ECTOPIC GENE CONVERSION IN MAMMALIAN SOMATIC CELLS

A Thesis

Presented to

The Faculty of Graduate Studies

of

The University of Guelph

by

PATRICIA LOUISE BELL

In partial fulfilment of requirements

for the degree of

Master of Science

August, 2001

ABSTRACT

ECTOPIC GENE CONVERSION IN MAMMALIAN SOMATIC CELLS

Patricia L. Bell
University of Guelph, 2001

Supervisor:
Dr. Mark D. Baker

An enhancer-trap vector was used to target a genetically-marked Cμ region into the endogenous Cα locus of murine hybridoma line igm482. The efficiency of targeting was similar to that of enhancer-trap targeting to the Cμ locus in igm482 (Ng and Baker 1998). Homologous recombination frequencies between the integrated vector-borne Cμ region and the endogenous haploid Cμ locus were determined for a targeted line and for lines in which the vector had integrated randomly. Conversion frequencies ranged from 10^{-6} to 10^{-8}, and were similar to previously established inter-chromosomal recombination rates in igm482. Five independent homologous recombination events were recovered and analyzed. Analysis of the donor Cμ region revealed that the transfer of information to the endogenous Cμ locus was uni-directional. Examination of the recipient Cμ revealed that gene conversion tract lengths ranged from ≥91 bp to ≥1469 bp. Conversion events are discussed in the context of the double-strand break repair model.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

LIST OF ABBREVIATIONS

Abs = antibodies

DSB = double strand break

DSBR = double strand break repair

En^- = enhancerless; enhancer-trap

GC = gene conversion

GT = gene targeting

hDNA = heteroduplex DNA

HR = homologous recombination

IGT = interrupted gene targeting

IMUT = Intron MUTation markers (RFLP markers of the Cμ region)

MEPS = minimal efficient processing segment

NHEJ = non homologous end joining

OSI = one sided invasion

PFC = plaque forming cell

RBC = red blood cells

RE = restriction enzyme

RES = recombination enhancing sequence

RFLP = restriction fragment length polymorphism

SDM = site directed mutagenesis

sRBC = sheep red blood cells

INTRODUCTION

Homologous recombination is the means by which genetic information is exchanged between two stretches of double stranded DNA with similar or identical genetic sequence. Gene conversion involves a non-reciprocal transfer of information from one DNA duplex to another, while cross-over is a reciprocal exchange of genetic information. Meiotic recombination generates genetic diversity by recombining pre-existing loci, or by creating deletions/duplications of genes. Such events can lead to species diversification or to speciation. Mitotic recombination is beneficial to the organism as it is used to repair damaged DNA, and to generate antibody diversity. However recombination can also generate deleterious DNA structures such as those associated with carcinogenesis (reviewed by Hoffmann 1994). Mutated tumour suppressor genes can behave as recessive cancer genes, and recombination can lead to the loss of the wildtype copy of such a gene through deletion, chromosome loss or gene conversion. Oncogenes can be activated by recombination through loss/alteration of a regulatory region, through duplication of the oncogene, or through a gene fusion event that creates a new gene. Because of the potential deleterious effects of ectopic recombination to an organism, a greater understanding of recombination mechanisms is required.

Most studies of mitotic recombination in mammalian cells have focused on the rates and mechanisms of intra-chromosomal homologous recombination. The low rates of inter-chromosomal recombination (often below the limits of detection in most assays) make it difficult to study recombination events between unlinked loci.

However the Cμ recombination assay has been successfully used to detect and measure inter-chromosomal homologous recombination between IgM constant regions (Cμ) (Shulman *et al.* 1995). Briefly, this assay detects homologous recombination events between Cμ regions that restore production of polymeric trinitrophenyl (TNP)-specific IgM. Recombinants are then detected based on complement-dependant lysis of TNP-coated red blood cells (the plaque assay). In the current study ectopic recombination rates were measured using the plaque assay, and information about gene conversion tract lengths was obtained. In addition, the efficiency of gene targeting to the Cα locus with an enhancer-trap vector was established.

1. LITERATURE REVIEW

1.1 MODELS OF HOMOLOGOUS RECOMBINATION

Homologous recombination (HR) is the exchange of genetic information between two identical or nearly-identical stretches of double-stranded DNA. This exchange may be reciprocal (a crossover event) or non-reciprocal (gene conversion) or both. Yeast and fungi provided a simple experimental system for early recombination studies, as all products of a meiotic recombination event could be recovered. Both cross-overs and aberrant segregation (gene conversion) were observed, and early recombination models had to account for both events. The first model proposed to explain these events was the Holliday model (Holliday 1964). In this model, recombination is initiated by two single stranded nicks at homologous sites on separate chromatids, followed by single strand invasion of the homologous DNA duplex to create a χ-shaped intermediate (the Holliday junction). Subsequent experimental data invalidated the proposed initiation mechanism, yet many key features of the Holliday model (most notably the Holliday junction) were incorporated into later models, and it remained the seminal work in the field of recombination. Meselson and Radding (1975) postulated that recombination was initiated by a single stranded nick on one chromatid, followed by single strand invasion of the homologous duplex to create a Holliday junction and to displace a D-loop from the invaded chromatid. Later experimental data demonstrated that double strand breaks (DSBs) were highly recombinogenic, and led to the proposal of the Double Strand Break Repair (DSBR) Model (Szostak *et al.* 1983). In this model,

recombination is initiated by a DSB in one of the two interacting chromatids. The 3' end of the broken duplex invades the homologous duplex to create a Holliday junction. The displaced D-loop from the unbroken chromatid pairs with complementary sequences on the other duplex to create a second Holliday junction.

In contrast to the above, some recombination models are non-conservative. The Single Strand Annealing Model (Lin *et al.* 1984) does not involve a Holliday junction intermediate. Instead, recombination is initiated by a double-stranded break that is enlarged by exonuclease activity to expose complementary single strands of DNA, which can then pair. In the One Sided Invasion model (Belmaaza and Chartrand 1994), a single strand liberated by a DSB invades the homologous duplex and primes DNA synthesis. The newly synthesized strand is then ligated to the non–invading end of the broken DNA duplex, and repair synthesis takes place. The copy choice model (d'Alençon *et al.* 1994, Tang 1994) is another example of non-conservative recombination. During DNA replication, the polymerase switches from one parental strand to another, to produce a recombinant daughter molecule. The above conservative and non-conservative recombination models are described in more detail below.

1.1.1 The Holliday Model

The Holliday model (Fig. 1.1) proposed that recombination iss initiated by two single stranded nicks made at homologous sites on separate chromatids (Fig. 1.1A, B) (Holliday 1964, 1968). The two DNA duplexes are unwound to produce single stranded ends that then exchange places by invading the other chromatid (Fig.

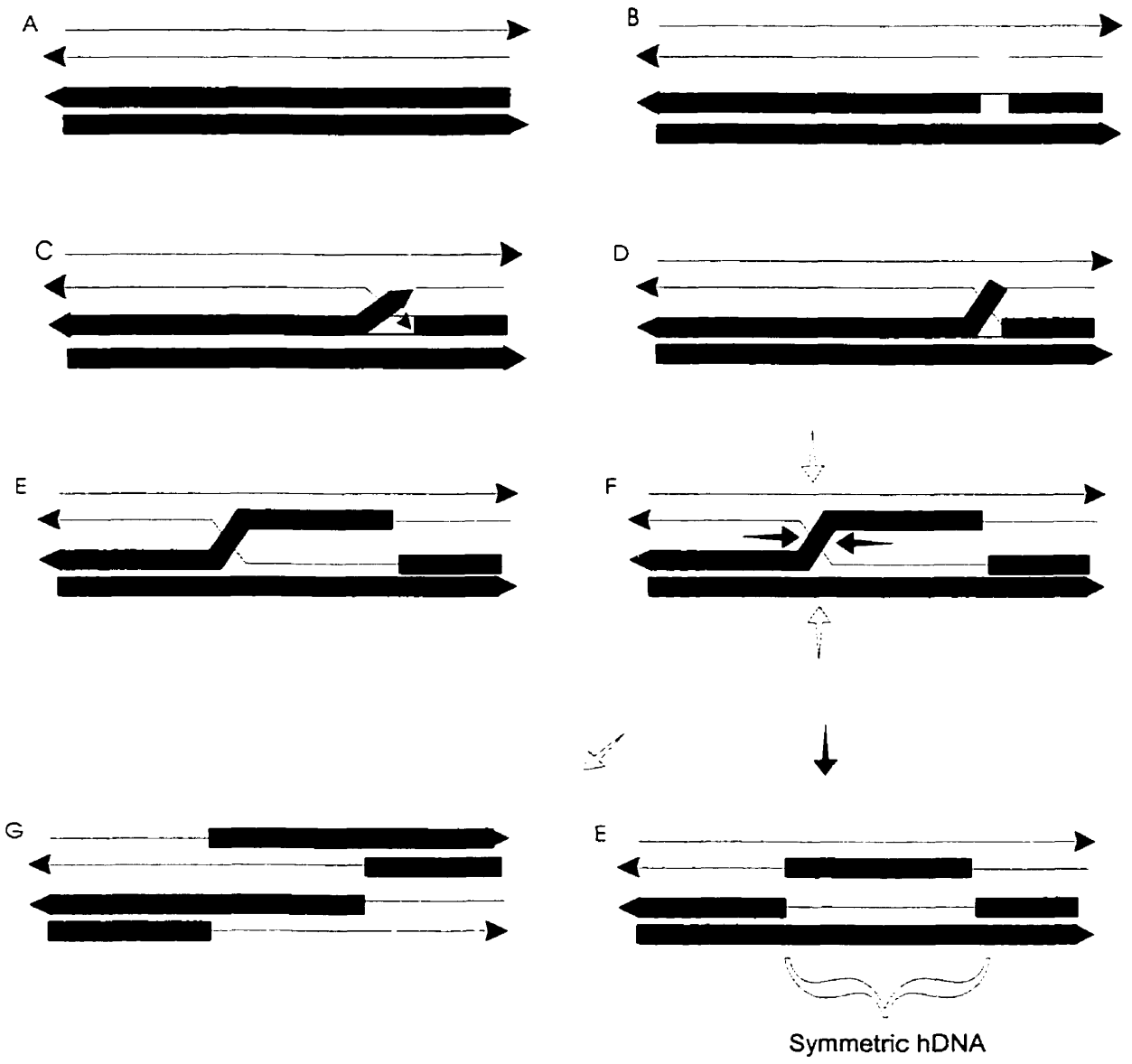


Figure 1.1 The Holliday Model. Recombination is initiated when homologous chromosomes align (A) and two single stranded nicks are made at homologous sits on separate chromatids (B). The two DNA duplexes unwind to produce single stranded ends that then invade the other duplex (C) to produce a Holliday junction (D). Branch migration can occur in either direction (E) to create tracts of symmetric hDNA. Resolution occurs by cutting either the outer non-crossed strands (indicated by the white arrows) or the inner crossed strands (grey arrows) (F) to produce a crossover (G) or non-crossover (H) product.

1.1C) and are ligated in place to produce a Holliday junction (Fig. 1.1D). This junction can move in either direction via branch migration (Fig. 1.1E), and is resolved by cutting in one of two ways (Fig. 1.1F) and ligated to recreate two separate strands (Fig. 1.1G, H). Resolution can produce a crossover product (Fig. 1.1G) if the outer, non-crossed strands (white arrows) are cleaved, or a non-crossover product (Fig. 1.1H) if the inner, crossed strands (gray arrows) are cleaved. If the two chromatids contain DNA sequence differences at the site of recombination, then symmetric heteroduplex DNA (hDNA) will be produced when the nicked strands exchange places. Repair of this mismatched hDNA results in gene conversion.

1.1.2 The Meselson-Radding Model

The Holliday model provided a conceptual paradigm for the molecular events underlying recombination and gene conversion, but it was limited in its prediction that only symmetric hDNA was formed during recombination. Many experimental results (for review see Petes *et al.* 1991) suggested that both symmetric and asymmetric hDNA were formed. Meselson and Radding (1975) proposed a revised model of recombination in which both asymmetric and symmetric hDNA were generated (Fig. 1.2). Recombination is initiated by a single stranded nick on one chromatid (Fig. 1.2A) followed by asymmetric invasion of a single strand into the other duplex (Fig. 1.2B), thereby generating asymmetric hDNA. The invading strand induces a single stranded break in the invaded duplex (Fig. 1.2C), and the invading strand is ligated into place. The single-stranded D-loop displaced from the invaded

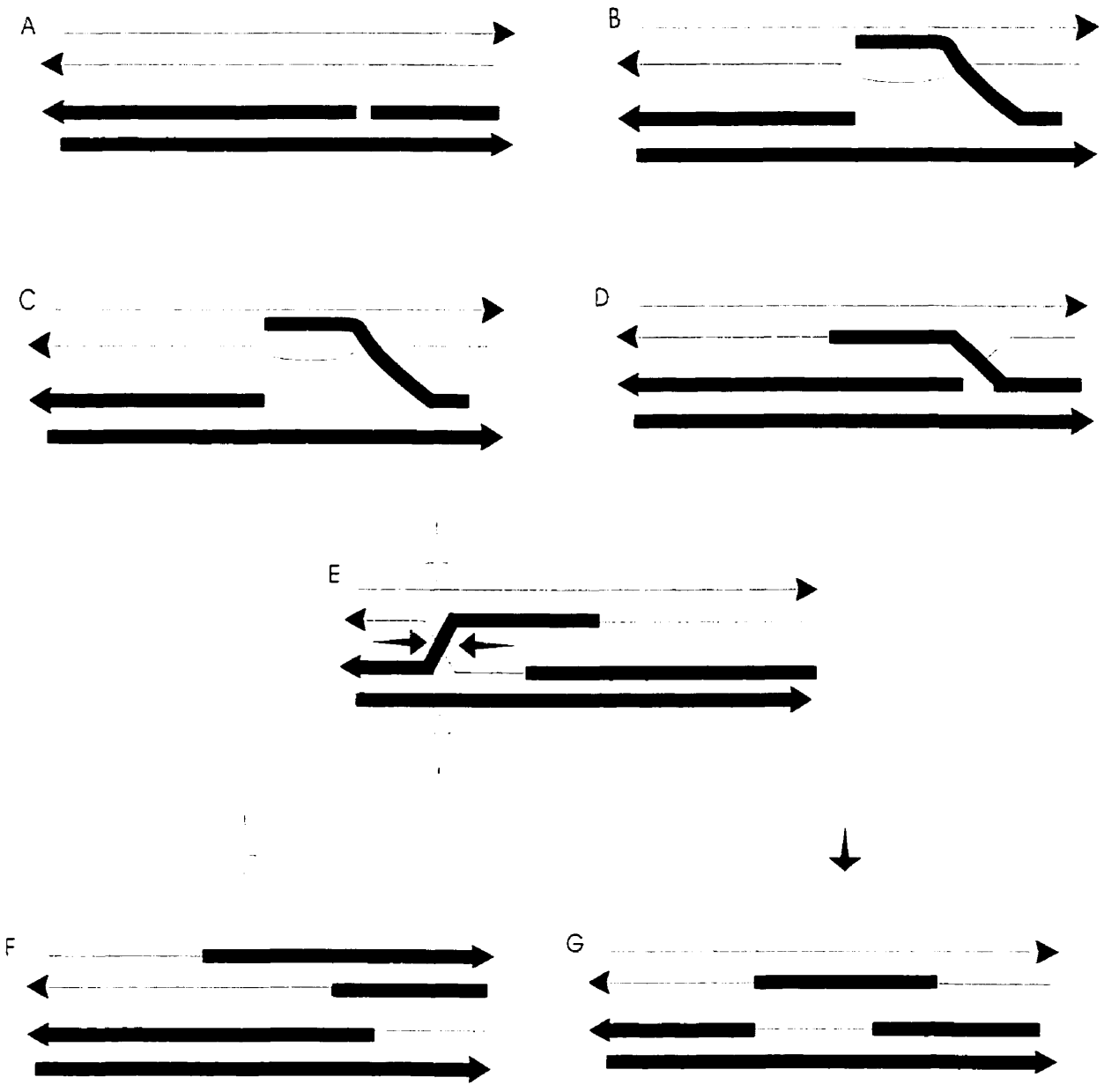


FIGURE 1.2 THE MESELSON-RADDING MODEL OF RECOMBINATION

Recombination is initiated by a single stranded nick on one chromatid (A) followed by asymmetric invasion of a single stranded end into the other duplex (B). The invading strand induces a single stranded break in the invaded duplex (C), allowing the invading strand to be ligated into place. DNA synthesis on the donor duplex replaces the invading strand and ligation of the 3' end of this newly synthesized DNA to the 5' end of the recipient strand on the invaded duplex results in the formation of a Holliday junction (D). This junction can move by rotary diffusion to generate tracts of symmetric hDNA (E). The Holliday junction can be resolved to produce either a crossover (F) or non-crossover event (G)

duplex is degraded by exonucleases, and extension of this degradation elongates the region of hDNA on the invaded duplex. DNA synthesis on the donor duplex replaces the invading strand and ligation of the 3' end of this newly synthesized DNA to the 5' end of the recipient strand on the invaded duplex results in the formation of a Holliday junction (Fig. 1.2D). This junction can move by rotary diffusion to generate tracts of symmetric hDNA (Fig. 1.2E). Resolution can produce either a crossover (Figure 1.2F) or non-crossover (Fig. 1.2G) event, depending on whether the uncrossed or crossed strands are cut. Thus the Meselson-Radding model accounts for the formation of both asymmetric and symmetric hDNA during recombination.

1.1.3 The Double Strand Break Repair Model

The Meselson-Radding model better explained yeast recombination data than the Holliday model but was unable to account for the recombinogenicity of double strand breaks. A break introduced into a homologous region contained on a plasmid increased gene targeting efficiency by as much as three orders of magnitude over that of the circular plasmid (Orr-Weaver *et al.* 1981). Furthermore, a circular plasmid containing two regions of homology to yeast DNA will integrate at either locus when transformed into yeast cells whereas a plasmid linearized within one of the two regions of homology will always integrate into the genomic locus homologous to the region of the plasmid containing the double strand break (DSB) (Orr-Weaver *et al.* 1981). Additionally, if a plasmid is cut with two restriction enzymes to create a gap in the region of homology, the gap is always replaced in

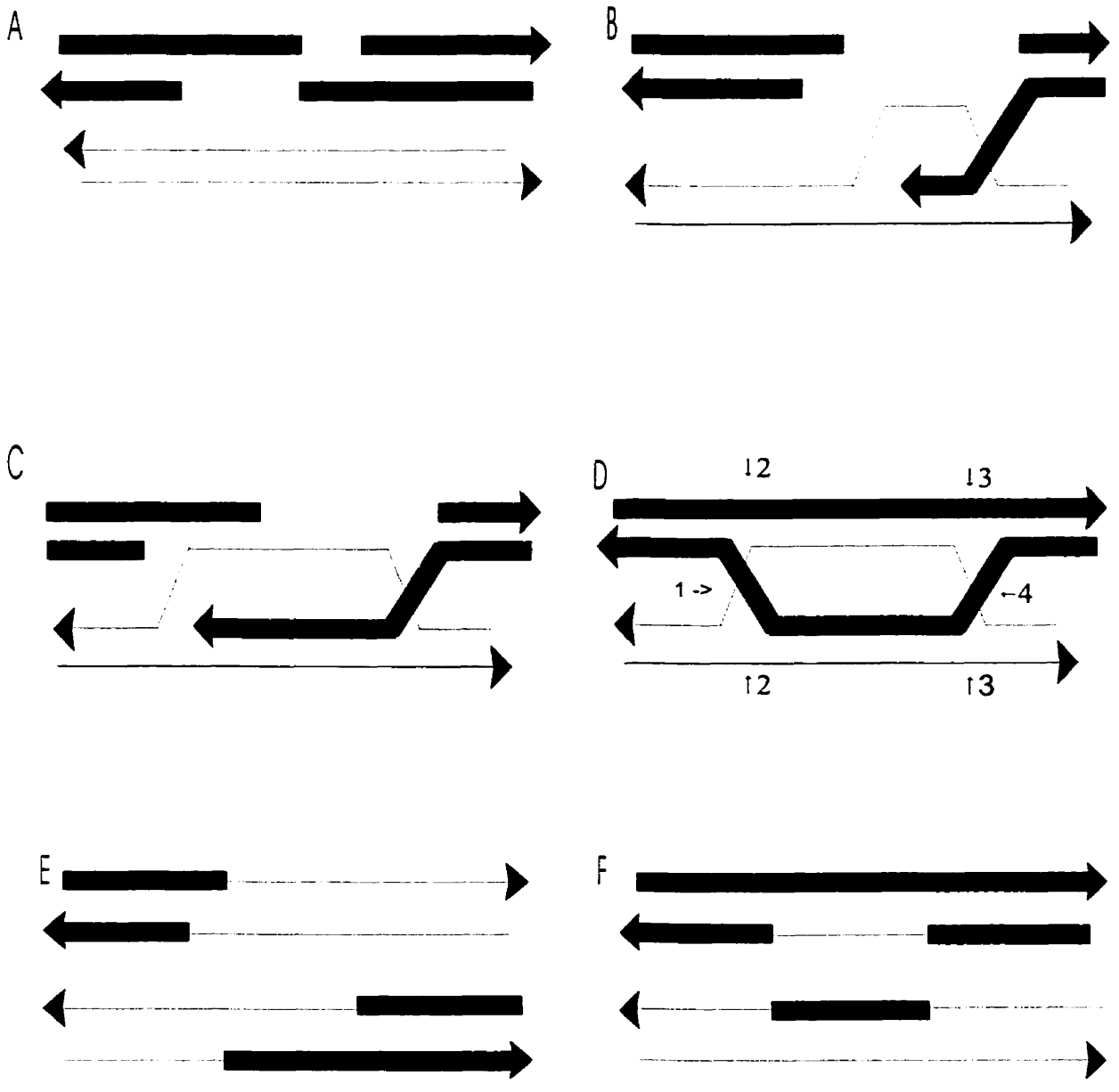


Figure 1.3 The Double Strand Break Repair Model of Recombination Recombination is initiated by a break produced by a double strand specific endonuclease in the recipient DNA duplex (A). One of the 3' termini produced by this break invades its homologous site on the donor duplex, dislodging a D-loop of single stranded DNA from the donor site (B). The invading 3' termini is then used to prime DNA repair synthesis, thereby enlarging the D loop which can then anneal to complementary sequences on the recipient chromatid (C). This creates a double Holliday junction intermediate (D) that can be resolved to produce either a crossover (E) or non-crossover (F) product.

targeted lines. Taken together, these results strongly imply that the DNA termini created by a DSB interact directly with the homologous region on the recipient (chromosomal) duplex. Accordingly, Szostak *et al.* (1983) proposed the double strand break repair (DSBR) model of recombination (Fig. 1.3). Recombination is initiated when a double strand-specific endonuclease produces a DSB in the recipient DNA duplex (Fig. 1.3A). One of the 3' termini produced by the DSB invades the homologous region on the donor duplex, dislodging a D-loop of single stranded DNA (Fig. 1.3B). The invading 3' termini is then used to prime DNA repair synthesis, thereby enlarging the D loop until it can anneal to complementary sequences on the recipient chromatid (Fig. 1.3C). The resulting recombination intermediate contains two Holliday junctions (Fig. 1.3D). Resolution of these junctions can produce a crossover (Fig. 1.3E) or non-crossover (Fig. 1.3F) event. Later research led to a modification to the DSBR model. Sun *et al.* (1991) found that DSBs were processed to yield 3' overhang single stranded ends. Invasion of the unbroken strand by one of these single-stranded tails led to the formation of hDNA. The recombination event then proceeded as described in the DSBR model (Fig. 1.1B, C, D, E, F).

1.1.4 The Single Strand Annealing Model

Although the DSBR model adequately explained intra- and inter-chromosomal HR, it could not account for certain experimental results in extrachromosomal HR. Lin *et al.* (1984) employed phage or plasmid vectors that contained two partial *tk* genes, each with a non-overlapping deletion. When

transformed into mouse L cells, intra-molecular homologous recombination between the two *tk* alleles generated a functional gene. If, prior to transformation, the *tk* vector was cut in the sequence between the duplicated *tk* genes, the frequency of tk^+ cells recovered was significantly increased. The frequency of transformation was greatly increased when the two partial *tk* genes were transformed on separate linear fragments versus transformation with two circular plasmids or with one linear and one circular plasmid.

Lin *et al.* (1984) proposed the single-strand annealing (SSA) model to explain the generation of wildtype *tk* genes in the above system. The two DNA termini created by the cut in the transformation vector are enlarged by a 5' exonuclease until complementary strands of the *tk* genes are available for pairing. An endonuclease then cuts away the single stranded tails that remain unpaired, thus allowing the *tk* gene to be corrected by DNA repair syntheis. If the location of the cut that linearizes the transformation vector is such that one of the *tk* alleles is degraded before the exonuclease exposes the second *tk*, then recombination efficiency will be low. If the exonuclease exposes complementary regions on the *tk* genes before either is completely degraded, then recombination efficiency will be high. The SSA model provides a simple explanation for extra-chromosomal recombination, but its role in either intra- or inter-chromosomal mammalian recombination between natural loci is unclear as degradation of intervening sequences between duplicated chromosomal genes might lead to deleterious deletion mutations. However SSA is a major pathway in yeast chromosomal recombination.

1.1.5 The One Sided Invasion Model

Another model that has been suggested as a mechanism of homologous recombination is the one sided invasion (OSI) model (Belmaaza and Chartrand 1994). Recombination is initiated by a DSB, followed by a single strand invasion of the unbroken duplex. The invading strand primes DNA synthesis, and this synthesis may extend past the region of homology. The newly synthesized strand is then ligated back into the break site to create an illegitimate junction. A Holliday junction can form if the D-loop (displaced by the invading strand) is nicked then ligated into the break site. This junction can then be resolved to produce crossover or non-crossover products, but an illegitimate junction may still be present.

1.1.6 The Copy Choice Model

The copy choice model is a non-conservative mechanism of recombination (d'Alençon *et al.* 1994, Tang 1994). In this model, DNA polymerase can switch from one parental strand to another during replication, to produce a recombinant daughter molecule. Evidence is available that this type of template switching can occur during the replication of repetitive sequences *in vitro* (Viguera *et al.* 2001).

1.2 Mitotic Recombination In Mammalian Cells

As the present study was an investigation of mitotic recombination in murine hybridomas, this review will concentrate on mitotic recombination in mammalian cells. The three main types of mammalian mitotic HR relevant to the current study

: [1] gene targeting (GT; an interaction between exogenous DNA and a homologous chromosomal locus), [2] intra-chromosomal recombination (including intra-chromatid and sister chromatid interactions), and [3] inter-chromosomal recombination (including inter-homolog HR). Extra-chromosomal homologous recombination can also occur, but is of little relevance to the present work.

1.2.1 Double Strand Breaks Stimulate Homology-Directed Repair in Mammalian Cells

DNA damage (including DSBs) stimulates HR in mammalian cells, but the mechanism of recombination is difficult to study when the site of the DNA damage is unknown. The discovery of the rare-cutting yeast restriction enzyme I-*Sce*I (with an eighteen base pair (bp) recognition sequence) provided a convenient mechanism for introducing site-specific DSBs in the mammalian genome. In most studies, the insertion of a I-*Sce*I recognition site into a marker allele was used to create a loss-of-function insertion mutation. DSBs introduced into this allele induced recombination with a second, non-functional, allele to generate a wild type allele. Choulika *et al.* (1995) observed that chromosomal DSBs were efficiently repaired using plasmid donor DNA (gene targeting) at a frequency at least 2 orders of magnitude above that of spontaneous homologous recombination . Moynahan and Jasin (1997) also observed that a site-specific DSB increased the frequency of inter-homolog recombination at least 2 orders of magnitude over background frequencies. Using tandemly duplicated APRT alleles, Sargent *et al.* (1997) demonstrated that a DSB in one of the two alleles stimulated intra-chromosomal recombination

approximately one hundred-fold while a similar study (Taghian and Nickoloff 1997) found that a DSB in one of two linked *neo* genes stimulated HR greater than one thousand-fold the level of spontaneous recombination events. As had been found for yeast (Orr-Weaver *et al.* 1981), a DSB in the transforming vector stimulated gene targeting events in mammalian cells (Smithies *et al.* 1985, Jasin and Berg 1988, Donoho *et al.* 1998). This data indicates that in mammalian cells (as in yeast cells) DNA ends are highly recombinogenic, and lends support for homology-directed repair as an important function of HR in mammalian cells. Yet despite this strong evidence, many studies also implicate the role of a non-homologous end joining mechanism in the repair of chromosomal DSBs.

1.2.2 Non-Homologous End Joining and Illegitimate Recombination in Mammalian Cells

DSBs greatly stimulate HR in *Saccharomyces cerevisiae*. As an example, gene targeting in yeast cells is remarkably efficient; most transformed lines represent targeting events while very few contain randomly inserted vector sequence (Orr-Weaver *et al.* 1981, Szostak *et al.* 1983, Hastings *et al.* 1993). In contrast, DSBs induce a much smaller increase in HR in mammalian cells, and illegitimate non-homologous end joining (NHEJ) recombination often predominates as a DSB repair mechanism (Godwin *et al.* 1994, Moynahan and Jasin 1997, Sargent *et al.* 1997).

Sargent *et al* (1997) used a loss of function assay and found that approximately 20% of APRT$^-$ cells recovered following the introduction of a DSB in

one of two linked APRT alleles were the result of illegitimate recombination. All of the APRT$^-$ cells recovered following the introduction of a DSB into a hemizygous copy of APRT represented NHEJ events. This led the authors to suggest that mammalian cells preferentially repair DSBs by illegitimate recombination, however their data suggests that when a second (donor) copy of the APRT gene is provided, intra-chromosomal homologous recombination occurs more frequently than does illegitimate recombination. Nevertheless, given the low frequency of inter-chromosomal HR, illegitimate recombination may predominate in the absence of a linked donor sequence. In support of this, a study of DSB-induced recombination found that all reconstituted wild type markers were produced by NHEJ rather than by inter-homolog HR (Godwin *et al.* 1994). These findings suggest that NHEJ is used to repair a DSB in the absence of a linked donor sequence.

1.2.3 Gene Targeting in Mammalian Cells

Gene targeting in mammalian cells is highly inefficient. Targeting frequencies range from 10^{-3} to 10^{-7} targeted cells per transfected cell, or 10^{-2} to 10^{-5} target cells per transformed cell (reviewed by Bollag *et al.*1989). Most transformed cells represent randomly integrated vector sequences rather than targeting events. This is in sharp contrast to yeast, in which gene targeting is a remarkably efficient process; most transformed lines are correctly targeted and very few contain randomly inserted transforming vector sequence (Orr-Weaver *et al.* 1981, Szostak *et al.* 1983, Hastings *et al.* 1993). The high number of random insertions found in mammalian cells may be due to the fact that NHEJ appears to be the more

predominant mechanism for incorporating transformed DNA into the genome. One of the anticipated benefits of achieving a greater understanding of gene targeting in mammalian cells is an improvement in gene targeting efficiency. Currently, the low frequencies of gene targeting to mammalian loci excludes gene therapy as a practical medicinal treatment. As such, many researchers have concentrated on strategies for improving the efficiency of gene targeting in mammalian cells.

Strategies such as oligonucleotide-mediated gene conversion have focussed on efficiently causing heritable changes in DNA without the insertion of marker DNA sequences into the genome. Other strategies have concentrated on reducing the background of non-targeted cells so that fewer transformed cells have to be screened in order to find the rare targeting events. The positive-negative selection strategy (Mansour *et al.* 1988) involves construction of a replacement targeting vector containing the selectable marker *neomycin* (*neo*) within the region of homology to the target locus and a second selectable marker, *thymidine kinase* (*tk*) outside the region of homology. A double crossover event at the target locus will lead to the integration of *neo* but not *tk*, while a random integration event will result in the incorporation of both *neo* and *tk* into the genome. Thus, the drug G418 can be used to select for all neo^R transformed cells, while selection with the drug gancyclovir will cull tk^+random insertion events. The promoter-trap (Joyner 1991) and enhancer-trap (En^-) (Jasin and Berg 1988) strategies are another approach used to reduce the background of non-targeted transformed cells, but are only effective if there is an enhancer/strong promoter located near the desired target locus. The targeting vector utilizes a selectable marker (i.e. *neo*) controlled by a

weak promoter. In a targeted line, the nearby promoter/enhancer will up-regulate transcription of *neo* such that the cells will be $G418^R$. If the targeting vector integrates randomly, *neo* will likely not be transcribed at a sufficiently high level to provide $G418^R$ to the cells. By happenstance, some random integrations will occur near an enhancer/strong promoter but the background of non-targeted lines is much reduced. Ng and Baker (1998) used an En^- vector to target the murine immunoglobulin heavy chain locus and found a 15-fold enrichment in targeted cells per transformed cell.

