Gene conversion during vector insertion in embryonic stem cells

Paul Hasty1,*, Jaime Rivera-Pérez1 and Allan Bradley1,2

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

Received December 22, 1994; Revised and Accepted April 10, 1995

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 SauI–EcoRI 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 NotI site was generated by ligating an oligonucleotide (5'-TTGCCGGCCGA-3') into an Accl 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. NotI linearizes in the mutation; XhoI linearizes 472 bp 5' to the mutation and NheI linearizes 3.2 kb 3' to the mutation.

IV6.8A86 has an 86 bp deletion that destroys the XhoI 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 EcoRI sites were destroyed. The site in the MC1neopA promoter was

* To whom correspondence should be addressed at present address: The University of Texas, M. D. Anderson Cancer Center, Department of Biochemistry and Molecular Biology, 1515 Holcombe Boulevard, Houston, TX 77030, USA
sites; *XhoI*, 132 bp 3' to the mutation on the short arm of the vector and *NheI*, 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'-TTGCGGCCGTACGGCCGCA-3') into the *XmnI* site in the *hprt* sequences (9) which destroys the *XmnI* site and generates *NotI-* *NheI-* *NotI* 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 *XhoI*, 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 *HindIII* sites by end filling and blunt end ligation. These sites were the most 5' *HindIII* site in the *hprt* sequences and a *HindIII* site in the plasmid polylinker. The filler DNA was a 2.3 kb fragment. The linearization sites were *HindIII* and *XhoI* which liberates the filler DNA from the vector. The gap in this vector lies 4.4 kb, 3' to the filled *HindIII* 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 *NheI* and *XhoI* sites separated by a 600 bp fragment of filler DNA. The vector was cut with *NheI* and *XhoI* restriction enzymes for the electroporation.

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

**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 *Sacl*--*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 boxed check indicates the *neo* selection cassette, MC1neoApA (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 IV6.8G2. 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, *HindIII*; N, *NheI*; R, *EcoRI*; S, *Sacl*; X, *XhoI*. The mutated restriction enzyme sites are in parentheses.

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 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

---

**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 *Sacl*--*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 boxed check indicates the *neo* selection cassette, MC1neoApA (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 IV6.8G2. 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, *HindIII*; N, *NheI*; R, *EcoRI*; S, *Sacl*; X, *XhoI*. The mutated restriction enzyme sites are in parentheses.
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 HincII and HpaII (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 BamHI 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, AccI digest; IV6.8BN and IV6.8Δ86, BamHI + XhoI digest; IV6.8XN3, XmnI and Nhel digests (re-probed with neo); IV6.8G0.5, HindIII and EcoRI digests; IV6.8G2.5, HindIII digest; IV6.8G4.2, Nhel and a HindIII 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’-CTCTGGATTATATTCTTAGGG-3’ (specific for the intron 2/exon 3 junction). The downstream antisense primer was 5’-TGGGATCAAGGCGATTAAGTT-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 AmpliTag polymerase (Perkin-Elmer Cetus); 50 ng of each primer; 250 μM of each dNTP; 1 x Cetus buffer; 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’-CCACATGGTGATGGCCTCCC-3’) by standard conditions. After exposure the filter was stripped and hybridized to a second oligonucleotide in exon 3 outside the deletion, z (5’-TTACATGATGGCCTCCC-3’) by standard conditions. The 1.2 and 1.1 kb EcoRI digests; IV6.8G2.5, BamHI, and HindIII digests were performed from clones generated with the following vectors: IV6.8AN, AccI digest; IV6.8BN and IV6.8Δ86, BamHI + XhoI digest; IV6.8XN3, XmnI and Nhel digests (re-probed with neo); IV6.8G0.5, HindIII and EcoRI digests; IV6.8G2.5, HindIII digest; IV6.8G4.2, Nhel and a HindIII digest. The probes were labeled with random primed labeling using manufacturer’s conditions (Boehringer Mannheim).

