The Molecular Basis of Multiple Vector Insertion by Gene Targeting in Mammalian Cells

Philip Ng* and Mark D. Baker*†

*Department of Molecular Biology and Genetics and †Department of Pathobiology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Received August 14, 1998
Accepted for publication November 10, 1998

ABSTRACT

Gene targeting using sequence insertion vectors generally results in integration of one copy of the targeting vector generating a tandem duplication of the cognate chromosomal region of homology. However, occasionally the target locus is found to contain > 1 copy of the integrated vector. The mechanism by which the latter recombinants arise is not known. In the present study, we investigated the molecular basis by which multiple vectors become integrated at the chromosomal immunoglobulin μ locus in a murine hybridoma. To accomplish this, specially designed insertion vectors were constructed that included six diagnostic restriction enzyme markers in the Cμ region of homology to the target chromosomal μ locus. This enabled contributions by the vector-borne and chromosomal Cμ sequences at the recombinant locus to be ascertained. Targeted recombinants were isolated and analyzed to determine the number of vector copies integrated at the chromosomal immunoglobulin μ locus. Targeted recombinants identified as bearing > 1 copy of the integrated vector resulted from a Cμ triplication formed by two vector copies in tandem. Examination of the fate of the Cμ region markers suggested that this class of recombinant was generated predominantly, if not exclusively, by two targeted vector integration events, each involving insertion of a single copy of the vector. Both vector insertion events into the chromosomal μ locus were consistent with the double-strand-break repair mechanism of homologous recombination. We interpret our results, taken together, to mean that a proportion of recipient cells is in a predetermined state that is amenable to targeted but not random vector integration.

Gene targeting, the homologous recombination between transferred and chromosomal DNA, allows for the precise introduction of predetermined modifications into chromosomal genes. This technology has important, wide-ranging applications in the study of gene structure and function, the creation of animal models of human genetic disease, and perhaps, ultimately, in areas of human gene therapy (Waldman 1992, 1995; Bertling 1995; Guénêt 1995; Vega 1995). Gene-targeting studies in mammalian cells that employ insertion vectors generate recombinants that, in the majority of cases, bear a single copy of the transfer vector integrated into the cognate chromosomal locus by homologous recombination. This generates a characteristic and predicted change in the target locus that the region of homology shared by the transfer vector and the chromosome is duplicated with the duplication being separated by the integrated vector sequences (Waldman 1992, 1995; Bertling 1995). However, in some cases, a fraction of recombinants are also generated in which > 1 copy of the insertion vector is present at the homologous chromosomal locus (Jasin and Berg 1988; Fell et al. 1989; Thompson et al. 1989; Schwartzberg et al. 1990; Hasty et al. 1991, 1994, 1995; Bautista and Shulman 1993; Ng and Baker 1998).

While the mechanism by which a single vector integrates within the target chromosomal locus is best understood in terms of the double-strand-break repair (DSBR) model of homologous recombination (Orr-Weaver et al. 1981; Orr-Weaver and Szostak 1983; Szostak et al. 1983; Valancius and Smithies 1991), the mechanism by which multiple vector copies become integrated at the homologous chromosomal locus has not been studied. However, the structure of these recombinants is consistent with three distinct mechanisms: (i) targeted integration of a vector concatamer; (ii) targeted vector insertion followed by a second homologous recombination event between replicated, unequally-paired sister chromatids [unequal sister chromatid exchange (USCE)] generating tandem copies of the vector; and (iii) multiple homologous recombination events each involving the targeted insertion of a single vector copy. Each of these mechanisms can be distinguished provided that the region of homology shared by the transfer vector and the chromosome bears unique diagnostic markers enabling the contribution of vector-borne and chromosomal sequences in the final recombinant product to be ascertained. However, in the numerous gene targeting studies cited above, neither the exact vector copy number nor the mechanism responsible for its placement at the target locus was determined.

Corresponding author: Mark D. Baker, Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada N1G 2W1. E-mail: mbaker@ovcnet.uoguelph.ca

Genetics 151: 1143-1155 (March 1999)
although, in some studies (Hast y et al. 1991, 1994; Bau tista and Shul man 1993), the authors had assumed that the mechanism was targeted integration of a vector concatamer. Elucidation of the mechanism has important implications with respect to understanding mammalian gene targeting and to the underlying goal of improving its efficiency in targeted-genome alteration.

We have described previously a gene-targeting assay based on the ability to detect homologous recombination events between the haploid, chromosomal immunoglobulin \( \mu \) gene locus in a murine hybridoma and a transferred DNA vector bearing a segment of homology to the chromosomal \( \mu \) gene constant (C\( \mu \)) region (Baker et al. 1988). Following vector transfer, the majority of targeted recombinants at the chromosomal \( \mu \) locus bear as expected, a single copy of the transfer vector correctly integrated at the chromosomal \( \mu \) gene target locus (Ng and Baker 1998). Consistent with earlier findings (Orr-Weaver et al. 1981, 1988; Orr-Weaver and Szostak 1983; Valancius and Smithies 1991) these recombinants could be explained by the DSBR mechanism of homologous recombination (Ng and Baker 1998). However, another class of recombinants was of the more interesting type, in which >1 copy of the transfer vector was inserted at the target locus and for which the mechanism was unknown.