Several studies support the DSBR model as the mechanism of gene targeting in mammalian cells. Linearization of the transforming DNA increases the efficiency of gene targeting (Jasin and Berg 1988, Valancius and Smithies 1991, Donoho *et al.* 1998), as does introducing a DSB into the target genomic locus (Choulika *et al.* 1995, Donoho *et al.* 1998). A restriction enzyme marker located immediately next to a DSB is usually lost during gene targeting with an insertion vector (Valancius and Smithies 1991, Ng and Baker 1999). This is consistent with the gap repair mechanism proposed in the DSBR model. Both restriction fragment length polymorphism (RFLP) and palindrome markers have been used to demonstrate that hDNA forms during gene targeting both with insertion-type vectors (Ng and Baker 1999, Li and Baker 2000a,b) and replacement vectors (Li and Baker 2000c, Li *et al.* 2001). Li *et al.* (2001) used a replacement targeting vector that contained palindrome markers on each flanking arm of homology. Several of the targeted lines generated with this marker contained the flanking palindromes on opposite DNA strands, as would be consistent with two separate crossover events during

vector integration. This data is in agreement with the DSBR model, which predicts the formation of two Holliday junctions, each of which can be resolved by crossing over. Resolution of the double Holliday intermediate appears to be biassed towards 2,4-cleavage (Fig. 1-3) instead of 1,3-cleavage (Baker and Birmingham 2001). This is consistent with recent data on the interaction of the resolvasome RuvABC with Holliday junctions in *E. coli* (Cromie and Leach 2000) and in yeast (Whitby and Dixon 1998).

The length of homology required for efficient gene targeting is dependant upon whether an insertion or replacement vector is used. Gene targeting with an insertion vector occurs approximately as frequently as targeting with an insertion vector (Deng and Capecchi 1992). Insertion vectors appear to require at least 1.3 Kb of homology, although the frequency of targeting increases dramatically with an increase in available homology (Hasty *et al.* 1991a, Deng and Capecchi 1992). Replacement vectors require a greater amount of homology (at least 1.9 Kb), and also display increased targeting efficiency with increased homology (Hasty *et al.* 1991a, Deng and Capecchi 1992). Interestingly, it is the total amount of homology available on replacement vector (i.e. on both arms) rather than the amount of homology available on the shorter arm that determines gene targeting efficiency (Hasty *et al.* 1991b).

1.2.4 The Search for Homology (Frequencies of Homologous Recombination)

In order for two homologous sequences to interact with one another, they must first locate one another within the genome. The extent and mechanism of this

search for homology have yet to be precisely determined. The frequency of each type of HR (gene targeting, intra-chromosomal, and inter-chromosomal recombination) provides valuable information on the extent of this search for homology. However, it is important to note that frequencies of HR may also be dependant upon cell type, amount of homology available, locus involved, and the transcriptional state of this locus. These confounding variables make it difficult to directly compare recombination frequencies from different studies. The majority of recombination studies in mammalian cells investigate recombination events between non-functional heteroallelic marker genes (most commonly *tk* or *neo*) that restore the wildtype phenotype. These marker genes are introduced into cultured cells through transformation and have stably integrated into the genome. A further complication in studies using transformed lines is that epigenic effects can influence gene expression. Thus the recombination frequencies that are inferred from the frequency of drug resistant cells may be inaccurate.

Most studies conclude that in mammalian cells intra-chromosomal recombination is more prevalent than inter-chromosomal interactions. Spontaneous intra-chromosomal recombination frequencies are generally around 10^{-5} to 10^{-6} recombinants per cell (see Table 1.1 for a review of measured frequencies), yet frequencies of 2.5×10^{-4} are reported in mouse embryonic stem (ES) cells (Donoho *et al.* 1998) and 0.13 to 8.0×10^{-2} in murine hybridomas (Baker and Read 1995, Baker 1989). Intra-chromosomal recombination frequencies increase ten- to a hundred-fold upon introduction of a DSB in one of the recombining alleles (Brenneman *et al.* 1996, Sargent *et al.* 1997, Taghian and Nickoloff 1997, Donoho

et al. 1998). There are two categories of intra-chromosomal recombination: intra-chromatid and sister chromatid (inter-chromatid). Bollag and Liskay (1991, 1992) devised a recombination system that allowed them to distinguish between the two types of interactions. Duplicated *tk* genes (each with a different point mutation) in two different configurations were used: if the mutations were proximal to the intergenic space between the two *tk* alleles, then both intra-chromatid and sister chromatid cross-overs would generate a single wildtype copy of *tk*. If the mutations were distal to the intergenic space, unequal sister chromatid exchange (USCE) would generate a triplication containing a wildtype copy of *tk*, while intra-chromatid reciprocal exchange could not reconstitute a wildtype *tk*. After comparing the frequency of tk^+ cells recovered from each configuration, the authors found that sister chromatid interactions predominated.

Table 1.1 Frequencies of Spontaneous Intra-chromosomal Recombination in Mammalian cells.

Recombination Frequency	**Cell Type (gene)**	**Reference**
$0.13\text{-}8.0\times10^{-2}$	murine hybridomas (Cμ)	Baker (1989), Baker and Read (1995)
2.5×10^{-4}	mouse ES (hprt)	Donoho *et al.* (1998)
6.7×10^{-5}	CHO (tk)	Godwin *et al.* (1994)
5.0×10^{-5}	CHO (tk)	Godwin *et al.* (1994)
8.7×10^{-6}	CHO (transcribed) (neo)	Nickoloff (1992)
5.0×10^{-6}	CHO (neo)	Nickoloff (1992)
3.7×10^{-6}	CHO (untranscribed)	Nickoloff (1992)
2.8×10^{-6}	mouse L cells (neo)	Bollag and Liskay (1991)
1.8×10^{-6}	mouse L cells (neo)	Bollag and Liskay (1991)
$1.2\text{X}10^{-6}$	mouse L cells (tk)	Bollag and Liskay (1988)

Often, the position or orientation of the marker genes can affect recombination frequencies. For example, investigation of recombination between closely-linked immunoglobulin μ heavy chain genes (Cμ) found a frequency of intra-chromosomal recombination of $0.13x10^{-2}$ when a wildtype Cμ was downstream and a mutant Cμ upstream (Baker and Read 1995) and $8.0x10^{-2}$ when the wildtype Cμ was upstream (Baker 1989). Insertion of a donor Cμ sequence ~1megabase (mb) away from the recipient endogenous locus resulted in an intra-chromosomal recombination frequency of $0.18x10^{-2}$ (Baker *et al.* 1999), comparable to that found when the donor Cμ was located immediately downstream of the recipient copy. Position effects can also influence recombination frequency, although usually only by a few fold difference (Liskay *et al.* 1984, reviewed in Bollag *et al.* 1989). Transcribed genes appear to recombine at higher frequencies than do untranscribed loci (Nickoloff 1992), however the frequency of HR is not increased by transcription if a DSB is present (Taghian and Nickoloff 1997).

Spontaneous inter-chromosomal recombination frequencies are much lower than intra-chromosomal recombination frequencies, and thus are usually measured as an absence of data. Godwin *et al.* (1994) directly investigated the search for homology by comparing the frequencies of intra-chromosomal and inter-chromosomal recombination between *tk* heteroalleles. Two potential donor *tk* genes were available for recombination: a copy located on the same chromosome, and a copy on the homolog. Spontaneous intra-chromosomal recombination events that restored *tk* occurred at a frequency of $7x10^{-5}$ while inter-chromosomal recombination events were below the limits of detection ($<4x10^{-9}$). Interestingly, this study also

indicated that illegitimate recombination (NHEJ) is preferentially used over inter-homolog recombination to used to repair DSBs in mammalian cells. Richardson *et al.* (1998) did not detect any spontaneous HR ($<1x10^{-9}$) between two unlinked *neo* alleles located on non-homologous chromosomes. A second study (Richardson and Jasin 2000) that used neo genes with less homology determined that the recombination frequency was $<8x10^{-8}$. In contrast to frequencies of inter-chromosomal HR that were below the limits of detection in other studies (Godwin *et al.* 1994, Richardson *et al.* 1998, Richardson and Jasin 2000), inter-chromosomal recombination has been measured in a few model systems. Murti *et al.* (1994) reported that inter-chromosomal recombination occurred in the mouse germline at a frequency of $1\text{-}7x10^{-3}$. This high recombination frequency may reflect differences between mitotic and meiotic HR. The murine Cμ recombination system has been successfully used to find detectable inter-chromosomal HR events. Recombination frequencies between unlinked sequences of 10^{-6} to 10^{-7} have been reported (Baker and Read 1992, Baker *et al.* 1996) with this system. Shulman *et al.* (1995) found a similar frequency of recombination (10^{-7}) between Cμ loci located on homogs. These values are approximately 10^{4}-fold higher than those reported for intra-chromosomal recombination (Shulman *et al.* 1995, Baker *et al.* 1996).

The frequencies of spontaneous mitotic recombination in mammalian cells indicate that recombination occurs preferentially within the same molecule (intra-chromosomal) followed by inter-chromosomal HR. It has been suggested (Godwin *et al.* 1994, Shulman *et al.* 1995) that spontaneous inter-chromosomal recombination may be infrequent in mammalian cells because the large genome and

nuclear volume preclude interactions between the two sequences. A second possibility is that each chromosome occupies a specific location in the interphase nuclei of mammalian cells, and it is this arrangement of chromosomes that determines whether unlinked homologous sequences can interact. Interphase nuclei in *Drosophila* (Hilliker 1985) and in mammalian cells (Lichter *et al.* 1988) have been reported to have such a highly structured chromosome arrangement. An third possibility is that mammalian cells have developed a mechanism to inhibit inter-chromosomal HR in order to prevent the formation of deleterious rearrangements following recombination between the highly repetitive sequences found in the mammalian genome (Godwin *et al.* 1994). Lastly, intra-chromosomal recombination may be preferred over inter-homolog recombination to prevent the loss of heterozygocity at the recombining loci. The above hypotheses imply that inter-chromosomal HR is inhibited in mammalian cells, however, two lines of evidence suggest that the search for homology is genome wide. First, introduced DSBs can be repaired using donor sequences from the same chromosome (Brenneman *et al.* 1996, Sargent *et al.* 1997, Taghian and Nickoloff 1997, Donoho *et al.* 1998), from the homolog (Moynahan and Jasin 1997), or from other chromosomal locations (Richardson *et al.* 1998). This data implies that the search for homology that ensues upon introduction of a DSB is genome-wide. Second, despite the fact that spontaneous intra-chromosomal recombination is far more common than inter-chromosomal, the latter has been detected in mammalian cells for recombination partners located in many different chromosomal positions (Baker and Read 1992, Shulman *et al.* 1995, Baker *et al.* 1996). That spontaneous recombination occurs

between unlinked sequences supports the idea that the search for homology encompasses the entire mammalian genome.

The fact that DSBs lead to an increase in HR emphasizes that mitotic recombination in mammalian cells is likely involved in DNA repair. An unrepaired DSB would have profound deleterious consequences to the cell, and it is advantageous to the cell to repair such a lesion using a homologous sequence as template so that no genetic information is lost. The distribution of recombination frequencies (intra-chromosomal > inter-homolog) reinforces the idea that mitotic HR is used primarily for homologous repair of DNA lesions. Sister chromatids are utilized preferentially over unlinked homologous loci (Bollag and Liskay 1992) as donor sequences for DSBR to prevent the loss of heterozygocity (and thus genetic information) at the recombining locus. Inter-chromosomal HR may be a rare event because of the possibility that crossovers could lead to deleterious chromosomal rearrangements. NHEJ appears to be another common repair mechanism in mammalian cells, and may represent an alternative pathway for the repair of DSBs.

1.2.5 Mitotic Recombination and Gene Conversion

Early studies in yeast found an association between crossover and gene conversion. The Holliday model and subsequent recombination models accounted for gene conversion as a consequence of gap repair or the repair of hDNA formed during HR (Holliday 1968, Meselson and Radding 1975, Szostak *et al.* 1983). Gene conversions are often associated with single crossovers in both yeast (Petes *et al.* 1991) and mammalian cells (Bollag and Liskay 1988). When considering gene

conversion (GC) it is important to clarify that a double crossover event will produce the same result as a GC event. Consequently, these two events can not be unequivocally differentiated, however the low frequency of single crossovers in mammalian cells leads to the assumption that double crossovers are extremely rare and that GC is a more common form of HR. As well, one crossover event is thought to inhibit nearby crossovers.

Gene conversion was first detected in mammalian cells by Liskay and Stachelek (1983) who found that approximately half of all HR events between two adjacent *tk* heteroalleles resulted in GC. Most subsequent work on spontaneous intra-chromosomal HR between closely linked sequences in mammalian cells found GC, rather than simple reciprocal exchange, to be the primary mechanism of HR (Liskay *et al.* 1984, Liskay and Stachelek 1986, Bollag and Liskay 1988, Baker 1989, Baker and Read 1995), however other data suggests that the structure of the recombining heteroalleles can influence the outcome of a HR event.

Mitotic GC involves continuous conversion tracts (Liskay and Stachelek 1986, Yang and Waldman 1997) and appears to operate with high fidelity (Stachelek and Liskay 1988). The length of homology required to efficiently initiate mitotic intra-chromosomal GC appears to be between 134 and 300 bp (Liskay *et al.* 1987, Waldman and Liskay 1988). Despite the fact that this research observed GC between heteroalleles sharing only 200 bp of homology, the conversion frequency was very low ($\sim 10^{-8}$), and no GC was observed when only 95 bp of homology was available (Liskay *et al.* 1987). The length of homology required to initiate ectopic recombination between unlinked sequences, is at least between 1.9 Kb (Baker *et*

al. 1996).

Several studies have investigated the effect of insertions, deletions and sequence divergence on GC frequencies. Waldman and Liskay (1987) found that intra-chromosomal GC between *tk* heteroalleles sharing only 81% sequence homology was reduced at least 1000 fold versus HR between completely homologous *tk* alleles. They later concluded (Waldman and Liskay 1988) that percent sequence divergence was not responsible for the sharp decrease in recombination frequency, but that a minimum length of homology (the minimal efficient processing segment (MEPS); approximately 232 bp) was required for efficient recombination to occur. Once recombination has initiated within a MEPS it can propagate through regions of sequence heterology (Waldman and Liskay 1988). An analysis (Yang and Waldman 1997) of the rare products recovered from intra-chromosomal HR between heterologous sequences found a strong bias towards simple crossover events; no GC events were recovered among 27 isolates. Godwin and Liskay (1994) found that large insertions (1.5 kb) inhibited GC while increasing the number of crossovers recovered. Furthermore GC involving large insertions was more likely to be associated with a crossover than GC involving small insertions (Godwin and Liskay 1994). In addition, some data (Lukacsovich and Waldman 1999) suggests that sequence heterologies leading to the formation of G:G or CC mismatches in the presumptive hDNA inhibit HR more so than other types of mismatches.

Recently, restriction fragment length polymorphism (RFLP) analysis has provided a more detailed examination of GC in mammalian cells, including

information on conversion tract lengths. Analysis of HR between adjacent *tk* heteroalleles artificially introduced into mammalian cells revealed GC tract lengths of 35 to ≥358 bp (Liskay and Stachelek 1986, Waldman and Liskay 1988, Yang and Waldman 1997). These studies were limited by the placement of the restriction enzyme polymorphisms. That is, conversion tracts longer than the distance between the two farthest RFLP makers could not be detected. An inherent bias in these studies is the fact that the presence of the markers themselves may constrain the size of the GC tract. Increased sequence divergence has been observed to shorten GC tract length (Lukacsovich and Waldman 1999). Gene conversion tracts resulting from repair of an introduced DSB have also been analysed. By measuring the conversion of RFLP makers, Elliott *et al.* (1998) found bidirectional tracts of length generally <58 bp, but ranging as high as 511 bp while another study found that bidirectional tracts of an average length of 230 bp (Taghian and Nickoloff 1997).

The use of palindromes as markers in GC studies has provided strong evidence for the presence of hDNA as an intermediate in HR, as well as supplying additional data on GC tract lengths. The Holliday and Meselson-Radding models postulate that GC is the result of mismatch repair acting on hDNA while in the DSBR model GC is the result of gap repair of a DSB (Szostak *et al.* 1983) or the result of mismatch repair acting on hDNA formed by extensive 5' single strand degradation from the break site (Sun *et al.* 1991). RFLP markers have been used to demarcate GC tracts, but can not confirm the presence of hDNA as an intermediate in the conversion event. Palindromes are thought to escape mismatch repair, and have been used to detect putative hDNA. Bollag *et al.* (1992) utilized

palindromic markers to monitor GC events between *tk* heteroalleles in mouse L cells, and found sectored tk^+ colonies that provided strong evidence for the presence of hDNA as a recombination intermediate. Donoho *et al.* (1998) also used palindromic markers to find sectored colonies subsequent to DSB-simulated GC. In contrast to the short conversion tracts (<500 bp) found using RFLP markers, this study found conversion tracts of up to 3 kb were generated during intra-chromosomal HR (Donoho *et al.* 1998). Gene targeting studies using palindrome markers found that hDNA tracts of ≥ 1.9 Kb were generated (Li and Baker 2000a).

Despite the fact intra-chromosomal GC has been extensively studied, little data is available on inter-chromosomal GC frequencies or tract lengths. The higher-than-average intra-chromosomal recombination frequency seen in the Cμ recombination system (10^{-2} to 10^{-3} (Baker 1989, Baker and Read 1995) versus 10^{-4} to 10^{-5} in other systems) provides an ideal system in which to examine ectopic GC frequencies and mechanisms.

1.3 Murine Immunoglobulin Structure and Function

1.3.1 Structure of the Murine IgH Locus

The murine immune system produces eight different types of antibodies (Abs): IgM, IgD, IgG1, IgG2a, IgG2b, IgG3, IgE and IgA. Antibodies can be monomeric (for example, IgH is usually found as a monomer) or polymeric (the active form of IgM is usually found as a hexamer) Nezlin 1998). An antibody monomer consists of 2 identical heavy chains (these determine the Ab isotype) and

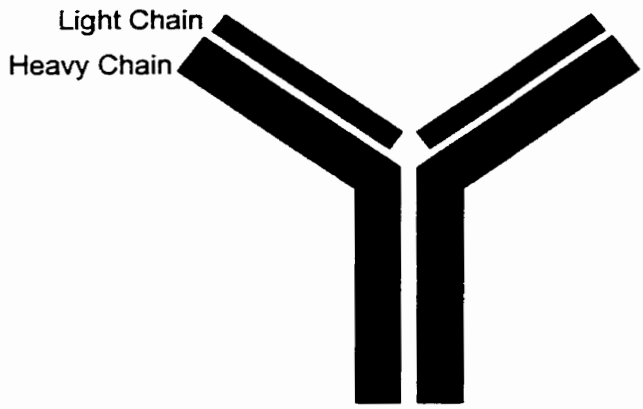
2 identical light chains arranged in a 'Y' shape (Fig. 1.4A) Monomeric antibodies can be joined to form polymeric antibodies (Fig. 1.4B). Antibody specificity is determined by V-D-J rearrangement in the antibody genes to produce the variable regions of both the heavy and light chains. The isotype of the Ab is determined by the choice of heavy chain constant region. Often different isotypes are required at different stages of the immune response or in different tissues. The mechanism of class switching involves a site-specific recombination event that maintains the same variable region while bringing a new IgH constant region into juxtaposition with the variable region (reviewed Zhang *et al.* 1995). Thus, despite the change in Ab isotype, the specificity of the Ab remains unchanged.

Shimizu *et al.* (1982) cloned the ~200 kb region of murine chromosome 12 that encompasses the constant region of the heavy chain genes (the IgH locus) (Fig. 1.5A). The organization of the IgH locus is as follows: 5'JH-Cμ-Cδ-Cγ3- Cγ1-Cγ2b-Cγ2a-Cε-Cα-3' (Shimizu *et al.* 1982). The IgM gene consists of four exons and spans approximately 2kb (Kawakami *et al.* 1980) (Figure 1.5B).

1.3.2 Murine IgM and IgA Structure and Function

Murine IgM is found in both a membrane-bound (monomeric) form and a secreted polymeric form (reviewed in Neslin 1998). Only IgG and polymeric IgM are capable of activating complement-mediated lysis (CML). Research by Arya *et al.* (1994) suggested that the Cμ3 domain of IgM was the key domain involved in CML, while later work by the same group (Chen *et al.* 1997) suggested that both the Cμ3 and Cμ4 domains were required. Certain key residues required for complement

A Monomeric antibody structure



B Pentameric antibody structure

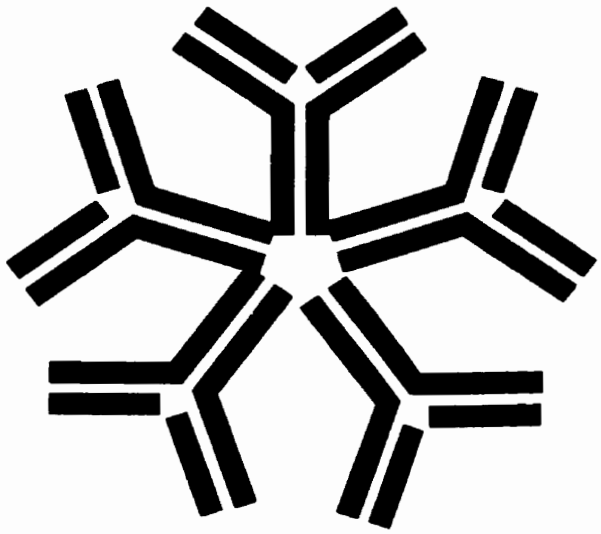
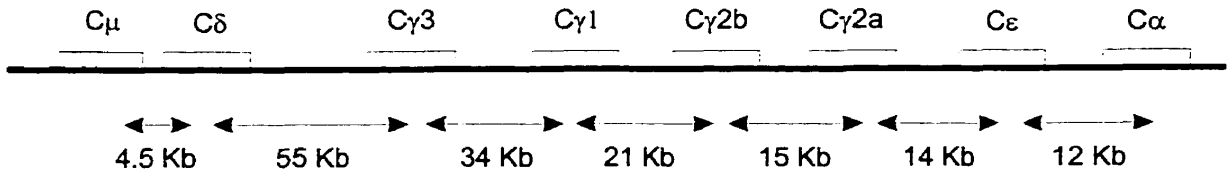
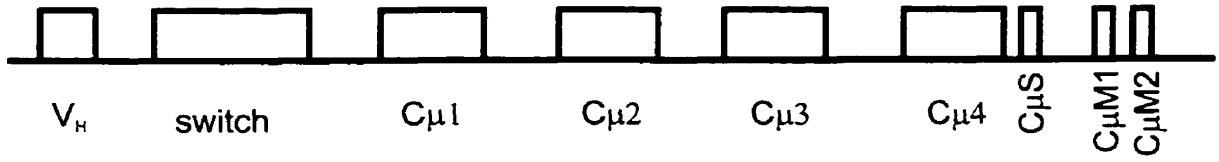


Figure 1.4 [A] Structure of a monomeric antibody showing the heavy and light chains. [B] A pentameric antibody.

A Structure of the murine IgH locus



B Structure of the murine Cμ locus



C Structure of the murine Cα locus

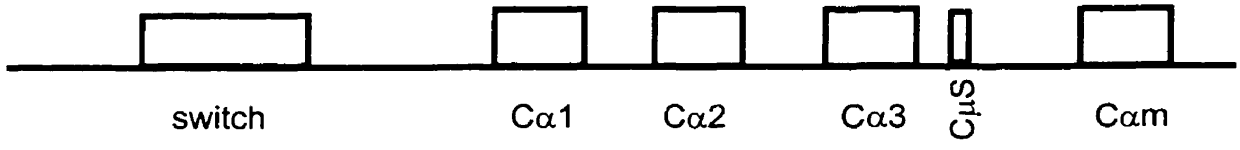


Figure 1.5 (A) Structure of the murine IgH locus showing approximate distances between the Ig constant regions. (B) A detailed structure of the murine IgM locus showing the variable region (V_H), switch region, the four exons (Cμ1, Cμ2, Cμ3 and Cμ4), the secreted exon (CμS), and the two membrane exons (CμM1 and CμM2). (C) A detailed structure of the murine IgA locus showing the switch region, three exons (Cα1, Cα2, Cα3), the secreted exon (CαS) and the membrane exon (Cαm).

activation have been identified in Cμ3 and Cμ4 (Shulman *et al.* 1986, Arya *et al.* 1994, Taylor *et al.*1994, Chen *et al.* 1997).

The murine Cα locus is located approximately 170 kb downstream of the Cμ locus (Shimizu *et al.* 1982). IgA is important in mucosal immunity, and is incapable of mediating CML. Cα consists of 3 exons as well as a single membrane exon ($C\alpha_m$; Fig. 1.5C) (Word *et al.* 1983). Until recently, it was thought that the only enhancer element in the IgH locus was the 5' enhancer located in the major intron 5' of the C_H region (Eμ). In 1990, Pettersson *et al.* found an enhancer element located 25 kb 3' of the rat Cα locus. Working separately, both Dariavach *et al.* (1991) and Lieberson *et al.* (1991) located the murine equivalent of this enhancer (3'αE or hs12) approximately 12 kb downstream of Cα. Both groups noted that the 3'αE enhancer appeared to be considerably weaker than Eμ. Matthias and Baltimore (1993) discovered another weak enhancer located ~1 Kb downstream of the $C\alpha_m$ exon which they called the Cα3' enhancer (Cα3'E). Michaelson *et al.* (1995) discovered a DNAseI hypersensitive site downstream of Cα, and determined that it was a third enhancer element located ~33kb downstream of Cα (the 3'α-hs4 enhancer). A transient transfection assay revealed that 3'α-hs4 and Eμ provided equal transcriptional upregulation in B cell lines, but 3'α-hs4 had only 25% the activity of Eμ in pre-B cell lines (Michaelson *et al.* 1995). The other DNAseI hypersensitive site (hs3) was also determined to be an enhancer (Madisen and Groudine 1994) , bringing the total number of known enhancers in the IgH locus to five: Eμ, Cα3'E, 3'αE (hs1,2), hs3, and 3'α-hs4.

Chauveau *et al.* (1998) used a transient expression assay to investigate the

interactions between Eμ and the four 3' enhancers. Individually, the 3' enhancer elements only weakly increased transcription, however a construct containing Eμ and all four 3' enhancers in the same orientation and structure as is found in the IgH locus produced 3-fold more transcriptional activity than was produced by a construct containing only Eμ. Constructs containing Eμ and combinations of the 3' enhancers in differing orientations/positions produced less transcriptional activity than the construct containing Eμ and the 3' enhancers in the same orientation as in the genome. In some cases these transient expression vectors had less activity than was seen using Eμ alone. Thus, each of the 3' enhancers alone is weak, but when acting together the four 3' enhancers along with Eμ can produce a synergistic increase in transcriptional activity (Chauveau *et al.* 1998).

2. MATERIALS AND METHODS

2.1 HYBRIDOMA LINES

The Cμ region homologous recombination system used in the present study is based on the detection of homologous recombination events between IgM gene constant regions (Cμ) that result in the restoration of normal trinitrophenyl (TNP) - specific IgM production. A cell producing wildtype IgM can be detected as plaque forming cells (PFCs) in a complement-dependent, TNP-specific plaque assay with an efficiency of 0.5-0.8 PFC/cell (Baker *et al.* 1988). The recipient for recombination is the haploid chromosomal IgM Cμ locus of murine hybridoma cell line igm482. The mutant igm482 was isolated from wildtype hybridoma cell line Sp6 which bears a single copy of the 2,4,6-trinitrophenyl (TNP)-specific chromosomal μ gene and synthesizes TNP-specific IgM (Köhler and Shulman 1980, Köhler *et al.* 1982). The Cμ region of igm482 contains a 2 base pair (bp) deletion in the third exon (Cμ3) (Baumann *et al.* 1985). This frameshift mutation results in the loss of the wildtype *Xmn*I restriction enzyme (RE) site and the creation of a novel *Tfi*I RE site as well as the production of a truncated heavy chain that lacks the Cμ4 domain (Köhler *et al* 1982, Baumann *et al.* 1985). It is this domain that is required to polymerize the IgM into a complement-activating form (Arya *et al.* 1994, Chen *et al.* 1997). The monomeric IgM produced by igm482 cannot effect the complement-dependent lysis of TNP-coated SRBCs in the TNP-specific plaque assay (efficiency $<10^{-7}$ PFC/cell; Baker *et al.* 1988). It has been previously demonstrated (Baker and Read 1992, Baker *et al.* 1996) that a wildtype Sp6 donor Cμ region inserted at various locations

in the igm482 genome can correct the 2 bp deletion in Cμ3 through homologous recombination. The recombinant cell can then mediate complement-dependent lysis in the plaque assay, and is detected as a PFC.

2.2 PLASMID USED FOR GENE TARGETING TO THE Cα LOCUS

The targeting vector used in the current study was an enhancer-trap vector designed to target an RFLP-marked donor Cμ locus into the Cα locus on murine chromosome 12, approximately 170 kb downstream of the Cμ locus. In order to ensure that the Cα sequence used in the vector was isogenic to the Cα locus of igm482, 5865 bp of Cα were amplified from igm482 genomic DNA using primers AB15676 (5'GCTGAGCTTGCTACACCAGACTGA) and AB15677 (5'GATCACAGACACGCTGACATTGGT). A 4377 bp *Hind*III/*Eco*RI fragment was inserted into the vector pUC19Not1,2 (engineered by site directed mutagenesis (SDM) to contain two *Not*I recognition sites flanking the polylinker region) to make the vector pUC19Not1,2α.

The Cμ-containing vector was constructed as follows. The backbone of this vector was the plasmid pSV2neoEn$^-$, which contained an enhancerless neo gene (*neo* confers kanamycin^r in bacteria and G418^r in mammalian cells). A 4.2 kb *Xba*I/*Spe*I Cμ fragment was inserted into pSV2neoEn$^-$ to create the vector pCμImutEn$^-$. The Cμ fragment had previously (Ng and Baker 1999) been engineered by SDM to contain five novel restriction enzyme sites (the Imut markers) in the introns of the Cμ. The Cμ fragment used in the targeting vector also differed

from the endogenous igm482 Cμ in that it contained a wildtype Cμ3 exon bearing the *Xmn*I diagnostic RE site. A 1.7 Kb *Stu*I/*Hin*dIII fragment containing *thymidine kinase* (*tk*) was inserted just upstream of the *Xba*I site to create the plasmid pCμImut*tk*En$^-$. This plasmid was then linearized at the *Sal*I site just 3' of the Cμ region and inserted into the *Xho*I site of pUC19not1,2α (described above). The resulting plasmid, pαCμImut*tk*En$^-$, was linearized with *Not*I to create a 15.4 kb replacement vector that contained 3595 bp and 782 bp (respectively) of 5' and 3' flanking homology to the Cα locus. All restriction enzymes were purchased from New England Biolabs (Mississauga, ON) and used according to manufacturers specifications.

