

## The Length of Homology Required for Gene Targeting 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*

Received 24 May 1991/Accepted 20 August 1991

**Homologous recombination has been used to introduce site-specific mutations into murine embryonic stem (ES) cells with both insertion and replacement vectors. In this study, we compared the frequency of gene targeting with various lengths of homology and found a dramatic increase in targeting with an increase in homology from 1.3 to 6.8 kb. We examined in detail the relationship between the length of homology and the gene-targeting frequency for replacement vectors and found that a critical length of homology is needed for targeting. Adding greater lengths of homology to this critical length has less of an effect on the targeting frequency. We also analyzed the lengths of homology necessary on both arms of the vector for gene replacement events and found that 472 bp of homology is used as efficiently as 1.2 kb in the formation and resolution of crossover junctions.**

Gene targeting, homologous recombination between introduced DNA and chromosomal DNA, has been used to make site-specific mutations in embryonic stem (ES) cells (3, 4, 6, 8, 9, 12, 14, 15). These mutations may subsequently be studied at the phenotypic level in chimeras and their progeny (9, 14). Despite the increasingly widespread use of gene-targeting schemes in ES cells, very little is known about the conditions that govern the frequency of gene targeting.

The essential feature of any gene-targeting vector is the sequence which is homologous to the target gene. Various lengths of homology may be important in the initial search, strand pairing, branch migration, and ultimate resolution of the recombination complex. The absolute length and distribution of homology in a targeting vector may have an impact on both the frequency of gene targeting and the integration pathway into the homologous target.

A previous investigation of extrachromosomal intramolecular recombination in mammalian cells illustrated a linear increase in the recombination frequency with the increased length of homology from 0.25 to 5 kb and a steep reduction in the frequency below 0.25 kb (11). These studies demonstrate that the recombination machinery may still function with limited homology, but they do not address vector/chromosome recombination which may be subject to different rate-limiting constraints. More recently, studies by Thomas and Capecchi have suggested an exponential relationship between target homology and targeting frequency with insertion and replacement vectors (15). However, another study observed a linear relationship between the length of homology and the targeting frequency (13).

We have described very different targeting frequencies between an insertion and replacement vector containing the same 6.8 kb of homologous sequences and alternative targeted integration patterns by replacement vectors (4). This suggests that aspects of vector configuration other than homology can greatly influence the frequency of the recombination event, and therefore the homology search may not be limiting. Targeting experiments at the *DHFR* locus with normal and amplified cell lines have also suggested that the

homology search may not be limiting with 4.6 kb of targeting homology (17).

In the context of the hypoxanthine phosphoribosyltransferase (*hprt*) locus (10) in XY ES cells, all of the targeted clones can be isolated by selection for *hprt* inactivity in 6-thioguanine (TG). We investigated the length of vector homology that is required for efficient gene targeting with both insertion and replacement vectors and found a dramatic increase in the absolute targeting frequency when the length of homology was increased from 1.3 to 6.8 kb. A more detailed analysis of replacement vectors showed that homology of less than 1.7 kb was insufficient to generate targeted events. An additional 225 bp increased the frequency by at least fivefold. Homology increases from 1.9 to 4.2 kb and from 4.2 to 6.0 kb resulted in 16- and 3-fold, respectively, increases in the targeting frequency. We have also determined that the targeting of replacement vectors with just 472-bp sequence homology on the short arm was as efficient as 1.2 kb for a double-crossover (gene replacement) event.

### MATERIALS AND METHODS

**Construction of vectors.** (i) **Insertion vectors.** IV1.3 is an *hprt* *EcoRI* fragment which contains exon 3. IV6.8 is a *SacI-EcoRI* fragment which overlaps the IV1.3 homology and contains exons 2 and 3. MC1neopA (Stratagene) was inserted into the pTZ (Pharmacia) polylinker.

(ii) **Replacement vectors that contain only exon 3.** RV1.3 is an *EcoRI* fragment, RV1.7 is a *HindIII-EcoRI* fragment, and RV1.9 is an *AccI-EcoRI* fragment.

(iii) **Replacement vectors that contain exons 2 and 3.** RV4.2 is a *PvuII-EcoRI* fragment, RV6.0 is a *BamHI-EcoRI* fragment, RV6.1 is a *SacI-AccI* fragment, and RV6.8 is a *SacI-EcoRI* fragment. MC1neopA was inserted into the *XhoI* site of exon 3. RV1.3PGK and RV6.8PGK are the same as RV1.3 and RV6.8 except that PGKneobpA (14) was inserted into the *XhoI* site in exon 3.