In the present study, we exploited our gene-targeting assay in the study of the latter recombinants. To accomplish this, the homologous vector-borne C\( \mu \) region was modified by site-directed mutagenesis making it genetically distinguishable from the target chromosomal C\( \mu \) region. This permitted contributions between the vector-borne and chromosomal C\( \mu \) sequences at the recombinant \( \mu \) locus to be determined. In conjunction with the specially designed insertion vector, we also modified our gene-targeting assay to enable the detection and isolation of all products of individual gene-targeting events in an unbiased manner. Following vector transfer, recombinants bearing >1 copy of the transfer vector integrated at the chromosomal \( \mu \) locus were identified. The structure of the chromosomal \( \mu \) gene in the recombinants was characterized and found to contain two tandem copies of the vector integrated by homologous recombination at the chromosomal \( \mu \) locus generating a triplication of the C\( \mu \) region bearing the various diagnostic markers. Analysis of the C\( \mu \) region marker pattern in the recombinants revealed that the C\( \mu \) region triplication in each recombinant was inconsistent with a mechanism involving integration of two vectors present as a concatamer [mechanism i (above)] or with a mechanism in which an initial gene targeting event involving a single vector was accompanied by a single USCE to generate the C\( \mu \) region triplication [mechanism ii (above)]. Rather, our results strongly suggested that recombinants containing two tandem vector copies at the target \( \mu \) locus were generated by two DSBR events, each involving a single copy of the transfer vector. The importance of these findings with respect to the mechanism of mammalian gene targeting is discussed.

### MATERIALS AND METHODS

**Recipient hybridoma and plasmids used in gene targeting:**

The mutant igm482 hybridoma (Bau mann et al. 1985) was used as the recipient for gene targeting. Mutant igm482 was derived from the wild-type Sp6 murine hybridoma that bears a single copy of the trinitrophenyl (TNP)-specific chromosomal immunoglobulin \( \mu \) gene and produces cytolytic, TNP-specific IgM (Köhler and Shul man 1980; Köhler et al. 1982). The haploid \( \mu \) gene in igm482 bears a 2-bp deletion in the third constant region exon (C\( \mu \)3), which results in synthesis of a truncated \( \mu \) chain and formation of noncytolytic IgM. Thus, whereas the wild-type Sp6 hybridoma can be detected as a plaque-forming cell (PFC) in a sensitive TNP-specific plaque assay (efficiency ∼0.8 TNP-specific PFC/cell), the mutant igm482 hybridoma cannot (efficiency <0.7 TNP-specific PFC/cell; Köhler et al. 1982; Bau mann et al. 1985; Baker et al. 1988). An additional feature of the 2-bp igm482 deletion is the deletion of an XmnI restriction enzyme site normally present at this position in the wild-type C\( \mu \)3 exon and the creation of a fill site. With the exception of the 2-bp deletion, the mutant igm482 and wild-type Sp6 chromosomal \( \mu \) genes are isogenic.

The 13.4-kb enhancer-trap vector pC\( \mu \)\( \mu \)En\( \mu \)16 and corresponding 13.7-kb enhancer-positive vector pC\( \mu \)\( \mu \)En\( \mu \)16 used in this study have been described in detail in the accompanying article (Ng and Baker 1999). In brief, each vector bears a 5.8-kb C\( \mu \) region segment from the wild-type Sp6 hybridoma that is homologous to the haploid mutant igm482 chromosomal \( \mu \) gene. The vector-borne C\( \mu \) region was modified by site-directed mutagenesis (Deng and Nickoloff 1992) to create six diagnostic restriction enzyme sites (Figure 1A) that can be distinguished from the restriction enzyme sites in the corresponding positions of the mutant igm482 chromosomal C\( \mu \) region (Figure 1B). Deletion of the SV40 early region enhancer driving \( \mu \) gene expression in the enhancer-trap vector pC\( \mu \)\( \mu \)En\( \mu \)16 enriches for gene targeting at the chromosomal immunoglobulin \( \mu \) locus (Ng and Baker 1998), a feature that facilitated isolation of recombinant hybridomas as described below.

**Vector transfer and transformant isolation:** Vector DNA (8.7 pmol) was linearized within the C\( \mu \) region at the unique XbaI site (Figure 1A) and transferred to \( 2 \times 10^7 \) recipient mutant igm482 hybridomas by electroporation as described (Baker et al. 1988). Vector linearization at XbaI provided 1.5 and 4.3 kb of homology to the haploid, mutant igm482 chromosomal C\( \mu \) region on the 5′ and 3′ sides of the vector cut site, respectively. Following electroporation, the hybridoma culture was resuspended in DMEM (Dulbecco’s modified Eagle medium containing 13% bovine calf serum and 5.3 \times 10^{-4} \text{m} 2-mercaptoethanol). An average of 30% of the hybridomas survived electroporation as determined by trypan blue staining. Screening of the surviving hybridomas for recombinants in which the transferred vector had integrated by homologous recombination was accomplished by two procedures as described in the accompanying article (Ng and Baker 1999) and summarized below.

**Procedure 1:** This method of isolating targeted hybridomas is based upon the ability of the wild-type C\( \mu \)3 exon carried on the insertion vector to correct the 2-bp deletion in the haploid, recipient mutant igm482 chromosomal C\( \mu \)3 exon (Baker et al. 1988). Homologous recombination generates a G418\( ^{\text{a}} \) hybridoma producing cytolytic, TNP-specific IgM, which can be identified as a PFC in a complement-dependent,
TNP-specific plaque assay (Baker et al. 1988) and recovered as described (Baker and Read 1992). In gene-targeting experiments involving the enhancer-trap vector pC\(_\mu\)E\(_\text{M1-6}\), an alternative procedure involved plating the hybridoma culture 2 days postelectroporation at densities of 10\(^4\) and 10\(^5\) cells well in microtiter plates in DMEM supplemented with G418 at an active concentration of 600 \(\mu\)g/ml. Following colony growth, culture supernatant was screened for TNP-specific IgM by complement-dependent lysis of TNP-coupled sheep erythrocytes in a spot test (Köhler and Shulman 1980; Baker 1989). All hybridomas producing cytoplasmic TNP-specific IgM were cloned at limiting dilution and saved for DNA analysis to verify the gene-targeting event.