2.3 Gene Targeting and Transformant Isolation

55 μg of linear targeting vector pαCμImut*tk*En$^-$ was phenol:chloroform cleaned, ethanol precipitated, then resuspended in 50 μl of phosphate-buffered saline (PBS). The vector DNA was mixed with $2x10^7$ igm482 hybridoma cells then electroporated twice (Gene Pulser, BioRad, Mississauga ON) using a 700 volt 25 μF pulse in a 0.4 cm gapped cuvette. The cells were then placed on ice for 10 minutes. Next, 1 mL of Dulbecco's modified H21 media (H21) supplemented with 0.5 mM 2-mercaptoethanol and containing 13% bovine calf serum (Hyclone, Logan Utah) was added to the cells. The cells were incubated in a humid 37°C CO_2 incubator for 20 minutes prior to resuspension in 90 mL H21 media. The resuspended cells were then plated at 0.75 mL/well (~$1.7x10^5$ cells/well) in 24-well

tissue culture plates (114 wells in total). Cells were grown at 37°C for 48 hrs prior to adding G418 to a final concentration of 600 µg/mL. Cells were maintained in 600 µg/ml G418 media at 37°C for ~14 days and genomic DNA was prepared using the method of Gross-Bellard *et al.* (1973) from any wells containing $G418^r$ colonies. In brief, the wells were resuspended and 0.75 ml of resuspended culture was washed twice in phosphate-buffered saline (PBS; 8 g/L NaCl, 0.2 g/L KCl, 1.14 g/L Na_2HPO_4, 0.2 g/L KH_2PO_4), then incubated overnight with 600µl lysis buffer (Applied Biosystems, Foster City, CA) and 35 µl proteinase K (Roche, Laval PQ). The DNA was then ethanol precipitated and transferred to 50 µl T.E. (Tris EDTA; 10 mM tris, 1mM EDTA pH 8.0).

2.4 DNA ANALYSIS OF $G418^R$ TRANSFORMANTS

The genomic DNA prepared from the $G418^R$ transformants was screened for gene targeting events using standard Southern blot analysis (Sambrook *et al.* 1989). In brief, the genomic DNA was digested with a diagnostic RE and electrophoresed through an ~0.9% agarose gel (restriction enzymes used were purchased from New England Biolabs and used according to manufacturers specifications). The DNA was then blotted onto nitrocellulose and screened by hybridization to ^{32}P-labelled probe F(an 870 bp *Xba*I/*Bam*HI fragment from pCµEn$^-$ that is specific to the Cµ locus). Targeted hybridomas were identified as containing a 2727 bp *Xmn*I fragment on the Southern blot. Putative targeted lines were cloned out at 0.1 cell/well to obtain subclones started from a single cell. Genomic DNA was

prepared from these subclones, and used to prepare confirmation Southern blots. The presence of a 6556 bp *Dra*III fragment with probe F, a 14511 bp *Dra*III fragment with probe G (a 762 bp *Pvu*II fragment from pSV2neo that is specific to *neo*) and a 9256 bp fragment with probe G were used to confirm that a transformed line contained a targeting event. Probes were labelled using the Random Primed Labelling Kit (Roche, Laval PQ).

2.5 MEASURING RECOMBINATION FREQUENCIES AND PFC ISOLATION

In the transformed lines produced in the current study, the recipient for recombination is the endogenous igm482 Cμ region. The vector-borne Cμ is the donor of genetic information, and HR events that restore the wildtype Cμ3 *Xmn*I site can be detected using the TNP-specific plaque assay (Baker *et al.* 1988). In brief, hybridomas (up to $2x10^7$) were washed once in 1X Hanks solution(8 g/L NaCl, 0.2 g/L $CaCl_2 \cdot 2H_20$, 0.2 g/L $MgSo_4 \cdot 7H_2O$, 0.4 g/L KCl, 0.1 g/L KH_2Po_4), then resuspended in 100 μL 1XHanks solution. The resuspended cells were mixed with 45°C 0.6% low gel temperature top agarose (ICN, Aurora OH) made with 1X Hanks solution, 700 μL 1XHanks, 100 μL TNP-coupled sheep red blood cells (sRBC) and poured over a bottom layer plate (0.5% electrophoresis-grade agarose in 1XHanks). The plates were incubated in a 37°C 7% CO_2 humid incubator at of 1-1.5 hours before flooding the plate with 2 mL of a 1:10 PBS dilution of guinea pig complement (Behring Diagnostics, Kanata Ontario) resuspended according to manufacturers instructions). After an additional 30 minute incubation in the 37°C 7% CO_2 humid

incubator, the plates were removed and the number of plaques was counted to determine the rate of recombination. The monomeric IgM produced by igm482 cannot activate the complement-dependant lysis of TNP-coupled sRBC, and is not detected on the plaque assay plates. Recombinants were detected as TNP-specific PFC and were recovered from the plaque assay plates using a sterile microcapillary pipette and transferred to 600 µg/ml G418 H21 media in individual wells of a 96 well plate. These cultures were re-tested for the presence of TNP-specific PFCs using either Cunningham chambers or spot tests (see below). PFCs were isolated from the background of non-PFC by either replating the well in another plaque assay or by plating the culture at a density of ~0.1 PFC/well, and retesting each new well for the presence of PFCs. This process was continued until a 'pure' PFC line was isolated from a 0.1cell/well plating.

2.5.1 CUNNINGHAM CHAMBERS

Cunningham chambers (Cunningham and Szenberg 1968) were used to confirm that plaques picked from plaque assay plates contained viable PFCs. Eighty µL of hybridoma cells were mixed with 10 µL TNP-sRBC and 5 µL undiluted complement and placed into a Cunningham chamber (a thin flat chamber created by attaching two glass slides together with double-sided tape). The chamber was sealed with wax, and plaques were counted after 30 minutes incubation at 37°C.

2.5.2 Spot Tests

Spot tests (Köhler and Shulman 1980) were also used to confirm that hybridoma cells picked from plaque assay plates contained viable PFCs. Two µL of supernatant was spotted onto a 5mL agarose plate (containing 200 uL TNP-sRBC and 200 µL complement per plate). The plate was scored for lysis after 30 minutes incubation at 37°C in a 7% CO_2 humid incubator.

2.6 Characterizing the Cµ Structure of PFCs

In order to characterize the PFC recovered in the current study, PCR was used to amplify both the endogenous (recipient) Cµ locus and the vector-borne (donor) Cµ locus. These PCR products were then digested with diagnostic RE to determine the pattern of Imut/wildtype RFLP markers.

2.6.1 PCR Amplification of the Endogenous Cµ Locus

Primers AB9703 (5'CTACTTGAGAAGCCAGGATCTAGG) and AB9438 (5'GTACCATCAGACTGCACTGTTCCA) were used to amplify a 4621 bp fragment from the endogenous Cµ locus. The PCR reaction used to amplify the endogenous Cµ contained (per 50 µl volume): 1µL of genomic DNA (0.5-1 µg/µL), 5 µL 2mM dNTPs, 1.5µL 6 µM primer AB9703, 1.5µL 6 µM primer AB9438, 2 µL 50mM $MgCl_2$, 5 µL PCR buffer (GibcoBRL, Gaithersburg, MD) and 33.5 µL dH_2O. These reagents were mixed, heated to 94°C for 2 minutes and held at 80°C while 0.5 µl of Taq DNA polymerase (GibcoBRL, Gaithersburg MD) was added. This was followed by 45

cycles of PCR using a GeneAmp PCR System 2400 (Perkin Elmer (Norwalk CT). Each cycle consisted of: 94°C for 15 seconds (s) (denaturing), 65°C for 30s (annealing), 70°C 2.5 minutes (extension). A final extension for 10 minutes at 70°C was included after the final cycle.

Two to 6 µl of PCR product was digested with diagnostic restriction enzymes (to determine if any gene conversion had taken place), electrophoresed through 0.8% agarose gel, and visualized by ethidium bromide staining. Restriction enzymes were purchased from New England Biolabs (Mississauga, ON) and used according to manufacturers specifications.

2.6.2 Amplifying the Donor Cµ Locus

The Expand™ long template PCR system (Roche Molecular Biochemicals, Indianapolis IN) was used with primers AB16073 (5'CCTTGTGGTCAGTGTTCATCTGCT) and Cµαoverlap (5'CATCTCGACTAGAGGCATCTCTCC) to amplify a 6326 bp fragment from the donor (vector-borne) Cµ locus. 25 µL master mix I (8.75 µL2mM dNTPs, 2 µL 6 mM primer AB16073, 2 µL 6 mM primer Cµαoverlap, 11.25 µL dH_2O) was mixed with 25µl master mix II (5 µL buffer 1, 1µL enzyme mix, 19 µL dH_2O), spun down and immediately placed into a thermocycler (GeneAmp PCR System 2400; Perkin Elmer, Norwalk CT). The sample was heated to 94°C for 2 minutes, followed by ten cycles of: 94°C 15s (denaturing), 65°C30s (annealing), 68°C 8 minutes (extension). This was followed by another 25 cycles consisting of: 94°C for 15s (denaturing), 65°C 30s (annealing), 68°C 8: minutes plus 20s per cycle (extension). A final

extension of 68°C for 10 minutes was included after the final cycle.

Two to 6 µl of PCR product was digested with diagnostic restriction enzymes (to determine if any gene conversion had taken place), electrophoresed through 0.8% agarose gel, and visualized by ethidium bromide staining. Restriction enzymes were purchased from New England Biolabs (Mississauga, ON) and used according to manufacturers specifications.

3. OBJECTIVES

The first phase of the current study involved using an enhancer-trap vector to target a genetically-marked Cμ region into the endogenous Cα locus of murine hybridoma line igm482. The goal of the gene targeting was twofold. First, it established the efficacy of gene targeting at the Cα locus using an enhancer-trap vector. This data was then compared to that of a previous study (Ng and Baker 1998) which determined that the use of an enhancerless vector increased targeting efficiency at the Cμ locus by approximately fifteen-fold. Second, targeted lines were used to provide a defined site of donor Cμ integration for the recombination assay (see below).

The second objective of the present study was to determine the frequencies of homologous recombination events between the donor (vector-borne) Cμ region and the recipient (endogenous) Cμ locus. Such events are detectable using the TNP-specific plaque assay described previously.

The final objective of the present study was to recover homologous recombination events and to characterize both the donor and recipient Cμ regions of these recombinant cell lines. Gene conversion involving the genetic markers contained within the endogenous Cμ locus was examined in order to determine the length of any conversion tracts. The genetic structure of the donor (vector-borne) Cμ region was examined to determine whether or not a reciprocal exchange had occurred.

3. MAPPING THE IGα REGION

3.1 INTRODUCTION

The first phase of this project involved targeting to the Cα region of murine chromosome 12. Transformed lines were screened for gene targeting events using Southern blotting, which required a restriction map of the Cα locus. A GenBank search uncovered several partial Igα sequences, but a complete nucleotide sequence of the region was unavailable. In order to construct a sequence map of the entire Cα region, the various GenBank entries containing partial Cα sequence were analyzed for regions of overlap. The resulting sequence was used to produce a restriction enzyme map of Cα (Fig. 3.1). Based on this restriction map, the enzymes *Xmn*I, *Dra*III and *Acc*I were chosen for diagnostic digests to screen for gene targeting events by Southern blotting. The presence of the recognition sites for these enzymes in the Cα region of hybridoma igm482 was verified using Southern blots probed with the Igα specific probe αXR. Based on this data, an adjusted map of Cα (as found in igm482) is presented.

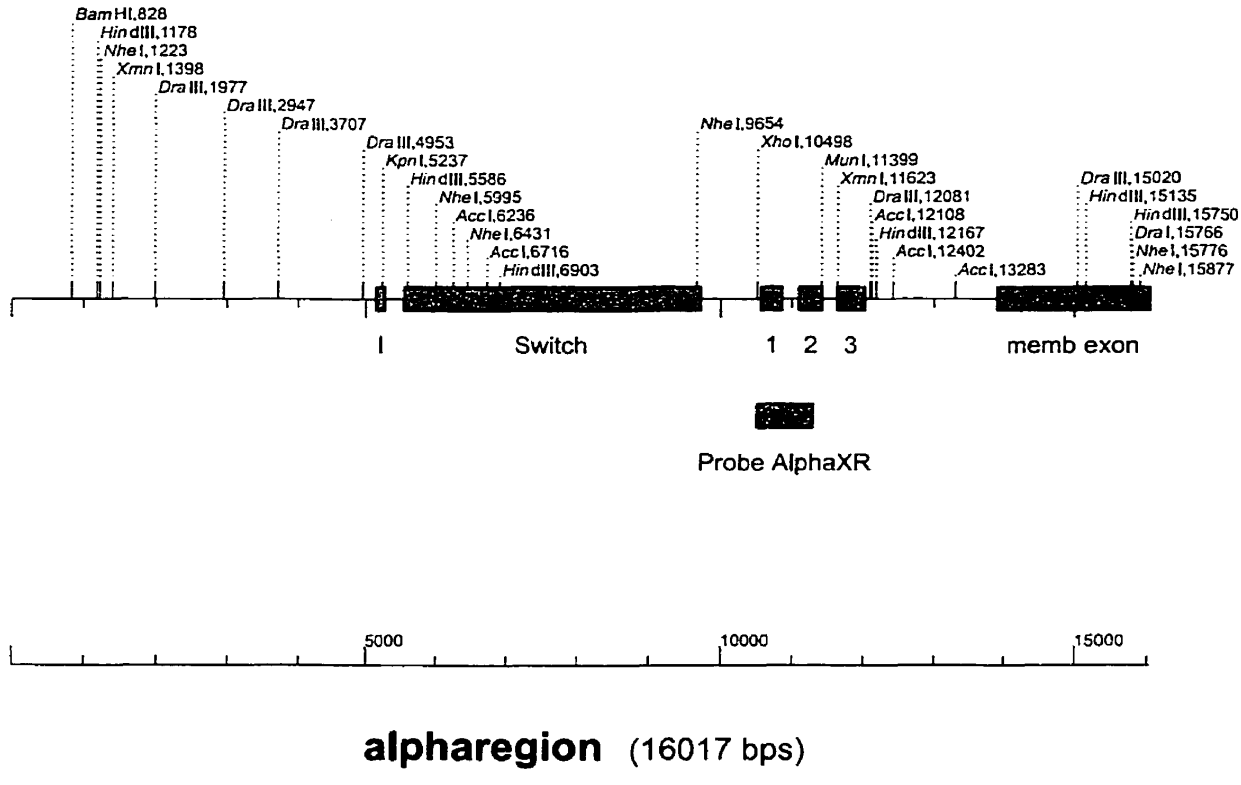


Figure 3.1 Expected structure of the Cα region of murine chromosome 12. The three exons (1,2 and 3) are indicated as well as the membrane exon (memb exon), switch region(switch) and promotor (I). The binding site of probe *αXR* (bp 10498 - bp 11280) is indicated. This map was constructed by analyzing Cα genbank sequences for regions of overlap, and was prepared using the computer program Clone Manager for Windows Version 4.1 (Copyright 1995-1996 Scientific and Educational Software). The following genbank entries were used (numbers are as indicated on the map below): U08933 (bp 1 - bp 5667); D11468 (bp 4911- bp 12166); X62548 (bp 6321 - bp 7465); J00474 (bp 7745 - bp 9180); K02101 (bp 10888 - 11283); J00475 (bp 10490 - bp 12166); U08816 (bp 12167 - 14056); K00691 (bp 13886 -16017). Please note that according to the authors of genbank accession U08816, this sequence adjoins that of genbank accession J00475. No region of overlap was available.

3.2 RESULTS

In order to confirm the presence of the *Xmn*I, *Dra*III and *Acc*I restriction enzyme sites in the Cα region of igm482, Southern blots were prepared using genomic DNA was digested with these enzymes both alone and in combination with other enzymes, and probed with the Cα-derived probe αXR (Fig. 3.2). This 782 bp probe spans the *Xho*I site at bp 10498 to the *Eco*RI site at bp 11280 (according to the numbering system of Fig.3.1), and provided a convenient method to confirm the presence of the RE sites in the Cα locus of igm482.

The *Acc*I and *Xmn*I bands are the expected sizes as predicted from the assembled Cα sequence map (5.4 Kb and 10.2 Kb respectively). However the *Dra*III, band on the blot (10.3 Kb) is larger than the predicted size of 7.1 Kb. Double digests involving the upstream *Dra*III site (*Eco*RI/*Dra*III and *Xmn*I/*Dra*III) yielded the predicted size bands (6.3 Kb and 6.7 Kb respectively). Double digests *Nhe*I/*Dra*III and *Xho*I/*Dra*III involving the downstream *Dra*III site yielded bands that were approximately 3 Kb larger than predicted from the putative map of the Cα region.

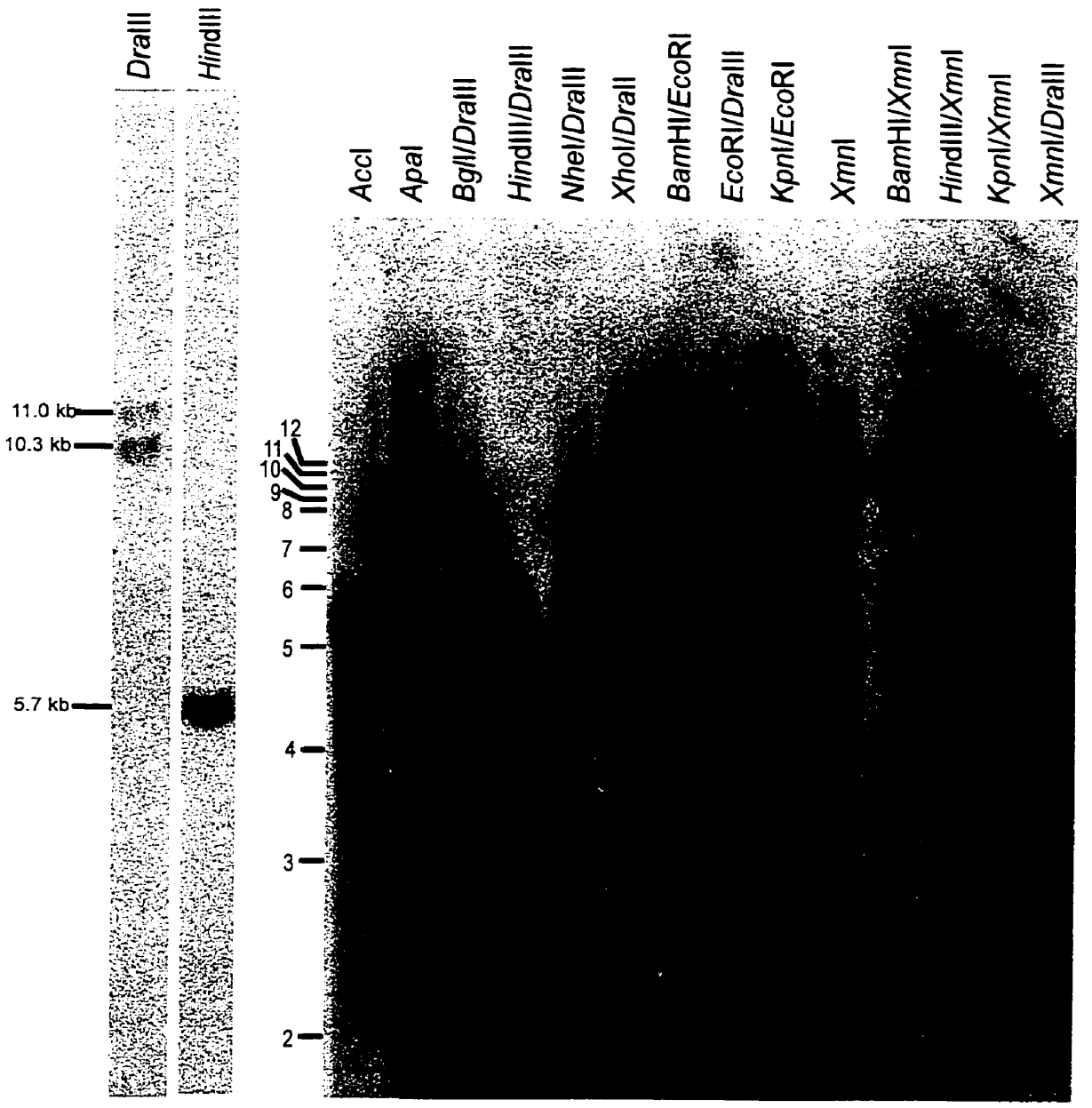


Figure 3.2 Southern Blot analysis of the Igα gene structure in murine hybridoma igm482. Hybridoma genomic DNA was digested with various restriction enzymes, as indicated above each lane. The sizes of the molecular weight marker bands are indicated to the left of the blot. The blot was probed with ^{32}P-labeled probe αXR. Expected band sizes (based on the presumptive genbank map in Fig. 3.1)are as follows: *Dra*III 7128 bp; *Hin*dIII 5624 bp; *Acc*I 5392 bp; *Apa*I 7541 bp; *Bgl*I/*Dra*III 2173 bp; *Hin*dIII/*Dra*III 5178 bp; *Nhe*I/*Dra*III 2427 bp; *Xho*I/*Dra*III 1583 bp; *Bam*HI/*Eco*RI 7069 bp; *Eco*RI/*Dra*III 6327 bp; *Kpn*I/*Eco*RI 6043 bp; *Xmn*I 10225 bp; *Bam*HI/*Xmn*I 7412 bp; *Hin*dIII/*Xmn*I 4720 bp; *Kpn*I/*Xmn*I 6387 bp; *Xmn*I/*Dra*III 6670 bp.

3.3 DISCUSSION

The goal of this phase of the current study was to verify the presence of various RE recognition sites in the Cα region of hybridoma igm482. Although a sequence map of the Cα region had been constructed from overlapping genbank entries, it is possible that sequence differences exist between this map and the Cα locus of igm482. These RE sites were to be used to screen the $G418^R$ transformants produced by gene targeting for the presence of targeted lines (chapter 4). Three RE sites (*Xmn*I, *Dra*III and *Acc*I) were chosen for their diagnostic value in distinguishing targeted lines from random transformants. In order to confirm the presence of these sites, both single and double digests of genomic DNA were used to prepare Southern blots, which were then probed with radio-labeled probe αXR. Several of the enzymes used in the double digests (*Eco*RI, *Xho*I, *Nhe*I, *Hin*dIII) had been previously used in vector cloning steps, and were thus known to be present in the Cα region of hybridoma line igm482.

3.3.1 PROBE αXR IS SPECIFIC TO THE Cα LOCUS

A Southern blot of digested genomic DNA was used to analyze the Igα gene structure of hybridoma line igm482 (Fig. 3.2). The blot was probed with radio-labeled probe αXR (the 782 bp *Xho*I/*Eco*RI fragment derived from the Cα PCR product). With the exception of a few lanes which show incomplete digestion (see discussion below), there is only a single band in each lane. This indicates that the αXR probe is specific to the Cα locus, and does not hybridize to other sequences.

3.3.2 THE *AccI*, *ApaI*, AND *XmnI* STES ARE INTACT

The expected sizes of the *Acc*I *Apa*I, and *Xmn*I fragments (respectively) are 5392 bp, 7541 bp and 10225 bp. Southern blot analysis (Fig. 3-2) of these sites reveals the expected band size in each case (5.3 Kb, 7.7 Kb, and 10.5 Kb respectively). Thus each of the three recognition sites is present in the Cα region of murine hybridoma line igm482.

3.3.3 THE 5' *DraIII* SITE IS INTACT

The expected band size for a *Dra*III digest was 7128 bp, yet Southern blot analysis (Fig. 3.2) revealed two bands at ~10.3 Kb and ~11.0 Kb. The faint 11 Kb band is likely the result incomplete digestion of genomic DNA, but may also be due to sequence polymorphisms between different Cα alleles present in the hybridoma line. As well, the presence of the larger than expected 10.3 Kb *Dra*III fragment suggested that one of the two *Dra*III sites expected to produce the 7128 bp band was missing. To determine which (if either) of these sites was missing, igm482 genomic DNA was cut with pairs of enzymes to confirm the presence or absence of the 5' and 3' *Dra*III sites (located at bp 4953 and bp 12081 respectively according to the numbering of Fig. 3.1). The lack of additional bands in the double digests (Fig. 3.2) confirmed that the 11 kb band was due to incomplete digestion rather than sequence polymorphisms between Cα alleles. Two double digests were used to confirm the presence of the 5'*Dra*III site. The *Eco*RI site was used to construct the targeting vector, and so was known to be intact. The *Xmn*I site was determined to be intact by Southern blot (Fig.3.2). If the 5' *Dra*III site was intact then the

*Dra*III/*Eco*RI and *Dra*III/*Xmn*I digests were expected to give 6.3 Kb and 6.7 Kb fragments respectively. Both double digests produced the expected bands (Fig. 3.2), indicating that the 5' *Dra*III map site was intact. The fainter 7.8 kb band seen in *Dra*III/*Xmn*I digest is probably due to partial digest of the genomic DNA, and is likely the 7919 bp fragment spanning the *Dra*III site at bp 3707 to the *Xmn*I site at bp 11623.

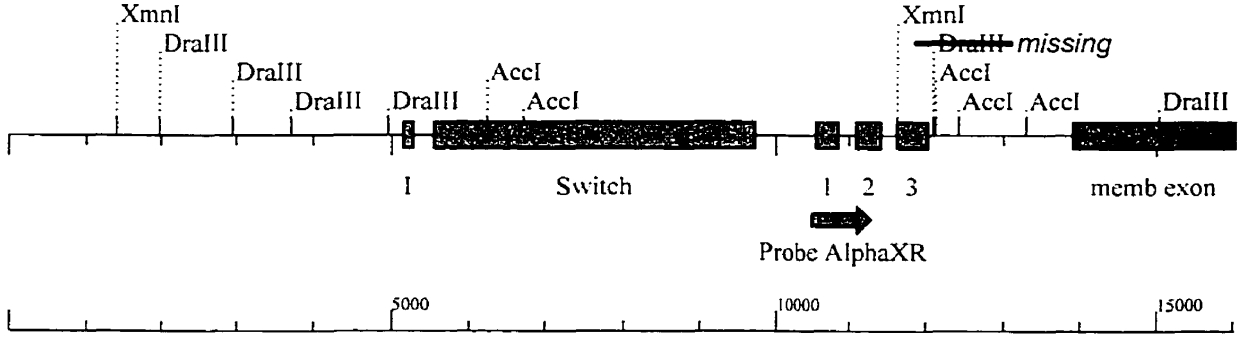
3.3.4 THE 3' *Dra*III SITE AT BP 12081 IS MISSING

The presence/absence of the 3' *Dra*III site was verified by digesting igm482 genomic DNA with *Dra*III and one of *Nhe*I, *Xho*I, or *Hin*dIII (the presence of the *Bgl*I recognition site in the Cα locus of igm482 was not established). The *Nhe*I and *Xho*I sites were used to construct the targeting vector, and so were known to be intact. The *Dra*II/*Xho*I digest was expected to produce a 1583 bp fragment, yet the blot reveals an ~4.8 Kb and ~10 Kb fragment (Fig 3.2). The 4.8 Kb fragment suggests that the 3'*Dra*III map site (located at bp 12081) is missing; the fragment seen in the blot would then extend from the *Xho*I site at bp 10498 to the *Dra*III site at bp 15020 (a total of 4522 bp in length). The fainter 10kb band likely represents incomplete digestion of the genomic DNA, and is presumably the 10067 bp fragment spanning the *Dra*III site at bp 4953 to the *Dra*III site at bp 15020. The *Dra*III/*Nhe*I digest was expected to yield a 2427 bp fragment, yet the blot reveals an ~5.2 kb band (Fig. 3.2), as would be expected if the *Dra*III site at bp 12081 was missing. This 5.2 Kb band is likely the 5366 bp fragment that spans the *Nhe*I site at bp 9654 to the *Dra*III site at bp 15020. As there is a *Hin*dIII site located only 86 bp away from the *Dra*III

map site at bp 12081, the *Hin*dIII/DraIII data is unusable. The data from the *Dra*II/*Xho*I and *Dra*III/*Nhe*I double digests suggests that the *Dra*III site located 3' of the α*XR* probe (at bp 12081) is missing in hybridoma line igm482.

3.3.5 A MODIFIED MAP OF THE Cα REGION

Based on the Southern blot analysis of the Cα locus of igm482, the *Xmn*I and *Acc*I RE recognition sites are intact. The 3' *Dra*III site (located at bp 4953) is intact however the 5' *Dra*III site (located at bp 12081) is missing in this hybridoma line. Based on this data, a modified map of the igm482 Igα locus is presented (Fig. 3.3).



igm482 Alpha (16017 bps)

Figure 3.3. Actual structure of the C" region of murine chromosome 12 in hybridoma line igm482. The three Ca exons (labeled 1,2 and 3), the Ca membrane exon (memb exon), switch region (switch) and promotor (I) are shown as grey boxes. The binding site of probe "XR (bp 10498 - bp 11280) is indicated by a grey arrow below the map. A scale showing bp intervals is located below the map. Only those restriction enzyme sites used to screen for targeted lines (XmnI, DraIII and AccI (see chapter 4)) are shown on the map. Note: This map was prepared using the computer program Clone Manager for Windows version 4.1 (Copyright 1995-1996 Scientific and Educational Software).

4. TARGETING TO Cα

4.1 INTRODUCTION

Gene targeting (GT) frequencies in mammalian cells generally range from 10^{-2} to 10^{-5} targeted cells per transformed cell (reviewed by Bollag *et al.*1989). The targeting frequency to the Cμ locus of murine hybridoma cells is approximately 1-$2x10^{-3}$ targeted cells per transformed cell (Ng and Baker 1998, Ng and Baker 1999). As the incidence of GT is so low in mammalian cells compared to yeast cells, several researchers have investigated methods of enriching for GT events in the population of transformed cells. For example, several groups (Bautista and Shulman 1993, Ng and Baker 1998) examined gene targeting at the Cμ locus of hybridoma line igm482 and found that the use of an enhancer-trap (En^-) vector resulted in a 15-fold enrichment of targeted lines relative to the number of transformed lines. One of the objectives of the current study was to determine whether a similar strategy could be used to enrich for gene targeting events at the murine Cα locus. The IgH locus contains five known enhancers, four of which are located near Cα (Fig. 4.1). The fifth (Eμ) is upstream of the C_H region. More is know about Eμ than the 3' enhancers, although individually each is known to be weaker than Eμ (Chauveau *et al.* 1998). The goal of the current study was to utilize an En^- vector to insert a donor copy of Cμ (along with the marker genes *neo* and *tk*) into the Cα locus of hybridoma line igm482 (Fig. 4.2). The efficiency of targeting to Cα with the enhancerless vector was compared to previous work in which an En^- vector was used to target the Cμ locus (Ng and Baker 1998). To date, all studies

that have examined gene targeting to the IgH locus of murine hybridoma lines have involved targeting to the Cμ (or Cδ) locus; no attempts had been made to target Cα with an enhancer-trap vector. The current study demonstrates that such targeting is possible.

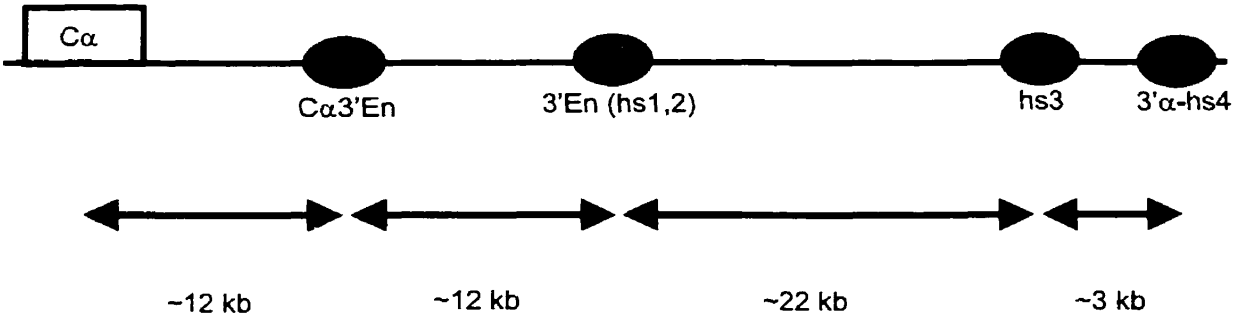


Figure 4.1 The alpha region enhancers (En) of the murine IgH locus (adapted from Michaelson *et al.* 1995). The hollow box represents the alpha constant region exons (labeled Cα) while the filled ovals represent the various enhancer elements located 3' of the IgAlpha constant region of murine chromosome 12 (Cα3'En, 3'En (hs1,2), hs3, 3'α-hs4). The approximate distances between the structural elements of the Igα region are indicated below the map.