**Electroporations and tissue culture.** The procedures have been described elsewhere (4, 9). Briefly, the DNA for electroporations was prepared by Triton lysis and banded once in CsCl. It was then linearized with the specified restriction enzyme and purified. The insertion vectors were

\* Corresponding author.

linearized in the homology (*XhoI*). RV1.3 was linearized in the polylinker on the 5' side of homology (*SacI*). RV1.7 was linearized in the polylinker on the 5' side of homology (*SacI*) or in the polylinker on the 3' side of homology (*SacI*). RV1.9, RV4.2, and RV6.0 were linearized in the polylinker on the 3' side of homology (*SacI*). RV6.1 and RV6.8 were linearized at the 5' edge of homology (*SacI*) or in the polylinker on the 3' side (*SacI*). RV1.3PGK was linearized in the polylinker on the 5' side of homology (*SacI*), and RV6.8PGK was linearized in the polylinker on the 3' side of homology (*SacI*). The location of the linearization site, on either the 5' or 3' side of homology, did not alter the targeting frequency (4). The linearized vector was compared with uncut vector by gel electrophoresis migration to ensure the vector was completely cut. Electroporations were performed on AB1 (9) cells resuspended in phosphate-buffered saline at a density of  $10^7$  cells per ml. DNA (25  $\mu$ g) was electroporated in 1 ml at 575 V/cm and 500  $\mu$ F with a Bio-Rad Gene Pulser. The same microgram quantity was electroporated for each vector; therefore, the molar quantity differs by as much as a factor of 2. Only the absolute frequencies are affected by this molar difference. The absolute frequencies in Tables 1 and 2 are not adjusted to compensate for the molar difference; however, comparisons of absolute transfection and targeting frequencies between vectors of different size are adjusted in the text (there is a linear increase in transfection with an increase in the DNA concentration within this range [2]). 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 (9).

Three different methods were used to determine G418<sup>r</sup> and TG<sup>r</sup> as described elsewhere (4). In all three methods, the cells were not replated before selection such that each clone represents a single event. Briefly, (i) 24 h after electroporation, 180  $\mu$ 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<sup>r</sup> 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 TG<sup>r</sup> colonies were counted. (ii) Plates were initially selected in 180  $\mu$ g of G418 (active ingredient) per ml, and after 5 or 6 days a final concentration of  $10^{-5}$  M TG was added; selection was maintained for another 10 days. (iii) Cells were scored by the incorporation of [<sup>3</sup>H]hypoxanthine. Transfected cells were selected in G418 for 10 days and then labeled for 24 h in 4 ml of medium containing 0.04  $\mu$ Ci of [<sup>3</sup>H]hypoxanthine. Cells that contained a functional *hprt* incorporated the labeled nucleoside and stained intensely black after fixation and exposure to a photographic emulsion. White colonies were scored as *hprt* negative.

All three methods were used for RV1.3 and RV6.8, and all three methods gave the same frequency of *hprt* inactivity for both constructs. For the other constructs, 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.** A *Bam*HI digest of genomic DNA was hybridized to a probe internal to the vector, *hprt* exons 2 and 3 (cDNA cut with *Hinc*II and *Hpa*II [10]); an *Nco*I digest of genomic DNA was 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). *Bam*HI-digested DNA (5  $\mu$ g) was separated by electrophoresis through a 0.7% agarose gel for 18 h at 60 V in 0.04 M Tris-acetate–0.001 M EDTA (TAE) buffer. The *Nco*I-digested DNA was