Procedure 2: A modification of the plating procedure described above was devised for use with the enhancer-trap vector pC\(_\mu\)E\(_\text{M1-6}\) that permitted independent, targeted G418\(^R\) recombinants to be isolated irrespective of their IgM phenotype. Following electroporation with pC\(_\mu\)E\(_\text{M1-6}\), the hybridoma culture was immediately resuspended in 1188 ml of DMEM and 0.1 ml (\(\sim 10^7\) cells) was distributed into individual wells of 96-well microtiter plates. Two days later, each culture well received 0.1 ml of DMEM supplemented with G418. Following outgrowth of G418\(^R\) colonies, the number of growth-positive wells was enumerated. Two electroporations were performed. Of the 11,616 wells plated from the first electroporation, 450 generated G418\(^R\) colonies. According to the Poisson distribution, this indicated a mean of 0.04 G418\(^R\) cells/well. From the second electroporation, 11,712 wells were plated and of these, 389 generated G418\(^R\) colonies. From the Poisson distribution, the mean number of G418\(^R\) cells/well was 0.03. Thus, in procedure 2, each G418\(^R\) growth-positive well originated from a single G418\(^R\) transformant. Also, as this procedure resulted in single hybridomas being segregated immediately after electroporation into individual culture wells, it ensured that the G418\(^R\) products of each individual gene-targeting event were retained for analysis in a single culture well. All G418\(^R\) transformants were saved for DNA analysis to identify targeted recombinants.

**DNA analysis:** DNA analysis procedures used in the present study include isolation of hybridoma genomic DNA, Southern blotting to nitrocellulose, and hybridization with \(^{32}\)P-labeled DNA probes. PCR analysis of C\(_\mu\) region DNA in recombinant hybridomas and the sequence of the PCR primers AB9703 and AB9745, specific for the 5'and middle C\(_\mu\) regions, and AB9703 and AB9438, specific for the 3' C\(_\mu\) region, have all been described in the accompanying article (Ng and Baker 1999).

### RESULTS

**Screening of G418\(^R\) transformants and frequency of class I and class II recombinants:** The G418\(^R\) transformants isolated by procedures 1 and 2 were screened by Southern blot and hybridization analysis using chromosome-and vector-specific probes to identify those in which the transferred vector had integrated into the recipient igm482 chromosomal \(\mu\) gene by homologous recombination. As shown in Figure 1B, the endogenous single copy mutant igm482 C\(_\mu\) region is present on a 12.5-kb EcoRI fragment. For recombinants in which \(>1\) copy of the transfer vector had integrated in tandem into the chromosomal \(\mu\) locus by homologous recombination (designated class II recombinants), the following fragments are expected (Figure 1C): for the integration of pC\(_\mu\)E\(_\text{M1-6}\), the EcoRI fragment bearing the 5' C\(_\mu\) region will be 16.2 kb, whereas it will be 16.5 kb for pC\(_\mu\)E\(_\text{M1-6}\) because of the presence of the SV40 early region enhancer. For both vectors, the 3' C\(_\mu\) will be present on a 9.6-kb EcoRI fragment. In addition to these EcoRI fragments, a vector repeat fragment(s), of size 13.4 kb for pC\(_\mu\)E\(_\text{M1-6}\) or 13.7 kb for pC\(_\mu\)E\(_\text{M1-6}\), bearing the middle C\(_\mu\) region(s) is also expected. In the case of recombinants in which only a single copy of the vector has integrated into the chromosomal \(\mu\) gene by homologous recombination (designated class I recombinants, not illustrated), the structure of the recombinant \(\mu\) gene will be identical to that of the class II recombinant except for the absence of the 13.4- or 13.7-kb EcoRI fragment(s) bearing the middle C\(_\mu\) region(s).

To determine the exact vector copy number integrated at the \(\mu\) locus in the targeted recombinants, hybridoma genomic DNA was analyzed with the restriction enzyme combination PacI/PaeR71, which did not cut within the targeting vectors. As shown in Figure 1B, the endogenous C\(_\mu\) region is present on a 14.8-kb PacI/PaeR71 fragment. Following targeted vector integration, the size of the PacI/PaeR71 \(\mu\) gene fragment will increase according to the number of integrated vector copies. For example, in the class I recombinants in which a single copy of the transfer vector had integrated into the chromosomal \(\mu\) locus (not illustrated), the endogenous 14.8-kb PacI/PaeR71 \(\mu\) fragment would be converted to a 28.2- or 28.5-kb PacI/PaeR71 \(\mu\) gene fragment following integration of the 13.4-kb pC\(_\mu\)E\(_\text{M1-6}\) or the 13.7-kb pC\(_\mu\)E\(_\text{M1-6}\) vector, respectively. Similarly, in the case where two vector copies have integrated, the endogenous 14.8-kb PacI/PaeR71 \(\mu\) fragment will be converted into a 41.6- or 42.2-kb PacI/PaeR71 \(\mu\) gene fragment for pC\(_\mu\)E\(_\text{M1-6}\) or pC\(_\mu\)E\(_\text{M1-6}\), respectively (Figure 1C).