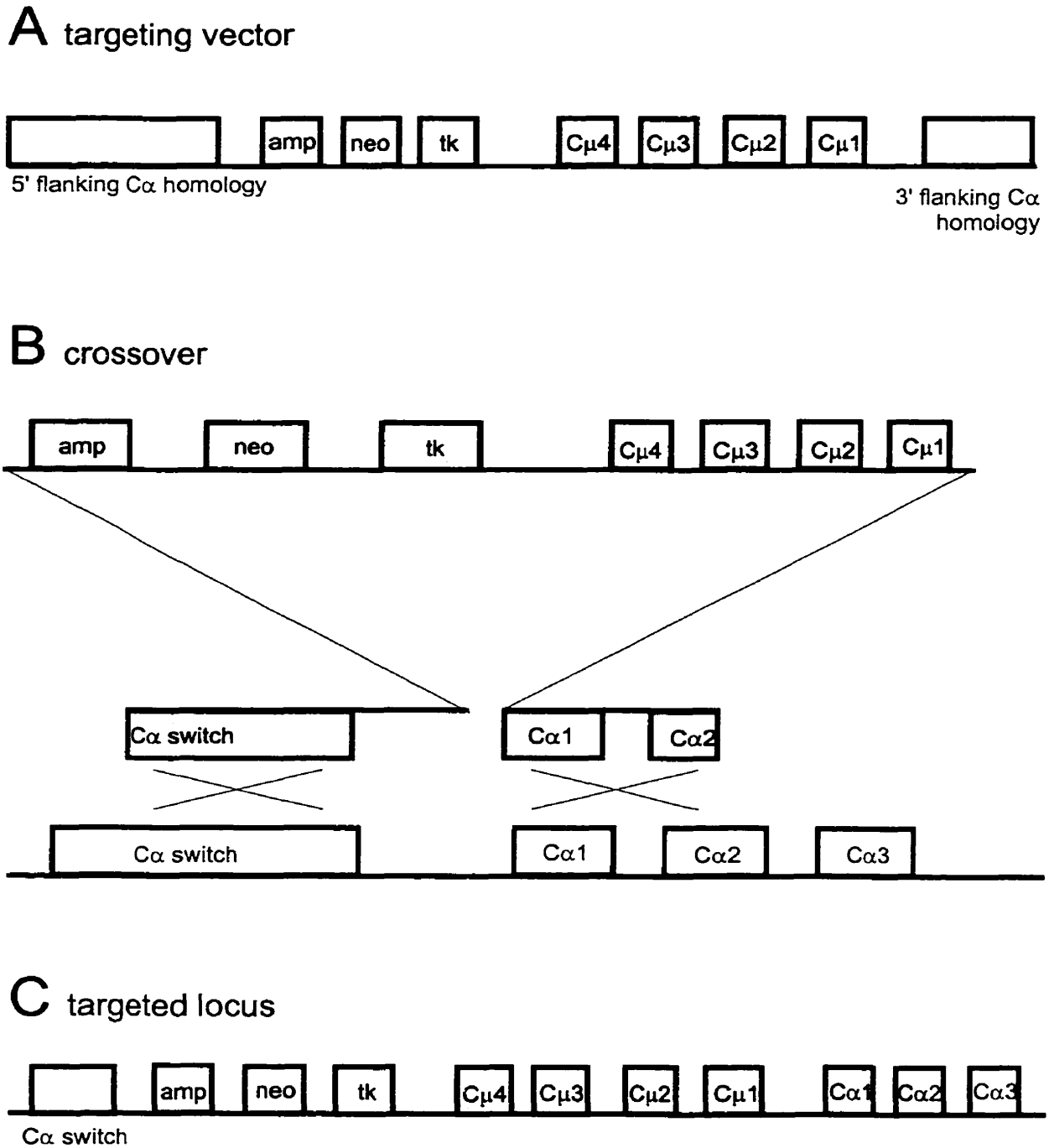


Figure 4.2 Gene targeting to the chromosomal Cα locus. The targeting vector (A) contains 3595 bp of homology to Cα on the 5' flank, and 782 bp homology on the 3' flank. This ends-out (replacement) vector recombines with the homologous chromosomal Cα locus (B) to generate a targeted Cα locus that contains a copy of Cμ (C). Vector sequences are shown in grey; chromosomal sequences are white.

4.2 RESULTS

4.2.1 CALCULATING THE TOTAL NUMBER OF TRANSFORMANTS

During the targeting, approximately 50 µg of vector DNA was electroporated into $2x10^7$ igm482 hybridoma cells. The cells were then distributed into 114 wells and grown in 600 µg/mL G418. Of the 114 wells, 76 were growth-positive and 38 growth-negative. From this information, the average number of $G418^R$ cells per pool can be calculated using the Poisson distribution:

$$P(s) = \frac{\chi^s}{s!}e^{-x}$$

$$P(0) = 38/114 = (\chi^0/0!)e^{-x}$$

$$\therefore \chi = -\ln(38/114) = 1.099$$

From this X value, the probability of having 1 cell in the well (or (P(1)) is 0.37, P(2) is 0.20, P(3) is 0.07 and P(4) is 0.02. Thus most wells contained one, two or three $G418^R$ cells and the total number of transformed lines was:

$$\sum P(1)76 + 2[P(2)]76 + 3[P(3)]76 + ..+n[P(n)]76$$

$$\approx \sum 1(0.37)114 + 2(0.20)114 + 3(0.07)114 + 4(0.02)114 + 5(0.004)114$$

$$\approx 123 \text{ transformed lines}$$

4.2.2 SCREENING POOLS OF TRANSFORMANTS FOR GENE TARGETING EVENTS

As the detection of one targeted line in a background of up to nine other $G418^R$ lines was within the resolution of a Southern blot, genomic DNA was prepared from each of the pools of cells and screened for targeted lines using *Xmn*I digestion. A targeted hybridoma line was expected to generate a 2727 bp fragment

when hybridized to the Cμ-specific Probe F (Fig. 4.3). All 76 growth-positive pools of cells were screened in this way (Figs. 4.4, 4.5, 4.6) and eight putative targeted lines (R1B4, R1B5, R1C1, R1D2, R2C1, R3A4, R3C5, and R3D4) were chosen for further analysis based on the presence of an ~2.7 kb *Xmn*I fragment on the Southern blots.

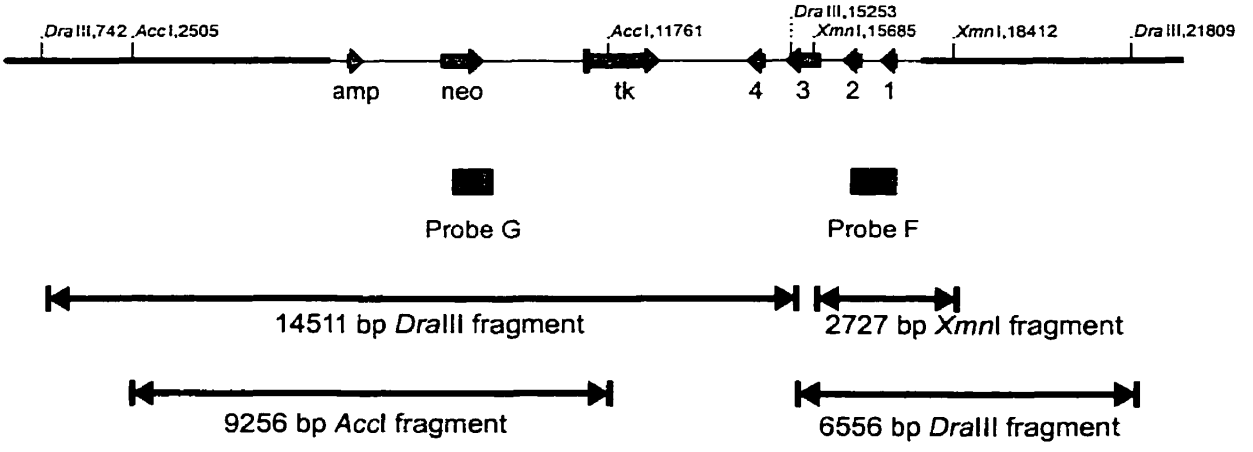


Figure 4.3 Structure of the targeted Cα locus. Non-vector sequences are shown with a thick line and vector-bourne sequences with a thin line. The marker genes *amp*^{R}, *neo*^{R} and *tk* are shown as grey arrows as are the four Cμ exons (labeled 1, 2, 3, 4). Transformed G418^R lines were screened by Southern blotting followed by hybridization with 32P-labeled probe F or probe G. The binding sites of probes F and G are indicated as grey rectangles beneath the map and the *Xmn*I, *Dra*III and *Acc*I diagnostic fragments expected in correctly targeted lines are shown below the map.

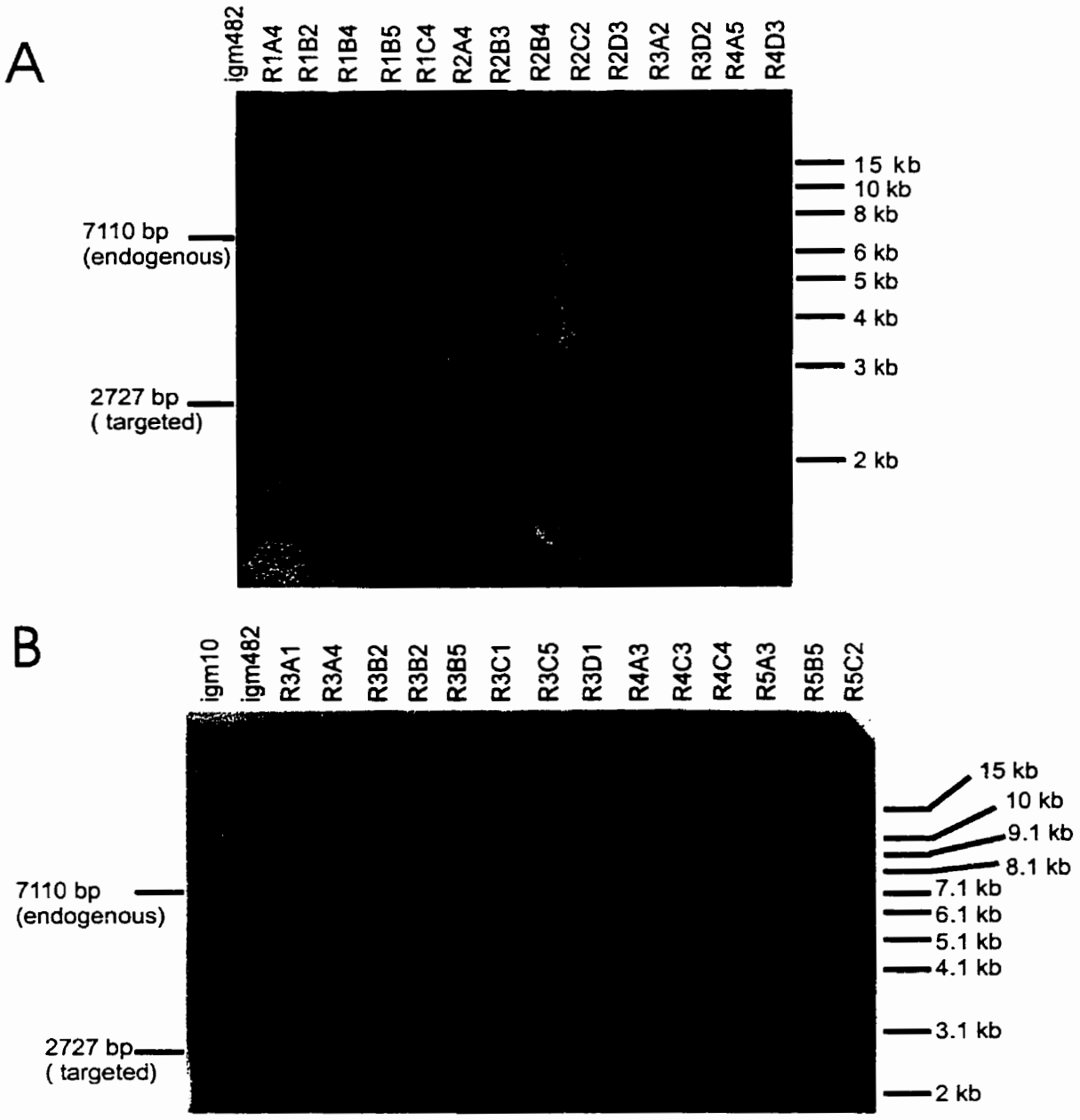


Figure 4.4 Analysis of the Cα gene stucture in G418^R lines recovered from a gene targeting experiment. Hybridoma genomic DNA was digested with *Xmn*I, electrophoresed through 0.9% agarose and blotted onto nitrocellulose and probed using ^{32}P-labeled probe F (specific to Cμ). In each figure, the sizes of the 7110 bp endogenous Cμ band and the 2727 bp fragment expected from a correctly targeted line are indicated to the left, while the molecular weight markers are indicated to the right. In blot A, hybridoma lines R1B4 and R1B5 have the 2.7kb band indicative of a targeting event, while in blot B lines R3A4 and R3C5 are putative targeted lines.

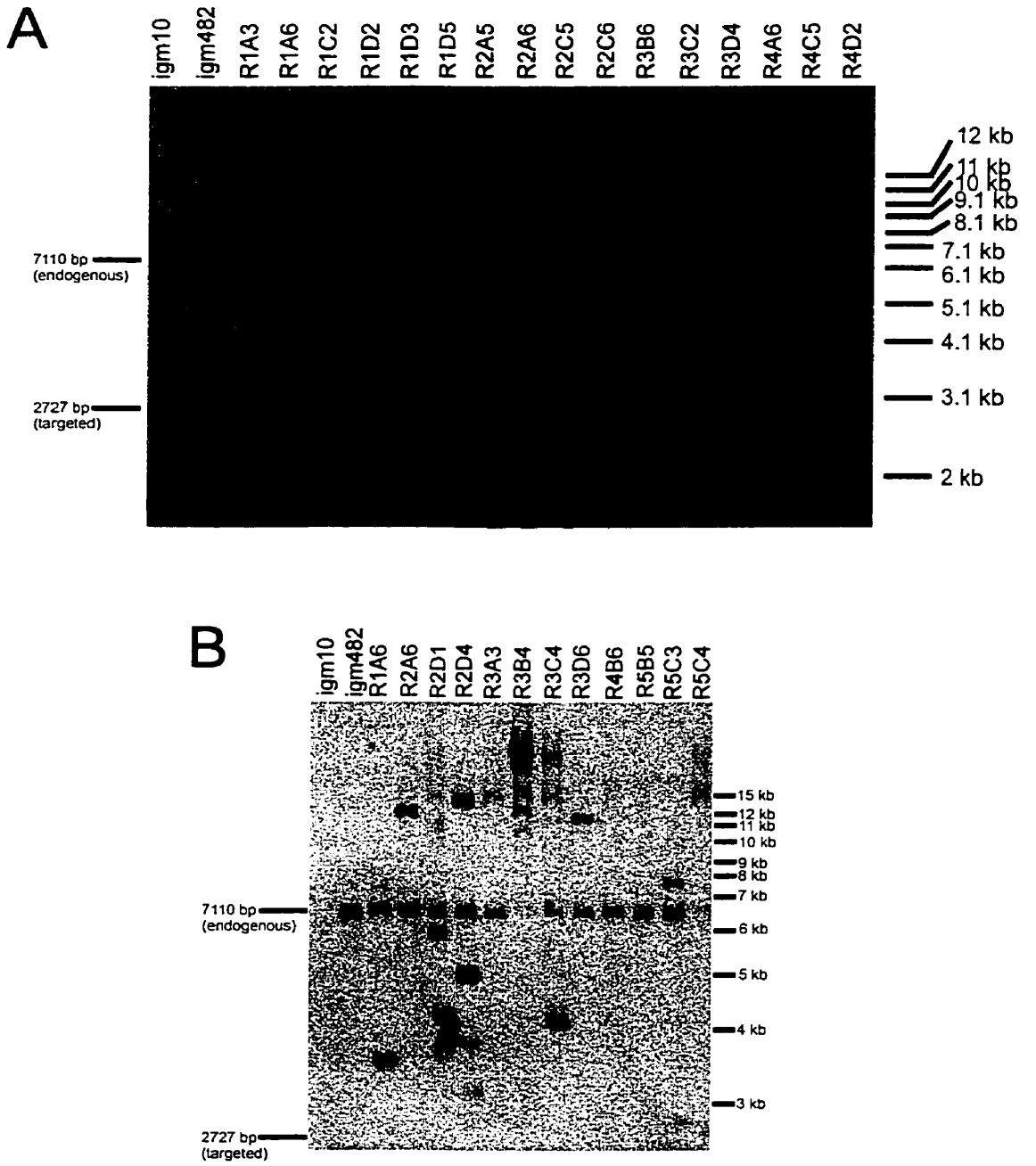


Figure 4.5 Analysis of the Cα gene stucture in G418^R lines recovered from a gene targeting experiment. Hybridoma genomic DNA was digested with *Xmn*I, electrophoresed through 0.9% agarose and blotted onto nitrocellulose and probed using ^{32}P-labeled probe F (specific to Cμ). In each figure, the sizes of the 7110 bp endogenous Cμ band and the 2727 bp fragment expected from a correctly targeted line are indicated to the left, while the molecular weight markers are indicated to the right. In blot A, hybridoma lines R1D2 and R3D4 have the 2.7kb band indicative of a targeting event, while blot B contains no putative targeted hybridoma lines.

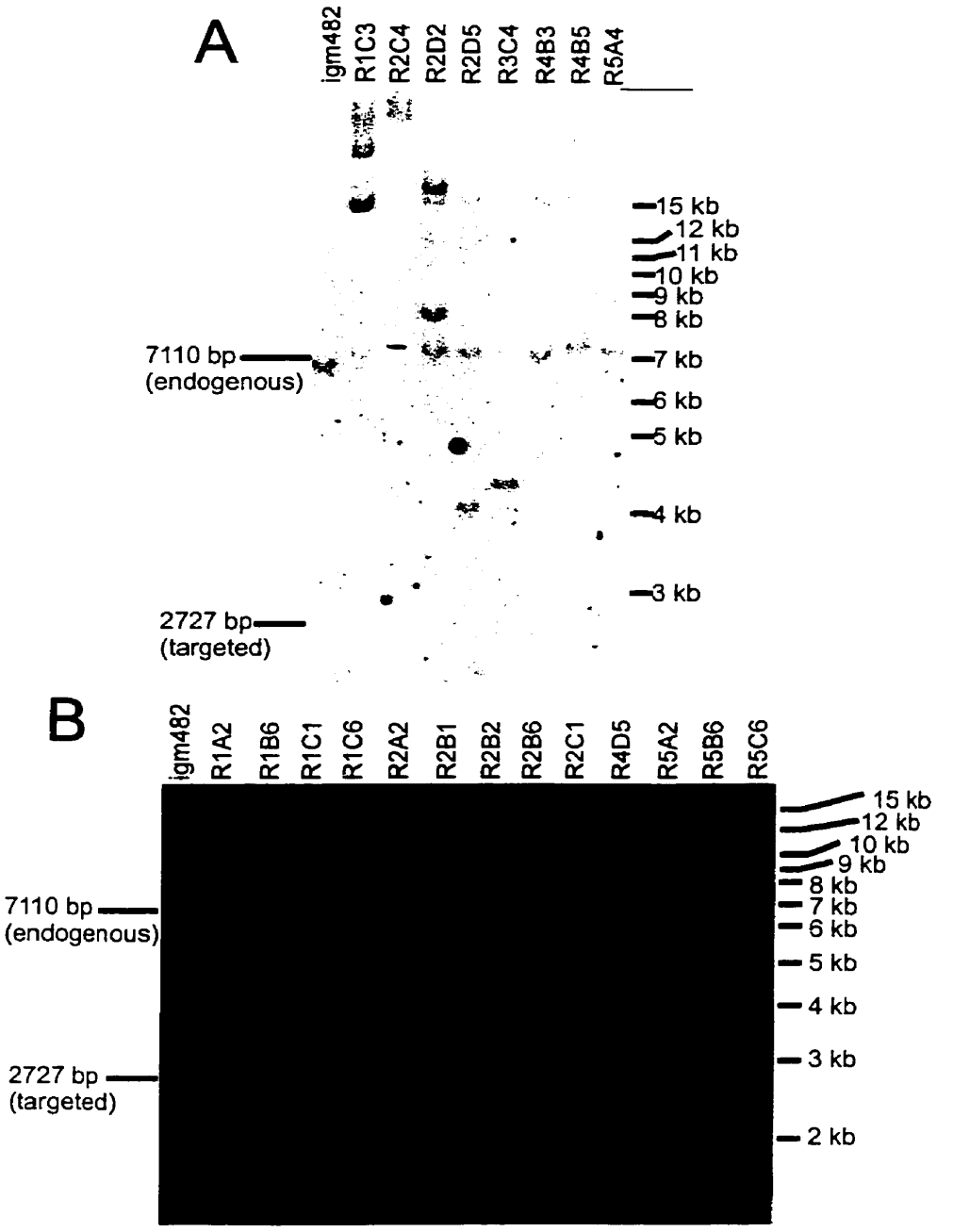


Figure 4.6 Analysis of the Cα gene stucture in G418^R lines recovered from a gene targeting experiment. Hybridoma genomic DNA was digested with *Xmn*I, electrophoresed through 0.9% agarose and blotted onto nitrocellulose and probed using ^{32}P-labeled probe F (specific to Cμ). In each figure, the sizes of the 7110 bp endogenous Cμ band and the 2727 bp fragment expected from a correctly targeted line are indicated to the left, while the molecular weight markers are indicated to the right. Blot A contains no putative targeted lines, while in Blot B, hybridoma lines R1C1 and R2C1 have the 2.7kb band indicative of a targeting event.

4.2.3 VERIFICATION OF PUTATIVE GENE TARGETING EVENTS

Verification was required to confirm that putative targeted lines R1B4, R1B5, R1C1, R1D2, R2C1, R3A4, R3C5, and R3D4 represented the outcome of gene targeting events. Each line was subcloned and genomic DNA from the subclones was digested with each of *Xmn*I, *Dra*III and *Acc*I, blotted onto nitrocellulose and probed using either Probe F or Probe G. Southern blotting of *Dra*III or *Acc*I digested genomic DNA in combination with the *neo*-specific probe G was used to verify the structure of the 5' end of the targeted Cα locus. These blots are expected to show a 14511 bp and a 9256 bp band for *Dra*III and *Acc*I respectively. The structure of the 3' end of the targeted Cα locus was verified using *Xmn*I or *Dra*III in combination with the Cμ-specific probe F (expected size of the novel bands was 2727 bp and 6556 bp respectively). As probe F binds to the endogenous Cμ locus as well, these blots will have a band representing the endogenous locus as well as a novel band representing the vector-borne Cμ. An *Xmn*I and a *Dra*III digest of igm482 will (respectively) produce a 7110 bp and a 10363 bp endogenous fragment. In summary, the expected fragment sizes for each of the restriction enzyme/probe combinations are as follows: *Xmn*I/Probe F - 2.7 kb (7110 bp endogenous band), *Dra*III/Probe F - 6.6 kb (10363 bp endogenous band), *Dra*III/Probe G - 14.5 kb, *Acc*I/Probe G - 9.3 kb.

Line R1B4 contained the expected bands for the *Xmn*I and *Acc*I digests (2.7 kb and 9.4 kb respectively), yet lacked the predicted 14.5 kb *Dra*III (probe G) and 6.6 kb *Dra*III (probe F) fragments expected for a targeting event (Fig. 4.7). Instead, the DraIII/probe G blot has an ~18 kb band and the DraIII/probe F blot has an ~5.6

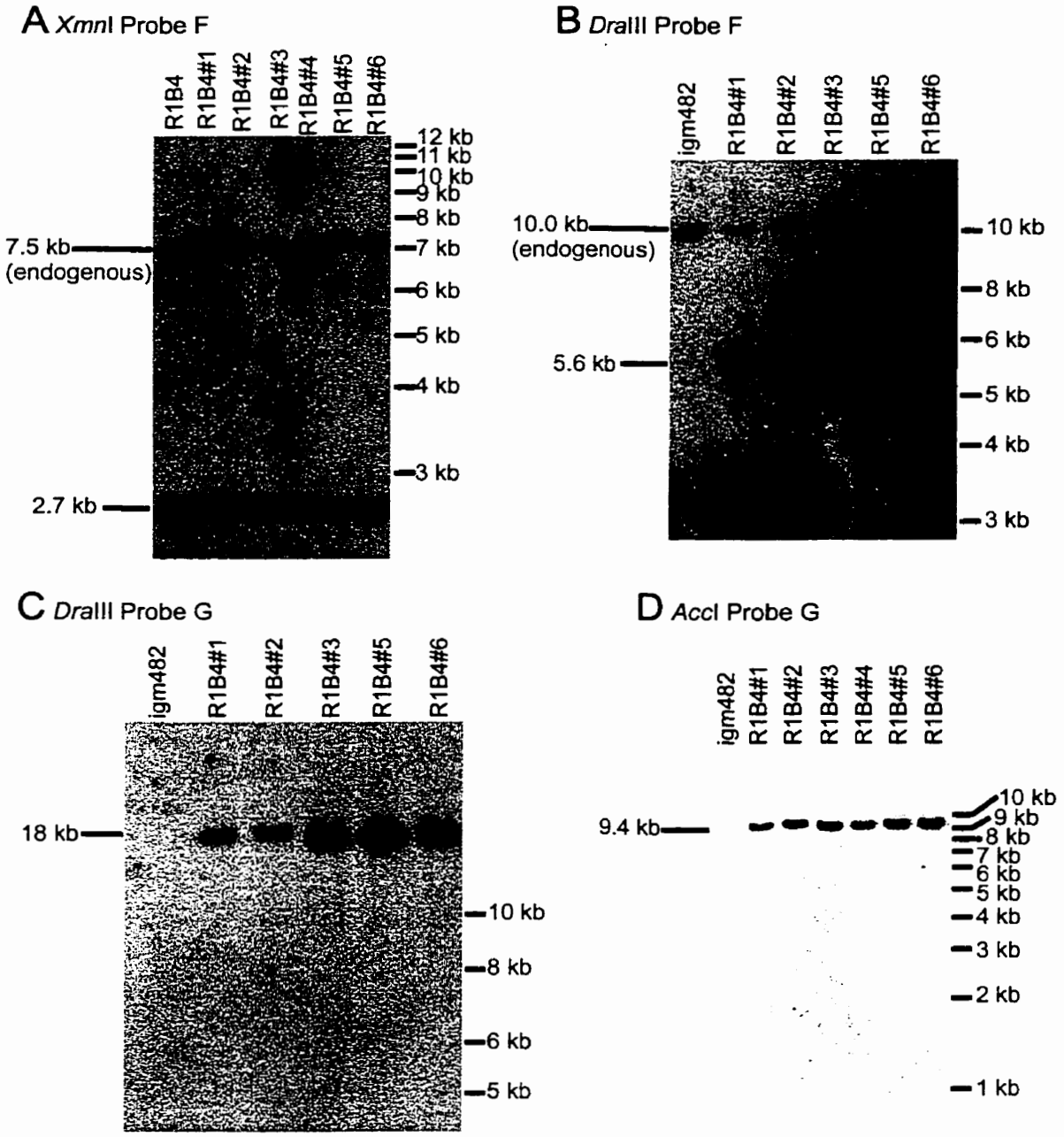


Figure 4.7 Southern blot analysis of putative targeted line R1B4 that contained the 2727 bp *Xmn*I fragment in the initial screening. This line was subcloned at 0.1 cells per well, and genomic DNA prepared from each of the subclones (#1, #2, #3, #4, #5, #6). This DNA was then digested with (A) *Xmn*I, (B) *Dra*III, (C)*Dra*III and (D) *Acc*I, electrophoresed through 0.9% agarose and blotted onto nitrocellulose. The blots were probed using either ^{32}P-labeled probe F (*Xmn*I, *Dra*III) or probe G (*Dra*III, *Acc*I). In each figure, the sizes of the bands of interest are indicated to the left of the blot while the molecular weight marker bands are indicated to the right. The expected sizes for a targeted line are 2727 bp (Probe F, *Xmn*I), 6556 bp (Probe F, *Dra*III), 14511 bp (Probe G, *Dra*III) and 9256 bp (Probe G, *Acc*I).

kb band. Thus R1B4 does not represent a cell line that has undergone gene targeting.

Line R1D2 contains fragments of the predicted size for all four enzyme/probe combinations (Fig. 4.8). The 5' structure of a line that has undergone gene targeting is expected to contain a 14.5 Kb *Dra*III and a 9.3 Kb *Acc*I fragment when probed with probe G. Southern analysis of R1D2 revealed a 14.2 Kb *Dra*III and a 9.4 Kb *Acc*I band (Fig. 4.8C, D). The 3' structure of a targeted line is expected to produce a 2.7 Kb *Xmn*I and a 6.6 Kb *Dra*III fragment when hybridized with probe F. Southern analysis revealed the expected bands for both digests (Fig. 4.8A, B); a 2.6 Kb *Xmn*I band and a 6.5 Kb *Dra*III band were seen. Accordingly, as all four digest/probe combinations yield the expected band size for a targeted line, R1D2 must represent the outcome of a gene targeting event.

Line R3D4 also has also undergone a gene targeting event as all four diagnostic digests contain the expected fragment sizes (2.9 Kb, 6.9 Kb, 14.2 Kb and 9.5 Kb respectively for each of *Xmn*I (probe F), *Dra*III (probe F), *Dra*III (probe G) and *Acc*I (probe G) respectively (Fig. 4.9A, B, C, D respectively).

R1C1 contains the predicted band sizes for the *Xmn*I and *Acc*I digests only (2.9 Kb and 9.4 Kb respectively) but the novel band is too low in the DraIII/Probe F blot (8.2 Kb) and too high in the *Dra*III/Probe G blot (19 Kb and 30 kb) (Fig. 4.10). There is also an additional ~20 Kb band on the *Acc*I blot (Fig. 4.10D). As the Southern analysis does not reveal the presence of the fragments expected for a cell line representing the outcome of a gene targeting event, R1C1 must represent a randomly integrated vector.