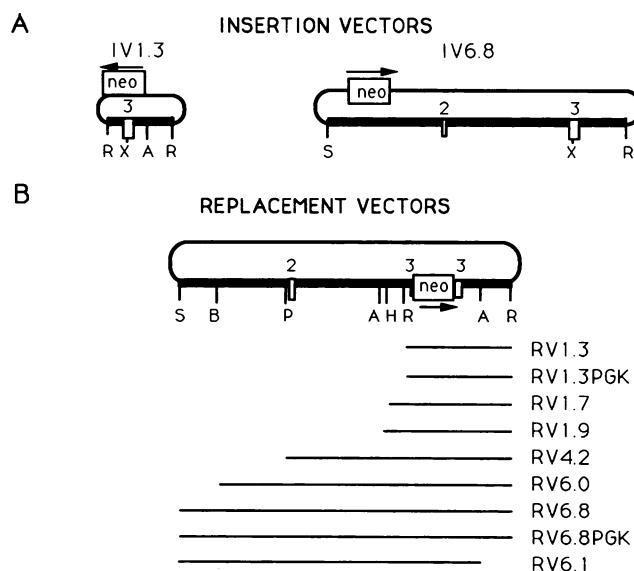


FIG. 1. *hprt* targeting vectors. The thick line is *hprt* homology in the vector, the thin line is plasmid, the box is the *neo* cassette (MC1neopA for all vectors except RV1.3PGK and RV6.8PGK, in which it is PGKneopA), and the arrow is the direction of transcription. Exons 2 and 3 are labeled rectangles. Length of DNA is in kilobases. Restriction sites are *Acc*I (A), *Bam*HI (B), *Eco*RI (R), *Hind*III (H), *Pvu*II (P), *Sac*I (S), and *Xho*I (X). (A) Insertion vectors IV1.3 and IV6.8; (B) replacement vectors RV1.3, RV1.3PGK, RV1.7, RV1.9, RV4.2, RV6.0, RV6.8, RV6.8PGK, and RV6.1.

separated by electrophoresis through a 0.7% gel in TAE buffer for 24 h, using field inversion gel electrophoresis at 4°C (the buffer was circulated). DNA was transferred onto a GeneScreen Plus filter (Dupont) and hybridized to the specified probe under standard conditions. The probe was labeled by the random primed method, using the conditions specified by the manufacturer (Boehringer Mannheim).

## RESULTS

**Insertion vectors: a 250-fold difference in the targeting frequency with vectors of 1.3 and 6.8 kb.** The insertion vectors IV1.3 and IV6.8 (Fig. 1A) were tested for their gene-targeting frequency. IV1.3 contains 1.3 kb of *hprt* DNA, and IV6.8 also contains the 5' contiguous 5.5-kb *Eco*RI fragment of *hprt* DNA. Both constructs contain a *neo* expression cassette (MC1neopA), outside the *hprt* homology. The constructs were cut within the *hprt* homology at the *Xho*I site in exon 3 and electroporated into ES cells. The transfected ES cells were selected for G418 and TG resistance. The relative targeting frequencies are presented in Table 1.

Southern analysis of the TG<sup>r</sup> colonies generated with IV1.3 confirmed that the vector had inserted into the *hprt* locus (Fig. 2). For these clones, the targeted events generated a 1.3-kb duplication of *hprt* homology separated by the *neo* gene and plasmid sequences. A *Bam*HI digest hybridized to exons 2 and 3 (exon 2 is external to the vector and exon 3 is internal to the vector) revealed that the 7.0-kb wild-type fragment was altered to fragments of 10.1 and 2.3 kb, as expected. An *Nco*I digest hybridized to a 3' external probe reduced the 23-kb wild-type fragment to 14 kb as a result of the introduction of the *Nco*I site in MC1neopA. Previous analysis on TG<sup>r</sup> clones generated with IV6.8 has

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

Vector	Homology (kb)		Linear site	Total electro-porations <sup>a</sup>	Total G418 <sup>r</sup>	G418 <sup>r</sup> /electro-poration	Total <i>hprt</i> <sup>-</sup>	<i>hprt</i> <sup>-</sup> /G418 <sup>r</sup>
	5'	3'						
IV1.3	0.132	1.2	<i>Xho</i> I	24	1,609	67	3	1/536
IV1.3	0.604	0.7	<i>Acc</i> I	10	2,055	205	1	1/2,055
IV6.8	5.6	1.2	<i>Xho</i> I	66	47,353	717	1,006	1/47

<sup>a</sup> Conditions for electroporation:  $10^7$  cells, 25  $\mu$ g of DNA per ml.

demonstrated similar predicted targeted alleles generated by vector insertion in 19 of 20 TG<sup>r</sup> colonies analyzed (4).

The increase in homology from 1.3 to 6.8 kb increased the absolute targeting frequency by at least 250-fold and the relative targeting frequency by 11-fold (Table 1). The relative targeting frequency for IV1.3 is artificially high compared with that of IV6.8 due to poor expression of MC1neopA in IV1.3. MC1neopA expression is known to be sensitive to position effects (9, 14). Sequences in intron 2 of the murine *hprt* gene enhance expression of MC1neopA, which increases the background of G418<sup>r</sup> colonies for IV6.8 by 21-fold. The low targeting frequency for IV1.3 may be due to rate-limiting steps in the homology search such that 1.3 kb of homology may be inadequate to find the target locus or

assemble a recombination complex. Alternatively, there was just 132 bp of homology on the 5' end, which may be too little homology on one end for efficient homologous recombination.

