GENE TARGETING IN EMBRYONIC STEM CELLS AND THE MANIPULATION OF MILK COMPOSITION

Satish Kumar

Ph. D.
University of Edinburgh
1993

"If we magnified a hen's egg to the size of the world (which would make atoms rather larger than eggs and electrons barely visible) we could still get a gene into a room and probably on to a small table. But such a magnification being impossible, the question how to interfere with a single gene without interfering with the others becomes serious, and some men have already spent their lives vainly on it; many more will. The two most hopeful methods seem to be to find chemical substances which will attack one gene and not another; and to focus ultraviolet rays on a fraction of a chromosome."

J.B.S. Haldane
(1892-1964)
'The Future of Biology',
In: On Being the Right Size and Other Essays.
J. M. Smith (ed.), Oxford University Press, 1985.

DECLARATION

I declare that I have written this thesis based on my own work, and the contribution of others has been clearly indicated.

ABSTRACT

Successful modification of milk composition through genetic manipulation requires an understanding of the role of milk constituents in determination of its properties. Gene targeting in the mouse is a powerful tool to analyse the *in vivo* function of gene products. β-casein is one of the most abundant proteins in milk, mainly existing as a component of casein-micelles. To find out the role of β-casein in lactation and in the properties of milk, mice carrying mutated β-casein genes were created through gene targeting in embryonic stem cells using a replacement-type vector. β-casein was absent from the milk of homozygous mutant mice, and no mRNA was detectable from the mutant allele. The mutation was compatible with viability, fertility and lactation. The total protein concentration of β-casein-deficient milk was reduced, although the loss of β-casein was partially compensated by enhanced synthesis and/ or secretion of other milk proteins. The pre-weaning growth of pups feeding on β-casein-deficient milk appeared to be slower, probably reflecting the reduction in the total milk protein concentration. The reduction (in heterozygotes) or absence (in homozygotes) of β-casein in milk resulted in smaller casein micelles, probably due to increases in the ratios of κ-casein to calcium-sensitive caseins. These results show that β-casein is dispensable in lactation.

Targeting of subtle mutations is likely to be important for many purposes, including the manipulation of milk structure. The 'hit and run'

method for targeting of subtle mutations would be facilitated by positive-negative selection during the first (targeted insertion) step.

To explore the feasibility of negative selection in conjunction with insertion-type gene targeting vectors, the effects of placing terminal heterologies on a hypoxanthine phosphoribosyltransferase (HPRT) insertion vector were determined. The targeting was generally accurate but the targeting efficiency was reduced by terminal heterologies. The magnitude of reduction was dependent on the length and position of the heterology. An increase in the length of heterology resulted in a more pronounced decrease in targeting efficiency. Heterology on the short (0.8 kb) arm was less detrimental than heterology on the long (4.8 kb) arm. Surprisingly, a vector with heterologies on both arms targeted with high fidelity, and with targeting efficiencies equal to or higher than those of a vector with heterology on the long arm only. These data imply that in principle, it should be possible to use positive-negative selection with insertion-type vectors; the actual advantage will depend on the level of enrichment obtained. Further, the results indicate that terminal sequences are removed simultaneously from both arms of the vectors, and that the removal of at least one strand of the DNA was effected by exonuclease(s) and was tightly coupled to the process of gene targeting.

ACKNOWLEDGEMENTS

I am grateful to Dr. J. Paul Simons, currently Lecturer, Department of Anatomy and Developmental Biology, Royal Free Hospital School of Medicine, London for his guidance, supervision and encouragement during my Ph.D studies at the Institute of Animal Physiology and Genetics Research, Edinburgh Research Station, Roslin.

I thank Dr. Martin L. Hooper, Reader, Department of Pathology, Medical School, University of Edinburgh, for a number of discussion sessions, the use of his laboratory facilities for the β-casein gene targeting experiments and for arranging the mouse blastocyst microinjections. All members of his laboratory, particularly Alan Clarke, A. Peter and L. Dobbie were very helpful in guiding me through the embryonic stem (ES) cell technology.

I thank Dr. A. J. Clark, Head Department of Molecular Genetics, Institute of Animal Physiology and Genetics Research, Edinburgh Research Station (now Roslin Institute) for providing resources for my project. All members of his Department were always ready to help me. Particularly, I thank Mr. John Bowman for making the stores available, many a times at very short notices!

I thank Professor Jeffrey Rosen, Baylor College of Medicine, Houston, Texas (USA) for the β-casein genomic clones, and Dr. D. Melton, Division of Biological Sciences, University of Edinburgh for the HPRT cDNA and genomic clones. Measurement of casein micelle size included in this thesis was performed by Drs. Jeffrey Leaver and David Horne of the Hannah Research Institute, Ayr, Scotland.

Ray Ansell arranged various supplies and solutions in the tissue culture laboratory. I also thank her for LIF and batch tested bovine foetal and newborn calf sera. Embryo transfers were performed by Roberta Wallace. Frances Thomson arranged for births, marriages and the well-being of my mice.

I thank Fiona Yull for the counselling sessions on 'How to cope with wrong bands on Southern blots'. I have greatly benefited from

discussions with Sam Al-Murrani, Bert Binas, Pamela Brown, Kenneth Dobie, Andrew Law, Maggie McClenaghan and Bruce Whitelaw. I thank Anthea Springbett for advice, and Norrie Russell, Roddy Field and Elliot Armstrong for photography and graphics.

I was financially supported by the Government of India, and the Committee of Vice-Chancellors and Principals of the United Kingdom. The Indian Council of Agricultural Research allowed me leave from my job.

John Webster kept my sense of humour (sarcasm to be precise!) alive. I am indebted to him and his family for keeping me a social being. I thank Dr. Alan Archibald for providing me daily transport from Edinburgh to Roslin. Dr. John Ansell has been a continuous source of valuable advice.

Pramod, my brother-in-law has been responsible to take care of all my worries. I am indebted to my parents for their inspiration. I thank Drs. B.R. Yadav and D.S. Balain for support.

To pursue my Ph.D. project, I had to leave my family thousands of miles away to fend for themselves. Any attempt on my part to express in some words my deep sense of indebtedness to my wife and friend Hemlata shall always be wanting. And of course, Babloo and Mannu had to take all the heat of this!