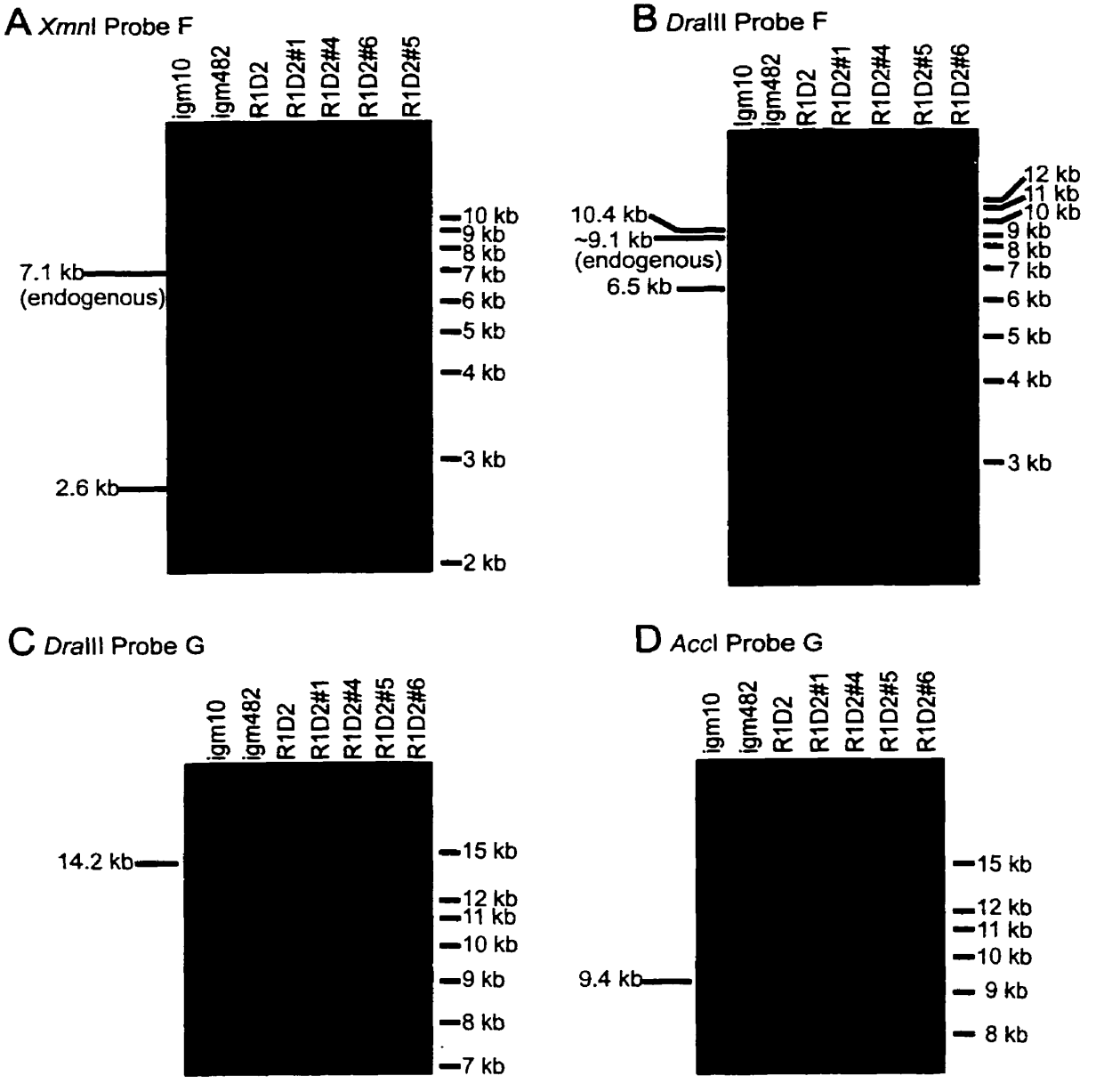


Figure 4.8 Southern blot analysis of putative targeted line R1D2 that contained the 2727 bp *Xmn*I fragment in the initial screening. This line was subcloned at 0.1 cells per well, and genomic DNA prepared from each of the subclones (#1, #2, #3, #4, #5, #6). This DNA was then digested with (A) *Xmn*I, (B) *Dra*III, (C)*Dra*III and (D) *Acc*I, electrophoresed through 0.9% agarose and blotted onto nitrocellulose. The blots were probed using either ^{32}P-labeled probe F (*Xmn*I, *Dra*III) or probe G (*Dra*III, *Acc*I). In each figure, the sizes of the bands of interest are indicated to the left of the blot while the molecular weight marker bands are indicated to the right. The expected sizes for a targeted line are 2727 bp (Probe F, *Xmn*I), 6556 bp (Probe F, *Dra*III), 14511 bp (Probe G, *Dra*III) and 9256 bp (Probe G, *Acc*I).

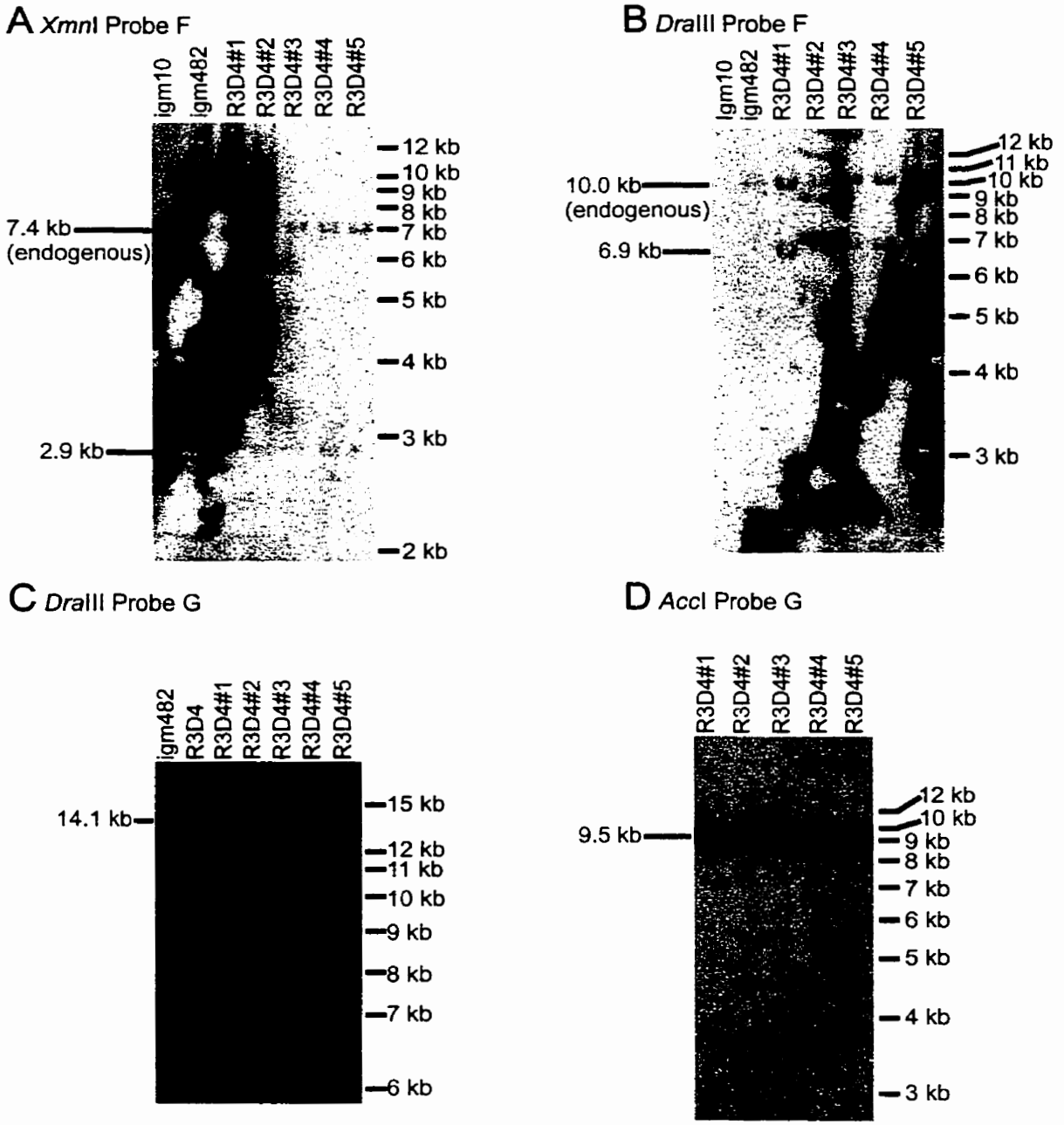


Figure 4.9 Southern blot analysis of putative targeted line R3D4 that contained the 2727 bp *Xmn*I fragment in the initial screening. This line was subcloned at 0.1 cells per well, and genomic DNA prepared from each of the subclones (#1, #2, #3, #4, #5). This DNA was then digested with (A) *Xmn*I, (B) *Dra*III, (C)*Dra*III and (D) *Acc*I, electrophoresed through 0.9% agarose and blotted onto nitrocellulose. The blots were probed using either ^{32}P-labeled probe F (*Xmn*I, *Dra*III) or probe G (*Dra*III, *Acc*I). In each figure, the sizes of the bands of interest are indicated to the left of the blot while the molecular weight marker bands are indicated to the right. The expected sizes for a targeted line are 2727 bp (Probe F, *Xmn*I), 6556 bp (Probe F, *Dra*III), 14511 bp (Probe G, *Dra*III) and 9256 bp (Probe G, *Acc*I).

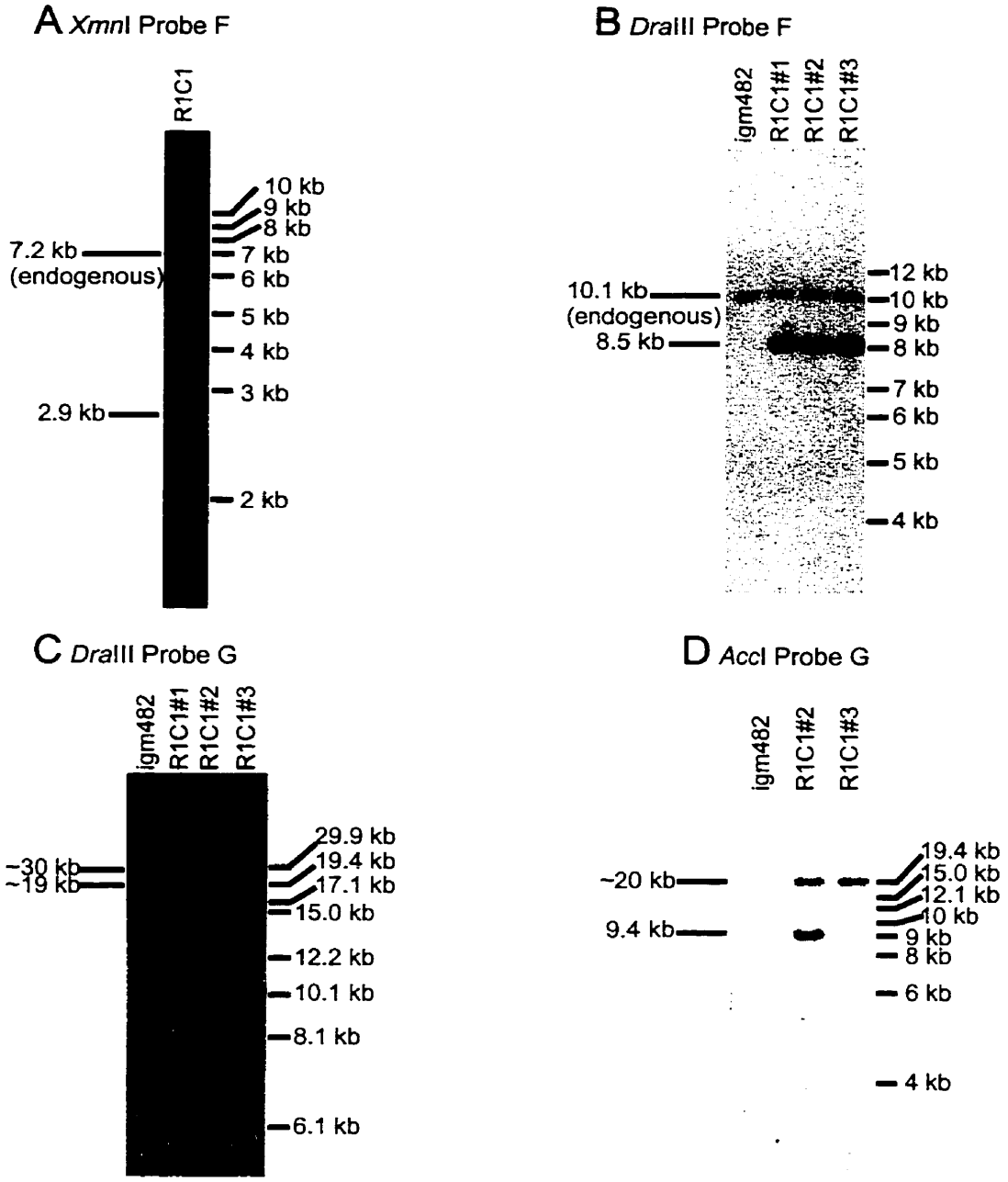


Figure 4.10 Southern blot analysis of putative targeted line R1C1 that contained the 2727 bp *Xmn*I fragment in the initial screening (A). This line was subcloned at 0.1 cells per well, and genomic DNA prepared from each of the subclones (#1, #2, #3). This DNA was then digested with (A) *Xmn*I, (B) *Dra*III, (C) *Dra*III and (D) *Acc*I, electrophoresed through 0.9% agarose and blotted onto nitrocellulose. The blots were probed using either ^{32}P-labeled probe F (*Xmn*I, *Dra*III) or probe G (*Dra*III, *Acc*I). In each figure, the sizes of the bands of interest are indicated to the left of the blot while the molecular weight marker bands are indicated to the right. The expected sizes for a targeted line are 2727 bp (Probe F, *Xmn*I), 6556 bp (Probe F, *Dra*III), 14511 bp (Probe G, *Dra*III) and 9256 bp (Probe G, *Acc*I).

Line R2C1 produced a 2.7 Kb band when digested with *Xmn*I (Fig. 4.11A). Interestingly, the *Dra*III and *Acc*I digests produced three different fragment patterns in the subcloned lines. Subclones R2C1#3 and R2C1#5 produced a 3.0 Kb *Dra*III/probe F fragment, a 14.4 Kb *Dra*III/probe G fragment and an 8.9 Kb *Acc*I fragment (Fig. 4.11B, C, D). Despite the fact that the *Dra*III (probe G) and *Acc*I fragments are of the expected sizes, these subclones likely do not represent the outcome of a targeting event as the *Dra*III/probe F fragment is much smaller than as is expected. Subclones R2C1#2 and R2C1#4 do not have a novel *Dra*III band on the blot probed with probe F (Fig. 4.11B). It is likely that the novel *Dra*III fragment was small enough to have run off the gel prior to blotting. Subclones R2C1#2 and R2C1#4 have a 20 Kb fragment on the *Dra*III digest probed with G (Fig. 4.11C) and a 9.7 Kb fragment on the *Acc*I digest (Fig. 4.11D). Subclone R2C1#12 also lacks a *Dra*III fragment on the probe F blot, has the same 20 Kb *Dra*III fragment seen on the probe G blots of subclones #2 and #4, but has a slightly higher *Acc*I fragment than these lines (9.7 Kb vs. 8.9 Kb) (Fig 4-11C,D). As all the subclones of R2C1 lack the correct *Dra*III (probe F) fragment expected for a line that has undergone gene targeting, they all represent random integrations of the targeting vector.

Line R1B5 contains the predicted *Xmn*I fragment (2.5 Kb; Fig. 4.12A) but lacks the expected fragments for the other digests. The novel *Dra*III fragment on the blot probed with probe F is 6.2 Kb (Fig. 4.12B) instead of the 6.6 Kb that would seen in a targeted line. The *Dra*III fragment on the probe G blot is 11.2 Kb (Fig. 4.12C) instead of the expected 14.5 kb and the *Acc*I fragment is only 8.8 Kb (Fig. 4.12 D)

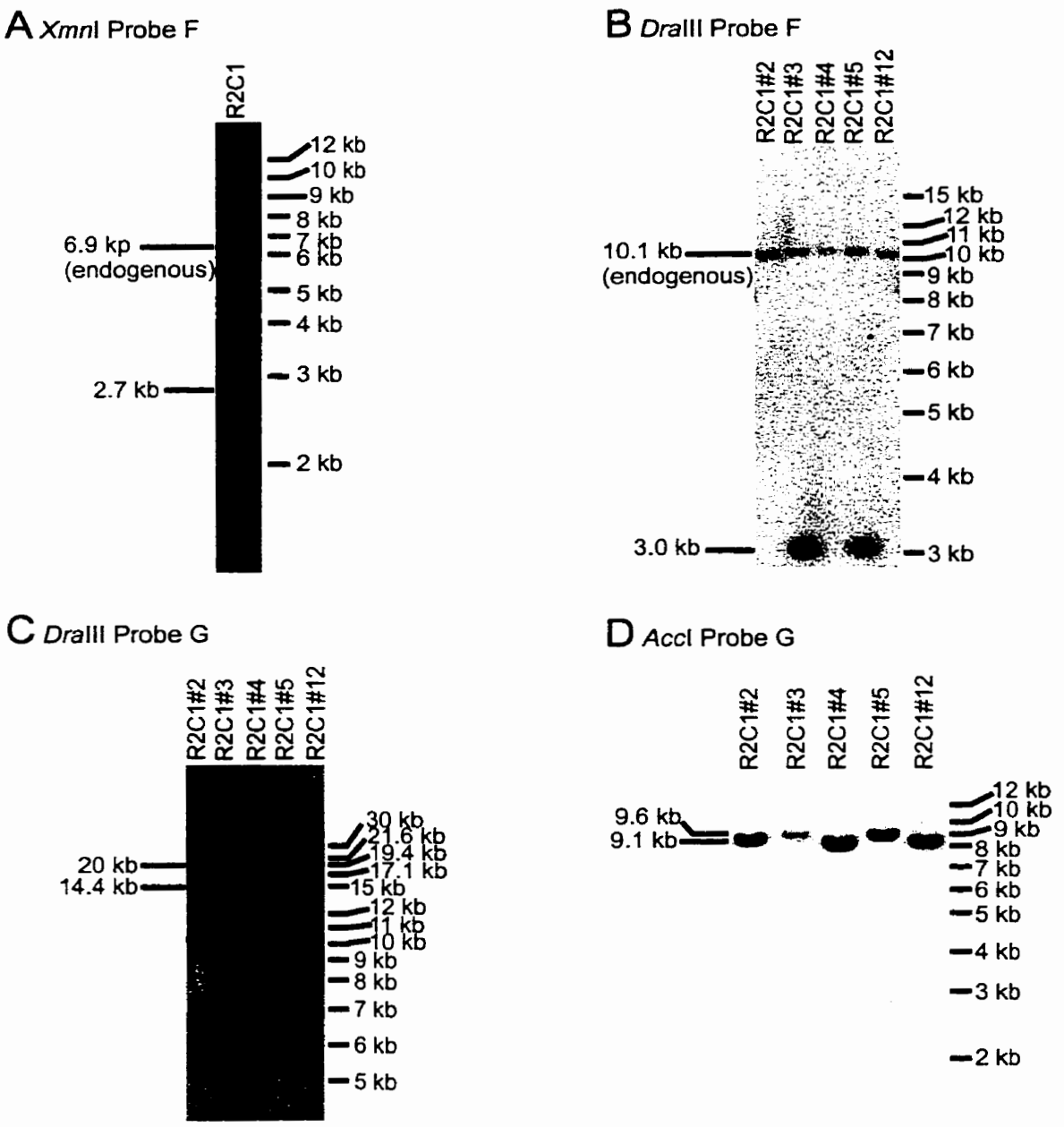


Figure 4.11 Southern blot analysis of putative targeted line R2C1 that contained the 2727 bp *Xmn*I fragment in the initial screening (A). This line was subcloned at 0.1 cells per well, and genomic DNA prepared from each of the subclones (#2, #3, #4, #5, #12). This DNA was then digested with (A) *Xmn*I, (B) *Dra*III, (C) *Dra*III and (D) *Acc*I, electrophoresed through 0.9% agarose and blotted onto nitrocellulose. The blots were probed using either ^{32}P-labeled probe F (*Xmn*I, *Dra*III) or probe G (*Dra*III, *Acc*I). In each figure, the sizes of the bands of interest are indicated to the left of the blot while the molecular weight marker bands are indicated to the right. The expected sizes for a targeted line are 2727 bp (Probe F, *Xmn*I), 6556 bp (Probe F, *Dra*III), 14511 bp (Probe G, *Dra*III) and 9256 bp (Probe G, *Acc*I).

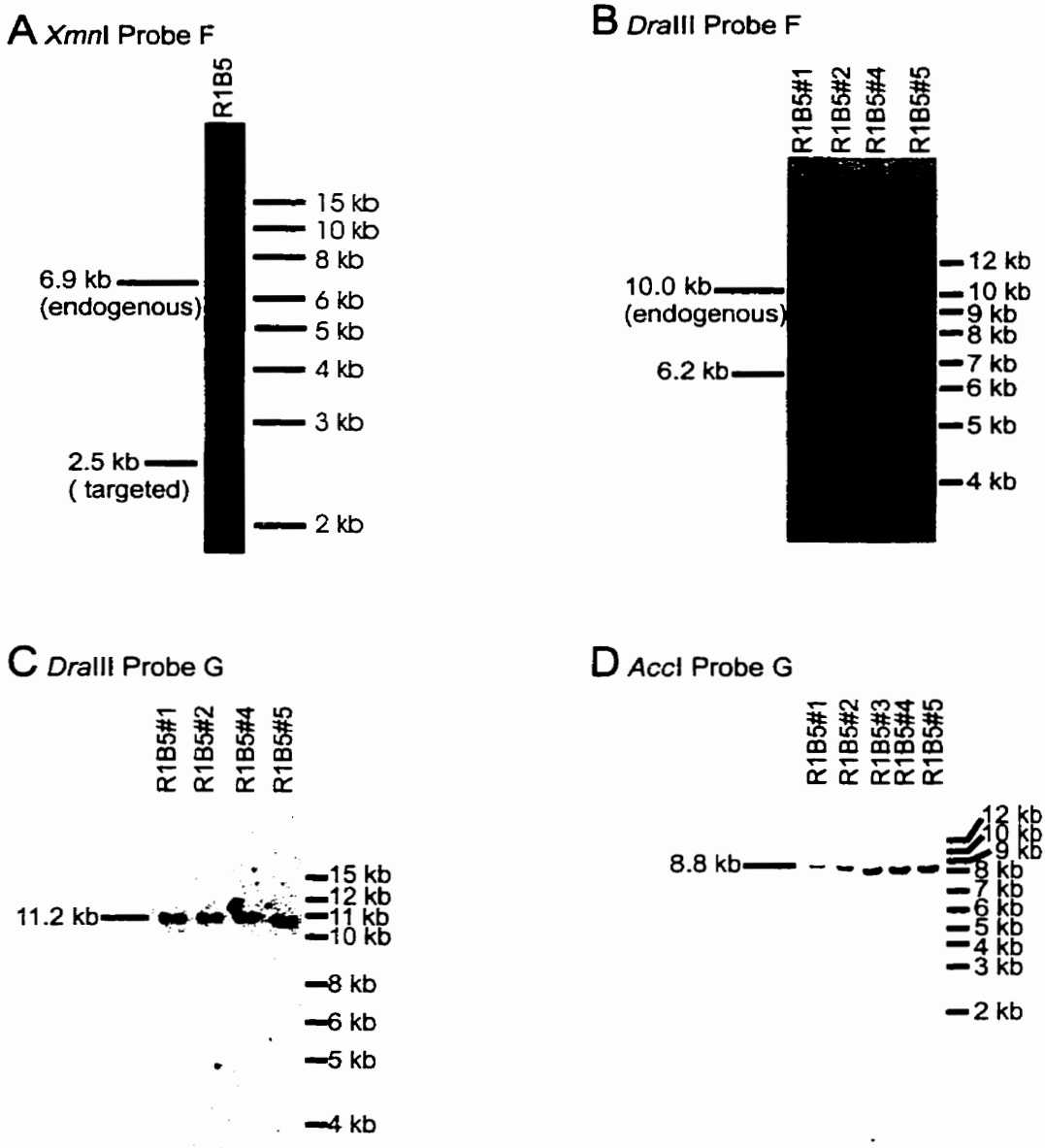


Figure 4.12 Southern blot analysis of putative targeted line R1B5 that contained the 2727 bp *Xmn*I fragment in the initial screening (A). This line was subcloned at 0.1 cells per well, and genomic DNA prepared from each of the subclones (#1, #2, #3, #4, #5). This DNA was then digested with (A) *Xmn*I, (B) *Dra*III, (C) *Dra*III and (D) *Acc*I, electrophoresed through 0.9% agarose and blotted onto nitrocellulose. The blots were probed using either ^{32}P-labeled probe F (*Xmn*I, *Dra*III) or probe G (*Dra*III, *Acc*I). In each figure, the sizes of the bands of interest are indicated to the left of the blot while the molecular weight marker bands are indicated to the right. The expected sizes for a targeted line are 2727 bp (Probe F, *Xmn*I), 6556 bp (Probe F, *Dra*III), 14511 bp (Probe G, *Dra*III) and 9256 bp (Probe G, *Acc*I).

instead of the expected 9.3 Kb. Thus hybridoma line R1B5 contains a randomly integrated vector.

Line R3A4 contained a 2.4 kb *Xmn*I fragment (Fig. 4.13A) but the *Dra*III fragment on the blot probed with F was much smaller than was expected for a targeted line (2.4 kb versus of 6.6 Kb; Fig. 4.13B). The *Acc*I fragment was also smaller than expected for a targeting event (5.4 Kb versus 9.3 Kb; Fig. 4.13B). Consequently, this line represents the outcome of random vector integration.

The final line tested, R3C5, had a 2.4 Kb *Xmn*I fragment as expected for a targeting event, but as it lacked the predicted 9.3 Kb *Acc*I fragment (a 5.0 Kb band was seen on the blot; Fig. 4.13E) it does not represent the outcome of a gene targeting. Thus, out of the eight putative targeted lines that contained the diagnostic 2.7 Kb *Xmn*I fragment, only two were the result of gene targeting (R1D2 and R3D4).

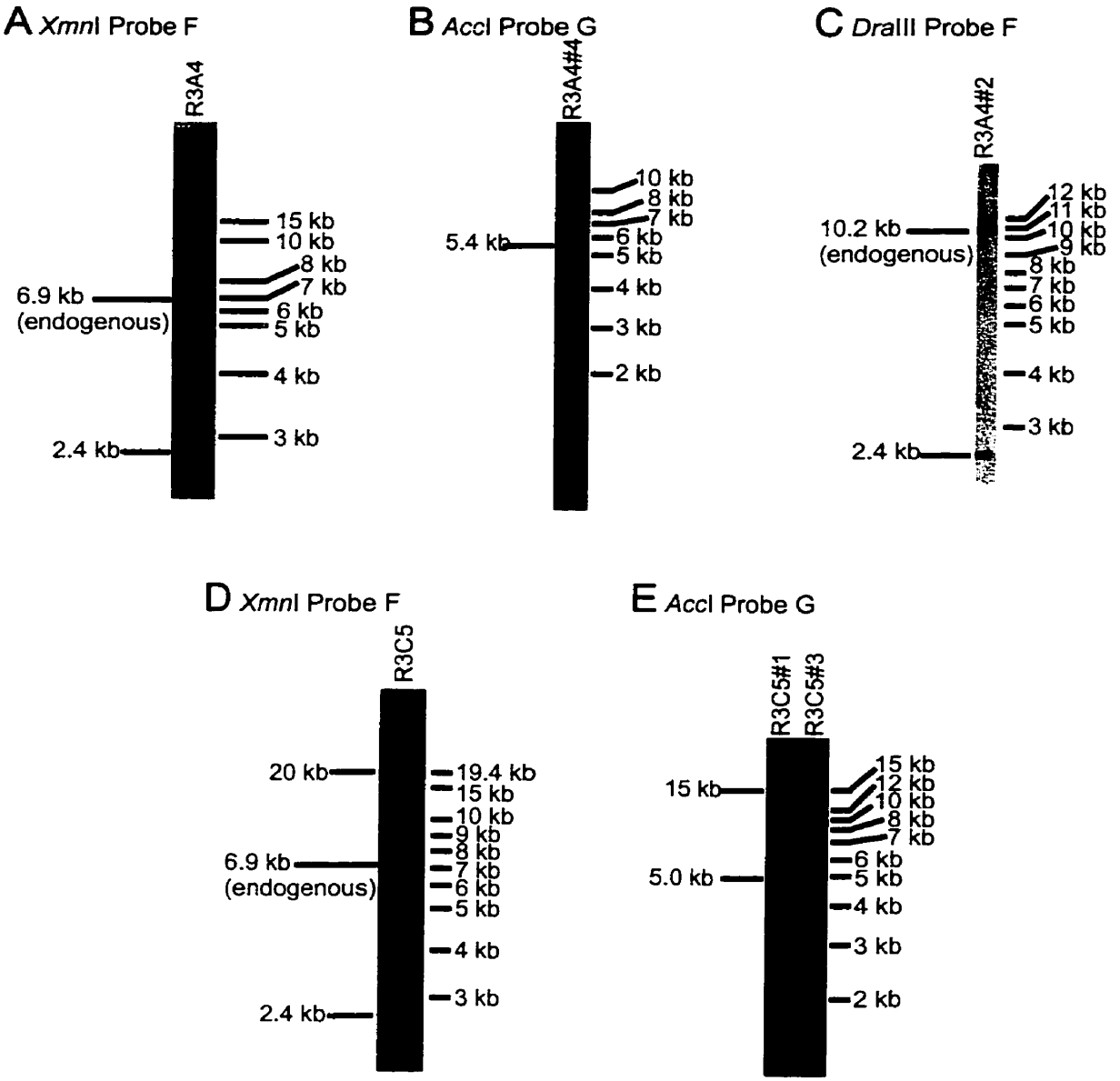


Figure 4.13 Southern blot analysis of putative targeted lines R3A4 and R3C5 that each contained the 2727 bp *Xmn*I fragment in the initial screening (A, D respectivly). These lines were subcloned at 0.1 cells per well, and genomic DNA was prepared from each of the subclones of line R3A4 (#2, #4) and line R3C5 (#1, #3). The R3A4 subclone DNA was then digested with either (B) *Acc*I or (C) *Dra*III while the R3C5 subclone DNA was digested with *Acc*I (E). The cut DNA was electrophoresed through 0.9% agarose and blotted onto nitrocellulose. The blots were probed using either ^{32}P-labeled probe F (*Xmn*I, *Dra*III) or probe G (*Acc*I). In each figure, the sizes of the bands of interest are indicated to the left of the blot while the molecular weight marker bands are indicated to the right. The expected sizes for a targeted line are 2727 bp (Probe F, *Xmn*I), 6556 bp (Probe F, *Dra*III) and 9256 bp (Probe G, *Acc*I).

4.3 DISCUSSION

4.3.1 THE ENHANCER TRAP VECTOR REDUCES THE BACKGROUND OF NON-TARGETED CELLS DURING GENE TARGETING TO THE Cα LOCUS

The Imut copy of Cμ was targeted to the Cα locus using an enhancer-trap replacement vector with 3595 bp and 782 bp flanking homology on the 5' and 3' flanks respectively (Fig. 4.2). When screened by Southern blotting, only two transformed lines represented targeting events: R1D2 and R3D4. The absolute frequency of GT events in the current study was low ($1x10^{-7}$ targeted lines per surviving cell (assuming a 50% hybridoma mortality rate following electroporation)) compared to the value measured using a CμEn$^-$ vector ($7.7x10^{-5}$; Ng and Baker 1998). Consequently, the Cα targeting generated fewer G418^R transformants per cell ($4x10^{-6}$) than was found in the En$^-$Cμ targeting study ($7.7x10^{-5}$; Ng and Baker 1998) However, while the absolute number of GT events was quite low (only two lines) the background of transformed lines was also reduced (a total of 123 G418^R lines). As such, the targeting efficiency ($2x10^{-2}$ GT events per transformed cell) was similar to reported targeting efficiencies for targeting to the Cμ locus with enhancer-trap vectors ($2.9\text{-}3.0x10^{-2}$; Ng and Baker 1998, 1999). This suggests that the enhancerless vector was able to efficiently target the Cα locus, and that the background of non-targeted G418^R transformants was greatly reduced. The targeting efficiency at Cα with the En$^-$ vector was higher than previously reported values for targeting at Cμ with an enhancer positive vector. These latter values range from $1.4x10^{-3}$ (Ng and Baker 1999) to $1.8x10^{-3}$ (Ng and Baker 1998) targeted lines per transformed line. This also implies that the Cα enhancer-trap vector used

in the current study has reduced the background of non-targeted G418^R lines. The enhancerless *neo* gene is less likely to be transcribed following random integration of the targeting vector, and thus the background of random integration events is reduced. The reduction in background improves the efficiency of gene targeting as fewer transformed cells need to be screened in order to find the targeting events. The current study used the data from a single gene targeting while the CμEn- studies used data from 4 separate electroporations (Ng and Baker 1998, 1999). Additional experiments would provide more accurate information on targeting efficiency of the CαEn$^-$ vector.

In addition, studies (Deng and Capecchi 1992) have shown that sequence polymorphisms in the targeting vector reduce targeting frequency. This should not be a factor in this study, as the Cα homology contained on the targeting vector was PCR amplified from igm482, and so was isogenic to the endogenous locus.