**IV1.3 targets inefficiently because of a lack of total homology.** The inefficient targeting of IV1.3 may be due to a lack of total homology (1.3 kb) or too little homology on the 5' end (132 bp). To test whether the distribution of homology or total homology was responsible for the poor targeting efficiency, IV1.3 was linearized near the center of homology with *Acc*I, leaving homologous arms of 604 and 700 bp (Fig. 1A). IV1.3 cut with *Acc*I generated one TG<sup>r</sup> colony in  $10^8$  cells electroporated (Table 1). IV1.3 cut with either *Xho*I or *Acc*I generated targeted colonies at a similar absolute frequency. This indicated that 1.3 kb of total homology was inadequate to target *hprt* efficiently. Presumably, the limited homology of 1.3 kb becomes rate limiting in the targeting pathway because of the search for the target sequences or formation of recombination intermediates.

**Replacement vectors: a 190-fold difference in targeting efficiency with vectors of 1.3 and 6.8 kb.** Replacement vectors (Fig. 1B) RV1.3 and RV6.8 (4) were tested for their gene-targeting frequency at the *hprt* locus. RV1.3 and RV6.8 contain the same *hprt* fragments as do IV1.3 and IV6.8, respectively. The MC1neopA expression cassette was inserted inside *hprt* exon 3 at the *Xho*I site. Prior to transfection, these vectors were linearized by cutting outside the targeting homology. ES cells were selected in G418 and TG. RV1.3 did not yield any TG<sup>r</sup> colonies in  $34 \times 10^7$  cells electroporated (Table 2). RV6.8 generated TG<sup>r</sup> colonies at an absolute frequency of 2.9 TG<sup>r</sup> colonies per  $10^7$  cells electroporated. We have previously shown that only a fraction of these TG<sup>r</sup> targeted clones have the predicted gene replacement event (4).

An increase in homology from 1.3 to 6.8 kb increased the targeting frequency. There was at least a 190-fold difference in the absolute targeting frequency between RV1.3 and RV6.8. We recovered TG<sup>r</sup> targeted clones with IV1.3 but not with RV1.3. The reason for this may reflect either our previous observations of the significantly higher targeting frequency with insertion vectors (4) or the possibility that low transfection efficiency of RV1.3 prohibits targeting.

To determine the effect of the transfection frequency on the targeting frequency, the MC1neopA selection cassette was replaced with the position-independent cassette PGK neobpA (14) in both RV1.3 and RV6.8 to make RV1.3PGK and RV6.8PGK (4) (Fig. 1B). Unlike RV1.3 and RV6.8, RV1.3PGK and RV6.8PGK have similar transfection efficiencies (Table 2). Despite the high transfection frequency for RV1.3PGK, there were no TG<sup>r</sup> colonies generated out of 46,144 G418<sup>r</sup> colonies. The absolute targeting frequencies for RV6.8PGK were the same as for RV6.8 (Table 2). Southern analysis for clones targeted with RV6.8PGK have been described previously (4).