**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 (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 Nhel and towards the 3’ end with XhoI 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 NotI site positioned 3.2 and 0.472 kb from the Nhel and XhoI linearization sites, respectively (Fig. 1). The

<table>
<thead>
<tr>
<th>Exp. Vector</th>
<th>Linearization</th>
<th>Electro-</th>
<th>Total</th>
<th>Total</th>
<th>Integration sites of TG resistant clones</th>
<th>Vector insertion sites modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>A IV6.8</td>
<td>Nhel</td>
<td>3</td>
<td>3240</td>
<td>110</td>
<td>1/29 ND ND</td>
<td>ND</td>
</tr>
<tr>
<td>IV6.8</td>
<td>Nhel</td>
<td>4</td>
<td>4144</td>
<td>429</td>
<td>1/18 ND ND</td>
<td>ND</td>
</tr>
<tr>
<td>IV6.8AN</td>
<td>Nhel</td>
<td>10</td>
<td>7560</td>
<td>110</td>
<td>1/71 32 3</td>
<td>ND</td>
</tr>
<tr>
<td>IV6.8AN</td>
<td>Nhel</td>
<td>10</td>
<td>7480</td>
<td>110</td>
<td>1/68 30 0</td>
<td>0</td>
</tr>
<tr>
<td>IV6.8AN</td>
<td>Nhel</td>
<td>9</td>
<td>6066</td>
<td>421</td>
<td>1/14 29 0</td>
<td>0</td>
</tr>
<tr>
<td>B IV6.8</td>
<td>Nhel</td>
<td>10</td>
<td>11270</td>
<td>282</td>
<td>1/40 52 4</td>
<td>4</td>
</tr>
<tr>
<td>NIV6.8BN</td>
<td>Nhel</td>
<td>7</td>
<td>6402</td>
<td>76</td>
<td>1/85 69 3</td>
<td>3</td>
</tr>
<tr>
<td>C IV6.8</td>
<td>Nhel</td>
<td>15</td>
<td>16950</td>
<td>511</td>
<td>1/33 9 0</td>
<td>0</td>
</tr>
<tr>
<td>IV6.8BN</td>
<td>EcoRI</td>
<td>13</td>
<td>20020</td>
<td>126</td>
<td>1/159 22 4</td>
<td>4</td>
</tr>
<tr>
<td>D IV6.8</td>
<td>Nhel</td>
<td>10</td>
<td>10500</td>
<td>108</td>
<td>1/97 2 ND ND</td>
<td>ND</td>
</tr>
<tr>
<td>IV6.8BN</td>
<td>EcoRI</td>
<td>13</td>
<td>20020</td>
<td>126</td>
<td>1/159 22 4</td>
<td>4</td>
</tr>
<tr>
<td>E IV6.8BN</td>
<td>Nhel</td>
<td>15</td>
<td>13440</td>
<td>1545</td>
<td>1/79 43 1</td>
<td>1</td>
</tr>
<tr>
<td>F IV6.8</td>
<td>Nhel</td>
<td>10</td>
<td>4590</td>
<td>167</td>
<td>1/24 ND ND</td>
<td>ND</td>
</tr>
<tr>
<td>IV6.8BN.5</td>
<td>Nhel+Nhel</td>
<td>10</td>
<td>9350</td>
<td>132</td>
<td>1/71 12 0</td>
<td>0</td>
</tr>
<tr>
<td>IV6.8BN.5</td>
<td>Nhel+Nhel</td>
<td>10</td>
<td>5870</td>
<td>61</td>
<td>1/96 11 1</td>
<td>1</td>
</tr>
<tr>
<td>IV6.8BN.4</td>
<td>Nhel</td>
<td>15</td>
<td>5940</td>
<td>11150</td>
<td>1/540 0</td>
<td>0</td>
</tr>
</tbody>
</table>