4.3.2 The Role of the Eμ and 3'α Enhancers in the Enhancer Trap Gene Targeting to the Cα Locus

It is unclear whether the high efficiency of GT to Cα with the enhancerless vector used in the current study was due to the action of the Eμ enhancer found 5' of the C_H locus, to the 3' enhancers or to a combination of some or all of these enhancers. Chauveau *et al.* (1998) found that the five C_H enhancers worked synergistically to upregulate transcription of a reporter gene. This suggests that the enhancer-trap vector required the presence of all the C_H enhancers. However, other research has suggested that certain Cα enhancers are only active at specific of the

cell cycle. For example, Ong *et al.* (1998) found evidence that Eμ is the most active enhancer in pre-B cells, but that the 3' enhancers contributed about equally to immunoglobulin gene expression in Ig-secreting cells. Other research indicates that Eμ is not required for Ig heavy chain production (Wabl and Burrows 1984), but that deletion of 3'αE abolished expression of Ig heavy chain (Liebeson *et al.* 1995). These results imply that it is the 3' enhancers, rather than Eμ, that are responsible for the upregulation of *neo* transcription in the cell lines in which a gene targeting event has occured. Thus the 3' enhancers are likely responsible for the efficacy of both the Cα and Cμ enhancer-trap targeting vectors. Further research in this area could include gene targeting with an enhancerless vector to the C_H locus of cell lines containing deletions of the various IgH enhancers (both singly and in combination). Such a study would determine the importance of the various enhancer elements in the efficacy of enhancer-trap gene targeting to C_H.

4.3.3 A Possible Interrupted Gene Targeting Event

Two of the subclones from line R2C1 (#3 and #5) have the *Xmn*I, *Dra*III (probe G) and *Acc*I fragments expected for a targeted line. However these subclones lack the expected 6.6 Kb band in the DraIII/Probe F Southern blot; a 2.8 Kb band is seen instead. These subclones may represent an example of an interrupted gene targeting (IGT) event. IGT occurs when a targeting vector interacts with the target locus, copies in endogenous sequences, and then goes on to integrate randomly elsewhere in the genome. The targeting vector used in the current study contains Cα sequence that begins at the *Hin*dIII site at bp 2692 and

ends at the *Eco*RI site at bp 18069 (numbering according to Fig. 4.3). If the targeting vector interacted with the endogenous locus and copied in ≥2kb of Cα sequence 5' of the *Hin*dIII site before randomly integrating elsewhere in the genome, than both the *Acc*I site at bp 2505 and the *Dra*III site at bp 742 would be present in the vector. Such an IGT line would have the same 5' structure as a targeted line. Subclones R2C1#3 and R2C1#5 both have a 14.4 kb band in the *Dra*III/probe G blot (a 14.5 kb band is expected in a targeted line) and a 9.6 kb *Acc*I band (a 9.3 kb band is expected in a targeted line). Thus the 5' structure of these lines is as expected for a targeting event. The 3' end of the Cα targeting vector contained Cα sequence up to the *Eco*RI site at bp 18069 (according to the numbering system of Fig. 4.3). The *Xmn*I site used to screen for targeted lines by Southern blotting lies only 343 bp 3' of *Eco*RI while the 3' *Dra*III site lies 3740 bp downstream of the *Eco*RI site. Despite the presence of the expected *Xmn*I band, the 3' structure of lines R2C1#3 and R2C1#5 does not match that expected for a targeting event as the 6.6 kb Probe F *Dra*III fragment is not present. This data suggests that the targeting vector in lines R2C1#3 and R2C1#5 (Fig. 4.11) may have acquired 3' sequence from the endogenous Cα locus. The vector likely copied in between 343 bp and 3740 bp of 3' Cα sequence (to include the *Xmn*I site at bp 18412 but not the *Dra*III site at bp 21809). An IGT event could be confirmed by using a Cα-specific probe that lies outside the vector-borne region of homology. Southern blot analysis would reveal an endogenous and a novel band if an IGT event had occurred. No such probe is yet available, but this is an avenue for future research.

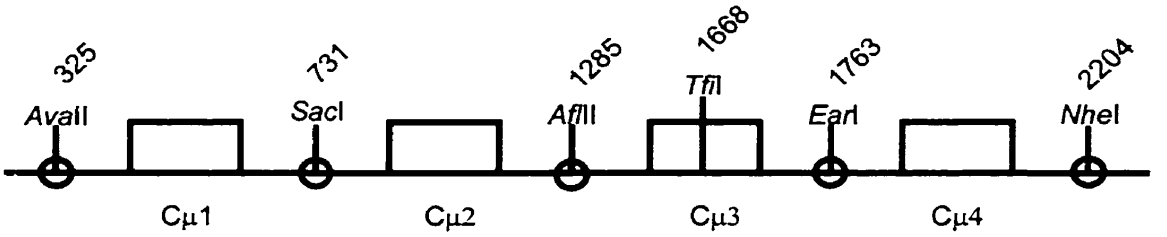
5. RECOMBINATION RATES AND PFC RECOVERY

5.1 INTRODUCTION

The recombination assay used in the present study (the Cμ recombination assay) is based upon the detection of homologous recombination events between Cμ loci that restore normal IgM production. The recipient for recombination in this system is the haploid chromosomal μ locus of the murine hybridoma cell line igm482. Hybridoma line igm482 was isolated from the wildtype hybridoma cell line Sp6 that bears a single copy of the 2,4,6-trinitrophenyl (TNP)-specific chromosomal μ gene and synthesizes normal polymeric TNP-specific IgM that can be detected using a complement-dependant, TNP-specific plaque assay (Baker *et al.* 1988). The Cμ region of igm482 includes the TNP-specific variable region but contains a 2 bp frameshift deletion in the third exon of the constant region (Cμ3) (Fig. 5-1A). This deletion results in the production of a truncated heavy chain that lacks the Cμ4 domain required to polymerize IgM. The resulting monomeric IgM cannot activate the complement-dependant lysis of TNP-coated sRBCs in the plaque assay. The deletion also results in the loss of the wildtype *Xmn*I RE site and the gain of a *Tfi*I site. The donor Cμ contained a wildtype Cμ3 exon and was marked with RFLP sites within the introns of the constant region (the Imut markers) but lacked the TNP-specific variable region (Fig. 5-1B).

Homologous recombination events between the endogenous igm482 Cμ and the vector-borne Cμ that restored the production of polymeric IgM were detectable using the TNP-specific plaque assay. The vector-borne Cμ acts as a donor of

A Endogenous (recipient) Cμ locus.



B Vector-bourne (donor) Cμ locus.

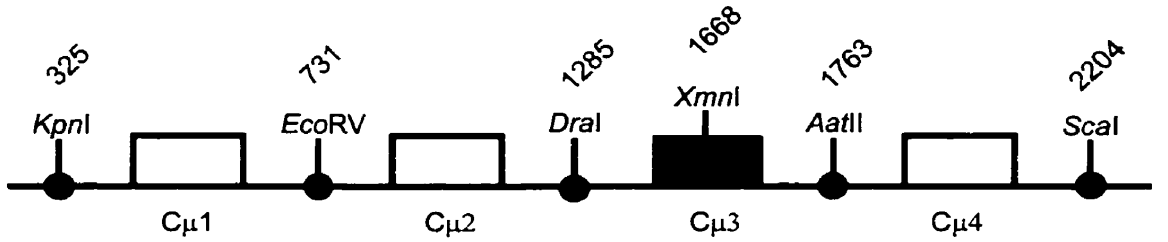


Figure 5.1 Diagnostic restriction enzyme sites found in the (A) endogenous (recipient) and (B) vector-borne (donor) Cμ loci. Endogenous (recipient) restriction enzyme site markers are indicated by open circles and vector-borne (donor) restriction enzyme site markers are indicated by black circles. Open boxes indicate exons. The diagnostic restriction enzyme fragments found in the 4.6 kb PCR product amplified from the endogenous (recipient) Cμ locus are as follows:
Intron mutation (Imut) marker 1 (Imut1) *Kpn*I digest: Endogenous: 1125 , 3497 bp. Conversion to donor site: 325, 1125, 3172 bp.
Imut2 (*Eco*RV): Endogenous: 1221, 3401 bp. Conversion: 73, 1221, 2670 bp.
Imut3 (*Dra*I): Endogenous: 4622 bp. Conversion: 1285, 3337 bp.
Cμ3 marker (*Xmn*I): Endogenous: 1892, 2730 bp. Conversion: 224, 1668, 2730 bp.
Imut4 (*Aat*II): Endogenous: 4622 bp. Conversion: 1763, 2859 bp.
Imut5 (*Sca*I): Endogenous site: 4622 bp. Conversion: 2200, 2422 bp.
The diagnostic restriction enzyme fragments found in the 6.3 kb PCR product amplified from the vector-bourne (donor) Cμ are as follows:
Imut1 (*Kpn*I digest): vector-bourne markers: 254, 1084, 1803, 3185 bp. Conversion to endogenous markers: 1084, 1803, 3435 bp.
Imut2 (*Eco*RV): vector-bourne: 104, 659, 788, 2091, 2684 bp. Conversion: 104, 788, 2091, 3343.
Imut3 (*Dra*I): vector-bourne: 290, 1219, 4817. Conversion: 290, 6036.
Cμ3 marker (*Xmn*I): vector-bourne: 228, 1606, 4492. Conversion: 1834, 4492.
Imut4 (*Aat*II): vector-bourne: 1701, 1498, 3127. Conversion: 1498, 4828.
Imut5 (*Sca*I): vector-bourne: 2143, 4183. Conversion: 6326.

genetic information during homologous recombination with the endogenous (recipient) Cμ locus. These events must restore the 2 bp Cμ3 deletion, and may include gene conversion events that result in the transfer of the Imut RFLP markers to the recipient (endogenous) Cμ locus. The plaque assay was used to measure recombination rates in one of the targeted lines recovered (R1D2) as well as in two lines with randomly integrated vector sequence (R2C1, R2B2). Recombinants from these lines were isolated for further study.

Recombination rates in cell lines R1D2, R2C1 and R2B2 were measured by plating out hybridoma cells and measuring the number of plaques produced by each line in the plaque assay. These plaques were then rescued and genomic DNA prepared from each. PCR was used to amplify a 4.6 Kb region of the recipient (endogenous) Cμ locus. The PCR product was then digested with diagnostic restriction enzymes to determine if there was any evidence of gene conversion from the vector-borne Imut Cμ. A second PCR was used to amplify a 6.3 Kb region of the donor (vector-borne) Cμ locus. This PCR product was digested with diagnostic restriction enzymes to determine if the Imut sites were intact in the vector-borne Cμ. This data is relevant to determining the type of homologous recombination that occurred, the extent of the search for homology within the mammalian genome and the extent of any hDNA that was formed during homologous recombination.

5.2 RESULTS

5.2.1 RATES OF RECOMBINATION

Five pools of cells were grown up from 50 cells for each of line R1D2 (a targeted line) and R2C1 (a random integration). These pools of cells were plated daily (1-2x10^7 cells from each pool) on plaque assay plates to determine the frequency of recombination. Line R2B2 (a random integration) was subcloned at 0.1 cells per well, to create 7 pools of cells. Recombination rates were also measured in these cell lines. The average frequency of recombination (PFC/cell) for each of these cell lines is given in Table 5.1. The average rate of recombination for the R2B2 pools is given in Table 5.2.

Table 5.1 Recombination Frequencies for cell lines R1D2, R2C1 and R2B2 as measured by the TNP-specific plaque assay.

Cell Line	Recombination Frequency	# of PFCs Recovered	Total Cells Plated
R1D2	1.78x10^{-7} PFC/cell	156	~ 9.0x10^8 cells
R2C1	0.12x10^{-7} PFC/cell	38	~ 15x10^8 cells
R2B2	26.4x10^{-7} PFC/cell	754	~ 6.5x10^7 cells

Table 5.2 Recombination Frequency for subclones of cell line R2B2 as measured by the TNP-specific plaque assay.

Subclone	Recombination Frequency	# of PFCs Recovered	Total cells plated
R2B2#1	$5.61x10^{-4}$	617	$1.1x10^{6}$
R2B2#2	$1.5x10^{-6}$	22	$1.5x10^{7}$
R2B2#3	$3.8x10^{-6}$	19	$5.0x10^{6}$
R2B2#4	$6.3x10^{-6}$	44	$7.0x10^{6}$
R2B2#5	$9x10^{-7}$	9	$1.0x10^{7}$
R2B2#6	$2.4x10^{-6}$	19	$8.0x10^{6}$
R2B2#7	$2.0x10^{-6}$	24	$1.2x10^{7}$

5.2.2 DNA Charaterization of Parental Lines

In order to ensure that the PFCs recovered in the current study were not the result of GC events that had occurred prior to subcloning, the Cμ structure of the parental line of each recombinant was characterized. PCR was used to amplify a 4.6 Kb fragment from the endogenous Cμ locus (Fig. 5.2A) and a 6.3 Kb fragment from the vector-borne Cμ locus (Fig. 5.2B). The resulting products were digested with the Imut enzymes *Kpn*I, *Eco*RV, *Dra*I, *Xmn*I, *Aat*II, and *Sca*I. Restriction digests of the 4.6 Kb endogenous locus of the parental lines (Fig. 5.3) revealed that all still had the endogenous Cμ sites. Restriction digests of the 6.3 Kb vector-borne locus of the parental lines (Fig. 5.4) revealed only vector-borne restriction enzyme sites

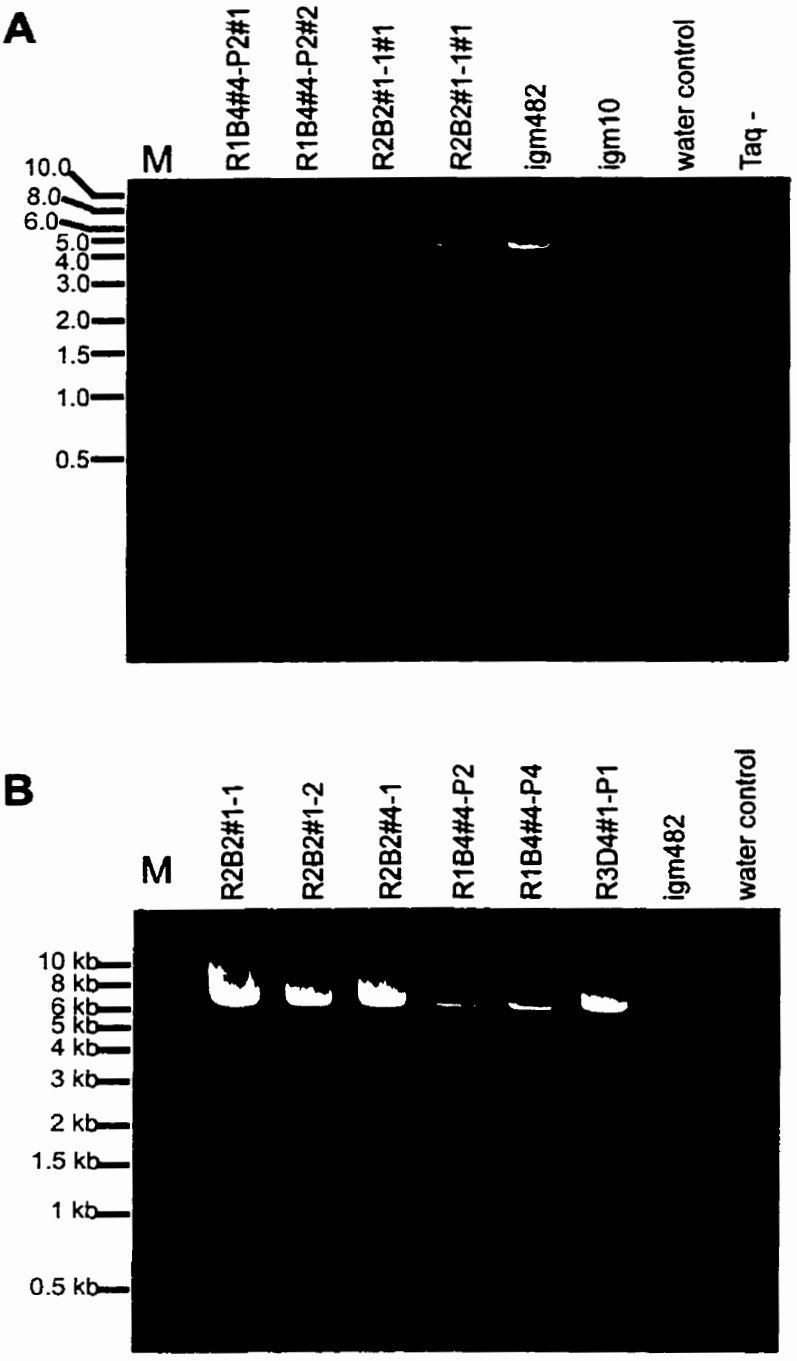


Figure 5.2 PCR amplification of (A) the 4.6 kb product from the endogenous Cu locus of PFC lines and (B) the 6326 bp product from the donor (vector-bourne) Cu locus. In figure A, amplification products of four PFC lines and three controls are presented. In figure B, amplification The sizes of marker bands (M) in kilobases are indicated to the left of the gel.

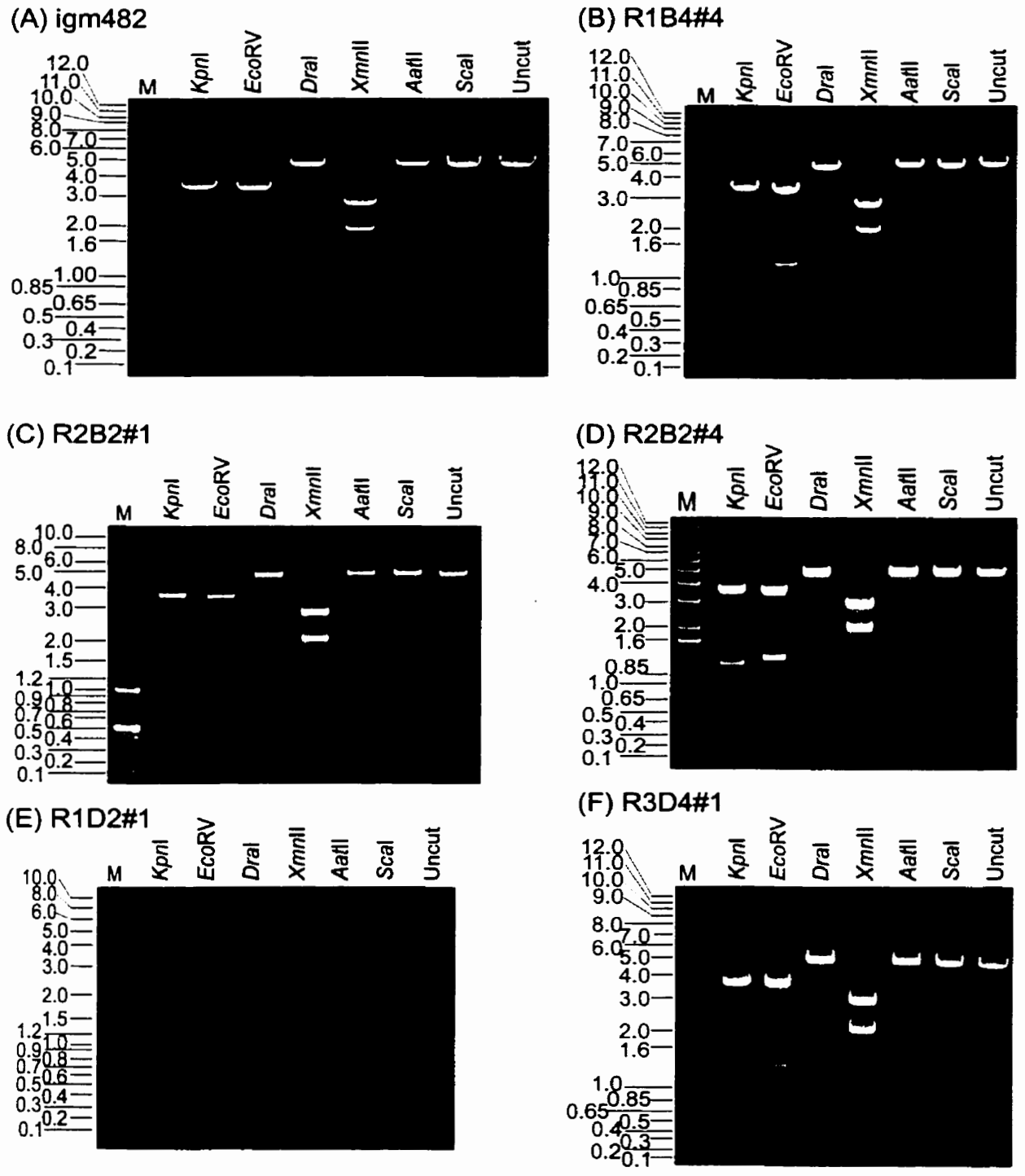


Figure 5.3 Restriction enzyme digests of the 4.6 kb endogenous (recipient) Cμ PCR product amplified from (A) hybridoma line igm482, parental lines (B) R1B4#4,(C) 2B2#1, (D)R2B2#4, (E) R1D2#1, and (F)R3D4#1 from which PFCs were recovered. Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5.1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

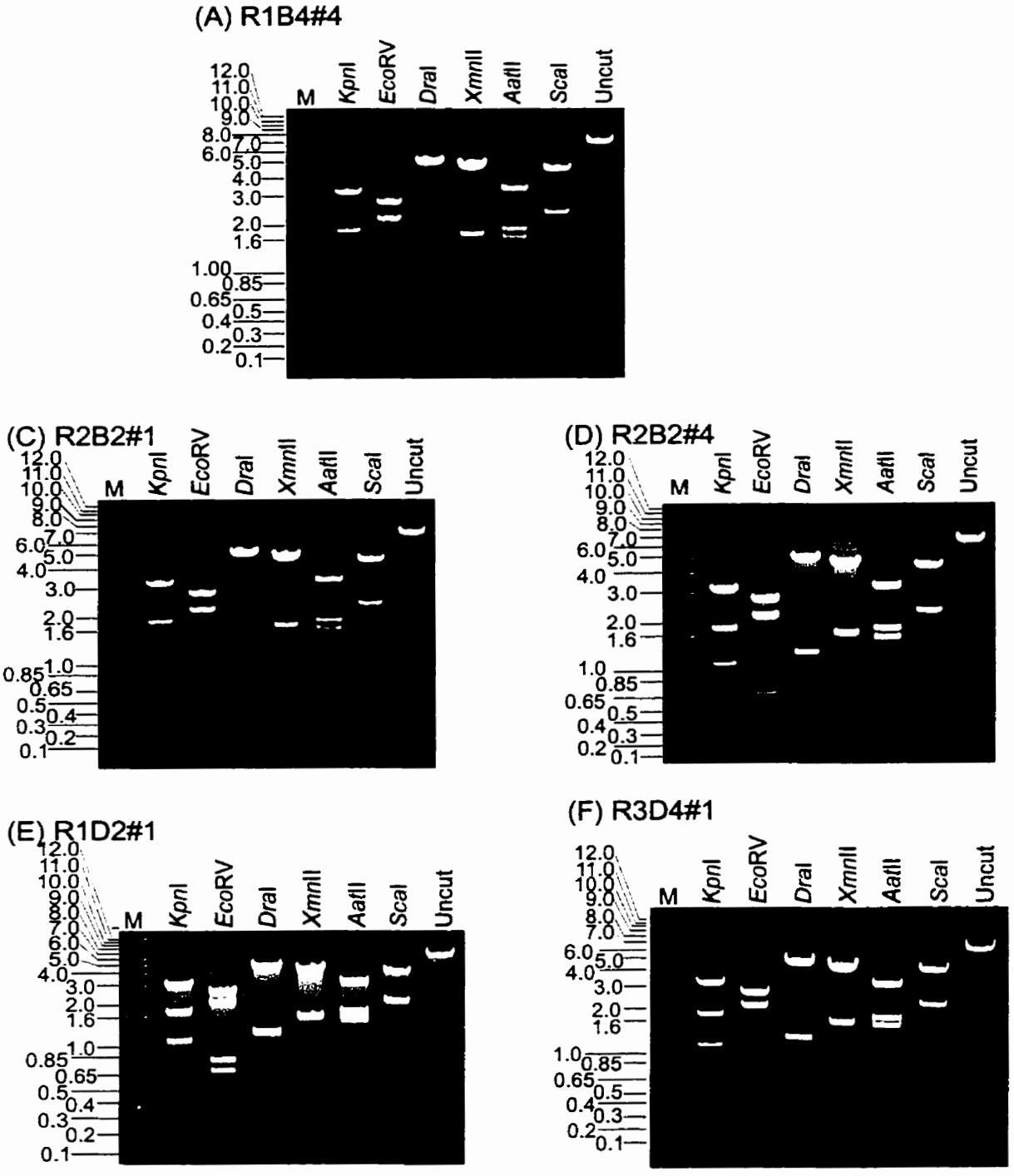


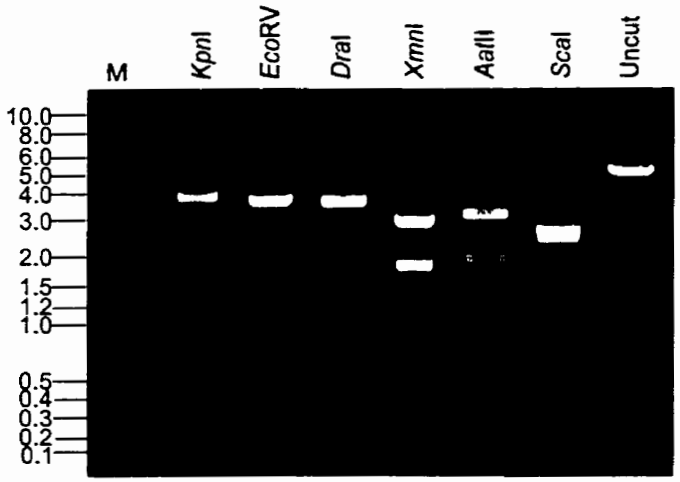
Figure 5.4 Restriction enzyme digests of the 6.3 kb vector-borne (donor) Cμ PCR product amplified from parental lines R1B4#4, R2B2#1, R2B2#4, R1D2#1 and R3D4#1 from which PFCs were recovered. Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

5.2.3 DNA Analysis of Recovered PFCs

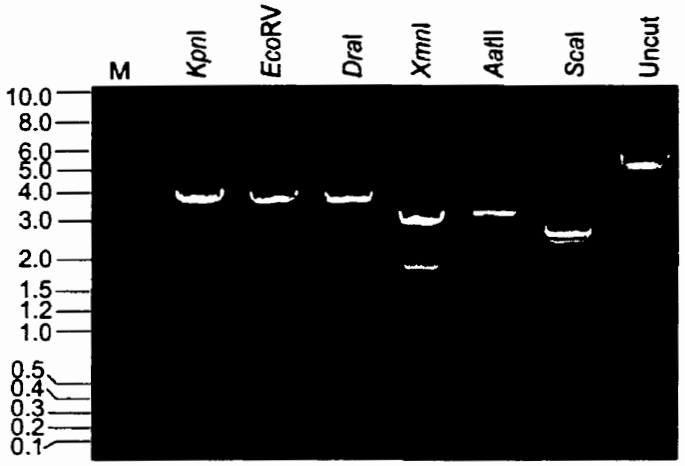
The gene targeting vector sent into igm482 contained a Cμ region marked with the Imut RFLP (Imut) markers but which contained wildtype Cμ exons (Fig. 5.1B). Recombination events between the endogenous (recipient) and vector-borne (donor) Cμ that converted the *Tfi*I site in Cμ3 to an *Xmn*I site restored polymeric IgM production, and were detected in the plaque assay. Such recombinants may display co-conversion of some or all of the Imut markers along with the *Xmn*I site in Cμ3. In order to determine the structure of the recipient (endogenous) Cμ locus in recombinants, PFCs were recovered, and a 4.6 kb product was amplified from the Cμ region using PCR (Fig. 5-2A) and digested with the diagnostic restriction enzymes *Kpn*I, *Eco*RV, *Dra*I, *Xmn*I, *Aat*II and *Sca*I (the Imut markers). A Cμ region bearing all endogenous sites will contain the endogenous intron sites *Ava*II, *Sac*I, *Afl*II, *Ear*I and *Nhe* (Fig. 5.1A) while a Cμ region in which all the endogenous restriction enzyme sites have been converted to vector-borne sites will contain the Imut markers *Kpn*I, *Eco*RV, *Dra*I, *Aat*II and *Sca*I (Fig. 5.1B). The structure of the donor (vector-borne) Cμ in the recombinant lines was determined by amplifying a 6.3 kb PCR product from each PFC line (Fig. 5.2B), and digesting the product with the RFLP enzymes *Kpn*I, *Eco*RV, *Dra*I, *Xmn*I, *Aat*II and *Sca*I.

Ten plaque forming cells (PFCs) were recovered from plaque assay plates and analyzed further. The three PFC lines recovered from parental cell line R1B4#4 (P2, P3, and P4) had the same RFLP pattern; the endogenous *Afl*II, *Tfi*I, *Ear*I and *Nhe*I sites had been converted to the vector sites *Dra*I, *Xmn*I, *Aat*II and *Nhe*I (Fig. 5.5A, B, C). PFC line R2B2#1-1 displayed conversion of the Cμ3 *Tfi*I site and all

(A) R1B4#4-P2



(B) R1B4#4-P3



(C) R1B4#4-P4

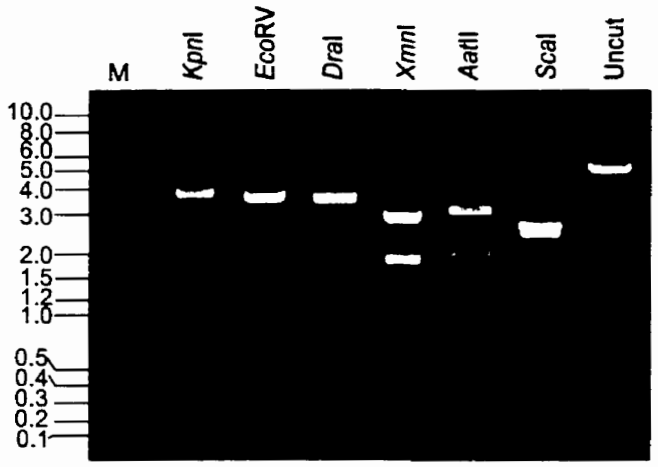


Figure 5.5 Restriction enzyme digests of the 4.6 kb endogenous Cmu PCR product amplified from PFCs recovered from parental cell line R1B4#4. PFC lines are designated Parental 2 (P2), Parental 3 (P3) and Parental 4 (P4). Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

Imut sites except the *Ava*II site located before Cμ1 (Fig. 5.6A). The other PFC lines recovered from the same parent line (R2B2#1-2, R2B2#1-3 and R2B2#1-4) each had the same conversion pattern: the *Tfi*I site in Cμ3 had been converted to *Xmn*I and the *Ear*I site located between Cμ3 and Cμ4 had been converted to the Imut site *Aat*II (Fig. 5.6B, C, D). PFC line R2B2#4-1, recovered from a different subclone of line R2B2, also displayed the same conversion pattern; the only vector-borne sites present in the recipient Cμ were the *Xmn*I site in Cμ3 and the *Aat*II Imut site located between exons Cμ3 and Cμ4 (Fig. 5.7). Only a single PFC line was recovered from line R1D2#1. This line, R1D2#1A-1, displayed conversion of exon Cμ3 to the wildtype *Xmn*I sequence as well as conversion of all the endogenous markers to Imut sites except the *Ava*II site located before Cμ1 (Fig. 5.8). The final PFC recovered, R3D4#1-P1, contained only endogenous sites (Fig. 5.9) and lacked any evidence of gene conversion events. A schematic representation of the RFLP marker patterns found in each of the ten PFC lines is provided in Figure 5.10.

In addition to examining the structure of the recipient (endogenous) Cμ locus in the PFC lines, the structure of the donor (vector-borne) Cμ locus was determined. A 6.3 kb product was amplified from the donor locus then digested with the Imut restriction enzyme set. In all ten PFC lines, the vector-borne *Xmn*I site in Cμ3 as well as the Imut sites were intact (Fig. 5.11, Fig. 5.12, Fig. 5.13, Fig. 5.14, Fig. 5.15).