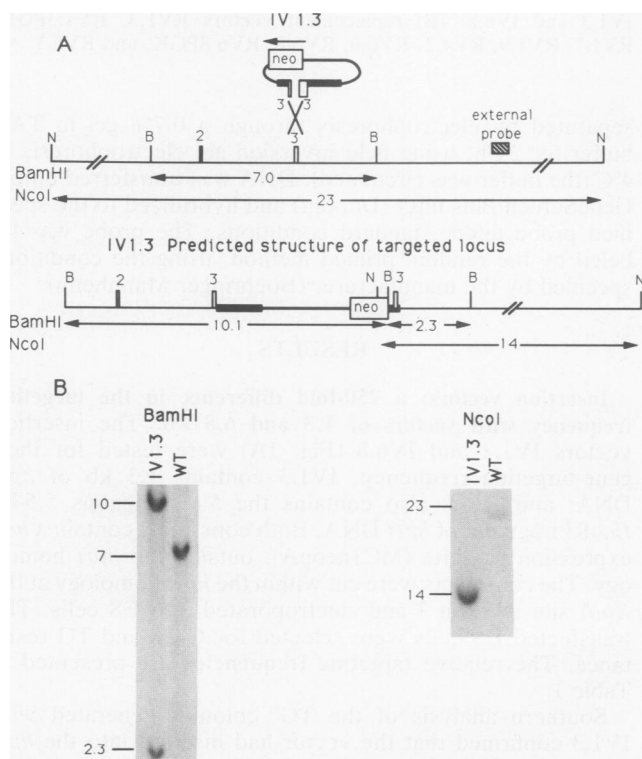


FIG. 2. Southern blot analysis for IV1.3. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid. MC1neopA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as a probe (taken from cDNA) for the *Bam*HI digest. The striped box is the external probe for the *Nco*I digest. The length of DNA is in kilobases. Shown are fragments and predicted size fragments for single-reciprocal recombination seen with a *Bam*HI (B) digest hybridized to exons 2 and 3 and an *Nco*I (N) digest hybridized to the external probe. WT, wild type.

TABLE 2. Replacement vectors which target the *hprt* locus

Vector	Homology (kb)		Total electro- porations	Total G418 <sup>r</sup>	G418 <sup>r</sup> /electro- poration <sup>a</sup>	Total <i>hprt</i> <sup>-</sup>	<i>hprt</i> <sup>-</sup> /electro- poration	<i>hprt</i> <sup>-</sup> / G418 <sup>r</sup>	Integration pattern <sup>b</sup>				
									Vector insertion		Gene replace- ment	VIM	UD
	5'	3'							Long arm	Short arm			
RV1.3	0.132	1.2	34	3,752	110	0							
RV1.3PGK	0.132	1.2	15	46,144	3,076	0							
RV1.7	0.536	1.2	28	4,057	145	0							
RV1.9	0.761	1.2	51	32,041	628	7	0.14	1/4,577	0	1	1	0	0
RV4.2	3.0	1.2	28	43,390	1,550	43	1.5	1/1,009	3	0	0	2	0
RV6.0	4.8	1.2	28	51,758	1,848	107	3.8	1/483	1	0	2	2	0
RV6.8	5.6	1.2	22	17,846	811	64	2.9	1/279	12	0	1	3	1
RV6.8PGK	5.6	1.2	21	28,702	1,367	58	2.8	1/495	5	0	2	1	1
RV6.1	5.6	0.472	29	15,547	536	37	1.5	1/420	12	1	7	3	1

<sup>a</sup> Conditions for electroporation:  $10^7$  cells, 25  $\mu$ g of DNA per ml.

<sup>b</sup> Vector insertion, integration of the entire vector, including the bacterial plasmid, with crossover junctions on only the long or short arm (4); gene replacement, a nonreciprocal exchange of information that replaces the target locus with only the mutation, the bacterial plasmid is not included (4, 15); VIM, vector integration modified, the integration pattern of the vector is not understood (4); UD, undetermined, the integration of the targeting vector could not be verified.

**Replacement vectors: the influence of total homology on the targeting frequency.** An overlapping set of replacement vectors was made to determine the relationship between the total length of homology and the frequency of homologous recombination. We have progressively removed 5' *hprt* homology from RV6.8 to make replacement vectors with 1.7 (RV1.7), 1.9 (RV1.9), 4.2 (RV4.2), and 6.0 (RV6.0) kb of *hprt* homology (Fig. 1B). The vectors were linearized outside the homology prior to transfection. The targeting frequencies are presented in Table 2 and Fig. 3. RV1.7 did not generate any TG<sup>r</sup> colonies in  $36 \times 10^7$  cells electroporated; RV1.9, RV4.2, and RV6.0 generated TG<sup>r</sup> colonies at absolute frequencies of  $0.14 \times 10^{-7}$ ,  $1.5 \times 10^{-7}$ , and  $3.8 \times 10^{-7}$  cells, respectively.

Southern analysis of TG<sup>r</sup> colonies generated with RV1.9, RV4.2, and RV6.0 showed the same alternative integration pathways that we described previously (4) (data not shown). In total, there were three gene replacement events (a nonreciprocal exchange that occurs via double-reciprocal recombination or gene conversion), five vector integration events

(integration of the entire vector, including the bacterial plasmid), and four modified integration events in which the vector integrated into the target locus in an undefined manner.

An increase of *hprt* homology from 1.3 to 1.7 kb did not result in the detectable generation of targeted colonies; however, a further increase of just 225 bp to 1.9 kb did generate targeted colonies. An addition of *hprt* homology from 1.9 to 4.2 kb resulted in a 16-fold elevation in the targeting frequency, and a further increase of homology from 4.2 to 6.0 kb resulted in an additional 3-fold increase in targeting. Greater homology did not increase the targeting frequency (Fig. 3).