From the Southern blot screenings of G418\(^R\) transformants isolated from seven electroporations, a total of 41 correctly targeted G418\(^R\) recombinants was identified. Of these, the majority (33/41) were of the class I type in which a single copy of the transfer vector had integrated into the chromosomal \(\mu\) locus by homologous recombination as reported in the accompanying article (Ng and Baker 1999). The remaining 8/41 G418\(^R\) recombinants were of the class II type in which the chromosomal \(\mu\) locus was found to contain exactly two copies of the integrated vector in tandem. Class II recombinants 2/1, 3/5, 19/9, and 4/2 were isolated according to procedure 1 while recombinants 26-1, 112-2, 29-1, and 42-2 were isolated by procedure 2. In addition, class I recombinants 2/1 and 3/5 were obtained following targeted integration of two copies of pC\(_\mu\)E\(_\text{M1-6}\) while the remaining six class II recombinants were derived following targeted integration of two copies of pC\(_\mu\)E\(_\text{M1-6}\). The results of the Southern blot analysis of genomic DNA from class II recombinants generated by gene targeting with pC\(_\mu\)E\(_\text{M1-6}\), digested with EcoRI and
Figure 1.—Gene targeting at the chromosomal \( \mu \) locus. (A) As described in materials and methods, the enhancer-positive vector \( pCmEn1 \) is identical to the enhancer-trap vector \( pCmEn2 \) except that the latter lacks the SV40 enhancer region. For both vectors, site-directed mutagenesis was used to create the indicated vector-borne restriction enzyme sites (shown as solid gray circles in A), which replace the endogenous restriction enzyme sites in the mutant igm482 chromosomal \( Cm \) (shown as open circles in B). The \( Cm \) nucleotide position of the various diagnostic restriction enzyme pairs relative to the \( XbaI \) site of vector linearization (defined as nucleotide position 0) is indicated. The wild-type \( Xmni \) site in the vector-borne \( Cm3 \) exon is shown. Although not relevant to this study, each vector bears an enhancerless Herpes Simplex Virus-1 thymidine kinase (\( tk \)) gene. (B) The structure of the haploid, recipient mutant igm482 chromosomal \( Cm \) gene. The positions of the restriction enzyme sites are indicated by the open circles. The \( TfiI \) restriction enzyme site denotes the position of the 2-bp deletion in the mutant igm482 \( Cm3 \) exon (indicated by the open triangle). In both A and B, all wild-type exons are denoted by the solid black squares. (C) Typical structure of the chromosomal recombinant \( \mu \) gene following targeted integration of two copies of either \( pCmEn1 \) or \( pCmEn2 \) vectors in tandem. The positions of the PCR primers AB9703, AB9745, and AB9438 and the sizes of the amplified products they produce are depicted. Each gene-targeting reaction has the potential to generate a different marker pattern in each recombinant. For this reason, the identity of the diagnostic restriction enzyme markers at each \( Cm \) region position is not indicated. Also, the \( Cm3 \) exon in each \( Cm \) region of the triplication is indicated as wild type (solid black square). The figures are not drawn to scale.

PaeR71 and analyzed with \( C\mu \)-specific probe fragment \( F \) are presented in Figure 2, A and B.

As the plating method of isolation procedure 2 permitted all G418\(^\text{R} \) transformants to be isolated immediately after electroporation as single cells before the first division, it allowed direct determination of the absolute frequency of both the class I and class II types of recombinant. From the total of 41 G418\(^\text{R} \) recombinants re-
ported above, 21 G418\(^{R}\) recombinants were isolated from two separate electroporations according to procedure 2. Of the 21 G418\(^{R}\) recombinants, 17 were of the class I type while 4 were of the class II type. Because the class I and class II recombinants were isolated from a total of 1.2 \(\times 10^{7}\) hybridomas that survived the two electroporations (as described in materials and methods; of the 2 \(\times 10^{7}\) hybridomas subjected to each electroporation, an average of 30\% or 6.6 \(\times 10^{6}\) hybridomas survived), the absolute frequency of the class I recombinants was 1.42 \(\times 10^{-6}\) recombinants/ cell while the absolute frequency of the class II recombinants was 3.33 \(\times 10^{-7}\) recombinants/ cell.

Class II recombinants do not contain random vector integrations: From the results presented in Figure 2, A and B, no other fragments other than those expected from the recombinant \(\mu\) locus were observed in the class II recombinants. This suggested the absence of random vector integration elsewhere in the genome of these targeted recombinants. Further direct confirmation of this result was obtained by Southern blot analysis following digestion of the class II genomic DNA with EcoRI and electrophoresis such that all DNA fragments \(\geq 300\) bp were retained for transfer onto nitrocellulose. Each random vector integration is expected to yield two novel EcoRI fragments bearing the vector-chromosome junctions. However, analysis of the blots with \(\Delta\mu\) probe fragment N and vector-specific neo probe fragment G failed to detect fragments other than those expected of the class II recombinant \(\mu\) locus (Figure 1C; data not shown). Therefore, random vector integration(s) elsewhere in the hybridoma genome was not present in the class II recombinants.

Examination of the site of vector linearization in the class II recombinants: An important question regarding the mechanism of homologous recombination was whether or not the XbaI site used for vector linearization was restored in the class II recombinants as depicted in Figure 1C. This was addressed by digestion of class II genomic DNA with the enzyme combination HaeIII/PstI and electrophoresis such that all DNA fragments \(\geq 300\) bp were retained for transfer onto nitrocellulose, and hybridized with \(^{32}\)P-labeled probe F. The blot presents representative class II recombinants in which two copies of the vector have integrated in tandem by homologous recombination into the mutant igm482 chromosomal \(\mu\) locus. In addition, representative class I recombinants, in which one copy of the transfer vector has integrated by homologous recombination into the mutant igm482 \(\mu\) locus, are also shown. The positions of bands of interest are presented on the left of the blot while the positions of relevant marker bands are indicated on the right.

![Figure 2](image-url)

**Figure 2.** Analysis of the \(\mu\) gene structure in targeted recombinants. Genomic DNA from representative hybridomas was digested with (A) EcoRI or (B) the combination PstI/ PaeRII, electrophoresed through 0.7% agarose gel, blotted to nitrocellulose, and hybridized with \(^{32}\)P-labeled probe F. The blot presents representative class II recombinants in which three copies of the vector have integrated in tandem by homologous recombination into the mutant igm482 chromosomal \(\mu\) locus. In addition, representative class I recombinants, in which one copy of the transfer vector has integrated by homologous recombination into the mutant igm482 \(\mu\) locus, are also shown. The positions of bands of interest are presented on the left of the blot while the positions of relevant marker bands are indicated on the right.
regions with primers AB9703 and AB9745 from each chromosomal EcoI sites of vector linearization in the class II recombinants. Genomic DNA from the indicated hybrids was digested with the combination Haell/Xbal, electrophoresed through a 0.7% agarose gel, blotted to nitrocellulose, and hybridized with 32P-labeled μ-specific probe N. The blot presents the eight class II recombinants. The positions of bands of interest are presented on the left of the blot while the positions of relevant marker bands are indicated on the right.

deletions (Figure 4). The significance of this finding will be addressed in the discussion.