Satish Kumar

to our ma

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

6-TG	6-thioguanine
BLG	β-lactoglobulin
bp	base pair
BRL	Buffalo rat liver
BSA	bovine serum albumin
CsCl	caesium chloride
CTL	cytotoxic T-lymphocyte
d.p.c.	*days post-coitus*
dCTP	2'-deoxycytidine 5'-triphosphate
DEPC	diethyl pyrocarbonate
DMEM	Dulbecco's modified Eagle's medium
DNAase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate
DSBR	double-strand break repair
DT-A	diphtheria toxin fragment A
EDTA	ethylenediamine tetra-acetic acid
EG	embryonic germ
EGTA	Ethylene Glycol -bis(β-amino-ethyl ether) N, N, N', N' -tetra acetic acid
ES	embryonic stem
FCS	foetal calf serum
GANC	gancyclovir
HAT	hypoxanthine/ aminopterin/ thymidine
HEPES	N-2-hydroethylpiperazine-N'-2-ethane sulphonic acid
HPRT	hypoxanthine-guanine phosphoribosyl transferase
HSV	herpes simplex virus
ICM	inner cell mass
IU	international unit
kb	kilobase pairs
kD	kilodalton

LB	Luria broth
LIF/DIA	Leukaemia inhibitory factor/ differentiation inhibitory factor
μF	micro-Farad
NCS	newborn calf serum
neo	neomycin
NPN	non-protein nitrogen
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PGC	primordial germ cells
PNS	positive negative selection
PVA	polyvinyl alcohol
Rb	retinoblastoma
RNAase	ribonuclease
RT	room temperature
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
TCA	trichloro acetic acid
tk	thymidine kinase
UTR	untranslated region
UV	ultraviolet
V	volt
WAP	whey acidic protein

CHAPTER ONE

INTRODUCTION

Milk production is an essential component of the reproduction strategy of mammals. The development of a zygote into an organism is accompanied by preparation of milk synthesis potential by the mammary gland of the mother. Milk is a complex mixture of more than a thousand ingredients. The main constituents of milk are water, proteins, lipids, carbohydrates and minerals, the relative concentrations of which have evolved in most species to meet the needs of the neonate during their pre-weaning period. However, in farm animals the milk composition is an outcome of artificial selection by man for higher milk yield to meet his nutritional and socio-economic needs. Milk is processed into a variety of products by altering its composition. The quality, yield and economics of various milk products are dependent upon the relative amounts and the biochemical and physical properties of individual constituents of milk. Thus, there is strong incentive to genetically modify milk composition (Bremel et al. 1989; Clark 1992; Hettinga 1989; Jimenez-Flores and Richardson 1988; Mercier 1986; Muysson and Gibbins 1989).

Advances in the understanding of genetic control of milk synthesis and the advent of transgenic technology have provided revolutionary methods to alter milk composition. It has been possible through gene transfer to produce novel proteins in the mammary gland of laboratory and farm animals, and similar approaches are currently being extended

to modify milk composition to improve the nutritional and physical properties of milk.

In this chapter the structure of milk, and the scope of conventional breeding in the modification of milk composition, with emphasis on bovine milk, are described. The progress in the production of novel proteins in milk is discussed. Finally, the current status of the available methods for manipulation of mouse genome through gene targeting in embryonic stem cells, and their potential application in the manipulation of milk composition are reviewed.

1.1. MILK COMPOSITION

There is a wide qualitative and quantitative variation among milks from different species (Table 1.1). In a comparison of milk composition of

Table 1.1. Average concentration of major constituents of milk of a number of species (modified from Davies et al. 1983).

Species	Fat (gl^{-1})	Protein (gl^{-1})	Lactose (mM)	Calcium (mM)
Cow	37	34	133	30
Goat	45	29	114	22
Sheep	74	55	133	58
Pig	68	48	153	104
Man	38	10	192	7
Mouse	NA	97*	NA	NA
Rat	103	84	90	80
Guinea Pig	39	81	83	41
Rabbit	183	136	60	214

* See Chapter 3. NA: not available.

30 species (Jenness 1986) protein concentration varied from 10-200 gkg^{-1}, fat from a trace to 500 gkg^{-1} and carbohydrate from a trace to 100 gkg^{-1}. It has been found that within a particular species genetic and non-genetic factors influence milk composition (reviewed by Davies et al. 1983; Mepham 1987). Non-genetic factors include the frequency of suckling, stage of lactation and plane of nutrition. The genetic variation in cow's milk composition is discussed in section 1.2.

1.1.1. Milk Proteins

Most of the protein secreted in milk is synthesised in the mammary gland from transcription of a small number of genes. Variable amounts of serum albumin, immunoglobulin and a range of enzymes are transferred from blood to the milk in the mammary alveoli. Broadly, milk proteins can be classified into two categories, viz. caseins which are precipitated at acidic pH (4.5-4.6 for bovine milk), and whey proteins which are left in the milk serum after acid precipitation of caseins. The relative proportion of these two groups of milk proteins varies among species (Jenness 1986). While in cow and mouse milk caseins constitute approximately 80% of the total protein, whey proteins are the more abundant proteins in human and Indian elephant milk.

1.1.1.1. Caseins

Caseins are a rich and balanced source of amino acids and serve as vehicles for the transfer of calcium and phosphorus from the mother to the neonate. Casein synthesis is initiated in the terminally differentiated mammary gland epithelium during mid-pregnancy in

response to peptide and steroid hormonal signals and continued throughout the lactation (see review: Vonderhaar and Ziska 1989).

The bovine casein fraction comprises four individual components, namely α_{s1}-, α_{s2}-, β- and κ-caseins. Comparisons of casein cDNA sequences from a number of species have shown that each milk has homologues of some or all of the four bovine caseins (Bonsing and Mackinlay 1987) and that β- and κ-caseins are present in all milks (Creamer 1991; Davies et al. 1983). The concentration of a particular casein varies between milks of different species (Table 1.2).