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 103 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).
targeting frequency of IV6.8AN was 1.7- and 2.4-fold lower when compared to IV6.8 linearized at the same site, NheI and XhoI 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 NheI was ~5-fold higher than when linearized with XhoI or NotI 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 IV6.8A86 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 NotI, respectively); 62 and 54% of the clones when the mutation was 0.123 kb away from the DSB (IV6.8RB cut with XhoI and IV6.8A86 cut with EcoRI, respectively); 40% of the clones when the mutation was 0.472 kb away from the DSB (IV6.8AN cut with XhoI). A detailed analysis of the clones generated with IV6.8A86 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 NheI) and in 16% of the clones when the mutation was 4.3 kb away from the DSB (IV6.8XN3 cut with XhoI). 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 NotI), 12 out of 175 clones targeted by a vector with the mutation 0.123 kb away from the DSB (IV6.8RB cut with XhoI) 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 XhoI). 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).
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+ 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 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 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.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Linearization site</th>
<th>Distance (kb)</th>
<th>No of clones analyzed</th>
<th>+/- (%)</th>
<th>+/- (%)</th>
<th>+/- (%)</th>
<th>+/- (%)</th>
<th>+/- (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV6.8BN</td>
<td>Xhol</td>
<td>0</td>
<td>72</td>
<td>65 (88%)</td>
<td>1 (14%)</td>
<td>0</td>
<td>3 (4%)</td>
<td>3 (5%)</td>
</tr>
<tr>
<td>IV6.8AN</td>
<td>Xhol</td>
<td>0</td>
<td>35</td>
<td>22 (69%)</td>
<td>0</td>
<td>0</td>
<td>10 (31%)</td>
<td>0</td>
</tr>
<tr>
<td>IV6.8BN</td>
<td>Xhol</td>
<td>0.132</td>
<td>175</td>
<td>104 (62%)</td>
<td>12 (7%)</td>
<td>17 (10%)</td>
<td>19 (11%)</td>
<td>23 (13%)</td>
</tr>
<tr>
<td>IV6.8BN</td>
<td>EcoRI</td>
<td>0.132</td>
<td>24</td>
<td>14 (54%)</td>
<td>0</td>
<td>0</td>
<td>6 (25%)</td>
<td>4 (16%)</td>
</tr>
<tr>
<td>IV6.8BN</td>
<td>Xhol</td>
<td>0.472</td>
<td>30</td>
<td>12 (40%)</td>
<td>0</td>
<td>0</td>
<td>18 (60%)</td>
<td>0</td>
</tr>
<tr>
<td>IV6.8BN</td>
<td>Xhol</td>
<td>2.6</td>
<td>44</td>
<td>1 (7%)</td>
<td>0</td>
<td>0</td>
<td>42 (95%)</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>IV6.8BN</td>
<td>Xhol</td>
<td>3.2</td>
<td>29</td>
<td>2 (7%)</td>
<td>0</td>
<td>1 (3%)</td>
<td>26 (90%)</td>
<td>0</td>
</tr>
<tr>
<td>IV6.8BN</td>
<td>Xhol</td>
<td>4.3</td>
<td>18</td>
<td>3 (15%)</td>
<td>2 (11%)</td>
<td>0</td>
<td>11 (61%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>IV6.8BN</td>
<td>HindIII+Xhol</td>
<td>4.4</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12 (100%)</td>
<td>0</td>
</tr>
</tbody>
</table>

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).
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 may form during strand invasion and may then be resolved to remove either the vector mutation or chromosomal sequence. 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 close to the DSB. There was not a strong...
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, gene conversion occurs at sites as distant as 4.3 kb from the DSB.

The targeting frequency by limiting either strand invasion or frequency. It is likely that mismatch repair of heteroduplexes is a bias.

Removal of the vector mutation may be lost due to ligation of the responsible for most of the gene conversion when the mutation is 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**

We would like to acknowledge Mike Crist and J. D. Wallace for technical assistance and thank our colleagues Richard Behringer, Mark Brenneman, David Roth and Geoff Sargent for critical review of the manuscript. This work was supported by grants from the NIH (AB), the Searle Scholars Program/Chicago Community trust (AB) and the Cystic Fibrosis Foundation (AB, PH, JR). AB is an associate investigator with the Howard Hughes Medical Institute.

**REFERENCES**


**Figure 5. Southern analysis for TG clones targeted with IV6.8G0.5, IV6 8G2.5 and IV6 8G4.2.** DNA was cut with HindIII for clones generated with IV6 8G0.5 and IV6 8G2.5 and NheI 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 HindIII and an NheI digest. The 5' HindIII site in the hprt homologue and the HindIII site in the poly linker was changed to an NheI site in IV6 8G0.5. The thick line is hprt DNA of vector origin, the thin line is hprt DNA of genome origin and the line of intermediate thickness is plasmid MC1neoA is shown as a box labeled neo. Exons 2 and 3 are numbered rectangles H, HindIII, N, NheI, X, XhoI (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)