Analysis of restriction enzyme markers in the tandem Cμ triplication: As indicated in Figure 1A, six positions in the vector-borne Cμ region can be distinguished from the corresponding sites in the recipient mutant igm482 chromosomal Cμ region (Figure 1B) by diagnostic restriction enzymes. To probe further the mechanism of homologous recombination, PCR was used to amplify a 4.8-kb product from the 5′ and middle Cμ regions with primers AB9703 and AB9745 and a 4.6-kb product from the 3′ Cμ region with primers AB9703 and AB9438 (Figure 1C). Following PCR amplification, the identity of all six markers in each of the three Cμ regions of every class II recombinant was determined by separately digesting the PCR products with each of the diagnostic restriction enzymes. The diagnostic fragments expected are presented for the 5′ and middle Cμ regions PCR products in Figure 5A and for the 3′ Cμ region PCR product in Figure 5B. As an example, we present next the analysis to determine whether the chromosomal SacI marker or the vector-borne EcoRV marker was present at Cμ nucleotide position 645 in each Cμ region of the triplication for the eight class II recombinants.

Following PCR amplification of the 5′ and middle Cμ regions with primers AB9703 and AB9745 from each of the class II recombinants, the PCR products were subjected to digestion with the restriction enzyme EcoRV and the results are presented in Figure 6. The Cμ region PCR products amplified from recombinants 4/2, 112-2, and 29-1 were completely sensitive to cleavage by EcoRV at Cμ nucleotide position 645, yielding the expected diagnostic 2.7- and 0.7-kb fragments. This indicated that the vector-borne EcoRV marker was present in both the 5′ and middle Cμ regions in these recombinants. This was further confirmed by the complete resistance of the PCR products to cleavage by SacI (data not shown). For recombinants 19/9, 2/1, and 3/5, the Cμ region PCR product was completely resistant to cleavage by EcoRV at Cμ nucleotide position 645 as judged by the continued presence of the diagnostic 3.4-kb fragment. Thus, the vector-borne EcoRV marker was not present in either the 5′ or middle Cμ regions in these recombinants. As expected, the PCR product was completely sensitive to cleavage by SacI yielding the diagnostic 0.7- and 4.0-kb fragments (data not shown), indicating that in these recombinants, the chromosomal SacI marker was present in both the 5′ and middle Cμ regions. In the case of recombinants 26-1 and 42-2, digestion of the PCR product with EcoRV revealed partial cutting at Cμ nucleotide position 645 (Figure 6), a result that was also obtained following digestion with SacI (data not shown). This suggested either that the identity of the marker at Cμ nucleotide position 645 in the 5′ Cμ region was different from that in the middle Cμ region or that the recombinants were each composed of two distinct subpopulations that differed with respect to the identity of the marker at Cμ nucleotide position 645 in the 5′ and/or middle Cμ region. The latter possibilities would result if symmetric heteroduplex DNA (hDNA), generated by homologous recombination, encompassed the marker at this position in the 5′ and/or middle Cμ region but was not repaired before DNA replication.

To distinguish between these possibilities, recombinants 26-1 and 42-2 were cloned by limited dilution and 10 subclones were isolated for Southern blot analysis to unambiguously determine the identity of the marker at Cμ nucleotide position 645 in the 5′ and middle Cμ regions.

As shown in Figure 1B, the presence of the chromosomal SacI marker in the endogenous igm482 Cμ region results in a 5.3-kb SacI fragment following digestion with SacI/EcoRI and analysis with the Cμ-specific probe fragment N. As shown in Figure 1C, the presence of the chromosomal SacI marker and the absence of the vector-borne EcoRV marker in the 5′ Cμ region are expected to

5′-GTGGTGGCTGAAGGGCCAGATCCACCTACCTCTAGAGGACTCTCTCCCT-3′

Figure 4.—The 26-bp deletion in the middle and 3′ Cμ region of recombinant 42-2. The 397-bp Stul/Xmal fragment normally encompassing the Xbal site of vector linearization was cloned from the PCR-amplified product of the 5′, middle, and 3′ Cμ regions of recombinant 42-2 into the EcoRV and Xmal sites of pBluescript (Stratagene, La Jolla, CA) and sequenced. Shown is the relevant portion of the sequence from the 5′ Cμ region with the Xbal site underlined. The 26-bp sequence that was deleted from the middle and 3′ Cμ region is indicated in italics.
Multiple Vector Insertion

Figure 5.—Restriction enzyme map of the 5', middle, and 3' Cμ region PCR products. The sizes (in base pairs) of the diagnostic fragments expected following digestion of (A) the 4765-bp 5' and middle Cμ region PCR products and (B) the 4621-bp 3' Cμ region PCR product with the indicated restriction enzymes are presented above each line. The size of the DNA fragment that would be produced if the indicated restriction enzyme site were not present is indicated below each line. The chromosomal restriction enzyme sites are denoted by an E while the vector-borne sites in the corresponding positions are denoted by a V. As indicated in Figure 1C and in this figure, primers AB9703 and AB9745 were used to specifically amplify the 5' and middle Cμ regions while primers AB9703 and AB9438 were used to specifically amplify the 3' Cμ region. In all cases additional nondiagnostic restriction fragments may be present but are not shown. The diagrams are not drawn to scale.
yield a 5.3-kb SacI/EcoRI fragment or an 11.1-kb EcoRV/NgoM1 fragment. Conversely, the absence of the SacI marker and the presence of the EcoRV marker in the 5′ Cμ region are expected to yield a 14.4-kb SacI/EcoRI fragment or a 8.4-kb EcoRV/NgoM1 fragment. In the case of the middle Cμ region, the presence of the SacI marker and the absence of the EcoRV marker are expected to yield a 3.8-kb SacI/EcoRI fragment or a 5.7-kb EcoRV/NgoM1 fragment. Conversely, the absence of the SacI marker and the presence of the EcoRV marker in the middle Cμ region are expected to yield a 12.9-kb SacI/EcoRI fragment or a 3.1-kb EcoRV/NgoM1 fragment.