Caseins are phosphoproteins and, in all species studied in detail, can be grouped in two broad classes depending upon the number of phosphorylated serine residues in them. The cow is the best characterised species in terms of its milk proteins and their properties. α_{s1}-, α_{s2}-, and β-caseins have up to 9, 13 and 5 phosphorylated serine residues, respectively, through which these proteins interact with calcium phosphate and aggregate to form colloidal particles called casein micelles. These proteins can be precipitated in the presence of low Ca^{2+} concentration hence, are known as Ca^{2+}-sensitive caseins. κ-casein lacks a cluster of phosphorylated serine residues, and therefore, is insensitive to Ca^{2+} and acts to stabilise the growing micelles by limiting their size (Schmidt 1982). Proteolysis of bovine κ-casein by chymosin at Phe (105)-Met (106) bond results in precipitation of caseins which leads to curd formation. In rat and mouse κ-casein, methionine 106 has been replaced by leucine (Thompson et al. 1985).

In bovine milk, the casein micelles vary in size from 15 nm to 600 nm. The casein composition of micelles changes with the micelle size (Dalgleish et al. 1989; Davies and Law 1983; Donnelly et al. 1984; McGann et al. 1980) and this information has been utilised to

predict the localisation of various caseins. McGann et al. (1980) and Donnelly et al. (1984) separated casein micelles by chromatography on controlled-pore-glass and reported that with the decreasing micelle size the proportion of κ-casein increased and that of α_s- and β-caseins decreased. A linear relationship between the κ-casein concentration and micelle surface area-to-volume ratio was interpreted as indicating surface

Table 1.2. Milk protein composition of various milks
(modified from Lathe et al. 1986).

Milk protein	Concentration in milk (g l^{-1})				
	Cow	Sheep	Mouse	Goat	Human
Caseins					
α_{s1}-casein	10.0	12.0	28.0†	0.7	0.4
α_{s2}-casein	3.4	3.8	NDA	4.0	
β-casein	10.0	16.0	21.0†	10.0	3.0
κ-casein	3.9	4.6	2.4†	6.0	1.0
Whey proteins					
α-lactalbumin	1.0	0.8	trace	1.2	1.6
β-lactoglobulin	3.0	2.8	none	2.3	none
Whey acidic protein	none	none	2.0		none
Serum albumin	0.4	NDA	NDA		0.4
Lysozyme	trace	NDA	NDA	trace	0.4
Lactoferrin	0.1	NDA	NDA	0.1	1.4
Immunoglobulins	0.7	NDA	NDA	0.5	1.4

† E. Stevenson, J. Leaver and A. J. R. Law (personal communication).

NDA: No data available.

N.B.: Nomenclature of caseins follows that of bovine caseins. Mouse homologues are α-, ε- and β-casein, respectively; see review by Bonsing and Mackinlay (1987).

localisation of κ-casein. Davies and Law (1983) fractioned casein micelles by differential centrifugation and found that κ-casein proportion increased with the decrease in micelle size. The relative proportion of α_{s1}-casein remained constant but β-casein fraction decreased as the micelle size decreased. Dalgleish et al. (1989) obtained data similar to those of Davies and Law (1983) using ultracentrifugation for the fractionation of casein micelles and photon correlation spectroscopy for the determination of micelle size. These authors concluded that the micelle surface has a predominance of κ-casein and α_{s1}-casein and small amounts of β-casein.

A number of models have been proposed for the structure of casein micelles (reviewed by Creamer 1991; Rollema 1992 Ruettimann and Ladisch 1987). Broadly, these can be grouped in three categories: coat-core models, internal structure models and sub-unit models. As per the coat-core models, casein micelles are formed by a core of one or more caseins and the remaining caseins form the coat around the core. *In vitro* spontaneous assembly of α- and β-caseins into micellar structures, and κ-casein concentration dependent size variability of these micelles was taken as evidence in support of these models. Internal structure models postulated even distribution of caseins inside the micelles. Neither of these two types of models accounted for the evidence obtained from electron microscopy that casein micelles consists of sub-micelles (see below); and the internal structure models were not consistent with the evidence suggesting surface localisation of κ-casein (see Ruettiman and Ladisch 1989 for the brief history of coat-core and internal structure models).

According to the sub-unit model (Schmidt 1982; Fig. 1.1), casein micelles are thought to be assembled from the sub-micelles of variable composition which are held together by clusters of calcium