**Crossover junctions can utilize just 472 bp of homology in gene replacement events.** One explanation for the low targeting frequency of vectors such as RV1.3 and RV1.7 is that the amount of homology on one vector arm is too short to form a productive recombination complex. This step would then become rate limiting in the formation of the final recombination product. To test this, we constructed RV6.1, which has just 0.472 kb on the short arm, 64 bp less than RV1.7, which did not generate targeted colonies at a detectable frequency. RV6.1, linearized outside the homology, generated TG<sup>r</sup> colonies at a frequency of  $1.3 \times 10^{-7}$  cells electroporated. This is a modest decrease in the targeting frequency compared with that of RV6.8, which has 1.2 kb on the short arm. It was important to verify that the 472-bp short arm was utilized for crossovers because the majority of targeted clones do not form crossover junctions on the short arm of homology for RV6.8 (4).

Southern analysis on 24 TG<sup>r</sup> colonies generated with RV6.1 (Fig. 4) determined that 7 were targeted by a gene replacement event (Fig. 4B, clones 9, 10, 12, and 13) and 12 were targeted by a vector insertion event on the long arm (Fig. 4B, clones 7, 11, and 15), 1 was targeted by a vector insertion event on the short arm (Fig. 4B, clone 8), and 3 were targeted by a modified integration event (Fig. 4B, clone 14). A *Bam*HI digest hybridized to an internal probe (exons 2 plus 3) revealed that the 7.0-kb wild-type band was altered to 5.9 and 2.3 kb for gene replacement events, 7.0, 5.9, and 4.2 kb for vector integration events that utilize the long arm, and 9.8, 6.6, and 2.3 kb for the vector integration events that utilize the short arm. An *Nco*I digest hybridized to an

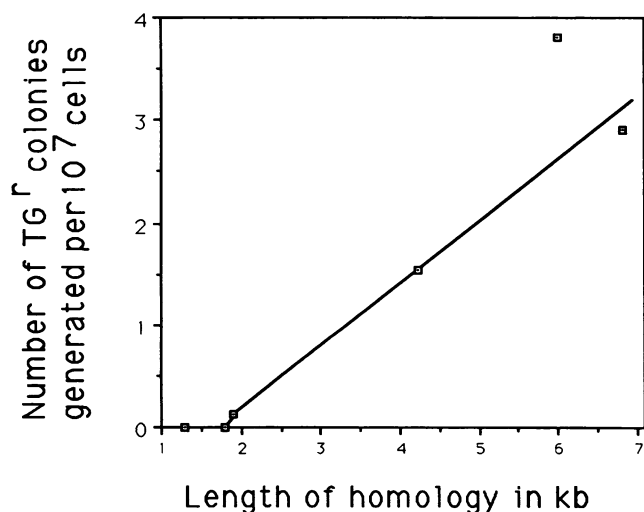


FIG. 3. Absolute frequency of recombination plotted against the length of homology to the target sequence with replacement vectors.

