Gene conversion during vector insertion in embryonic stem cells

Paul Hasty[1],*, Jaime Rivera-Pérez[1] and Allan Bradley[1,2]

[1]Department of Human and Molecular Genetics and [2]Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

ABSTRACT

Recombination of an insertion vector into its chromosomal homologue is a conservative event in that both the chromosomal and the vector sequences are preserved. However, gene conversion may accompany homologous recombination of an insertion vector. To examine gene conversion in more detail we have determined the targeting frequencies and the structure of the recombinant alleles generated with a series of vectors which target the *hprt* gene in embryonic stem cells. We demonstrate that gene conversion of the introduced mutation does not significantly limit homologous recombination and that gene conversion occurs without a sequence specific bias for five different mutations. The frequency of the loss of a vector mutation and the gain of a chromosomal sequence is inversely proportional to the distance between the vector mutation and the double-strand break. The loss of a chromosomal sequence and the gain of a vector mutation occurs at a low frequency.

INTRODUCTION

Homologous recombination between introduced and chromosomal DNA is referred to as gene targeting in mammalian cells and the models used to describe gene targeting mechanisms were principally derived from fungal systems (1–3). In mammalian cells the mechanism and final structure of the recombinant allele depends upon the configuration of the targeting vector. Replacement vectors (4) will effectively replace host sequences by either double reciprocal recombination (an exchange of sequences between the two substrates) or by gene conversion (loss of sequences from one substrate and gain of sequences from the other substrate). The observable consequence of either mechanism results in identical mutant alleles because the non-integrated components are degraded; therefore, the recombination mechanism cannot be determined. By contrast, the recombinant allele generated with an insertion vector (4,5) can yield considerably more information because the recombination substrates are preserved. Insertion vectors are characterized by a double-strand break (DSB) inside the target homology and integrate into the

host target by a mechanism which is similar to the double-strand break repair (DSBR) model (3–12).

In this report we examined gene conversion during homologous recombination of an insertion vector into the hypoxanthine phosphoribosyltransferase (*hprt*; 17) locus in embryonic stem (ES) cells. The insertion vector contains a disruption in the *hprt* homology when compared to the chromosomal sequence. This disruption is referred to as a vector mutation when describing the vector sequence. The chromosomal sequence is considered to be wild-type. Gene conversion was not found to significantly reduce the targeting frequency and occurred without a sequence specific bias for five different mutations. A progressively higher frequency of gene conversion that resulted in loss of the vector mutation (gain of the chromosomal sequence) was observed as the distance between the mutation and the DSB decreased. Gene conversion that resulted in loss of the wild-type chromosomal sequence (gain of the vector mutation) was observed at a low frequency. We also demonstrate that large gaps present in the vector sequence were accurately repaired during gene targeting.

MATERIALS AND METHODS

Vector construction

All vectors consist of a *Sac*I–*Eco*RI *hprt* fragment which contains exons 2 and 3 inserted into a pTZ cloning vector (Pharmacia) containing MC1neopA (Stratagene). IV6.8 (10) contains the *hprt* fragment without modifications which was mutated to make all the other vectors (Fig. 1).

IV6.8AN: A *Not*I site was generated by ligating an oligonucleotide (5′-TTGCGGCCGCAA-3′) into an *Acc*I site, previously filled in with nucleotides which divides the homologous sequences into a 6.2 kb long arm and a 0.7 kb short arm. *Not*I linearizes in the mutation; *Xho*I linearizes 472 bp 5′ to the mutation and *Nhe*I linearizes 3.2 kb 5′ to the mutation.

IV6.8Δ86 has an 86 bp deletion that destroys the *Xho*I site in exon 3. The mutation divides the homologous sequences into a 5′, 5.6 kb long arm and a 3′, 1.2 kb short arm. The deletion was constructed by utilizing the 3′ to 5′ exonuclease activity of Klenow (Boehringer Mannheim) in the presence of limiting dNTPs for 30 min at room temperature. To generate a unique linearization site in the vector two of the three existing *Eco*RI sites were destroyed. The site in the MC1neopA promoter was
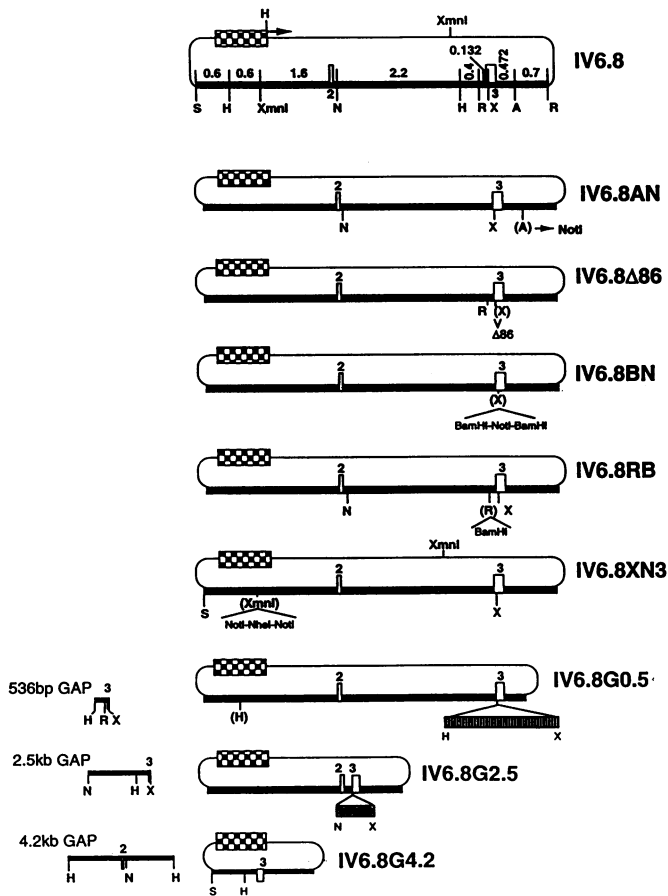