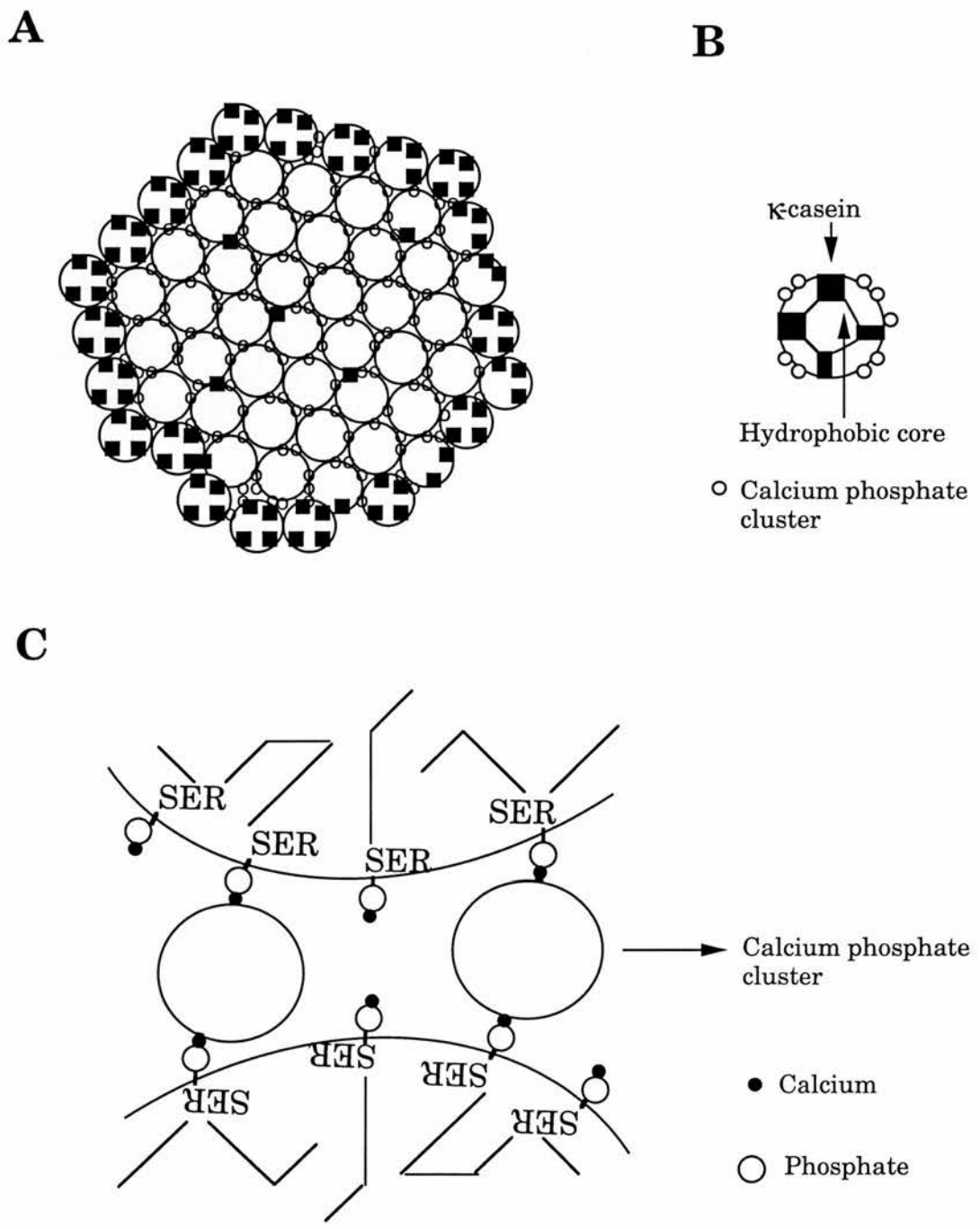


Figure 1.1. Sub-unit casein micelle model.
Redrawn from Schmidt (1982) and Mepham (1987).
A) Micelle.
B) Sub-micelle.
C) Binding between two sub-micelles.

phosphate [$Ca_9(PO_4)_6$]. This model is an extension of an earlier model proposed by Slattery and Evard (1973). The latter workers introduced the concept of sub-micelles of variable composition and argued that κ-casein was located on the surface of micelles. However, they did not postulate any primary role of calcium phosphate in the interactions between sub-micelles and maintained that hydrophobic binding between caseins was responsible for such interactions. In the model of Schmidt (1982), sub-micelles comprise α_{s1}-, α_{s2}-, β- and κ-caseins, the amount of κ-casein being variable. Further, most sub-micelles with κ-casein are present on the surface of the micelles and hence, the micelle size is determined by the fraction of κ-casein in micelles (see above). Ono and Obata (1989) have proposed two types of sub-micelles in their model. By gel chromatography these authors isolated two main types of sub-micelles; one consisting of α_s- and β-caseins and the other one included α_s- and κ-caseins with equimolar ratio. Ono and Obata (1989) could assemble artificial micelles of varying sizes by modulating the relative ratio of the above two types of sub-micelles. The implied importance of α_s-caseins in this model is not corroborated by the fact that goat milk has a very small proportion of α_s-caseins (see Creamer 1991). The precise structure of sub-micelles and the relative roles of the various Ca^{2+}-sensitive caseins remain unclear.

1.1.1.2. Whey proteins

β-lactoglobulin and α-lactalbumin are the two major whey proteins synthesised in the ruminant mammary gland and secreted in milk. β-lactoglobulin may be important in vitamin A transport (Papiz et al. 1986). This protein is absent from rodent milk. However, rodent mammary gland synthesises another whey protein i.e whey acidic protein

(WAP, Hennighausen and Sippel 1982). The function of this protein in mammary gland or milk is not known. α-lactalbumin has a role in lactose synthesis (Kuhn 1983; see section 1.1.3 below). Lactoferrin and transferrin are iron binding proteins in milk and these are synthesised in the mammary gland (Davies et al. 1983). Human milk is rich in lactoferrin; rodents lack this protein and instead have transferrin. Both of these proteins are present in ruminant milk.

Immunoglobulins in milk are derived mainly from the blood and provide passive immunity to the offspring (Mepham 1987). In ruminants the transfer of immunoglobulin from mother is after birth via milk, and IgG is the main class of immunoglobulin in milk. In humans the placenta allows prenatal transfer of IgG and IgA is the main form of immunoglobulin present in milk. Albumin is derived from the blood serum and is present in low levels in human and bovine milk in contrast to rat milk where albumin is the major whey protein (reviewed by Davies et al. 1983).

1.1.2. Milk Fat

Although milk fat levels vary considerably between species, there is little qualitative variation: triglycerides are the predominant component of the lipid fraction in all species examined so far (reviewed by Davies et al. 1983 and Dils 1986). Other minor fractions include di- and monoacylglycerols, cholesterol and its esters, phospholipids and free fatty acids. Long-chain fatty acids utilised for synthesis of triglycerides are derived either from blood lipids or from *de novo* synthesis in the mammary gland. Short- and medium-chain fatty acids are more abundant in most milks in contrast to other body tissues in the respective

species. The relative abundance of short-chain fatty acids is a characteristic feature of ruminant milks. Milk fat is secreted as membrane bound spherical globules (from 0.1 to 20 μm in diameter in bovine milk). Milk lipid concentration and protein concentration are positively correlated in many species (Mepham 1987).

1.1.3. Milk Carbohydrates

Lactose, a disaccharide of glucose and galactose is the major carbohydrate in the majority of milks. It is synthesised within the Golgi vesicles. α–lactalbumin, a whey protein acts as a cofactor in the synthesis of lactose by the enzyme, galactosyltransferase. Lactose complements the role of diffusible ions in maintaining the osmotic balance of milk. Consistent with this role of lactose, its concentration is inversely proportional to that of diffusible ions (Jenness 1986). Milk fat and protein concentrations are also negatively correlated with the lactose level in various species (Mepham 1987).