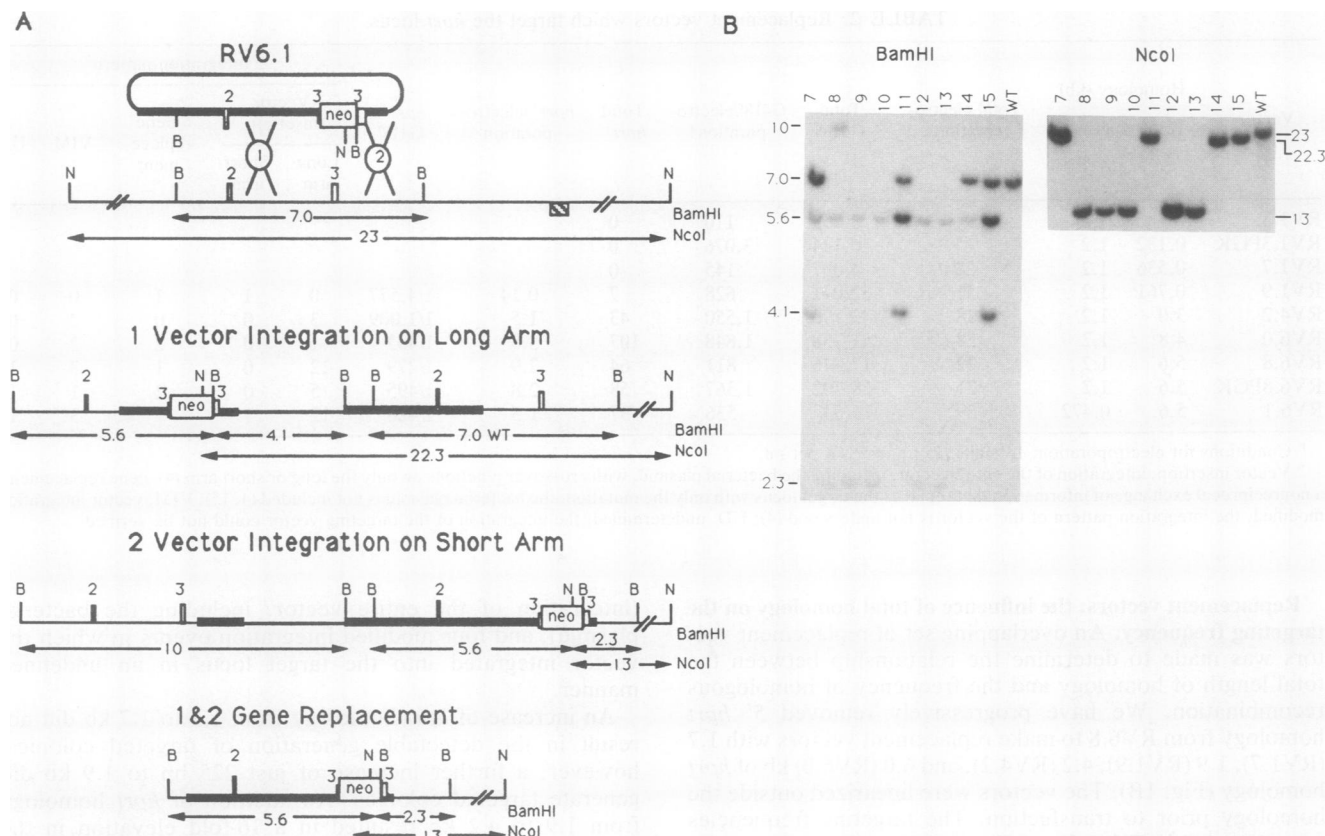


FIG. 4. Southern blot analysis for RV6.1. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid. MC1neopA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as the internal probe (taken from cDNA). The striped box is the external probe. The encircled number 1 represents the crossover point on the long arm, and the encircled number 2 represents the crossover point on the short arm. The length of DNA is in kilobases. Shown are predicted size fragments seen with *Bam*HI (B) digests hybridized to exons 2 and 3 and *Nco*I (N) digests hybridized to the external probe. WT, wild type.

external probe altered the 23-kb wild-type band to 13 kb for gene replacement events and the vector integration events on the short arm and 22.3 kb for vector integration events on the long arm. The modified integration event seen in clone 14 in Fig. 4B appears to be a vector integration event on the long arm on the *Nco*I digest. The *Bam*HI digest shows the 4.1-kb band to be absent and a 3.7-kb band to be present. This may have occurred as a result of exonuclease activity associated with the end-to-end joining prior to the integration (4, 16). The other two modified integration events also appear to be vector integrations on the long arm in association with exonuclease activity (data not shown).

These results demonstrate that just 472 bp of sequence homology is adequate to generate a functional crossover event. This implies that the rate-limiting step that reduces the gene targeting frequency for RV1.3 and RV1.7 to below detectable limits is the total length of homology, not the formation and resolution of the crossover junction on the short arm of homology.

## DISCUSSION

The gene-targeting frequency is dependent on the length of the target homology. We have shown there is at least a 200-fold increase in gene-targeting efficiency associated with an increase of homology from 1.3 to 6.8 kb for replacement and insertion vectors.