Figure 1. Vectors which target the *hprt* locus. The thick line represents the homologous *hprt* sequences and the thin line represents the plasmid. All of the vectors were derived from the same 6.8 kb *SacI–EcoRI* fragment of *hprt* with exons 2 and 3 (labeled rectangles) as described in the experimental procedures. IV, insertion vector (*neo* outside the *hprt* homology). The checked box indicates the *neo* selection cassette, MC1neopA (Stratagene), and the arrow illustrates the direction of transcription. The striped box represents 'filler' DNA for IV6.8G0.5. The stippled box represents filler DNA for IV6.8G4.2. For the constructs with gaps, the filler DNA replaces the indicated fragment of *hprt* that is displayed to the left. IV6.8G4.2 does not contain filler DNA. H, *Hind*III; N, *Nhe*I; R, *Eco*RI; S, *Sac*I; X, *Xho*I. The mutated restriction enzyme sites are in parenthesis.

destroyed by inserting the oligonucleotide 5′-AATTGCCGGAT-CCGGCCGGATCCGGC-3′ which also generates a *Bam*HI site. The *Eco*RI site at the 3′ end of *hprt* was destroyed by end filling and blunt end ligation. The remaining *Eco*RI site is the linearization site which is 132 bp 5′ to the mutation.

IV6.8BN was generated by inserting a 26 bp oligonucleotide (5′-TCGATGGATCCGCGGCCGCGGATCCA-3′) into the *Xho*I site in exon 3 (9) which destroys the *Xho*I site and generates *Bam*HI–*Not*I–*Bam*HI sites. The *Not*I site serves as the linearization site and divides the homologous sequences into a 5′, 5.6 kb long arm and a 3′, 1.2 kb short arm.

IV6.8RB was generated by inserting a 26 bp oligonucleotide (5′-AATTGCCGGATCCGGCCGGATCCGGC-3′) into the *Eco*RI site of the *hprt* sequences (18) which destroys the *Eco*RI site, generates a *Bam*HI site and divides the homology into a 5′, 5.5 kb long arm and a 3′, 1.3 kb short arm. This vector has two linearization

sites; *Xho*I, 132 bp 3′ to the mutation on the short arm of the vector and *Nhe*I, 2.7 kb 5′ to the mutation on the long arm of the vector.

IV6.8XN3 was generated by inserting of a 22 bp oligonucleotide (5′-TTGCGGCCGCTAGCGGCCGCAA-3′) into the *Xmn*I site in the *hprt* sequences (9) which destroys the *Xmn*I site and generates *Not*I–*Nhe*I–*Not*I sites and divides the homology into a 5′, 1.2 kb short arm and a 3′, 5.6 kb long arm. The linearization site is *Xho*I, 4.3 kb, 3′ to the mutation on the long arm.

IV6.8G0.5 has a 0.5 kb gap. To ensure that ends of the vector DNA which constitute the gap were homologous with the chromosome, a piece of 'non-homologous filler DNA' with compatible ends was utilized. To construct the 0.5 kb gap it was necessary to destroy two additional *Hind*III sites by end filling and blunt end ligation. These sites were the most 5′ *Hind*III site in the *hprt* sequences and a *Hind*III site in the plasmid polylinker. The filler DNA was a 2.3 kb fragment. The linearization sites were *Hind*III and *Xho*I which liberates the filler DNA from the vector. The gap in this vector lies 4.4 kb, 3′ to the filled *Hind*III site in *hprt*. No attempt was made to purify the filler DNA from the vector prior to the electroporation.

IV6.8G2.5 has a 2.5 kb gap between the endogenous *Nhe*I and *Xho*I sites separated by a 600 bp fragment of filler DNA. The vector was cut with *Nhe*I and *Xho*I restriction enzymes for the electroporation.

IV6.8G4.2 has a gap of 4.2 kb between the two endogenous *Hind*III sites. An existing *Hind*III site in the polylinker was destroyed by end filling and blunt end ligation. The vector was linearized with *Hind*III for the electroporation.

Electroporation and tissue culture

The experiments described in Table 1 have been grouped into experiments performed on the same day (experiments A–F). This is necessary since there is some variation in the relative targeting frequencies from experiment to experiment. However, within a group the ratios of the targeting frequencies from one vector to the next are always consistent (9).

Twenty-five µg of cut DNA was electroporated into 10^7 AB1 cells (19) at 575V/cm as previously described (10) and cultured as previously described (20). A single electroporation was plated onto a 90 mm SNL76/7 feeder plate (19).

Two methods were used to determine the ratio of 6-thioguanine (6-TG) resistant clones as a fraction of G418 resistant clones. In both methods the cells were not replated before selection so that each clone represents a single targeting event as previously described (10). In the first procedure 180 µg/ml of G418 (active ingredient) was added to each 90 mm plate 24 h after electroporation. After 10 days the G418 resistant clones were counted. Then 6-TG was added to each plate at a final concentration of 10 µM. After 21 days the 6-TG + G418 resistant colonies were counted. In the alternative protocol the plates were initially selected in 180 µg/ml G418 (active ingredient) and after 5–6 days a final concentration of 10 µM TG was added. The selection was maintained for another 10 days. The first procedure was used for 10–20% of the plates to control for transfection efficiencies and the alternative for 80–90% of the plates. The different protocols always gave the same absolute targeting frequency for each construct.

