mutation of the *Hprt* gene in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **85**:8583–8587.
- Holliday, R. 1964. A mechanism for gene conversion in fungi. *Genet. Res.* **5**:282–304.
- Johnson, R. S., M. Sheng, M. E. Greenberg, R. D. Kolodner, V. E. Palaioannou, and B. M. Spiegelman. 1989. Targeting of nonexpressed genes in embryonic stem cells via homologous recombination. *Science* **245**:1234–1236.
- Joyner, A. L., W. C. Skarnes, and J. Rossant. 1989. Production of a mutation in mouse *En-2* gene by homologous recombination in embryonic stem cells. *Nature (London)* **338**:153–156.
- Koller, B. H., P. Marrack, J. W. Kappler, and O. Smithies. 1990. Normal development of mice deficient in β_2 M class I proteins, and CD8⁺ T cells. *Science* **248**:1227–1230.
- Koller, B. H., and O. Smithies. 1989. Inactivating the β_2 -microglobulin locus in mouse embryonic stem cells by homologous recombination. *Proc. Natl. Acad. Sci. USA* **86**:8932–8935.
- 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.
- 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.
- 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.
- 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.
- Meselson, M. S., and C. M. Radding. 1975. A general model for genetic recombination. *Proc. Natl. Acad. Sci. USA* **72**:358–361.
- 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.
- Pfarr, D. S., L. A. Rieser, R. P. Woychik, F. M. Rottman, M. Rosenberg, and M. E. Reff. 1986. Differential effects of polyadenylation regions on gene expression in mammalian cells. *DNA* **5**:115–122.
- Schwartzberg, P. L., S. P. Goff, and E. J. Robertson. 1989. Germ-line transmission of a *c-abl* mutation produced by targeted gene disruption in ES cells. *Science* **246**:799–803.
- Schwartzberg, P. L., E. J. Robertson, and S. P. Goff. 1990. Targeted gene disruption of the endogenous *c-abl* locus by homologous recombination with DNA encoding a selectable fusion protein. *Proc. Natl. Acad. Sci. USA* **87**:3210–3214.
- 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.
- 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.
- Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstien, and F. Wong-Stahl. 1983. The double-strand-break repair model for recombination. *Cell* **33**:25–35.
- Thomas, K. R., and M. R. Capecchi. 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**:503–512.
- Thomas, K. R., and M. R. Capecchi. 1990. Targeted disruption of the murine *int-1* proto-oncogene resulting in severe abnormalities in the midbrain and cerebellar development. *Nature (London)* **346**:847–850.
- Thompson, S., A. R. Clarke, A. M. Pow, M. L. Hooper, and D. W. Melton. 1989. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic

- stem cells. *Cell* **56**:313–321.
25. **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.
26. **Wong, E. A., and M. R. Capecchi.** 1986. Analysis of homologous recombination in cultured mammalian cells in transient expression and stable transformation assays. *Somat. Cell Mol. Genet.* **12**:63–72.
27. **Zijlstra, M., E. Li, F. Sajjadi, S. Subramani, and R. Jaenisch.** 1989. Germ-line transmission of a disrupted β_2 -microglobulin gene produced by homologous recombination in embryonic stem cells. *Nature (London)* **342**:435–438.