1.2. MODIFICATION OF MILK COMPOSITION BY CONVENTIONAL BREEDING

There are a number of reasons for envisaging modification of milk composition. Milk is processed into a variety of milk products; the nutritive value and economics of these products are determined by milk composition. The physical behaviour of various milk constituents during milk processing determines the technological properties of milk. Milk may not be a suitable food in some circumstances; for example many human adults are intolerant of lactose because of lactase deficiency.

Historically, conventional animal breeding has been used to increase the level of fat in dairy cattle but the current trend of lower fat consumption in dairy products may require a change in breeding objectives. Physical separation of fat from milk is a readily available option but in the longer term if the requirement of fat is likely to be reduced it may be desirable to breed cattle for reduced milk fat.

1.2.1. Quantitative Genetic Variation

Naturally existing genetic variation between and within breeds is the raw material to modify a quantitative trait. Assuming that there is sufficient economic incentive for change, the rate of genetic gain is dependent upon the heritability (or accuracy of selection), the generation interval, the genetic correlation among various traits under selection and the intensity of selection (Falconer 1989).

Gibson (1987, 1989) reviewed the literature to examine the feasibility of altering milk composition through selection in dairy cattle. Protein and fat concentrations differ among breeds of cattle and there exists a positive correlation between these two traits. Holstein is the most widely used breed for milk production and if the objective is to reduce the fat:protein ratio, there is little to gain from replacement of this breed (Table 1.3). In the Holstein breed, fat and protein concentrations are moderately heritable and these traits have a low coefficient of variation (Table 1.4). Gibson (1989) argued that direct selection for change in the fat concentration would allow alteration of the fat:protein ratio. It may however, be noted that the low coefficient of variation in fat concentration would preclude any major modification, and the positive genetic

correlation between fat and protein concentrations would be antagonistic to such a selection objective (Table 1.4).

No separate estimations of genetic parameters of casein and whey protein levels are available, but a positive genetic correlation between these two traits would not be unexpected. Assuming that is the case, direct selection for increased milk protein or casein concentration would be accompanied by correlated response in whey protein level. In

Table 1.3. Milk composition of five dairy breeds (from Gibson 1987).

Breed	Lactation milk yield (kg)	Concentration (gkg^{-1})			Fat: protein
		Fat	Protein	Lactose	
Holstein-Friesian	7073	37.0	31.1	46.1	1:1.19
Ayrshire	5247	39.9	33.4	46.3	1:1.20
Jersey	4444	51.3	38.0	47.0	1:1.35
Guernsey	4809	48.7	36.2	47.8	1:1.35
Brown Swiss	5812	41.6	35.3	48.0	1:1.18

Table 1.4. Genetic correlations, heritabilities and coefficients of variation of major milk constituents (from Gibson 1987).

Trait	Coefficient of variation	Y	F	P	L
Liquid yield (Y)	0.22	0.27	-0.27	-0.18	0.01
Fat concentration (F)	0.09		0.47	0.55	0.22
Protein concentration (P)	0.08			0.48	0.02
Lactose concentration (L)	0.07				0.28

The underlined values are heritabilities and the remaining values are genetic correlations.

milk production systems where caseins are the main basis for economic returns such as in cheese manufacture; there would be little to gain from increased whey proteins in milk. Similarly, the low coefficient of variation in lactose concentration would not allow significant change in lactose concentration of milk. Due to the sex-limited nature of milk composition traits, it is not possible to apply direct selection in males. In conclusion, while conventional selection has been extremely successful for increasing milk yield, it is likely to be slow and offers limited choice in altering milk composition.

1.2.2. Qualitative Genetic Variation

Genetically determined polymorphic variants of milk proteins have been studied by protein gel electrophoresis to identify variants with desirable effects on the properties and composition of milk (reviewed by Grosclaude and Martin 1992; Lin et al. 1992). Recently, it has become possible to identify these polymorphisms in both sexes using DNA based-techniques (Damiani et al. 1990; Denicourt et al. 1990; Medrano and Aguilar-Cordova 1990).

A large number of workers have investigated the effect of various genetic variants of milk proteins on the composition and physical properties of milk (reviewed by Lin et al. 1992). Three consistent associations have been noted: (1) the κ-casein B allele has been associated with a higher level of κ-casein and a shorter rennet time (Aleandri et al. 1990; McLean et al. 1984; Schaar et al. 1985; Van den Berg et al. 1992). (2) the β-lactoglobulin A allele has been associated with higher β-lactoglobulin and lower casein, α-lactalbumin and fat levels (Aleandri et al. 1990; Hill 1993; McLean et al. 1984; Ng-Kwai-Hang et al. 1990;

Schaar et al. 1985; Van den Berg et al. 1992). (3) In goats, the A, B and C alleles at the α_{s2}-casein locus have been associated with higher milk casein concentration than the D, E and F alleles (Grosclaude et al. 1987; Mahé and Grosclaude 1989). The A, B and C alleles have also been correlated with high fat content and cheese yields, and a selection programme is underway to increase the frequency of these alleles in the Alpine breed (Grosclaude and Martin 1992). It may be noted that selection for A, B and C alleles at the α_{s2}-casein locus in goats to increase the casein content of milk would be expected to be accompanied by increased fat content of milk which may be contrary to long-term breeding objectives. Similarly, decreasing the frequency of the A allele at the β-lactoglobulin in cattle may increase the casein concentration but not without an undesirable effect on the fat contents of milk.

It remains a possibility that the above described associations may not be the primary effects of the milk protein genes themselves but could be due to some other linked genes. Particularly, it must be noted that all the caseins genes are linked within 200 kb of DNA in cattle (Ferretti et al. 1990; Threadgill and Womack 1990) and therefore, it is difficult to unequivocally dissect the effects of various casein alleles on qualitative and quantitative properties of milk.