Figure 7 presents the analysis of genomic DNA from recombinants 26-1 and 42-2 along with their respective subclones digested with SacI/EcoRI and probed with Cμ fragment N. The results revealed that 26-1 was composed of two subpopulations in about equal proportions. Six of the 10 subclones of 26-1 were of subpopulation A and bore the 14.4- and 12.9-kb SacI/EcoRI fragments indicating that the vector-borne EcoRV marker was present in both the 5′ and middle Cμ regions. The remaining 4 of the 10 subclones of recombinant 26-1 were of subpopulation B and contained the 5.3- and 12.9-kb SacI/EcoRI fragments indicating that the chromosomal SacI marker was present in the 5′ Cμ region while the vector-borne EcoRV marker was present in the middle Cμ region. As shown in Figure 7, the results of this analysis for the parental 26-1 recombinant are consistent with it being composed of these two subpopulations. These results were confirmed following Southern blot analysis with Cμ probe fragment N of EcoRV/NgoM1-digested genomic DNAs from recombinant 26-1 and its subclones (data not shown). The 3′ Cμ region of the parental recombinant 26-1 bears the endogenous SacI marker (see below), thus generating the 3.8-kb SacI/EcoRI fragment (Figure 1C) that is present in all subclones (Figure 7). As indicated above, the finding that 26-1 was composed of two subpopulations in approximately equal proportions that differed with respect to the marker at Cμ nucleotide position 645 in the 5′ Cμ region suggested that hDNA, generated during homologous recombination, had encompassed at least this marker position and that this mismatched site was not repaired before DNA replication.

For recombinant 42-2 and its subclones, the results revealed the presence of 5.3- and 12.9-kb SacI/EcoRI fragments (Figure 7) indicating that, in all cases, the 5′ Cμ region bears the chromosomal SacI marker while the middle Cμ region bears the vector-borne EcoRV marker. This was confirmed by Southern blot analysis with probe fragment N following digestion of genomic DNA from recombinant 42-2 and its subclones with EcoRV/NgoM1 (data not shown). The parental recombinant 42-2 bears the EcoRV marker in the 3′ Cμ region (see below), thus resulting in the presence of the 9.3-kb SacI/EcoRI (Figure 1C) fragment in all subclones (Figure 7).

The PCR primers AB9703 and AB9438 specifically amplify the 3′ Cμ region (Figures 1C and 5B). Therefore, the identity of the restriction enzyme marker at Cμ nucleotide position 645 in the 3′ Cμ region of the class II recombinants could be determined unambiguously following digestion of the PCR product with SacI or EcoRV. This analysis revealed that marker position 645 in the 3′ Cμ region PCR product of recombinants 19/9, 26-1, and 112-2 was sensitive to SacI but resistant.

![Figure 6](image6.png)

**Figure 6.** Restriction enzyme analysis of Cμ nucleotide position 645 in the 5′ and middle Cμ region PCR products. The cleavage products resulting from digestion with EcoRV are presented for the eight class II recombinants. The positions of the fragments of interest are located on the left of the gel while the 1-kb marker bands (denoted M) are indicated on the right of the gel. In all cases a nondiagnostic 1.4-kb fragment is generated.

![Figure 7](image7.png)

**Figure 7.** Analysis of Cμ nucleotide position 645 in the tandem Cμ triplication of recombinants 26-1 and 42-2. Genomic DNA from the indicated hybridomas was digested with the combination SacI/EcoRI, electrophoresed through 0.7% agarose gel, blotted to nitrocellulose, and hybridized with 32P-labeled probe N. The blot presents class II recombinants 26-1 and 42-2 along with their respective subclones. The positions of the bands of interest are presented on the left of the blot while the positions of relevant marker bands are indicated on the right.
TABLE 1

Analysis of Cm region markers in Class II recombinants

<table>
<thead>
<tr>
<th>Recombinant</th>
<th>5' Cm nucleotide position (bp)</th>
<th>Middle Cm nucleotide position (bp)</th>
<th>3' Cm nucleotide position (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>239 645 1199 1586 1677 2114</td>
<td>239 645 1199 1586 1677 2114</td>
<td>239 645 1199 1586 1677 2114</td>
</tr>
<tr>
<td>19/9</td>
<td>Avall Sac Dral Xmr1 Earl Scal</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
</tr>
<tr>
<td>4/2</td>
<td>Avall EcoRV AflI Xmr1 Aat1 Scal</td>
<td>Avall EcoRV AflI Xmr1 Aat1 Scal</td>
<td>Avall EcoRV AflI Xmr1 Aat1 Scal</td>
</tr>
<tr>
<td>2/1</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
</tr>
<tr>
<td>3/5</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
</tr>
<tr>
<td>26-1</td>
<td>Avall EcoRV Dral Xmr1 Aat1 Scal</td>
<td>Avall EcoRV Dral Xmr1 Aat1 Scal</td>
<td>Avall EcoRV Dral Xmr1 Aat1 Scal</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>112-2</td>
<td>Avall EcoRV Dral Xmr1 Aat1 Scal</td>
<td>Avall EcoRV Dral Xmr1 Aat1 Scal</td>
<td>Avall EcoRV Dral Xmr1 Aat1 Scal</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>29-1</td>
<td>Kmnl EcoRV AflI Tfl Earl Nhel</td>
<td>Kmnl EcoRV AflI Tfl Earl Nhel</td>
<td>Kmnl EcoRV AflI Tfl Earl Nhel</td>
</tr>
<tr>
<td>42-2</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
<td>Avall Sac Dral Xmr1 Aat1 Scal</td>
</tr>
</tbody>
</table>

Chromosomal markers from the endogenous mutant igm482 Cm region are indicated in normal type face while those in boldface indicate vector-borne markers from the Cm region of the transfer vector. The indicated Cm nucleotide position defines the location of the various markers relative to the Xmr1 site of vector linearization.
regions would have already undergone DNA replication before USCE. In contrast, the persistence of mismatches is permitted in the middle Cµ region in the absence of mismatch repair because it is the product of the USCE event that occurred following DNA replication. The four recombinants isolated by procedure 1 (19/9, 4/2, 2/1, and 3/5) cannot be used to address this issue because, as described in materials and methods, this procedure did not permit for the recovery and analysis of all products of gene targeting. However, this was not the case for recombinants isolated by procedure 2. Of the four recombinants isolated by this method, two (26-1 and 112-2) revealed direct evidence of incomplete hDNA repair and, in both cases, all unrepaird mismatches were located in the 5’ Cµ region making their generation entirely inconsistent with USCE mechanism ii. The third mechanism postulates that the class II recombinants were generated by two single vector integration events that each occurred by DSBR. Unlike the first two models, mechanism iii does not place any constraints on the fate of the restriction enzyme markers in the Cµ region triplication and thus best explains all of the recombinants.