Southern blot analysis of targeted colonies

The DNA from the 6-TG + G418 resistant clones was digested with a restriction enzyme, separated through a 0.7% TAE gel by

electrophoresis, transferred to Gene Screen filters (Dupont) and hybridized to a probe internal to the vector, *hprt* exons 2 and 3 from the cDNA cut with *Hinc*II and *Hpa*II (17). Some Southern blots were re-probed with *neo* sequences. The use of internal probes for these vectors has been shown to be reliable and result in an accurate determination of the recombinant allele structure (10,18,21). A *Bam*HI digest was performed on DNA from 6-TG + G418 resistant clones generated from all the vectors. Other digests were performed from clones generated with the following vectors: IV6.8AN, *Acc*I digest; IV6.8BN and IV6.8Δ86, *Bam*HI + *Xho*I digest; IV6.8XN3, *Xmn*I and *Nhe*I digests (re-probed with *neo*); IV6.8G0.5, *Hin*dIII and *Eco*RI digests; IV6.8G2.5, *Hin*dIII digest; IV6.8G4.2, *Nhe*I and a *Hin*dIII digest. The probes were labeled with random primed labeling using manufacturer's conditions (Boehringer Mannheim).

PCR and oligonucleotide hybridization

The clones targeted with IV6.8Δ86 were analyzed by PCR with primer combinations specific for the 5′ duplicate. The upstream sense primer was 5′-CCTGATTTTATTTCTATAGG-3′ (specific for the intron 2/exon 3 junction). The downstream antisense primer was 5′-TGCTGCAAGGCGATTAAGTT-3′ and is complementary to the plasmid and thus amplifies the 5′ duplicate. The following conditions were used in 20 μl total volume: 150 ng genomic DNA; 1 U AmpliTaq Polymerase (Perkin-Elmer Cetus); 50 ng of each primer; 250 μM of each dNTP; 1 × Cetus buffer (Perkin-Elmer Cetus), 5% dimethylsulfoxide (DMSO); 16.7 mM MgCl₂. Amplification was performed in a DNA Thermal Cycler (Perkin-Elmer Cetus) for 1 cycle at 93°C for 6 min followed by 38 cycles of denaturing at 93°C for 30 s, annealing at 37°C for 30 s and extending at 70°C for 1 min. The final cycle was extended at 70°C for 15 min.

Oligonucleotide hybridization. The 1.2 and 1.1 kb PCR products were separated by electrophoresis through a 1% TAE gel at 50 V for 4 h. The DNA was transferred on to a Gene Screen Plus filter and hybridized to an oligonucleotide within the 86 bp deletion, y (5′-CCACAATGTGATGGCCTCCC-3′) by standard conditions. After exposure the filter was stripped and hybridized to a second oligonucleotide in exon 3 outside the deletion, z (5′-TTACAT-TAAAGCACTGAATA-3′).

RESULTS

The affect of the position of the double-strand break and vector mutation on the targeting frequency

This study focuses on gene conversion during homologous recombination of an insertion vector. The gene targeting frequency and the integration pattern after homologous recombination was analyzed for a series of vectors that contain the same homologous sequences; however, each vector has a unique mutation. The targeting frequency was determined after the vector DNA was transfected into ES cells by electroporation and selected in G418 for neomycin phosphotransferase activity and in 6-TG for the absence of Hprt function. Southern analysis was performed to confirm that the 6-TG + G418 resistant clones were targeted and to observe gene conversion. The relative targeting frequencies are expressed as the ratio of 6-TG + G418 resistant clones to G418 resistant clones. Absolute targeting frequencies

Table 1. The targeting frequencies for *hprt* vectors

Exp.	Vector	linearization	electro-porations	Total G418ʳ	Total TGʳ	(TGʳ+G418ʳ)/G418ʳ	Integration analysis of TG resistant clones Vector insertion	modified
A	IV6.8	XhoI	3	3240	110	1/29	ND	ND
	IV6.8	NheI	4	4144	489	1/8	ND	ND
	IV6.8AN	NotI	10	7860	110	1/71	32	3
	IV6.8AN	XhoI	10	7480	110	1/68	30	0
	IV6.8AN	NheI	9	6066	421	1/14	29	0
B	IV6.8	XhoI	10	11270	282	1/40	52	4
	IV6.8BN	NotI	7	6482	76	1/85	69	3
C	IV6.8	XhoI	15	16980	511	1/33	9	0
	IV6.8	NheI	15	23265	1599	1/14	16	0
	IV6.8RB	XhoI	14	17332	233	1/74	152	23
	IV6.8RB	NheI	15	13440	1545	1/9	43	1
D	IV6.8	XhoI	10	10500	108	1/97	ND	ND
	IV6.8Δ86	EcoRI	13	20020	126	1/159	22	4
E	IV6.8XN3	XhoI	10	5500	96	1/57	16	2
F	IV6.8	XhoI	10	4590	167	1/27	ND	ND
	IV6.8G0.5	HindIII+XhoI	10	9350	132	1/71	12	0
	IV6.8G2.5	NheI+XhoI	10	5870	61	1/96	11	1
	IV6.8G4.2	HindIII	15	5940	11	1/540	8	0

A series of vectors were electroporated into AB1 ES cells and the number of G418 resistant and 6-TG + G418 resistant clones was scored. Since there is some variation in the targeting frequencies between experiments, each experimental group indicates an experiment in which all electroporations were performed at the same time and with a common batch of ES cells under the same conditions. Under these conditions the relative relationship between the targeting frequencies obtained with each vector are constant (9). Linearization refers to the restriction enzyme used to linearize the construct before the electroporation. All of the constructs are linearized in the homology. Each electroporation consists of 25 μg DNA transfected into 10⁷ ES cells. A representative sample of the 6-TG resistant clones were analyzed to ascertain if the recombinant allele had the expected structure. Modified insertion events, where the exact structure of the targeted allele has not been elucidated were previously described (5,8).