1.3. MODIFICATION OF MILK COMPOSITION BY GENE TRANSFER

Successful introduction of functional exogenous genes into the mouse germ line through microinjection of cloned genes into fertilised eggs (Palmiter et al. 1982; reviewed by Palmiter and Brinster 1986) led to the proposal that extension of similar approaches to milk protein genes

would allow germ line modification of milk composition in farm animals (Clark et al. 1987; Lathe et al. 1986). These authors argued that regulatory sequences of milk protein genes fused with coding sequences of a human protein gene of pharmaceutical importance would target the expression of the cognate protein into the secretory epithelium of the mammary gland, and the pharmaceutical protein could readily be harvested from milk. Jimenez-Flores and Richardson (1988) proposed alteration of the nutritional and physical properties of milk by over-expression of the native milk protein genes or by expression of modified versions of some of these genes in transgenic cows. In the following sections the progress in the application of gene transfer to modification of milk composition has been reviewed.

1.3.1. Expression of Heterologous Proteins in Transgenic Mouse Milk

A number of milk protein genes and fusion genes driven by the regulatory sequences of milk protein genes have been transferred to the mouse genome to target expression in the mammary gland (reviewed by Clark et al. 1993; Hennighausen et al. 1990; Rosen et al. 1989; Wilmut et al. 1990). These studies have been undertaken to locate the putative regulatory sequences in the rat (Lee et al. 1988), rabbit (Bühler et al. 1990) and goat (DiTullio et al. 1992; Persuy et al. 1992) β-casein, bovine α_{s1}-casein (Meade et al. 1990; Rijnkels et al. 1993), mouse whey acidic protein (McKnight et al. 1992 and references therein), and sheep β-lactoglobulin genes (reviewed by Clark et al. 1993), and to optimise the transgene design for synthesis of foreign proteins and secretion into the milk of transgenic farm animal.

Milk protein genes from various species when transferred to the mouse genome are expressed in the mammary gland (Bayna and Rosen 1990; Lee et al. 1988; Simons et al. 1987; Vilotte et al. 1989). Position-independent high level expression of sheep β-lactoglobulin gene in mice (Whitelaw et al. 1992) has suggested that this gene may have most of the sequences necessary for overriding the position-effect (Al-Shawi et al. 1990) generally encountered with transgenes (Jaenisch 1988; Palmiter and Brinster 1986). Casein genes, except the goat β-casein gene (Persuy et al. 1992), are generally expressed inefficiently in transgenic mice even when several kb of upstream and downstream sequences are included in the transgene (Lee et al. 1988; Rijnkels et al. 1993). Thus, expression data on the casein transgenes do not exclude the possibility of co-ordinated regulation of casein locus by elements shared amongst linked casein genes (Rosen et al. 1989) as has been demonstrated for β-globin locus (reviewed by Kollias and Grosveld 1992). It is possible to achieve a high level of WAP gene expression but the levels are variable and the element(s) responsible for temporal regulation and copy number-dependent expression are yet to be identified (Bayna and Rosen 1990). Matrix attachment region sequences from chicken lysozyme locus (reviewed by Sippel et al. 1992) can restore the endogenous regulation pattern of the WAP transgene but are not sufficient to overcome the position-effect (McKnight et al. 1992).

Heterologous cDNAs under the control of milk protein gene regulatory sequences are expressed at extremely low levels (Bühler et al. 1990; McClenaghan et al. 1991; Pittius et al. 1988; Whitelaw et al. 1991). Incorporation of introns from the heterologous genes into the transgenes may enhance the level of expression (Archibald et al. 1990) as has been found with a number of other transgenes (Brinster et al. 1988; Choi et al.

1991; Palmiter et al. 1991) but it may be at the cost of tissue specific expression. An α_1-antitrypsin minigene, lacking its first intron and driven by 4 kb of the β-lactoglobulin gene upstream sequences was expressed in 10 out of 15 lines analysed (Archibald et al. 1990). The expressing lines had variable levels of mRNA in the mammary gland or salivary gland or both of these tissues. Very poorly expressed cDNA-based transgenes can be rescued by co-integration with an efficiently expressed intact BLG gene (Clark et al. 1992). It is worth noticing that the level of expression of the rescued transgene was at least 20 fold lower than that of the BLG gene.

Insertion of heterologous cDNAs into the 5' untranslated region of the WAP gene may be compatible with a reasonable level of expression (Velander et al. 1992) in contrast to fusion genes promoted by 2.6 kb of WAP gene upstream sequences (Pittius et al. 1988). However, expression of analogous constructs derived from the BLG gene was extremely inefficient (McClenaghan et al. 1991). Given the very high level of expression of the intact BLG gene (Simons et al. 1987) it is intriguing that disruption of exon 1 of this gene by various cDNAs is detrimental to the expression. It seems unlikely that insertion of cDNA destroys some vital element in the BLG gene responsible for efficient expression as intermission of exon 1 of this gene in the hybrid construct of Archibald et al. (1990) was compatible with high level expression. Most workers have assayed the transgene expression by examining the level of heterologous proteins and/ or steady state mRNA levels. These data alone are insufficient to infer that the poor expression of the cDNA based transgenes is due to the failure of transcription. It is possible that in some of these examples the problems may lie downstream of transcription.

1.3.2. Production of Heterologous Proteins in Milk of Farm Animals

Manipulation of the genome by pronuclear injection has been successfully applied to produce transgenic sheep, pigs, cattle and goats; although the efficiency of transgenesis is low and numerous logistic problems remain (reviewed by Clark et al. 1990; Wall and Seidel Jr. 1992). A number of transgenes have been transferred to the genome of livestock species to produce biomedical proteins (Clark et al. 1989; Ebert et al. 1991; McClenaghan et al. 1991; Velander et al. 1992; Wright et al. 1991) or heterologous milk proteins (Krimpenfort et al. 1991; Wall et al. 1991) in milk.