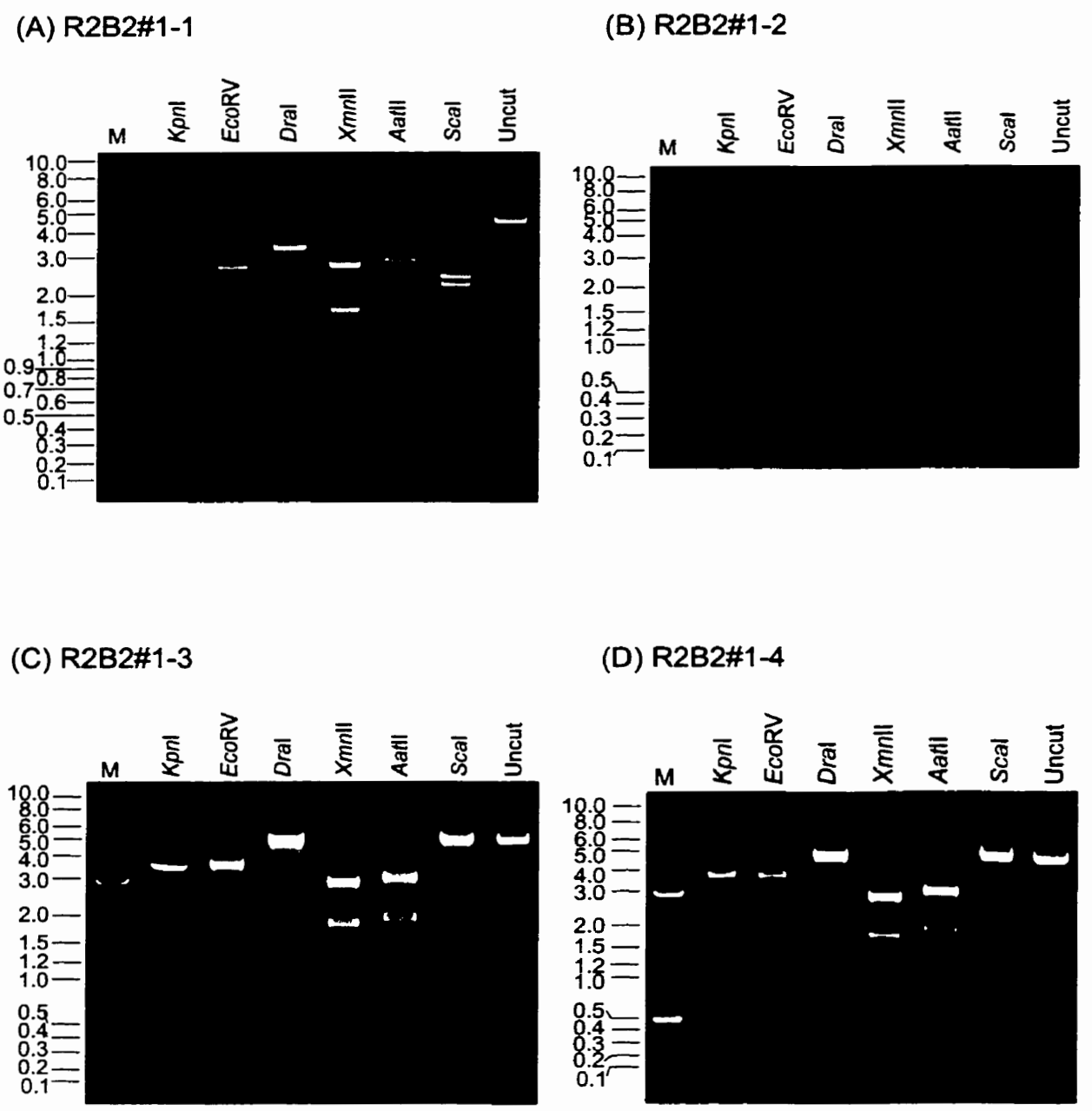


Figure 5.6 Restriction enzyme digests of the 4.6 kb endogenous Cμ PCR product amplified from PFCs recovered from parental cell line R2B2#1. PFC lines recovered are designated -1, -2, -3 and -4. Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

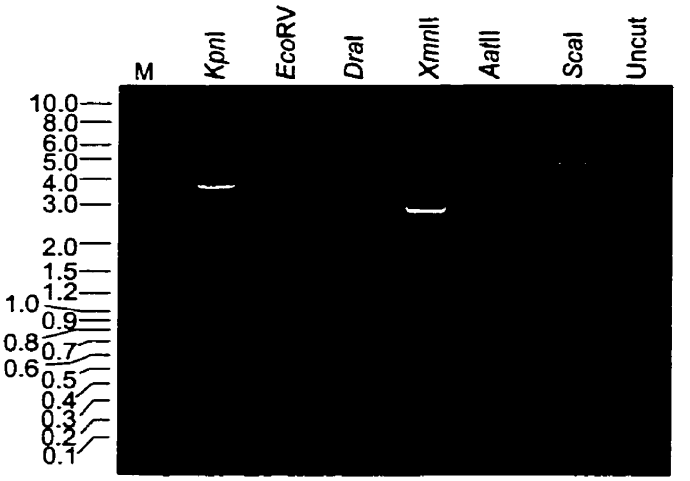


Figure 5.7 Restriction enzyme digests of the 4.6 kb endogenous Cμ PCR product amplified from the PFC recovered from parental cell line R2B2#4 (R2B2#4-1). Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

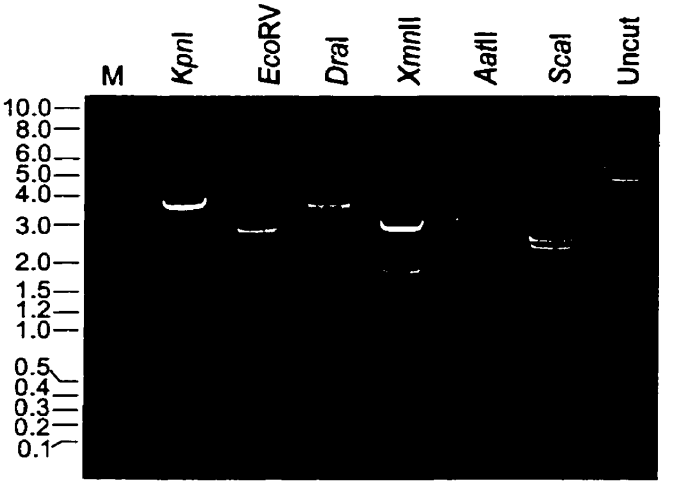


Figure 5.8 Restriction enzyme digests of the 4.6 kb endogenous Cμ PCR product amplified from the PFC recovered from parental cell line R1D2#1 (R1D2#1A-1). Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

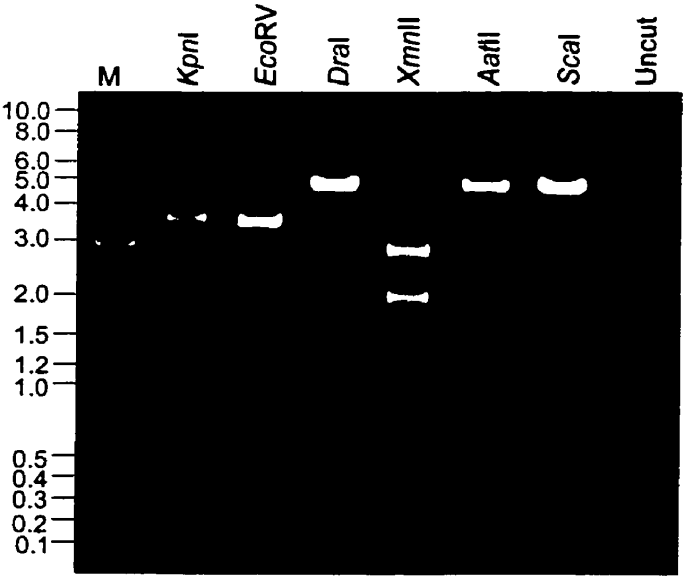


Figure 5.9 Restriction enzyme digests of the 4.6 kb endogenous Cμ PCR product amplified from the PFC recovered from parental cell line R3D4#1 (R3D4#1-P1). Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

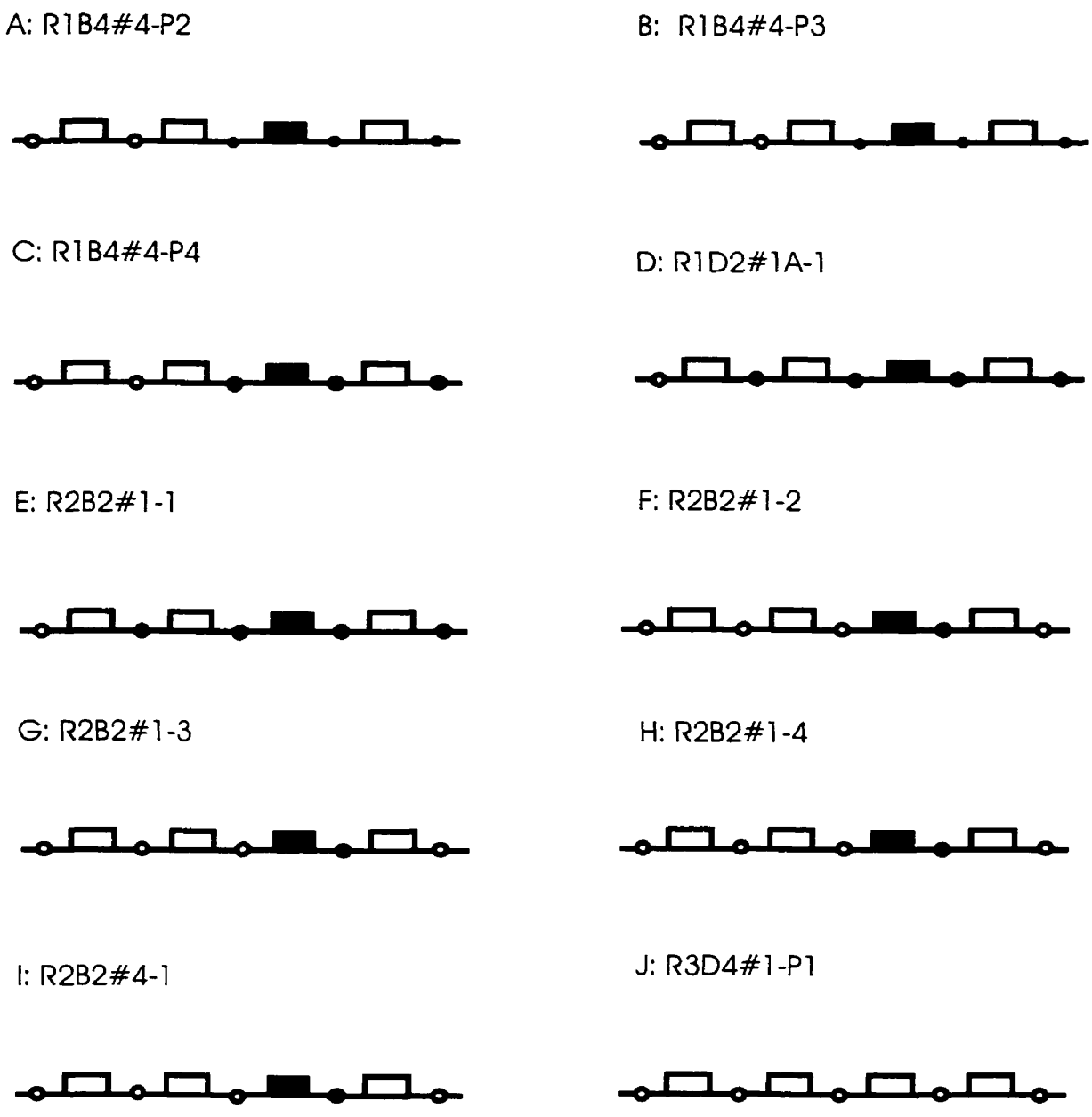


Figure 5.10 Schematic representation of restriction fragment length polymorphism patterns found in the recipient Cmu region of plaque forming cells recovered from plaque assay plates. Open circles/boxes indicate endogenous sequence. Dark circles/boxes indicate conversion to the donor sequence.

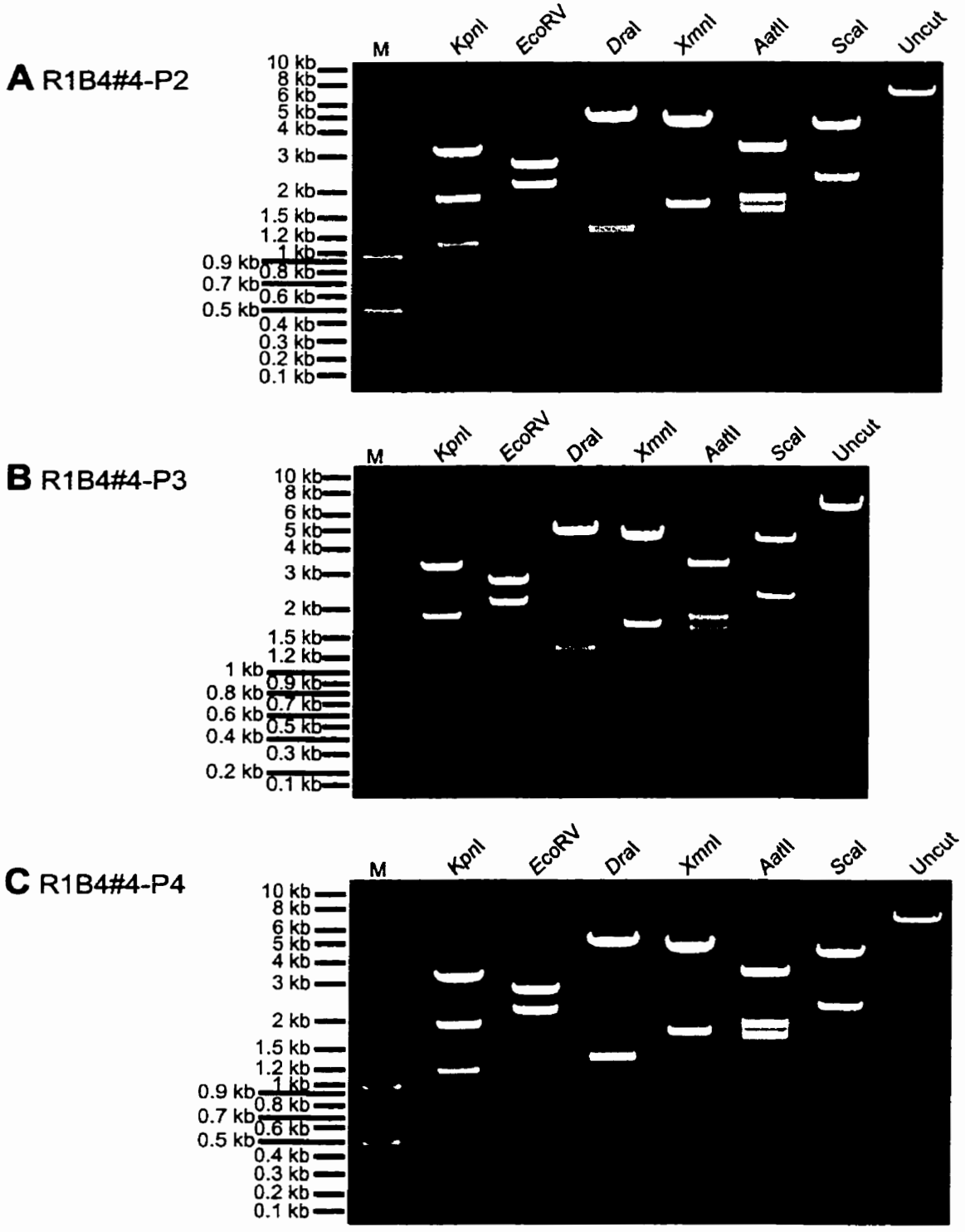


Figure 5.11 Restriction enzyme digests of the 6.3 kb donor Cμ PCR product amplified from PFCs recovered from parental cell line R1B4#4. PFC lines recovered are designated parental 2 (P2), parental 3 (P3) and parental 4 (P4). Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

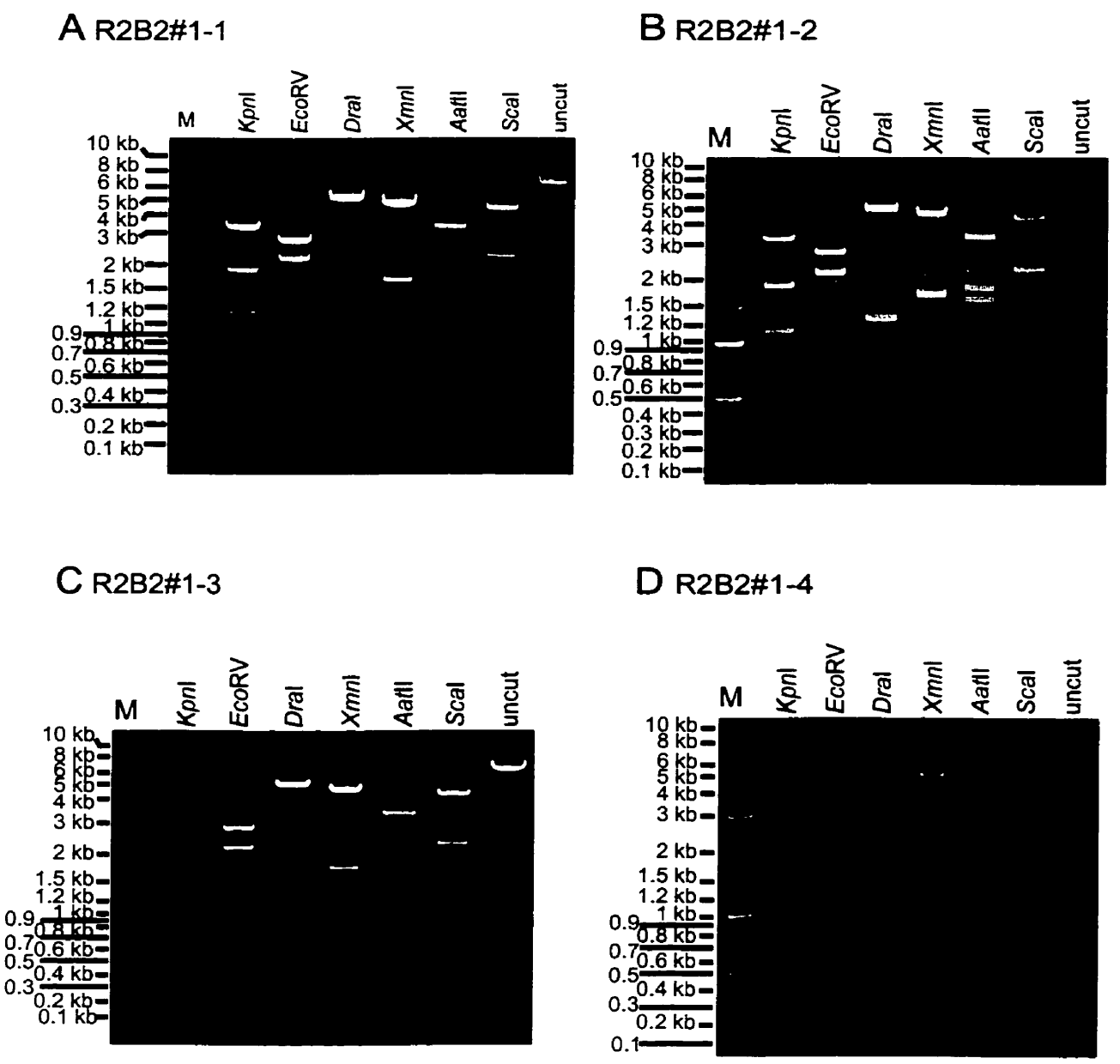


Figure 5.12 Restriction enzyme digests of the 6.3 kb donor Cμ PCR product amplified from PFCs recovered from parental cell line R2B2#1. PFC lines recovered are designated -1, -2, -3, and -4. Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

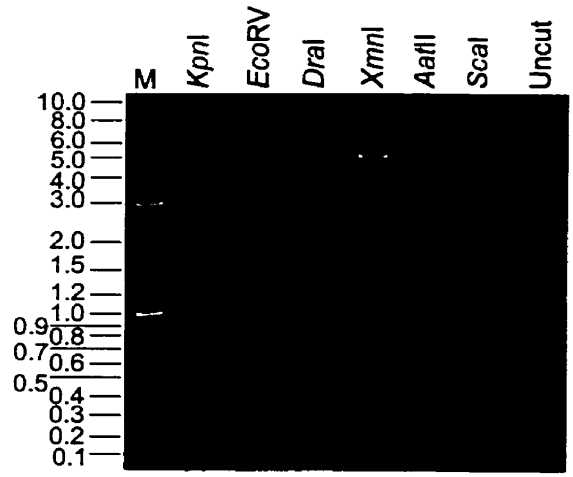


Figure 5.13 Restriction enzyme digests of the 6.3 kb donor (vector-bourne) Cμ PCR product amplified from the PFC recovered from parental cell line R2B2#4 (R2B2#4-1). Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

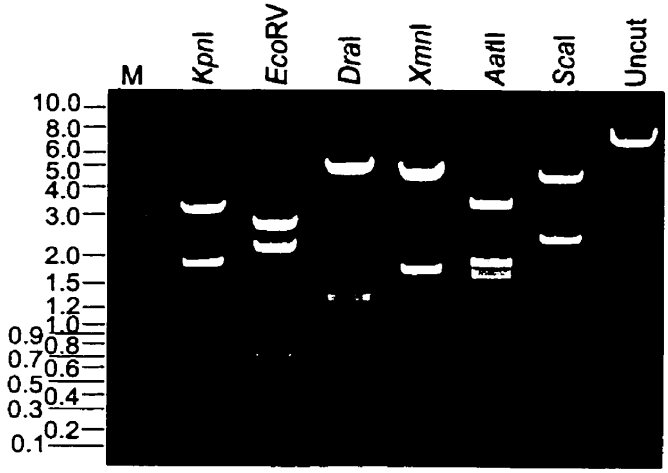


Figure 5.14 Restriction enzyme digests of the 6.3 kb donor (vector-bourne) Cμ PCR product amplified from the PFC recovered from parental cell line R1D2#1 (R1D2#1A). Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

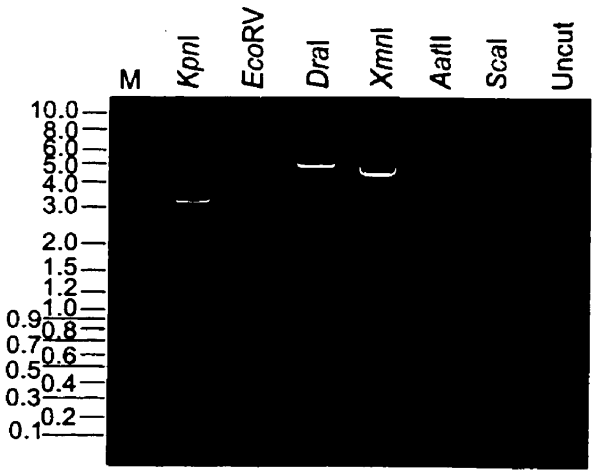


Figure 5.15 Restriction enzyme digests of the 6.3 kb donor (vector-bourne) Cμ PCR product amplified from the PFC recovered from parental cell line R3D4#1 (R3D4#1-P1). Diagnostic fragment sizes for each digest are as indicated in the title of Figure 5-1. The sizes of the marker bands (lane designated M) are indicated to the left of the gel.

5.3 DISCUSSION

There is little data available on the extent of the search for homology in mammalian cells. Intra-chromosomal homologous recombination occurs more frequently than does recombination between unlinked loci (inter-chromosomal). As such, most studies have investigated HR between closely-linked sequences (Bollag and Liskay 1988, Baker 1989, Bollag and Liskay 1991, Nickoloff 1992, Godwin *et al.* 1994, Baker and Read 1995, Donoho *et al.* 1998). Most studies of inter-chromosomal ectopic recombination in cell lines failed to detect any spontaneous recombination events between unlinked sequences (Richardson *et al.* 1998, Richardson and Jasin 2000). Shulman *et al.* (1995) was one of the few groups to detect such recombination events; recombination between unlinked Cμ loci occurred at a frequency of 10^{-7}. Inter-chromosomal ectopic recombination was also observed in the mouse germline at a rate of $\sim 10^{-3}$ (Murti *et al.* 1994).

The Cμ locus of murine hybridoma line igm482 has several features that facilitate the study of HR. First, as the Cμ locus in igm482 is haploid, there is only one potential recipient for homologous recombination events. Second, gene conversion events in which the endogenous Cμ is the recipient of information can be detected based on the reconstitution of a 2 bp deletion in exon Cμ3 (the plaque assay). Third, most studies examine recombination between artificially introduced marker gene heteroalleles such as *tk* or *neo* (Bollag and Liskay 1991, Bollag and Liskay 1992, Taghian and Nickoloff 1997, Richardson *et al.* 1998, Richardson and Jasin 2000). The Cμ recombination system permits the study of recombination between two mammalian genes, with the recipient of information the endogenous

Cμ loci located in a natural chromosomal context. Last, HR occurs at a higher frequency at the Cμ locus than elsewhere in the genome (Table 1.1); this is likely caused by the presence of a recombination enhancing sequence (RES) within the Cμ locus (Raynard *et al.* unpublished). Little published data is available on the types of recombination events occurring during inter-chromosomal recombination. The higher rates of HR within the Cμ locus facilitate the recovery of inter-chromosomal recombination events.

5.3.1 The Search for Homology: Frequencies of Recombination

Previous studies (Baker 1989, Baker and Read 1995) determined that intra-chromosomal recombination between closely-linked Cμ loci occured at a frequency of ~10^{-3} while inter-chromosomal recombination between Cμ genes occurred at a frequency of ~10^{-7} (Shulman *et al.* 1995). An additional study in which the recombining loci were separated by a distance of ~1 MB on the same chromosome found the frequency of recombination to be 10^{-3} (Baker *et al.* 1999), similar to that found between closely linked genes. The Cμ locus in hybridoma line igm482 is haploid, but it is not known whether multiple copies of the Cα locus are present. If, during the gene targeting, the vector integrated into a Cα locus on the same chromosome as the endogenous Cμ, the donor and recipient Cμ loci would be separated by a distance of ~170 kb (Shimizu *et al* 1982). In this study, the frequency of recombination in the targeted line R1D2 is ~$2x10^{-7}$ PFC/cell. As this frequency of recombination is similar to that measured for HR between unlinked Cμ loci (10^{-7}; Shulman *et al.* 1995), the PFCs produced by this line were likely

generated by inter-chromosomal and not intra-chromosomal recombination. As such, the gene targeting event that created R1D2 was likely a targeting event at a Cα locus located on a separate chromosome than the recipient (endogenous) Cμ. This supposition is based solely on the frequencies of recombination observed in targeted line R1D2; further investigation is required for verification. Fluorescence *in situ* hybridization (FISH) mapping of R1D2 with a Cμ-specific probe would indicate the chromosomal location of both the endogenous Cμ locus and the vector-borne copy. This would indicate whether the donor and recipient Cμ loci were located on the same or on separate chromosomes, and thus would reveal whether PFCs were generated by intra- or inter-chromosomal recombination.

The frequency of recombination in hybridoma line R2C1 was twenty-fold lower than that found in R1D2 ($1x10^{-8}$ vs $2x10^{-7}$) (Table 5.1). As with R1D2, the low recombination frequency of line R2C1 (the outcome of a random vector integration) suggests that it is inter-chromosomal recombination is occurring. Again, FISH mapping of the hybridoma line would be required to verify this hypothesis.

Line R2B2 had a recombination frequency fifteen-fold higher than that of R1D2 ($3x10^{-6}$ vs $2x10^{-7}$) and three hundred-fold higher than R2C1 ($3x10^{-6}$ vs $1x10^{-8}$) (Table 5.1). The recombination frequency for line R2B2 ($3x10^{-6}$) is lower than the frequency of 10^{-2} reported for intra-chromosomal recombination between Cμ loci (Baker 1989, Baker and Read 1995), but is higher than the rate of 10^{-7} reported for inter-chromosomal recombination between Cμ loci (Shulman *et al.* 1995). Line R2B2 likely represents a vector integration event on a separate chromosome than the endogenous Cμ locus, but FISH mapping would be required to verify the location

of the donor Cμ. The higher rate of recombination seen in this cell line versus previously measured rates of inter-chromosomal recombination may be due to the site of vector integration. An integration event next to a RES might increase HR between the vector-borne Cμ and the endogenous Cμ locus (already thought to contain a RES). An alternate possibility is that the vector has integrated next to a strong enhancer or promoter. Transcribed genes recombine at a higher frequency than do non-transcribed genes (Nickoloff 1992) and thus if the vector-borne Cμ were inserted next to a strong enhancer/promoter, it may be more likely to undergo HR with the endogenous Cμ locus.

If indeed the donor Cμ sequence in these three cell lines is located on a different chromosome than the endogenous Cμ sequence, then the fact that recombinants are generated by these cell lines indicates that the search for homology is genome-wide. However, the previously established frequencies of intra- and inter-chromosomal recombination indicate that linked sequences are preferentially used as recombination partners over sequences located on separate chromosomes.

5.3.2 Effect of Imut Markers on Homologous Recombination

Rates of recombination are influenced both by the degree of heterology between two recombining loci (Liskay *et al.* 1987, Lukacsovich and Waldman 1999) and by the length of the largest uninterrupted stretch of homology (Waldman and Liskay 1988). The 4.2 Kb donor (vector-borne) Cμ contained five intron mutation (Imut) markers (Fig. 5.1) that were a source of sequence heterology between the

donor and recipient Cμ loci. Initiation of HR is thought to require a minimum amount of uninterrupted homology, measured at ~232 bp for intra-chromosomal HR in mammalian cells (Waldman and Liskay 1988). Once initiation has occurred the HR event can involve regions of sequence heterology (Waldman and Liskay, 1988). In the Cμ recombination system, the distance separating the Imut markers ranged from 91bp (from the *Tfi*I/*Xmn*I site in Cμ3 to the *Ear*I/*Aat*II site) to 554 bp (from the *Sac*I/*Eco*RV site to the *Afl*II/*Dra*I site) (Fig. 5.1). Thus, several tracts of homology of the required length (≥232 bp) exist in which homologous recombination could initiate. Previous studies (Liskay *et al.* 1987) have determined that no detectable intra-chromosomal GC occurs when only 95 bp of homology is available. Accordingly, initiation of recombination likely did not occur in the 91bp between the *Tfi*I/*Xmn*I site in Cμ3 and the *Ear*I/*Aat*II site in intron3. That the homology tract lengths were sufficient for HR is evidenced by the fact that gene targeting with a Imut Cμ vector produced targeted lines at a similar frequency to that found with a completely isogenic Cμ vector (Ng and Baker 1999). The length of homology required for inter-chromosomal HR is at least 1.9 Kb (Baker *et al.* 1996), but the effect of sequence heterologies on recombination between unlinked loci is unclear. As GC events were detected in this study, there must be a sufficient amount of homology to initiate HR in this system. The frequencies of recombination in the cell lines R1D2, R2B2 and R2C1 (2×10^{-7}, 26×10^{-7}, and 0.1×10^{-7} respectively) are similar to previously published frequencies of ectopic inter-chromosomal recombination in the Cμ system (1×10^{-7}; Shulman *et al.* 1995) which suggests that the presence of the RFLP markers in the donor Cμ did not affect the efficiency of HR.

The frequency of HR is also influenced by the overall degree of sequence heterology between the recombining loci (Liskay *et al* 1988, Yang and Waldman 1997, Elliott *et al.* 1998). The percent sequence divergence between the endogenous Cμ and the vector-borne Cμ is minimal (<1%) and likely does not influence the rate of HR. This is evidenced by the fact that Ng and Baker (1999) found that there was no significant difference in targeting frequencies between gene targeting at the Cμ locus with a Imut Cμ vector versus with a isogenic Cμ vector. As well, the frequency of inter-chromosmal HR in a similar study in which the only sequence heterology between the two recombining loci was the 2 bp deletion found in exon Cμ3 (Shulman *et al.* 1995) found a frequency of recombination similar to that found in the current study.

5.3.3 Gene Conversion Events

In the present study, homologous recombination events that resulted in the restoration of the wildtype *Xmn*I site in exon Cμ3 were detected as plaque forming cells. The primers used to amplify the endogenous Cμ locus are specific to the this locus. As such a single crossover event would produce a recombinant Cμ structure in which primers AB9703 and AB9438 (used to amplify the endogenous locus) were no longer linked. As such, the endogenous Cμ locus could not be amplified using PCR. Thus the HR events produced by targeted lines are gene conversion events not associated with single crossovers. During gene conversion events, the haploid endogenous Cμ was the recipient for homologous recombination while the RFLP-marked vector-borne Cμ locus was the donor. It should be noted that a gene

conversion event and a double crossover will result in the same recombinant structure. However, as double crossovers are thought to be extremely rare, gene conversion is the more plausible explanation for such events. Thus, while the HR events in this study are referred to as GC events, it is important to keep in mind that a double crossover is another potential mechanism.