(6-TG + G418 resistant clones per electroporation) are not presented but may be derived from Table 1.

Will the relative position of the double-strand break and the mutation affect the targeting frequency of an insertion vector? An alteration in the targeting frequency may indicate that the position of the DSB and the mutation might impose a rate limiting constraint of gene conversion. To determine the affect the location of the DSB has on the targeting frequency, the parental vector IV6.8 (Fig. 1) was linearized centrally with *Nhe*I and towards the 3′ end with *Xho*I and electroporated into ES cells. Linearization at the central site resulted in a 3.6-fold higher targeting frequency than in the 3′ site (Table 1A). To determine the affect of the introduced mutation on the targeting frequency, a vector with a 14 bp oligonucleotide mutation was analyzed. IV6.8AN has a 14 bp insertion with a unique *Not*I site positioned 3.2 and 0.472 kb from the *Nhe*I and *Xho*I linearization sites, respectively (Fig. 1). The
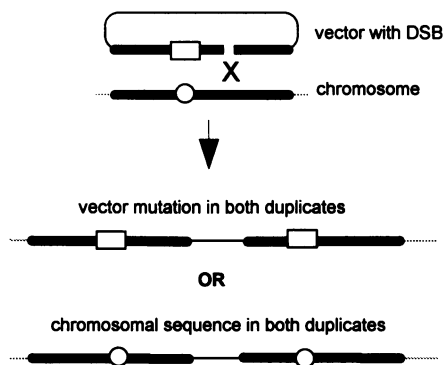
Figure 2. Gene conversion during gene targeting with an insertion vector. The thick line represents the homologous sequences in the vector and chromosome. The thin line represents the bacterial plasmid in the vector. The dotted line represents chromosomal sequences not included in the vector. The rectangle represents the vector mutation and the circle represents the chromosomal sequence that is different from the vector. The X represents the location of the crossover. After gene targeting with an insertion vector there is a duplication of homology separated by the bacterial plasmid. Gene conversion has occurred when the vector or the chromosomal sequence is located in both duplicates. Clones that have the vector sequence in one duplicate and the chromosomal sequence in the other duplicate are not considered here.

targeting frequency of IV6.8AN was 1.7- and 2.4-fold lower when compared to IV6.8 linearized at the same site, *Nhe*I and *Xho*I respectively (Table 1A). Thus, the presence of the 14 bp does not greatly alter the targeting frequency. The targeting frequency of IV6.8AN linearized with *Nhe*I was ~5-fold higher than when linearized with *Xho*I or *Not*I which is about the same difference observed with the control IV6.8; therefore, the distance between the DSB and the vector mutation did not affect the targeting frequency (Table 1A).

Gene conversion was observed during vector insertion events

Does gene conversion accompany homologous recombination with an insertion vector? A series of insertion vectors were designed to study gene conversion during homologous recombination in ES cells (Fig. 1) which differ only in the position and type of mutation present in the *hprt* sequences. Gene conversion of different mutations was compared in vectors linearized at the same site and gene conversion events of the same mutation was analysed from vectors linearized at different sites. This will help determine if gene conversion has a sequence specific bias or if the gene conversion frequency is dependent upon the distance between the mutation and the double-strand break (linearization site).

Insertion vectors recombine into the chromosomal target to form a duplication of homology separated by the bacterial plasmid. Gene conversion can be scored if there is a disruption in the homology between the vector and chromosomal sequences. Only those clones in which either the vector mutation or chromosomal sequence is present in both duplicates can be unequivocally identified as derived from gene conversion events (Fig. 2). It is not possible to determine if the clones in which the vector and chromosomal sequences are each represented in one duplicate arose by gene conversion; therefore, these clones are not presented here.

The IV6.8 derivative constructs (Fig. 1) with mutations (between 12 and 86 bp) that include both the addition and the deletion of nucleotides compared to the chromosomal template were linearized at different distances from the vector mutation and introduced into ES cells. The targeting frequency of the vectors IV6.8BN, IV6.8AN, IV6.8RB and IV6.8Δ86 was reduced by about 1.5–2-fold compared with IV6.8 linearized at a comparable location in the homology (Table 1A–D); however, a reduction in the targeting efficiency was not always observed (Table 1C).

Southern analysis of 6-TG + G418 resistant clones demonstrated that loss of the vector mutation and gain of the chromosomal sequence occurred at a high frequency when the mutation was close to the DSB for all mutations observed (Table 2). The loss of the vector mutation occurred at a frequency inversely proportional to the distance between the mutation and the DSB. Loss of the vector mutation occurred in: 88 and 69% of the clones with terminal heterologies (IV6.8BN and IV6.8AN cut in the mutation with *Not*I, respectively); 62 and 54% of the clones when the mutation was 0.123 kb away from the DSB (IV6.8RB cut with *Xho*I and IV6.8Δ86 cut with *Eco*RI, respectively); 40% of the clones when the mutation was 0.472 kb away from the DSB (IV6.8AN cut with *Xho*I). A detailed analysis of the clones generated with IV6.8Δ86 is presented in Figure 3. Detailed analysis of the targeted clones generated with IV6.8BN and IV6.8RB is presented elsewhere (9,18).