Generally, the performance of hybrid genes in livestock species reflects the expression levels achieved in the transgenic mice. Thus a fusion gene derived from sheep BLG gene and genomic sequences from the human α_1-antitrypsin (hα_1-AT) gene has been expressed at high level in sheep milk (Wright et al. 1991). The concentration of hα_1-AT in sheep milk ranged from 1.5-37.5 gl^{-1}; in one animal this protein represented approximately 50% of the total milk protein. The hα_1-AT or human anti-haemophilic factor IX cDNA inserted into exon I of BLG gene was expressed at extremely low levels in the milk of transgenic sheep (Clark et al. 1989) as was the case in mice (McClenaghan et al. 1991). An analogous hybrid construct carrying human protein C cDNA within exon 1 of the mouse WAP gene directed the synthesis of protein C and secretion into the milk of transgenic pigs (Velander et al. 1992). The level of protein C varied from 1-1000 μgml^{-1} in six founder animals, and the protein C level varied five fold between two successive lactations in the two animals analysed. Expression of the mouse WAP gene in some

transgenic pig lines has resulted in agalactia (Shamay et al. 1992; Wall et al. 1991), as was the case in mice (Burdon et al. 1991). The authors have suggested that precocious expression of WAP may be responsible for this phenotype. However, it cannot be concluded whether this phenotype is due to the WAP transcript or to the protein produced. This information may have implications for the design of fusion genes incorporating WAP genomic sequences. WAP regulatory sequences fused to human-tissue type plasminogen activator (hTPA) cDNA and SV40 polyadenylation site have been used to direct the synthesis of hTPA and secretion into goat milk (Ebert et al. 1991). Consistent with the results obtained in mice with this construct (Pittius et al. 1988) the level of hTPA in goat milk was low.

In conclusion, the literature on the synthesis of heterologous proteins in the mammary gland and secrertion into the milk of farm animals suggests that low level of expression in many early studies might have been due to non-optimum design of fusion genes (Clark et al. 1989; Ebert et al. 1991; McClenaghan et al. 1991), and the example of α_1-AT (Wright et al. 1991) clearly demonstrates that it is possible to modify the milk composition by gene transfer. However, the low efficiency of transgenesis in farm animals, and the unpredictable performance of fusion genes resulting from the interaction of the transgene design and lack of control of the site and structure of transgene locus are likely to limit any rapid progress.

1.3.3. Modification of Properties of Milk

Unlike conventional animal breeding, genetic engineering has offered the possibility of creation of new genetic variation in milk composition (section 1.2). Existing gene transfer methods allow addition

of novel proteins in bovine milk (see section 1.3) or an increase in the level of a native milk protein. The emerging gene targeting technology (discussed below in section 1.4) holds the promise for deletion of individual constituents of milk and precise modification in endogenous milk proteins to improve their nutritional or physical properties.

Approximately 80% of the proteins in bovine milk exist as a colloidal complex of casein micelles (see section 1.1.1.1). The physical behaviour of milk during its processing is determined to a great extent by the properties of these micelles and their interactions with whey proteins (reviewed by Dalgleish et al. 1993). Although the precise organisation of the casein micelles is not understood, there is sufficient evidence to suggest that quantitative and qualitative genetic variation in caseins may alter the physical properties of casein micelles. Thus, a number of proposals have been made in the literature to modify the properties of casein micelles either by manipulating the relative proportions of various caseins or by changing the primary structure of a given protein (Bremel et al; 1989; Clark 1992; Jimenez-Flores and Richardson 1988; Kang and Richardson 1985; Muysson and Gibbins 1989).

The smaller casein micelles have a higher proportion of κ-casein (Dalgleish et al. 1989; Davies and Law 1983; Donnelly et al. 1984; McGann et al. 1980), and are more heat stable (Fox 1982). Kang and Richardson (1985) have proposed that an increase in κ-casein in bovine milk would result in heat stable milk and this would avoid unwanted gelation during the thermal processing of milk products. Jimenez-Flores and Richardson (1988) have suggested that the addition of 20-30% more β-casein may enhance curd firmness, rennet clotting time and the rate of whey expulsion. An increase in the level of β-casein is likely to alter the ratio of the calcium-sensitive casein to κ-casein, and in

turn, the micelle size may also change. Thus, over-expression of the β-casein gene in bovine germ line may not result in all the desired changes in milk properties. Alternatively, Clark (1992) has argued that since curd firmness and renneting properties are determined by the extent of phosphorylation of caseins, the creation of an additional phosphorylation site in the β-casein would achieve the same objective. The advantage of such a modification may be that the ratio of various caseins would remain the same. However, it is not possible to predict that the proposed phosphorylated centre (substitution of Val with Glu in the sequence Ser_{164}-Leu-Ser-Gln-Ser-Lys-Val_{170}) would be available for phosphorylation; there is also an assumption that the mammary gland would have the capacity for this additional phosphorylation. Similarly, the extra phosphorylation may interfere with the organisation and/ or secretion of micelles. Construction of additional phosphorylation sites has also been suggested for α_{s1}-casein to increase its thermal stability (Kang and Richardson 1985). Ripening of Cheddar cheese is thought to occur by proteolytic disintegration of micelles by the action of chymosin on α_{s1}-casein. Kang and Richardson (1985) have suggested that addition of another chymosin cleavage site may enhance the rate of cheese maturation.

Cow's milk even after considerable processing is not suitable for many human infants. Some of the changes to make it suitable for consumption may include removal of β-lactoglobulin (Mercier 1986) and addition of human lactoferrin (Krimpenfort et al. 1991). Bovine milk has relatively a low level of lactoferrin whereas in human milk this protein is one of the major whey proteins. Krimpenfort et al. (1991) have reported production of a transgenic bull carrying a human lactoferrin transgene driven by bovine α_{s1}-casein gene regulatory sequences. It will be some

time before the milk from the daughters of this bull would be available for analysis.

Lactose intolerance is a problem in Indo-European populations because some adults are unable to metabolise lactose present in milk. It has been suggested that the lowering of the expression of α-lactalbumin, a essential cofactor in the biosynthesis of lactose might reduce the concentration of lactose in milk (Clark et al. 1992; Mercier 1986). Complete removal of lactose may be ideal for nutritional as well as milk processing reasons but this may not be compatible with its role in milk secretion, as lactose is responsible for the secretion of water into the milk (section 1.1.3).

The preference for less fat in diet is growing and it may be commercially desirable to reduce the level of fat in the milk. Bremel et al. (1989) have suggested that the reduction in the level of acetyl CoA carboxylase (ACC), a key enzyme responsible for *de novo* synthesis of fat in the mammary gland, may lead to the reduction in the level of fat in milk. As ACC is central to the fat metabolism of all cells it may not be possible to interfere with the expression of this enzyme. One potential way to circumvent this problem may involve deleting the expression of ACC specifically in the mammary gland without affecting its expression in any other tissue. This could be achieved by targeting two *lox*P sites flanking an exon of ACC gene and mammary specific expression of bacteriophage PI Cre recombinase (Orban et al. 1992).