The plaque assay was designed to recover recombinants in which the 2bp deletion in Cμ3 has been converted into the wildtype *Xmn*I site. Spontaneous reversions of this deletion will also be detected, however HR events which do not result in the restoration of the *Xmn*I site will not be recovered. With the exception of R3D4#1-P1, all PFCs recovered in this study represented gene conversion events.

Three PFC lines were recovered from parental cell line R1B4#4 (R1B4#4-P2, R1B4#4-P3, and R1B4#4-P4). These lines each display the same pattern of conversions: exon Cμ3 has been converted to the wildtype *Xmn*I site and vector-borne Imut markers *Dra*I, *Aat*II and *Sca*I are present in the endogenous Cμ (Fig. 5.3). The vector-borne Cμ contains only Imut markers. These three PFC lines were isolated from the same parental culture of R1B4#4, which had not been subcloned prior to their isolation. Given the low rate of HR in the parental line, it is likely that these three PFC lines are all daughter cells (i.e. sibs) of a single GC event.

Four PFC lines were recovered from parental cell line R2B2#1. R2B2#1-1 contains all Imut (vector) sites except for the *Kpn*I site located before exon Cμ1 (Fig. 5.4A). Lines R2B2#1-2, R2B2#1-3 and R2B2#1-4 have identical conversion patterns (Fig. 5.4A, B, C); exon Cμ3 contains the wildtype *Xmn*I site and intron 3

contains the vector-borne Imut marker *Aat*II. Prior to measuring recombination rates in cell line R2B2, the parental culture was subcloned at 0.1cell/well to produce the lines R2B2#1, R2B2#2 etc. Seven subcloned lines were generated in this manner, and plaque assays were used to measure recombination in each line (Table 5.2). Subcloned line R2B2#1 had a recombination rate of $5.6x10^{-4}$ while the other lines had an average rate of recombination of $1.8x10^{-6}$. The higher rate of recombination in line R2B2#1 is likely due to a 'jackpot' event. That is, a GC event early on in the growth of the subcloned culture resulted in a higher number of PFCs in that culture. Because of this jackpot event, there is a possibility that cell lines R2B2#1-2, R2B2#1-3 and R2B2#1-4 (which all have the same RFLP pattern) are sibs. PFC line R2B2#4-1 also has the same conversion pattern (the vector-borne *Xmn*I and *Aat*II sites are present in the endogenous recipient Cμ locus) (Fig. 5.5) but as this line was isolated from a different subcloned line (R2B2#4) it represents an independent isolate.

Cell line R1D2#1 produced 156 plaques but only one PFC was recovered. In this line, the endogenous (recipient) Cμ contained all vector sites except for the *Kpn*I site located before exon Cμ1 (Fig. 5.6).

Cell line R3D4#1-P1 contained all chromosomal markers in the recipient Cμ locus; no Imut markers were present. As the 2 bp deletion in exon Cμ3 had not been corrected to the wildtype *Xmn*I sequence (Fig. 5.7), this line must represent a reversion, not a conversion, event. Reversion of the 2bp deletion in hybridoma line igm482 occurs at a rate of $<1x10^{-7}$ (Baker and Read, 1995), and presumably occurs at the same rate in lines derived from igm482. The reversion event was

likely a single bp deletion or a 2 bp insertion that corrected the 2 bp deletion frameshift mutation in Cμ3. That a reversion event was recovered in an assay for HR events emphasizes that the recombination rates found in these cell lines (10^{-6} to 10^{-8}) are not that much higher than background reversion rates.

Thus, out of the ten PFC lines recovered in the present study, there are five independent GC events and one reversion event. The independent isolates R1B4#4-P2/P3/P4, R2B2#1-1, R2B2#1-2/-3/-4, R2B2#4-1, and R1D2#1A represent conversion events while R3D4#1-P1 is a reversion event.

5.3.4 Gene Conversion and the Double Strand Break Repair Model

Studies involving introduced DSBs have noted that the broken locus is usually the recipient of genetic information during gene conversion, while the unbroken locus acts as a donor of information (Taghian and Nickoloff 1997, Elliott *et al.* 1998). The plaque assay used to capture GC events has an inherent bias towards events in which the 2 bp deletion in exon Cμ3 is corrected to the wildtype *Xmn*I site. Thus, in the context of the DSBR model, initiation of HR between the Cμ loci in the present study probably occurred when a spontaneous DSB appeared in the endogenous (recipient) Cμ locus. Gene conversion would occur if the DSB was enlarged to a gap prior to repair from the unbroken Imut Cμ locus, or if the DSB was resected to form 3' single strands and hDNA was formed as a recombination intermediate. Recent work suggests that the gap formed during gene targeting is relatively small, and that mismatch repair of hDNA is responsible for most gene conversion events (Li and Baker 2000a, b). If intra- and inter-chromosomal HR has

the same mechanism, then the GC seen in the PFC lines was more likely the result of mismatch repair of a hDNA intermediate rather than gap repair.

The endogenous *Ava*II site located before exon Cμ1 was not converted to the Imut *Kpn*I site in any of the six independent GC events. This implies that the DSB that initiated HR occurred some distance away from this site. If GC occurred due to gap repair, then the DSB was enlarged to a gap that did not encompass the AvaII site. If, as is more likely, GC occurred due to mismatch repair of hDNA, then the hDNA did not encompass the *Ava*II/*Kpn*I RFLP site. As stated previously, the plaque assay used in this study has an inherent bias towards GC events in which exon Cμ3 is converted to the wildtype *Xmn*I sequence (these recombinants can be detected as PFCs). Thus a DSB that initiates HR must be close enough to the Cμ3 site that the *Tfi*I to *Xmn*I conversion occurs. GC events in which the endogenous Cμ3 site is not converted to the wildtype *Xmn*I will not be recovered. The *Ava*II*Kpn*I site is located 1829 bp away from *Tfi*I/*Xmn*I site in Cμ3 and is the farthest Imut marker from Cμ3 (Fig. 5.1). The *Ava*II site may be too far away from the initiating DSB to be encompassed in hDNA and thus is not converted to the Imut *Kpn*I sequence.

5.3.5 Transfer of Information During Inter-Chromosomal Gene Conversion is Unidirectional

By definition, gene conversion is the non-reciprocal transfer of genetic information during homologous recombination. In the present study, PCR was used to amplify a 6326 bp fragment from the donor Cμ locus (Fig. 5.2). Diagnostic

restriction enzyme digests of the PCR product were used to determine whether the Imut markers were still present in the vector-borne Cμ locus. In all ten PFC lines recovered, the vector Cμ locus contained only vector-borne Imut markers and the vector-borne wildtype *Xmn*I site in exon Cμ3 (5.11, 5.12, 5.13, 5.14, 5.15). The ability to recover both homologous recombination products provides the opportunity to conclusively demonstrate that the PFC were generated by GC. There is a uni-directional transfer of information from the donor (vector-borne) Cμ, which remained unchanged, to the donor (endogenous) Cμ, in which one or more chromosomal sites were converted to the Imut markers.

5.3.6 Gene Conversion Tracts are Continuous

A least five independent GC events were recovered in the present study: R1B4#4-P1(-P2, -P3), R2B2#1-1, R2B2#1-2(-3, -4), R2B2#4-1 and R1D2#1A-1 (see section 5.3.3). Each of the five isolates contained a continuous, uninterrupted incorporation of the Imut markers. This is in agreement with other studies in which GC tracts were found to be continuous (Liskay and Stachelek 1986,Yang and Waldman 1997, Donoho *et al.* 1998, Elliott *et al.* 1998), although there is a report of a single, apparently discontinuous GC tract (Taghian and Nickoloff 1997). The continuity of the GC tract may be due to either gap repair of a DSB or to mismatch repair of hDNA formed during DSBR. It has been suggested (Elliott *et al.* 1998) that the continuity of GC tracts is due to extensive degradation of DNA ends produced by a DSB or that there is a strong bias towards correction of hDNA in the direction of the unbroken strand. Recent work examining GC during gene targeting (Ng and

Baker 1999, Li and Baker 2000a) suggests that the processing of a DSB usually yields only a small (<645 bp) double stranded gap. If intra- and inter-chromosomal recombination utilize a similar mechanism to gene targeting, then the GC events recovered in the current study are likely due to mismatch repair of hDNA in the direction of the donor (vector-borne) Cμ.

5.3.7 Gene Conversion Tract Length

Several studies have used RFLP markers to examine tract length in the GC products of intra-chromosomal recombination (Liskay and Stachelek 1986, Taghian and Nickoloff 1997, Donoho *et al.* 1998, Elliott *et al.* 1998) but due to the low rate of inter-chromosomal HR, few products of an GC involving unlinked sequences have been recovered. Murti *et al.* (1994) detected GC events between unlinked sequences present in the mouse germline, but did not obtain any information about tract continuity or length. The frequencies of HR in the present study suggest that the recombining Cμ loci are unlinked. If so, it represents the first study in which both products of ectopic inter-chromosomal GC have been recovered. If indeed the recipient and donor Cμ loci are unlinked, the current study would demonstrate that GC (a uni-directional transfer of genetic information) between unlinked sequences occurs in mammalian somatic cells.

The GC tract inR1B4#4 PFCs extends at least 915 bp from the *Afl*II/*Dra*I site to the *Nhe*I/*Sca*I site (Fig. 5.8A, B, C). The tract length in line R1D2#1A-1 and line R2B2#1-1extends from the *Sca*I/*Eco*RV site to the NheI/ScaI site and is at least 1469 bp in length (Fig. 5.8D, E). Lines R2B2#1-2/ -3/-4 and line R2B2#4-1 have

shorter GC tract of at least 91 bp that includes only the *Xmn*I site in Cμ3 and the *Aat*II site between exons Cμ3 and Cμ4 (Fig. 5.8F, G, H, I). The endogenous *Ava*II site was not converted to the Imut *Kpn*I site in any of the PFC lines. The spontaneous DSB that initiated the GC in these lines was likely close to exon Cμ3, as the plaque assay only detects those recombinants in which the wildtype *Xmn*I site is present (see above). As the *Ava*II site was not converted in any of the GC events, this suggests an upper limit to the length of a hDNA recombination intermediate. More data on inter-chromosomal GC tract lengths may lend support to this hypothesis.

Thus the tract lengths observed in the current study range from ≥91 to ≥1469 bp in length, and are longer than those reported for intra-chromosomal recombination. Elliot *et al.* (1998) used RFLP markers and found that the majority of conversion tracts were <58 bp in length with the largest tract spanning only 511 bp. A similar study (Taghian and Nickoloff 1997) found GC tracts from 26 to 979 bp in length. Both studies used artificially introduced DSBs to induce HR. The GC tracts found in the current study are longer than the tracts generated by intra-chromosomal HR. This may reflect a mechanistic difference between the two processes, or may be due to differences inherent in each recombination assay. Lukacsovich and Waldman (1999) have suggested that increased sequence heterologies result in shorter gene conversion tracts. In the current study, there is >1% sequence divergence between the recipient and donor Cμ loci. Both Taghian and Nickoloff (1997) and Elliott *et al.* (1998) found shorter GC tracts than those in the current study, yet both had approximately the same amount of sequence

divergence (0.9% and 0.8% respectively). An examination of DSB-induced inter-chromosomal recombination (Richardson *et al.* 1998) found that most GC tracts were short (<20bp) with 1-3% of GC tracts extending ≥2.5 Kb into regions of non-homology and that inter-homolog GC events produced tracts of ≥2.8Kb in length. These results, taken together with those of the current study, suggest that GC tracts generated by inter-chromosomal HR are larger than those generated by intra-chromosomal recombination. However there is a report of longer GC tracts (up to 3 Kb) involving linked sequences (Donoho *et al.* 1998). This latter study used palindromes, which are thought to escape mismatch repair, rather than RFLP markers. The conflicting data on conversion tract lengths measured with palindrome markers suggests that hDNA may be more extensive than was determined using RFLP markers. Palindromes may produce longer GC tracts if some of the RFLP markers are corrected in the direction of the recipient sequence and some (the conversion tract) are corrected in the direction of the donor sequence. If so, this would imply that the hDNA undergoes discontinous mismatch repair. However most studies (Liskay and Stachelek 1986,Yang and Waldman 1997, Donoho *et al.* 1998, Elliott *et al.* 1998) report that GC tracts are continuous. A more plausible explanation is that GC tract length is dependant upon the recombination assay system utilized.

Further studies to investigate the differences (if any) between intra- and inter-chromosomal GC tracts lengths and between RFLP and palindrome tract lengths could be done using the Imut Cμ recombination system. This would allow direct comparison of any differences in tract lengths.

6. CONCLUSIONS

The goals of the current study were twofold: to determine whether an enhancer-trap vector could be used to target an RFLP-marked donor Cμ to the Cα locus of murine hybridoma line igm482, and to study ectopic homologous recombination between the donor Cμ and the endogenous, haploid, Cμ locus.

The data suggests that enhancer-trap gene targeting to the Cα locus occurred with the same efficiency as targeting to the Cμ locus. In both cases, there was a reduction in the background of non-targeted $G418^R$ cells. The efficacy of the enhancerless vector is likely due to the 3' IgH enhancers, but further research is needed to verify that this is the case.

Recombination frequencies were measured for cell lines R1D2, R2C1 and R2B2. The frequencies of recombination in these lines (10^{-6} to 10^{-8}) are comparable to frequencies measured for inter-chromosomal recombination in the assay system (10^{-7}). This suggests that the gene conversion events in the present study were the result of interactions between unlinked loci. This could be verified by using FISH to determine the chromosomal location of both the endogenous and vector-bourne Cμ loci.

At least five of the PFC lines recovered represented independent recombination events. Characterization of both the endogenous and vector-bourne Cμ loci in these lines demonstrated that there was a unidirectional transfer of information from the donor (vector-bourne) Cμ locus to the recipient (endogenous) locus. These PFCs probably represent gene conversion events, but could also be

the outcome of a double crossover. The tract lengths observed in these PFCs range from ≥91 to ≥1469 bp in length, and are longer than those reported for intra-chromosomal recombination. More examples of GC between unlinked Cμ loci would aid in determining a more accurate measure of tract lengths. The Imut marker located furthest from exon Cμ3 (the presumptive location of an initiating DSB) is never found in the PFCs. This suggests an upper limit to the length of a GC tract, but again, more data is needed to verify this hypothesis.

7. LITERATURE CITED

d'Alençon, E., M. Petranovic, B. Michel, P. Noirot, A. Aucouturier, M. Uzest, and S.D. Ehrlich. 1994. Copy-choice illegitimate DNA recombination revisited. EMBO **13**: 2725-2734.

Arya, S., F. Chen, S. Spycher, D.E. Isenman, M.J. Shulman, and R.H. Painter. 1994. Mapping of amino acid residues in the C mu 3 domain of mouse IgM important in macromolecular assembly and complement-dependent cytolysis. J. mmunol.**152**:1206-1212.

Baker, M.D., N. Pennell, L. Bosnoyan, and M.J. Shulman. 1988. Homologous recombination can restore normal immunoglobulin production in a mutant hybridoma cell line. Proc. Natl. Acad. Sci. USA **85**: 6432-6436.

Baker, M. D. 1989. High-frequency homologous recombination between duplicate chromosomal immunoglobulin μ heavy-chain constant regions. Mol. Cell. Biol. **9**:5500-5507.

Baker, M.D., and L. R. Read. 1992. Ectopic recombination within homologous immunoglobulin μ gene constant regions in a mouse hybridoma cell line. Mol. Cell. Biol. **12**:4422-4432.

Baker, M. D., and L. R. Read. 1995. High-frequency gene conversion between repeated Cμ sequences integrated at the chromosomal immunoglobulin μ locus in mouse hybridoma cells. Mol. Cell. Biol. **15**:766-771.

Baker, M.D., L.R. Read, B.G. Beatty, and P. Ng. 1996. Requirements for ectopic homologous recombination in mammalian somatic cells. Mol. Cell. Biol. **16**:7122-7132.

Baker, M.D., L.R. Read, P. Ng, and B.G. Beatty. 1999. Intrachromosomal recombination between well-separated, homologous sequences in mammalian cells. Genetics. **152**:685-697.

Baker, M.D., and E. C. Birmingham. 2001. Evidence for biased Holliday junction cleavage and mismatch repair directed by junction cuts during double-strand-break repair in mammalian cells. Mol. Cell. Biol. **21**:3425-3435.

Baumann, B., M.J. Potash, and G. Köhler. 1985. Consequences of frameshift mutations at the immunoglobulin heavy chain locus of the mouse. EMBO J. **4**:351-359.

Bautista, D. and M.J. Shulman. 1993. A hit-and-run system for introducing

mutations into the IgH locus of hybridoma cells by homologous recombination. J. Immunol. **151**:1950-1958.

Belmaaza, A., and P. Chartrand. 1994. One-sided invasion events in homologous recombination at double-strand breaks. Mutat. Res. **314**:199-208.

Bollag, R.J., and R.M. Liskay. 1988. Conservative intrachromosomal recombination between inverted repeats in mouse cells: association between reciprocal exchange and gene conversion. Genetics **119**:161-169.

Bollag, R.J., Waldman, A.S., and R.M. Liskay. 1989. Homologous recombination in mammalian cells. Annu. Rev. Genet. **23**:199-225.

Bollag, R.J., and R.M. Liskay. 1991. Direct-repeat analysis of chromatid interactions during intrachromosomal recombination in mouse cells. Mol.Cel.Biol. **11**:4839-4845.

Bollag, R.J., and R.M. Liskay. 1992. Chromatid interactions during intrachromosmal recombination in mammalian cell, p.3-13. *In* Gottesman, M.E. and H.J. Vogel (ed.) Mechanisms of eukaryotic DNA recombination, Academic Press Inc., Toronto, ON.

Bollag, R.J., D.R. Elwood, E.D.Tobin, A.R. Godwin, and R.M. Liskay. 1992. Formation of heteroduplex DNA during mammalian intrachromosomal gene conversion. Mol. Cell. Biol. **12**:1546-1552.

Brenneman, M., F.S. Gimble, and J.H. Wilson. 1996. Stimulation of intrachromsomal homologous recombination in human cells by electroporation with site-specific endonucleases. Proc. Natl. Acad. Sci. USA. **93**:3608-3612.

Chauveau, C., E. Pinaud, and M. Cogne. 1998. Synergies between regulatory elements of the immunoglobulin heavy chain locus and its palindromic 3'locus control region. Eur. J. Immunol. **28**:3048-3056.

Chen F.H., S.K., Arya, A. Rinfret, D.E. Isenman, M.J. Shulman, and R.H. Painter. 1997. Domain-switched mouse IgM/IgG2b hybrids indicate individual roles for C mu 2, C mu 3, and C mu 4 domains in the regulation of the interaction of IgM with complement C1q. J. Immunol. **159**:3354-3363.

Choulika, A., A. Perrin, B. Dujon, and J.Nicolas. 1995. Induction of homologous recombination in mammalian chromosomes by using the I-*Sce*I system of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **15**: 1968-1973.

Cromie, G.A., and D.R. Leach. 2000. Control of crossing over. Mol. Cell. **6**:815-826.

Cunningham, A.J., and A. Szenberg. 1968. Further improvements in the plaque technique for detecting single antibody-forming cells. Immunology **14**:599-600.

Dariavach, P., G.T. Williams, K. Campbell, S. Pettersson, and M.S. Neuberger. 1991. The mouse IgH 3'-enhancer. Eur. J. Immunol. **21**:1499-1504.

Deng, C., and M.R. Capecchi. 1992. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. Mol. Cell. Biol. **12**:3365-3371.

Donoho, G., M. Jasin, and P. Berg. 1998. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. Mol. Cell. Biol. **18**:4070-4078.

Elliott, B., C. Richardson, J. Winderbaum, J.A. Nickoloff, and M. Jasin. 1998. Gene conversion tracts from double-strand break repair in mammalian cells. Mol. Cell. Biol. **18**:93-101.

Godwin, and R.M. Liskay. 1994. The effects of insertions on mammalian intrachromosomal recombination. Genetics **136**:607-617.

Godwin, A.R., R.J. Bollag, D. Christie, and R.M. Liskay. 1994. Spontaneous and restriction enzyme-induced chromosomal recombination in mammalian cells. Proc. Natl. Acad. Sci. USA. **91**:12554-12558.

Gross-Bellard, M.,P. Qudet, and P. Chambon. 1973. Isolation of high-molecular weight DNA from mammalian cells. Eur. J. Biochem. **36**:32-38.

Hasty, P., J. Rivera-Pérez, C. Chang, and A. Bradley. 1991a. Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. Mol. Cell. Biol. **11**:4509-4517.

Hasty, P., J. Rivera-Pérez, and A. Bradley. 1991b. The length of homology required for gene targeting in embryonic stem cells. Mol. Cell. Biol. **11**:5586-5591.

Hastings, P.J., C. McGill, B. Shafer, and J.N. Strathern. 1993. Ends-in *vs.* ends-out recombination in yeast. Genetics **135**:973-980.

Hilliker, A.J. 1985. Assaying chromosome arrangement in embryonic interphase nuclei of *Drosophila melanogaster* by radiation inducted interchanges. Genet. Res. Camb. **47**:13-18.

Hoffmann, G.R. 1994. Induction of genetic recombination: consequences and model systems. Environ Mol Mutagen. **23** Suppl 24:59-66.

Holliday, R. 1964. A mechanism for gene conversion in fungi. Genet. Res. Camb. **5**:282-304.

Holliday, R. 1968. Genetic recombination in fungi. In Replication and Recombination of Genetic Material pp.157-174.

Inbar, O., and M. Kupiec. 1999. Homology search and choice of homologous partner during mitotic recombination. Mol. Cell. Biol. **19**:4134-4142.

Jasin, M. and P. Berg. 1988. Homologous integration in mammalian cells without target gene selection. Genes Dev. **2**:1353-1363.

Joyner, A.L. 1991. Gene targeting and gene trap screens using embryonic stem cells: new approaches to mammalian development. Bioessays **13**:649-656.

Lichter, P., T. Cremer, J. Borden, L. Manuelidis and D.C. Ward. 1988. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. Hum. Genet. **80**:224-234.

Kadyk, L. C., and L. H. Hartwell. 1992. Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. Genetics **132**:387-402.

Kawakami, T., N. Takahashi and T. Honjo. 1980. Complete nucleotide sequence of mouse immunoglobulin μ gene and comparison with other immunoglobulin heavy chain genes. Nucleic Acids Res. **8**:3933-3945.

Köhler, G., M.J. Potash, H. Lehrach, and M.J. Shulman. 1982. Deletions in immunoglobulin mu chains. EMBO J. **1**:555-563.

Köhler, G., and M. J. Shulman. 1980. Immunoglobuline M mutants. Eur. J. Immunol. **10**:467-476.

Li, J., and M.D. Baker. 2000a. Use of a small palindrome genetic marker to investigate mechanisms of double-strand-break repair in mammalian cells. Genetics **154**:1281-1289.

Li, J., and M.D. Baker. 2000b. Formation and repair of heteroduplex DNA on both sides of the double-strand break during mammalian gene targeting. J. Mol. Biol. **295**:505-516.

Li, J., and M.D. Baker. 2000c. Mechanisms involved in targeted gene replacement in mammalian cells. Genetics **156**:809-821.

Li, J., L. R. Read, and M.D. Baker. 2001. The mechanism of mammalian gene replacement is consistant with the formation of long regions of heteroduplex DNA associated with two crossing over events. Mol. Cell. Biol. **21**:501-510.

Lichten, M., and J. E. Haber. 1989. Position effects in ectopic and allelic mitotic recombination in *Saccharomyces cerevisiae*. Genetics **123**:261-268.

Lieberson, R., S.L. Giannini, B.K. Birshtein, and L.A. Eckhardt. 1991. An enhancer at the 3' end of the mouse immunoglobulin heavy chain locus. Nucleic Acids Res. **19**:933-937.

Lieberson, R., Ong, J., Shi, X., and L.A. Eckhardt. 1995. Immunoglobulin gene transcription ceases upon deletion of a distant enhancer. EMBO **14**:6229-6238.

Lin, F., K. Sperle, and N. Sternberg. 1984. Model for homologous recombination during transfer of DNA in mouse L cells: role for DNA ends in the recombination process. Mol. Cell. Biol. **4**:1020-1034.

Liskay, R.M., and J.L. Stachelek. 1983. Evidence for intrachromosomal gene conversion in cultured mouse cells. Cell **35**:157-165.

Liskay, R.M., J.L. Stachelek and A. Letsou. 1984. Homologous recombination between repeated chromosomal sequences in mouse cell. Cold Spring Harb. Symp. Quant. Biol. **49**:183-189.

Liskay, R.M., and J.L. Stachelek. 1986. Information transfer between duplicated chromosomal sequences in mammalian cells involved contiguous regions of DNA. Proc. Natl. Acad. Sci. USA **83**: 1802-1806.

Liskay, R.M., A. Letsou, and J.L. Stachelek. 1987. Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics **115**:161-167.

Lukacsovich, T., and A.S. Waldman. 1999. Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. Genetics **151**:1559-1568.

Madisen, L. and M. Groudine. 1994. Identification of a locus control region in the immunoglobulin heavy-chain locus that deregulates c-*myc* expression in plasmacytoma and Burkitt's lymphoma cells. Genes Dev. **8**:2212-2226.

Mansour, A.M., K.R. Thomas, and M.R. Capecchi. 1988. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature **336**:348-352.

Matthias, P. and D. Baltimore. 1993. The immunoglobulin heavy chain locus contains another B-cell-specific 3' enhancer close to the α constant region. Mol. Cell. Biol. **13**:1547-1553.

Meselson, M.S., and C.M. Radding. 1975. A general model for genetic recombination. Proc. Natl. Acad. Sci. USA **72**:358-361.

Michaelson, J.S., S.L. Giannini, and B.K. Birshtein. 1995. Identification of 3'α-hs4, a novel Ig heavy chain enhancer element regulated at multiple stages of B cell differentiation. Nucleic Acids Res. **23**:975-981.

Moynahan, M.E., and M. Jasin. 1997. Loss of heterozygosity induced by a chromosomal double strand break. Proc. Natl. Acad. Sci. USA. **94**:8988-8993.

Murti, J.R., M. Bumbulis and J.C. Schimenti. 1994. Gene conversion between unlinked sequences in the germline of mice. Genetics **137**:837-843.

Nezlin, R. S. 1998. Immunoglobulin Classes. In The Immunoglobulins structure and function. Academic Press, New York, NY. pp. 4-9.

Ng, P., and M.D. Baker. 1998. High efficiency site-specific modification of the chromosomal immunoglobulin locus by gene targeting. J. Immun. Meth. **214**:81-96.

Ng, P., and M.D. Baker. 1999. Mechanisms of double-strand break repair during gene targeting in mammalian cells. Genetics **151**:1127-1141.

Nickoloff, J.A. 1992. Transcription enhances intrachromosomal homologous recombination in mammalian cells. Mol. Cell. Biol. **12**:5311-5318.

Ong, J., S. Stevens, R.G. Roeder, and L.A. Eckhardt. 1998. 3' IgH enhancer elements shift synergistic interactions during B cell development. J. Immunol. **160**:4896-4903.

Orr-Weaver, T. L., J. W. Szostak, and R. J. Rothstein. 1981. Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. USA **78**:6354-6358.

Petes, T.D., R.E. Malone, and L.S. Symington. 1991. Recombination in Yeast, p. 407-521. *In* J.R. Broach, J.R.Pringle and E.W.Jones (ed.), The Molecular and Cellular Biology of the yeast *Saccharomyces*: Genome Dynamics, Protein Synthesis, and Energetics. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

Pettersson, S., G.P. Cook, M. Brüggemann, G.T. Williams, and M.S. Neuberger.

1990. A second B cell-specific enhancer 3' of the immunoglobulin heavy-chain locus. Nature **344**:165-168.

Raynard, S., and M.D. Baker. 2001. upublished.

Richardson, C., and M. Jasin. 2000. Coupled homologous and non-homologous repair of a double-strand break preserves genomic integrity in mammalian cells. Mol.Cell.Biol. **20**:9068-9075.

Richardson, C., M.E. Moynahan, and M. Jasin. 1998. Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. Gen.Dev. **12**:3831-3842.

Sambrook, J., E. F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sargent, R.G., M.A. Brenneman, and J.H. Wilson. 1997. Repair of site-specific double-strand breaks in a mammalin chromosome by homologous and illegitimate recombination. Mol. Cell. Biol. **17**:267-277.

Shimizu, A., N. Takahashi, Y. Yaoita, and T. Honjo. 1982. Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. Cell **28**:499-506.

Shulman, M.J., N. Pennell, C. Collins, and N. Hozumi. 1986. Activation of comlement by immunoglobulin M is impaired by the substitution serine-406 → asparagine in the immunoglobulin μ heavy chain. Proc. Natl. Acad. Sci. USA **83**:7678-7682.

Shulman, M.J., C. Collins, A. Connor, L.R. Read, and M.D. Baker. 1995. Interchromosomal recombination is suppressed in mammalian somatic cells. EMBO. **14**:4102-4107.

Smithies, O., R.G. Gregg, S.S. Boggs, M. A. Koralewski, and R. S. Kucherlapati. 1985. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature **317**:230-234.

Stachelek, J.L. and Liskay, R.M. 1988. Accuracy of intrachromosomal gene conversion in mouse cells. Nucl. Acids. Res. **16**: 4069-4076.

Sun, H., D. Treco, and J.W. Szostak. 1991. Extensive 3'-overhanging single-stranded DNA associated with the meisosi-specific double-strand breaks at the ARG4 recombination initiation site. Cell **64**: 1155-1161.

Szostak, J. W., T. L. Orr-Weaver, and R. J. Rothstein. 1983. The double-strand-break repair model for recombination. Cell **33**:25-35.

Taghian, D. G. , and J. A. Nickoloff. 1997. Chromosomal double-strand breaks induce gene conversion at high frequency in mammalian cells. Mol. Cell. Biol. **17**:6386-6393.

Tang, R. S. 1994. The return of copy-choice in DNA recombination. BioEssays **16**:785-788.

Taylor, B., J. F. Wright, S. Arya, D. E. Isenman, M. J. Shulman, and R. H. Painter. 1994. C1q binding properties of monomer and polymer forms of mouse IgM μ-chain variants. J. Immunol. **145**:5303-5313.

Valancius, V., and O. Smithies. 1991. Double-strand gap repair in a mammalian gene targeting reaction. Mol. Cell. Biol. **11**:4389-4397.

Viguera, E., D. Canceill, and S. D. Ehrlich. 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO J. **20**:2587-2595.

Wabl, M. R., and P. D. Burrows. 1984. Expression of immunoglobulin heavy chain at a high level in the absence of a proposed immunoglobulin enhancer element in cis. Proc. Natl. Acad. Sci. USA **81**:2452-2455.

Waldman, A.S., and R.M. Liskay. 1987. Differential effects of base-pair mismatch on intrachromosmal versus extrachromosomal recombinaiton in mouse cells. Proc. Natl. Acad. Sci. USA. **84**:5340-5344.

Waldman, A.S., and R.M. Liskay. 1988. Dependance of Intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol. Cell. Biol. **8**:5350-5357.

Whitby, M. C., and J. Dixon. 1998. Substrate specificity of the SpCCE1 holliday junction resolvase of Schizosaccharomyces pombe. J. Biol. Chem. **273**:35063-35073.

Word, C.J., J. F. Mushinski, and P.W. Tucker. 1983. The murine immunoglobulin α gene expresses multiple transcripts from a unique membrane exon. EMBO **2**:887-898.

Yang, D., and A.S. Waldman. 1997. Fine-resolution analysis of products of intrachromosomal homeologous recombination in mammalian cells. Mol. Cell. Biol. **17**:3614-3628.

Zhang, J., F.W. Alt, and T. Honjo. 1995. Regulation of class switch recombination

of the immunoglobulin heavy chain genes. In Immunoglobulin Genes, Second ed. Honjo, T. and F.W. Alt, eds. (Academic Press, Toronto, Canada) pp. 235-265.