The frequency of loss of the vector mutation and gain of the chromosomal sequence dropped sharply when the vector mutation was further away from the DSB. Loss of the vector mutation occurred in just 7% of the clones when the mutation was 2.6 and 3.2 kb away from the DSB (IV6.8RB and IV6.8AN cut with *Nhe*I) and in 16% of the clones when the mutation was 4.3 kb away from the DSB (IV6.8XN3 cut with *Xho*I). In one experiment loss of the vector mutation was not observed (IV6.8G0.5 with 4.4 kb between the DSB and the vector mutation).

Gain of the vector mutation and loss of the chromosomal sequence occurred at a low frequency (Table 2). Gain of the vector mutation was not observed in most of the experiments; however, it was observed in 1 out of 72 clones targeted by a vector with terminal heterologies (IV6.8BN cut in the mutation with *Not*I), 12 out of 175 clones targeted by a vector with the mutation 0.123 kb away from the DSB (IV6.8RB cut with *Xho*I) and in 2 out of 18 clones targeted by a vector with the mutation 3.2 kb away from the DSB (IV6.8XN3 cut with *Xho*I). Southern analysis that demonstrates gene conversion during homologous recombination with IV6.8XN3 is presented in Figure 4.

Gaps of up to 4.2 kb are accurately repaired

To determine if there were any length constraints in the repair mechanism, a set of vectors with large double-strand gaps in the *hprt* DNA were constructed and linearized at the location of the deletion. Each gapped construct was cloned so that the missing target DNA was not co-electroporated into the ES cells; thus, eliminating the possibility of the two fragments being ligated back together before targeting (22). Filler DNA was inserted into vectors where the sites delimiting the gaps were non-compatible to ensure the ends were perfectly homologous. The vectors IV6.8G0.5, IV6.8G2.5 and IV6.8G4.2 have gaps of 0.536, 2.5 and 4.2 kb, respectively (Fig. 1).

Table 2. Gene conversion during vector insertion

Construct	Linearization site	Distance (kb)	No of clones analyzed	-/- (%)	+/+ (%)	-/+ (%)	+/- (%)	modified (%)
IV6.8BN	NotI	0	72	65 (88%)	1 (1.4%)	0	3 (4%)	3 (4%)
IV6.8AN	NotI	0	35	22 (69%)	0	0	10 (31%)	0
IV6.8RB	XhoI	0.132	175	104 (62%)	12 (7%)	17 (10%)	19 (11%)	23 (13%)
IV6.8Δ86	EcoRI	0.132	24	14 (54%)	0	0	6 (25%)	4 (16%)
IV6.8AN	XhoI	0.472	30	12 (40%)	0	0	18 (60%)	0
IV6.8RB	NheI	2.6	44	1 (7%)	0	0	42 (95%)	1 (2%)
IV6.8AN	NheI	3.2	29	2(7%)	0	1 (3%)	26 (90%)	0
IV6.8XN3	XhoI	4.3	18	3 (16%)	2 (11%)	0	11 (61%)	2 (1%)
IV6.8G0.5	HindIII+XhoI	4.4	12	0	0	0	12 (100%)	0

Targeted 6-TG resistant clones confirmed by Southern analysis were analyzed to determine the position of the vector mutation and the chromosomal sequences in both duplicates of homology. All the vectors are linearized in the homologous sequences at the site indicated. Distance refers to the separation between the vector mutation and the DSB. +, indicates the presence of a vector mutation (absence of the chromosomal sequence). –, indicates absence of the vector mutation (presence of the chromosomal sequence).

The targeting frequency (Table 1F) and integration pattern (Fig. 5) were determined for each vector. The targeting efficiency for IV6.8G0.5, IV6.8G2.5 and IV6.8G4.2 was reduced by 2.6-, 3.5- and 20-fold compared to IV6.8 linearized in the homology (XhoI), respectively (Table 1). The reduction in the targeting frequencies are due to either rate limiting constraints of repair or the reduced length of homology (4,21,23). The gaps were shown to be repaired for 12/12, 11/12 and 8/8 6-TG^r clones targeted with IV6.8G0.5, IV6.8G2.5 and IV6.8G4.2, respectively (Table 1F and Fig. 5). The single clone that gave an alternative pattern did not resemble recombination without repair; we classify this as a modified insertion event (9). Repair was demonstrated by restoration of the fragment length and restriction enzyme sites so that the final recombinant allele resembles one generated with the parental IV6.8 vector.

In control experiments, the same vectors were linearized at the edge of the homologous sequences. Although 6-TG resistant clones were generated at lower frequencies than when linearized in the homology, none of the clones repaired the deletion and the filler DNA integrated into the recombinant structure (data not shown).

DISCUSSION

In this report we examined gene conversion during homologous recombination of insertion vectors into a chromosomal target in ES cells. Gene conversion was scored by the presence of either the vector mutation or chromosomal sequence in both duplicates. We made the following observations. (i) The presence of the vector mutation had only a marginal impact on the targeting frequency and the distance between the mutation and the DSB did not affect the targeting frequency. This indicates that gene conversion is not rate limiting for the vector mutation since gene conversion was frequently observed in some cases. (ii) A sequence specific bias in gene conversion was not observed. The same pattern of gene conversion was observed for five different vector mutations that included the addition of nucleotides and the deletion of nucleotides. (iii) Loss of the vector mutation and gain of the chromosomal sequence was bi-directional and occurred with an increased frequency as the distance between the vector mutation and DSB decreased. (iv) Gain of the vector mutation and loss of the chromosomal sequence occurred at a low frequency. (v) Repair of up to 4.2 kb gaps is accurate.