Additional evidence supporting the mechanism whereby two DSBR events each involving a single copy of the vector generated the eight class II recombinants comes from examination of the Cµ region marker patterns in the recombinants (Table 1). In six of the eight class II recombinants analyzed (19/9, 4/2, 2/1, 3/5, 26-1, and 112-2), the vector-borne KpnI marker located only 239 bp from the XbaI site of vector linearization was lost and it was replaced in all three Cµ regions by the chromosomal AvaI marker. This result is best explained by two gap repair events according to the DSBR model of homologous recombination (Orr-Weaver et al. 1981; Orr-Weaver and Szostak 1983; Szostak et al. 1983). Another interesting feature of the Cµ region triplication in the class II recombinants was that in each recombinant, a pair of adjacent Cµ regions was identical. This information suggested a possible target Cµ region for the second DSBR event involved in generating the Cµ region triplication. That is, the middle and 3’ Cµ regions in four of the eight recombinants (19/9, 4/2, 29-1, and 42-2) were identical, suggesting that in these recombinants, the second DSBR event might have occurred in the 3’ Cµ region generated by the first DSBR event. In the remaining four recombinants (2/1, 3/5, 26-1, and 112-2), the 5’ and middle Cµ regions were identical, suggesting that in these recombinants, the second DSBR event might have occurred in the 5’ Cµ region that was generated by the first DSBR event. The apparent equivalent usage of the 5’ and 3’ Cµ regions in the duplication that was generated by the first DSBR as a target for the second DSBR event suggests a lack of bias concerning where the second DSBR event occurred.

Examination of the Cµ region marker patterns in the class II recombinants revealed evidence of extensive gene conversion of markers adjacent to a region of gap repair. That is, gene conversion of many other markers, mostly toward vector sequences, was also evident, suggesting that gene conversion resulted from mismatch repair of hDNA generated during gene targeting. Direct evidence for the presence of hDNA in the DSBR process was obtained following analysis of recombinants isolated by procedure 2, where all products of individual targeted vector integration events were available for analysis in a single culture well. Of the four recombinants isolated by this procedure, two, 26-1 and 112-2, were mitotically sectored, suggesting that there was incomplete repair of hDNA before DNA replication. Evidence of hDNA in the recombinants is consistent with gene targeting occurring by DSBR (Szostak et al. 1983; Orr-Weaver et al. 1988).

Before electroporation, the targeting vector was linearized at the unique XbaI site within the Cµ region of homology. As indicated above, with the exception of recombinant 42-2, the XbaI site of vector linearization was restored in all class II recombinants, a result consistent with the predictions of the DSBR model (Szostak et al. 1983) and, where examined, with previous gene targeting studies (Smithies et al. 1985; Baker et al. 1988; Baker and Shulman 1988; Kang and Shulman 1991; Ng and Baker 1998, 1999). However, our results suggest that on rare occasions the site of vector linearization may be lost. In recombinant 42-2, identical 26-bp deletions in both the middle and 3’ Cµ regions resulted in the loss of the XbaI site. That 42-2 had suffered two rare, identical deletions strongly suggested that they were not independent mutations. Rather, they were explained more readily by two DSBR events as follows: during the first vector integration event, the DSB at XbaI was enlarged to form a gap. Because gap repair may not always occur with fidelity (Strathern et al. 1995), it is possible that some part of this process generated the 26-bp deletion in 42-2, removing the XbaI site from one of the two recombinating Cµ regions. Resolution of the recombination intermediate resulted in the first vector integration in which the correctly repaired sequence was present in the 5’ Cµ region while the XbaI-deleted Cµ region was in the 3’ Cµ region. Following this, a second vector gapped at XbaI integrated by DSBR into the XbaI-deleted 3’ Cµ region. As a consequence of “correct” gap repair, the 26-bp deletion was copied into the newest recombining Cµ region, generating a tandem Cµ region triplication with the identical 26-bp deletion in the middle and 3’ Cµ regions. By this mechanism, the rare deletion resulting in the loss of XbaI is directly associated with vector integration by DSBR and need occur only once. The second “loss” is copied as a consequence of normal gap repair and thus easily explains the presence of two identical rare deletions.

In summary, our results strongly suggest that the class II recombinants harboring two tandem vector copies at
the target chromosomal μ locus arise predominantly, if not exclusively, as a consequence of two vector integration events, each occurring by DSBR. This conclusion agrees with gene-targeting studies in yeast (Orr-Weaver et al. 1981). In the yeast studies, following introduction of plasmids bearing a gap in the region of homology to the target locus, about half of the targeted recombinants were found to bear a tandem array of the vector integrated at the target locus. Because each copy of the vector in the tandem array had repaired the gap, the authors concluded that they arose as a result of multiple vector integration events, each occurring by DSBR.