A more detailed analysis of replacement vectors was unable to detect targeting in a total of  $3.4 \times 10^8$  cells electroporated with 1.3 kb of homology and  $2.8 \times 10^8$  cells electroporated with 1.7 kb of homology. This low targeting frequency was not associated with the low transfection efficiency of the vectors, since the 1.3 kb of homology, modified to contain a *neo* cassette subject to fewer positional effects and with a stronger promoter, did not yield TG<sup>r</sup> colonies in  $46 \times 10^3$  G418<sup>r</sup> cells analyzed. Within the ranges of homology from 1.7 to 1.9, 1.9 to 4.2, and 4.2 to 6.0 kb, there are increases in the targeting frequency of >21-fold, 7-fold, and 1.6-fold per kb, respectively. The largest change in the targeting efficiency compared with a proportional change in homology was the >21-fold per kb seen with the 225-bp increase from RV1.7 to RV1.9. It is possible that the added 225 bp of homology contains a hot spot for recombination. Previously reported data showed that a 2-fold increase in homology resulted in a 20-fold increase in the targeting frequency (15). These added lengths of homology may also have incorporated hot spots for gene targeting. However, a more likely explanation is that a critical length of homology is required for targeting, and once this requirement is met, then further increases in homology have less effect. This has been observed in extrachromosomal recombination in mammalian cells (1, 11). RV1.3 and RV1.7 may not target efficiently because of inadequate total homology or inadequate homology on the short arm.

The total length of homology has a greater impact on the targeting efficiency than does the length of homology on the short arm. Colonies targeted with RV6.1, which only has 472 bp on the short arm, were analyzed to map the length of homology necessary to form and resolve a Holliday (5) junction or initiate and resolve a gene conversion event. The 472-bp short arm was shown to be utilized as efficiently as the 1.2-kb short arm in RV6.8 (both vectors have 5.6 kb of homology on the long arm). This implies that the rate-limiting step that reduces the gene-targeting frequency for RV1.7 (short arm of 536 bp) to below detectable limits is the inadequate total homology and not the amount on the short arm. We have also shown that the targeting frequency of IV1.3 does not detectably change when the vector is cut more centrally.

Replacement vectors usually integrate the entire vector into the target locus, including the bacterial plasmid, with the crossover occurring on the long arm (4). The bias in crossover formation with respect to the long arm indicates a relationship between homology and the formation of a crossover. However, 472 bp is sufficient and apparently not rate limiting for the formation and resolution of crossover junctions. This observation suggests that it is beneficial to have a greater length of homology to search for the target locus and form recombination intermediates than is absolutely necessary to productively recombine and form a crossover junction.

These data also support our previous observation on the difference in recombination frequencies with insertion and replacement vectors (4). Here the 1.3-kb insertion vector is significantly more efficient at targeting than is the 1.3-kb replacement vector. This observation indicates that in the context of the same amount of homologous sequences, the targeting mechanism can also be a rate-limiting step. The formation and resolution of recombination intermediates is probably more efficient with the two adjacent homologous ends presented by an insertion vector as has been demonstrated in yeast cells (7).

There are many facets that affect the frequency of gene targeting. In the context of *hpri*, we have illustrated the relationship between the targeting frequency and the length of homology in a targeting vector. We have previously demonstrated that the targeting frequency can be manipulated by the vector configuration (4). Targeting efficiency is also dependent on the location of the target in the genome. The appreciation of the relationship between vector design and the targeting frequency along with an understanding of the parameters that affect the availability of any locus for gene targeting will ultimately enable the targeting of any genomic region at high efficiency.

#### ACKNOWLEDGMENTS

We acknowledge J. D. Wallace for technical assistance and thank our colleagues Ann Davis, Marty Matzuk, Ramiro Ramirez-Solis, 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

1. Ayares, D., L. Chekuri, K. Song, and R. Kucherlapati. 1986. Sequence homology requirements for intermolecular recombination in mammalian cells. *Proc. Natl. Acad. Sci. USA* 83:5199-5203.
2. Davis, A. C. Personal communication.
3. Hastly, P., R. Ramirez-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.
4. Hastly, P., J. Rivera-Perez, C. Chang, and A. Bradley. 1991. Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. *Mol. Cell. Biol.* 11:4509-4517.
5. Holliday, R. 1964. A mechanism for gene conversion in fungi. *Genet. Res.* 5:282-304.
6. Johnson, R. S., M. Sheng, M. E. Greenberg, R. D. Kolodner, V. E. Papaioannou, and B. M. Spiegelman. 1989. Targeting of nonexpressed genes on embryonic stem cells via homologous recombination. *Science* 245:1234-1236.
7. 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.
8. 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.
9. 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.
10. 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.
11. Rubnitz, J., and S. Subramani. 1984. The minimum amount of homology required for homologous recombination in mammalian cells. *Mol. Cell. Biol.* 4:2253-2258.
12. 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.
13. 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.
14. 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.
15. Thomas, K. R., and M. R. Capecchi. 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51:503-512.
16. 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.
17. Zheng, H., and J. Wilson. 1990. Gene targeting in normal and amplified cell lines. *Nature (London)* 344:170-173.