What are the gene conversion mechanisms responsible for these observations? Two forms of gene conversion have been proposed: gap formation and repair and mismatch repair of heteroduplexes. Gap formation and repair describes removal of nucleotides from the strand with a DSB and subsequent repair on a homologous DNA template (3,7). Mismatch repair of heteroduplexes describes the repair of a mismatch after the formation of a heteroduplex during strand invasion or after branch migration of the Holliday junctions (3).

Why was loss of the vector mutation more common when it was close to the DSB? Both forms of gene conversion could explain this observation. Gap formation and repair predicts removal of vector sequences adjacent to the DSB and repair that utilizes an intact chromosomal template. A vector mutation within the range of the gap is replaced by chromosomal sequence. If this mechanism was used, then gaps of ≥0.472 kb are commonly formed but rarely extend to 2.6 kb. Mismatch repair of heteroduplexes may also explain the data. Both vector ends invade the homologous template without significant degradation of the ends. We have previously presented data that shows the ends of the vector do not always undergo gap formation (9). After strand invasion a heteroduplex would be formed with a nick present at the end of the vector strand. The mismatch could be resolved to either the vector mutation or chromosomal sequence; however, the nick may stimulate removal of the vector mutation. A nick was shown to stimulate bidirectional repair preferentially directed toward the nicked strand (25). If this mechanism was used, then it is likely that heteroduplexes of 0.472 kb are frequently formed during strand invasion but rarely extend to 2.6 kb. By either mechanism, loss of the vector mutation in favor of the chromosomal sequence increases as the distance between the vector mutation and the DSB decreases.
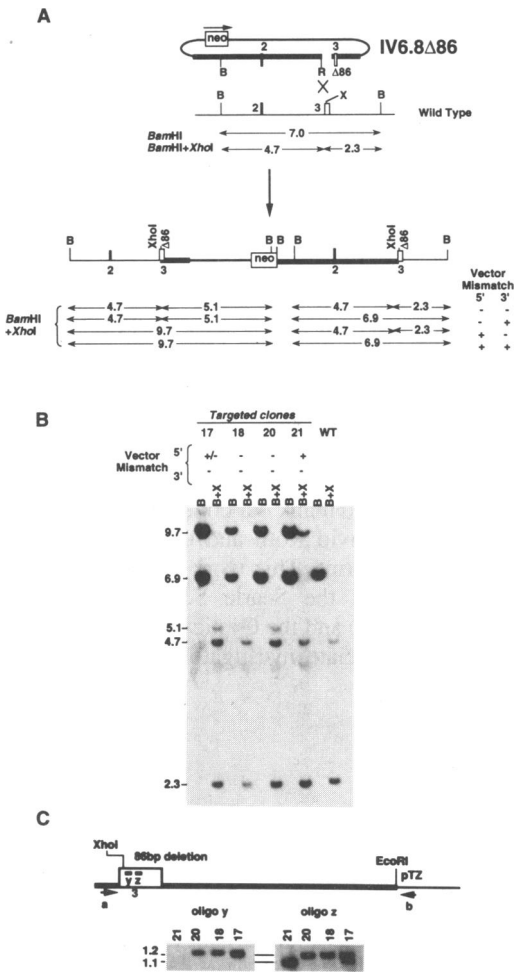
A