**DISCUSSION**

In this study, our analysis of the class II recombinants suggested that, during gene targeting, an individual hybridoma has the ability to perform two independent DSBR events resulting in the integration of two tandem copies of the insertion vector into the chromosomal μ locus by homologous recombination. Importantly, the two independent targeted vector insertion events occurred in the complete absence of random vector integration elsewhere in the hybridoma genome. The lack of random vector integration in hybridomas bearing a single targeted vector integration event at the chromosomal μ locus was found in other studies (Ng and Baker 1998, 1999). Generally, this is also the case for gene targeting at other chromosomal loci in different mammalian cells (reviewed in Bollag et al. 1989; Waldman 1992, 1995; Bertling 1995), although on a few occasions, random integration elsewhere in the genome was observed (Lin et al. 1985; Thomas et al. 1986; Adair et al. 1989; Sedivy and Sharp 1989; le Mouellic et al. 1990; Reid et al. 1991). The general lack of random vector integration in targeted cells is consistent with the notion that the pathways of targeted and of random vector integration are not directly competitive. However, in view of our own work (Baker and Shulman 1988; Ng and Baker 1998) and that of others (reviewed in Bollag et al. 1989; Waldman 1992, 1995; Bertling 1995) indicating that the frequency of random vector integration is typically \( \sim 10^{-3} \) while that of targeted vector integration is \( \sim 10^{-6} \), targeted cells bearing a random integration would be expected only at the low frequency of \( \sim 10^{-9} \). On the basis of this frequency argument, it was pointed out (Waldman 1992) that the lack of random integration in targeted mammalian cells might be explained simply on the basis of the low probability of detecting the simultaneous occurrence of these two rare events in a single cell rather than as a result of any mutual exclusion in the random and homologous recombination pathways in a cell. However, the results of the present study showing that an individual mammalian cell is more likely to perform two targeted vector integration events in the complete absence of random integration rather than the more frequently expected events of single gene targeting accompanied by random vector integration is a finding that, to our knowledge, is not only unprecedented but provides support that, indeed, targeted vector integration does not directly compete with random vector integration. The following frequencies indicate just how striking this conclusion is. As indicated in the results, the absolute frequency of a single targeted vector insertion event into the chromosomal immunoglobulin μ locus in the hybridoma is \( 1.42 \times 10^{-6} / \text{cell} \). Accordingly, class II recombinants bearing two independent targeted vector insertion events at the chromosomal immunoglobulin μ locus would be expected at a frequency of \( 2.02 \times 10^{-12} / \text{cell} \). However, this was not the case. In fact, the class II recombinants were observed at the exceptionally high frequency of \( 3.33 \times 10^{-7} / \text{cell} \), a value that is \( 1.65 \times 10^{5} \)-fold higher than expected. We interpret these results to mean that a small proportion of the recipient cells exist in a state that favors targeted but not random vector integration.

What might the nature of the recipient population be that makes it so proficient for gene targeting? Perhaps in the general population, a factor(s) responsible for random integration is/are abundant while a factor(s) responsible for gene targeting is/are scarce/absent. Conversely, a small fraction of the recipient population might possess a factor(s) that makes it proficient for gene targeting but lack a factor(s) responsible for random vector integration. Whatever the nature of this state, it appears transient because gene-targeting efficiencies are similar in cells regardless of whether they have been targeted previously or not (te Riele et al. 1990; Cruz et al. 1991; Askew et al. 1993; Detloff et al. 1994; Stacey et al. 1994; Wu et al. 1994).

In this study, the question is raised as to whether or not the second (cotargeting) event occurs with high efficiency only at the loci where the first gene targeting event has occurred or might occur with a similar efficiency elsewhere in the genome. The ability to conduct high-efficiency cotargeting at unlinked chromosomal loci would have several practical applications with respect to the modification of chromosomal genes in cultured mammalian cells and animals. In one study where this issue was addressed (Reid et al. 1991), cotargeting at the hprt and β2-microglobulin loci was not observed. However, the authors attributed this to the fact that the absolute frequency of gene targeting at the two target loci differed by 200-fold. Therefore, it is possible that cotargeting might indeed occur at unlinked chromosomal loci provided that they shared a similar absolute gene targeting efficiency.

The ability to specifically modify chromosomal sequences by gene targeting is clearly a very powerful technology. However, in mammalian cells, this power is tempered by the low absolute frequency of the initial event which, despite over a decade of research, has remained unchanged at \( \sim 10^{-6} / \text{cell} \) for most loci in a
variety of cell lines (reviewed in Bollag et al. 1989; Waldman 1992, 1995; Bertling 1995). For gene targeting to attain its full potential, the absolute frequency of gene targeting will have to be improved. On the basis of the assumption that the pathways of random and targeted integration were competitive, strategies were proposed to improve the absolute frequency of gene targeting by blocking random integration so that more vector molecules would be available for entry into the targeting pathway (Waldman 1992, 1995; Bertling 1995). For example, because random integration involves DNA end-joining (Chang and Wilson 1987; Roth and Wilson 1988), one suggested approach was to modify the termini of the vector with dideoxynucleotides to effectively block random integration (Chang and Wilson 1987). Another approach was based on the observation that poly(ADP-ribosylation) plays an essential role in random integration (Farzaneh et al. 1988) and it was demonstrated that by inhibiting poly(ADP-ribosylation) during transfection, random vector integration in mammalian cells was effectively blocked (Farzaneh et al. 1988; Waldman and Waldman 1990). However, the results of this study suggest that the pathways of targeted and random vector integration are not directly competitive. Thus, it is highly unlikely that such strategies of blocking random integration will increase the absolute frequency of gene targeting. Indeed, Waldman et al. (1996) have investigated the effect of blocking random integration by inhibiting poly(ADP-ribosylation) on the efficiency of gene targeting at the aprt locus in CHO cells and found that this treatment did not increase the absolute frequency of gene targeting, a result consistent with the prediction presented above. Our results suggest that achieving an increase in the absolute frequency of gene targeting will first require determination of the nature of the transient state amenable to targeting. Perhaps then, the absolute frequency of gene targeting can be increased by enriching for cells in the transient state by or manipulating the recipient population to enter this state before or during gene transfer.

This work was supported by an Operating Grant from the Medical Research Council of Canada to M.D.B.

LITERATURE CITED

Multiple Vector Insertion


Communicating editor: C. Kozak