IV6.8Δ86

Figure 3. Southern analysis for TG^r + G418^r clones targeted with IV6.8Δ86. Analysis of 24 clones showed that a single unit of the vector integrated into the target locus for 22 clones and multiple units for two clones. (**A**) Correction of the deletion was analyzed by the restoration of an XhoI site in both duplicates with a BamHI and a BamHI + XhoI digest hybridized to exons 2 and 3; 54% corrected the deletion (representative clones 18 and 20). Two additional clones contained an integration of at least two units in the target locus in which at least one unit was corrected and the other(s) was not (representative clone 17). The panel to the right of the predicted fragments indicates the presence (+) or absence (−) of the vector mutation in exon 3 of both duplicates. The thick line represents hprt DNA of vector origin, the thin line represents hprt DNA of chromosomal origin and the line of intermediate thickness represents plasmid. MC1neopA is shown as a box labeled neo. Exons 2 and 3 are numbered rectangles. The Δ86 represents the 86 bp deletion. B, BamHI; R, EcoRI; X, XhoI. (**B**) Southern analysis. B, BamHI and B+X, BamHI + XhoI digests were done on TG^r + G418^r clones (clones 17, 18, 20 and 21 are represented). The presence (+) or absence (−) of the vector mutation for both 5′ and 3′ duplicates is indicated in the upper panel. WT, wild type (AB1 cell DNA). The length of DNA is in kb. (**C**) Primers a and b, small arrow heads, were used to amplify exon 3 (large box) of the 5′ duplicate by the polymerase chain reaction (PCR) and hybridized to oligo y, inside the deletion and oligo z, outside the deletion. The sizes of the PCR products are in kb. The 1.2 kb fragment represents correction of the deletion and hybridized to both oligonucleotides (representative clones 18 and 20). The 1.1 kb fragment represents the deletion and hybridizes to only oligo Z (representative clone 21). Clones with multiple units in the target locus (some corrected and some not) or clones which segregated without repair of the heteroduplex, amplified a 1.2 kb fragment which hybridized to both oligonucleotides and a 1.1 kb fragment which hybridized just to oligonucleotide z (representative clone 17).
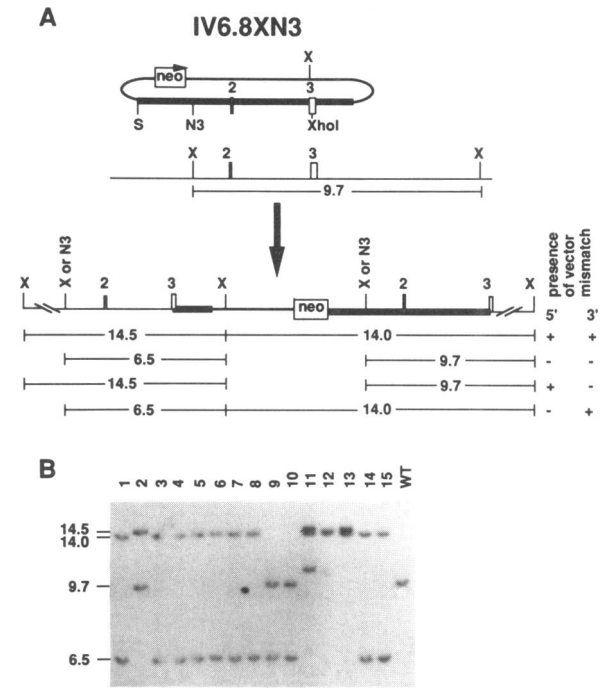
Figure 4. Southern analysis for TG^r + G418^r clones targeted with IV6.8XN3. An XmnI digest was used to confirm the presence of the chromosomal sequence. An NheI digest was performed and confirmed the presence of the vector mutation when XmnI did not cut (not shown). Either the vector mutation or the chromosomal sequence was always present. Exons 2 + 3 are shown as numbered rectangles and were used as the probe. The thick line represents hprt DNA of vector origin, the thin line represents hprt DNA of chromosomal origin and the line of intermediate thickness indicates plasmid sequences. The MC1neopA cassette is shown as a box labeled neo. The length of DNA is shown in kb. WT, wild type. (**A**) Predicted fragment size for a XmnI digest. X, XmnI. (**B**) Representative clones for Southern blot analysis WT, wild type (AB1 DNA). Clones 11 and 13 are +/+, where + indicates presence of the vector mutation. There is an additional band in clone 11 that does not match any expected band which is probably the result of a random integration of a second vector unit elsewhere in the genome. Clones 9 and 10 are −/−, where − indicates absence of the vector mutation. Clone 2 is +/−. Clones 1, 3–8, 14 and 15 are −/+. Clone 12 is a modified insertion event in which the exact integration structure was not determined.

Why was there gain of the vector mutation and loss of the chromosomal sequence in a few clones when the vector mutation was close to the DSB? Gap formation and repair predicts sequence removal from only the strand with the DSB (the vector) and therefore, can not be responsible for removal of the chromosomal sequence. Mismatch repair of heteroduplexes predicts gene conversion directed to either strand and is a likely mechanism to describe these rare clones. This may reflect a preference but not an absolute requirement to repair the vector strand due to the presence of a nick after strand invasion or this may reflect ligation of the nicked strand prior to mismatch repair. Alternatively, there may be competing gene conversion mechanisms. Gap formation may frequently remove the vector mutation but not in every instance. For these rare events a heteroduplex may form during strand invasion and may then be resolved to remove either the vector mutation or chromosomal sequence.

A small percentage of clones contain only the vector mutation or chromosomal sequence in both duplicates when the vector mutation was ≥2.6 kb away from the DSB. There was not a strong
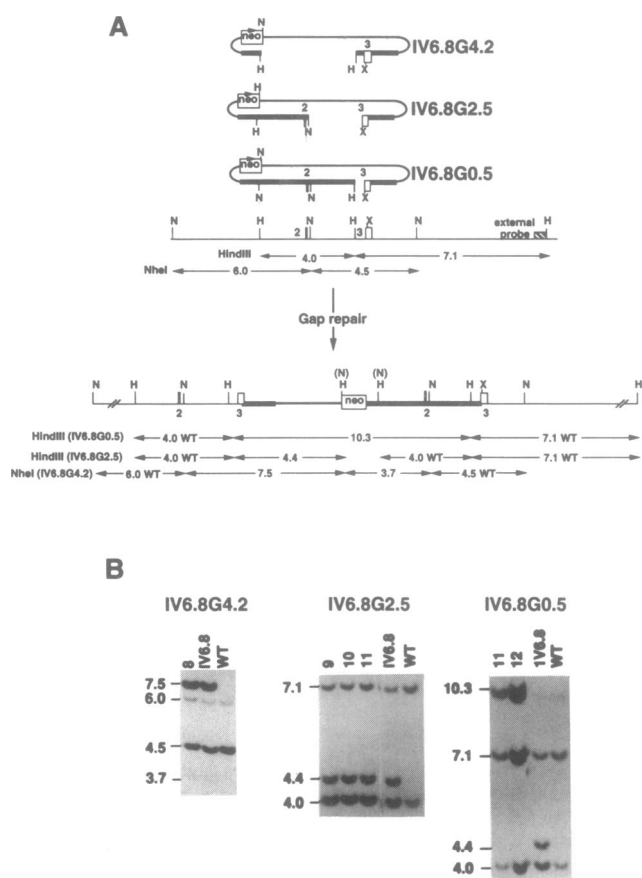
Figure 5. Southern analysis for TG^r clones targeted with IV6.8G0.5, IV6.8G2.5 and IV6.8G4.2. DNA was cut with *Hind*III for clones generated with IV6.8G0.5 and IV6.8G2.5 and *Nhe*I for clones generated with IV6.8G4.2. All digests were hybridized to exons 2 and 3. The length of DNA is in kb. WT, wild type. (**A**) Predicted fragment size for a *Hind*III and an *Nhe*I digest. The 5' *Hind*III site in the *hprt* homology and the *Hind*III site in the polylinker was changed to an *Nhe*I site in IV6.8G0.5. 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 shown as a box labeled *neo*. Exons 2 and 3 are numbered rectangles. H, *Hind*III; N, *Nhe*I; X, *Xho*I. (**B**) Representative clones for Southern blot analysis: clone 8 for IV6.8G4.2; clones 9, 10 and 11 for IV6.8G2.5; clones 11 and 12 for IV6.8G0.5. IV6.8, DNA from a clone targeted with IV6.8 which does not contain a gap and represents a full length restriction pattern. WT, wild type (AB1 DNA).

bias in either direction in this small data set. It is unlikely that gene conversion occurred by gap formation and repair because there would be a strong bias to remove the vector mutation. However, it remains possible that gaps of >2.6 kb are generated at a low frequency. It is likely that mismatch repair of heteroduplexes is responsible for most of the gene conversion when the mutation is separated from the DSB by >2.6 kb. Heteroduplexes may form after branch migration of the Holliday junctions. Preferential removal of the vector mutation may be lost due to ligation of the nicked strand. Thus, repair of the mismatch would no longer have a bias.

The gapped constructs, where the gaps existed in the vector prior to the transfection, were accurately repaired. The gaps were never transferred to the chromosomal sequences. The reduced targeting frequencies for these vectors indicates the gaps inhibit the targeting frequency by limiting either strand invasion or

repair. However, the shorter length of homology will contribute to a reduction in the targeting frequency (21); therefore, the affect the gaps have on the gene targeting efficiency is difficult to determine.

Gene targeting by vector insertion is considered to be a conservative recombination event since both the vector and chromosomal DNA components are retained in the recombinant allele. However, there is a non-conservative component that can be scored when there are sequence differences between the two partners. In this study we showed that gene conversion was frequent when the vector mutation was close to the DSB. However, gene conversion occurs at sites as distant as 4.3 kb from the DSB.

ACKNOWLEDGEMENTS

REFERENCES

1 Rossignol, J. L. (1969) *Genetics* **63**, 795–805.
2 Rossignol, J. L. and Paquette, N. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2871–2875.
3 Szostak, J. W., Orr-Weaver, T. L., Rothstien, R. J. and Stahl, F. W. (1983) *Cell* **33**, 25–35.
4 Thomas, K. R. and Capecchi, M. R. (1987) *Cell* **51**, 503–512.
5 Deng, C., Thomas, K.R. and Capecchi, M.R. (1993) *Mol. Cell. Biol.* **13**, 2134–2140.
6 Hinnen, A., Hicks, J.B. and Fink, G. R. (1978). *Proc. Natl. Acad. Sci. USA* **75**, 1929–1933.
7 Orr-Weaver, T. L., Szostak, J. W. and Rothstein, R. J. (1981) *Proc. Natl. Acad. Sci.* **78**, 6354–6358.
8 Resnick, M. A. (1976) *J. Theor. Biol.* **59**, 97–106.
9 Hasty, P., Rivera-Perez, J. and Bradley, A. (1992) *Mol. Cell. Biol.* **12**, 2464–2474.
10 Hasty, P., Rivera-Pérez, J., Chang, C. and Bradley, A. (1991) *Mol.Cell. Biol.* **11**, 4509–4517.
11 Pennington, S.L., Wilson, J.H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9498–9502.
12 Valancius, V. and Smithies, O. (1991) *Mol. Cell. Biol.* **11**, 4389–4397.
13 Brenner, D. A., Smigocki, A. C. and Camerini-Otero, R. D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1762–1766.
14 Jasin, M., and Berg, P. (1988) *Genes Dev.* **2**, 1353–1363.
15 Jessberger, R., and Berg, P. (1991) *Mol. Cell. Biol.* **11**, 445–457.
16 Holliday, R. (1964) *Genet. Res.* **5**, 282–304.
17 Melton, D. W., Konecki, D. S., Brennand, J. and Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2147–2151.
18 Hasty, P., Ramírez-Solis, R., Krumlauf, R. and Bradley, A. (1991) *Nature* **350**, 243–246.
19 McMahon, A. P. and Bradley, A. (1990) *Cell* **62**, 1073–1085.
20 Robertson, E. J. (1987) In E. J. Robertson (ed.) *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, IRL Press, Oxford, pp. 71–112.
21 Hasty, P., Rivera-Pérez, J. and Bradley, A. (1991) *Mol. Cell. Biol.* **11**, 5586–5591.
22 Wilson, J. H., Berget, P. B. and Pipas, J. M. (1982) *Mol. Cell. Biol.* **2**, 1258–1269.
23 Shulman, M. J., Nissen, L. and Collins, C. (1990) *Mol. Cell. Biol.* **10**, 4466–4472.
24 Subramani, S. (1986) *Mol. Cell. Biol.* **6**, 1320–1325.
25 Umar, A. Boyer, J.C., and Kunkel, T.A. (1994) *Science* **266**, 814–816.