It should be emphasised that none of the potential applications discussed above has been realised yet. In addition to the logistic problems associated with the production of transgenic livestock, it is not feasible to modify endogenous genes with the existing gene transfer methods in livestock species. Milk is a complex system and thus, the

consequences of even the modifications which are possible with the current technology cannot be predicted beforehand. Understanding of the functions of individual milk proteins in determination of properties of milk and its behaviour during processing will be necessary before attempting modifications in milk of farm animals. The mouse has been serving as a useful model to optimise transgene design to target expression to the mammary gland of farm animals (section 1.3). The availability of gene targeting technology in the mouse has offered a powerful tool to find out the *in vivo* function of any gene (section 1.4). In the short term, application of this technology to the casein genes should help to address the role of these proteins in the organisation of casein micelles, and in determination of physical properties of milk. In the longer term, it may be possible to replace the various mouse milk proteins genes with those of cattle. Such an animal model system would allow quick and reliable evaluation of the effects of various modification intended in the bovine caseins (Clark 1992). The current status of the gene targeting technology has been reviewed in the next section.

1.4. MANIPULATION OF THE MOUSE GENOME BY GENE TARGETING

Transgenesis by pronuclear microinjection of cloned DNA has made a tremendous impact on our understanding of biology and its potential is being realised in solving some practical problems (see section 1.3.3). However, some of the limitations associated with this technology are: 1). There is no control over the site of integration of the injected DNA, and generally multiple copies of transgene are integrated in tandem arrays; both of these factors contribute to the uncertainty of the

phenotype. 2). Normally, only gain of function mutations are feasible and the endogenous gene expression cannot be manipulated, although it is possible to make dominant negative modifications of multimeric proteins (Herskowitz 1987; Levin et al. 1993) and the antisense RNA approach has been used successfully in a few cases (Katsuki et al. 1988; Pepin et al. 1992). The above limitations of transgenic technology have been overcome in mouse genetics by the availability of embryonic stem (ES) cells and the ability of these cells to support homologous recombination between chromosomal genes and extraneously introduced DNA fragments. ES cells are pluripotential cells derived from the inner cell mass (ICM) of the blastocyst (Evans and Kaufman 1981; Martin 1981) and can be propagated *in vitro* under conditions which inhibit their differentiation, at the same time preserving pluripotency. When injected into developing mouse blastocysts these cells can resume normal development and contribute to most or all cell lineages including the germ line (Bradley et al. 1984). While in culture these cells can be genetically manipulated and rare clones with pre-determined changes in endogenous genes can be isolated and be used to reconstitute animals (reviewed by Baribault and Kemler 1989; Bradley et al. 1992; Capecchi 1989a; 1989b; Evans 1989; Koller and Smithies 1992; Robertson 1991).

During the last decade it was shown that DNA fragments when introduced into mammalian cells could participate in homologous recombination with the endogenous genes (reviewed by Bollag et al. 1989). Application of homologous recombination in embryonic stem cells has meant that pre-defined mutations can be introduced in the germ line of the mouse. Since Thompson et al. (1989) showed the germ line transmission of the targeted correction of a null mutation in the mouse HPRT gene in embryonic stem cells, application of gene targeting

technology has pervaded almost every aspect of animal biology research (reviewed by Bradley et al. 1992; Koller and Smithies 1992; Robertson 1991).

1.4.1. Mouse Embryonic Stem Cells

In 1981, two laboratories independently reported isolation of pluripotential cells *in vitro* from the mouse embryo (Evans and Kaufman 1981; Martin 1981). Evans and Kaufman (1981) induced implantational delay on day 2.5 of pregnancy by ovariectomy and hormonal (Depo-Provera) treatment to increase the cell number and to recover blastocysts representing early post-implantation stage. Initially, groups of embryos were seeded on plastic dishes in drops of culture medium under paraffin oil. The ICM cells grew out as egg-cylinder-like structures; the latter were taken off the dishes, and seeded on mitomycin C-inactivated STO cell feeder layers in gelatinised dishes after dispersal of cells by trypsinisation. Subsequent passaging of cells led to the establishment of a number of ES cell lines (originally called EK cell lines) with a normal karyotype. The pluripotency of these cell lines was demonstrated by their capacity to differentiate *in vitro* after withdrawal of feeder layers, and the formation of teratocarcinomas in syngeneic mice (Evans and Kaufman 1981).

Martin (1981) established an ES cell line directly from the ICM released by immunosurgery from the blastocysts. The ICMs were grown on mitomycin C-treated STO feeder layer in culture medium conditioned by a teratocarcinoma-derived cell line. An embryonic stem cell line was derived from the colonies of cells grown from the ICMs. After five passages the stem cell line could be maintained on feeder layers in

the absence of the conditioned medium. Martin (1981) argued that the conditioned medium contained a factor essential for the promotion of growth of stem cells/ inhibition of their differentiation *in vitro* or both. The ES cells were able to generate teratocarcinomas *in vivo* and differentiated into embryoid bodies *in vitro*. Subsequently, Axelrod (1984) showed that the use of conditioned medium and immunosurgery (Martin 1981) or implantational delay (Evan and Kaufman 1981) was not necessary to isolate diploid and pluripotential ES cell lines. The author reasoned that the use of small culture volume obviated the need for the conditioned medium. Eistetter (1989) reported isolation of pluripotential ES cells from blastomeres cultured on mitomycin C-treated embryonic fibroblasts. The blastomeres were obtained by removal of zonae pellucidae and disaggregation of morulae of approximately 16 cells.

Bradley et al. (1984) demonstrated that three independently derived male ES cell lines could contribute extensively to the soma and germ line when injected into the host blastocysts. Approximately 30% of the male chimaeras tested produced ES cell derived progeny. In an extensive study the same laboratory showed that all the 17 independent ES cell lines tested could contribute to chimaerism and a distortion in sex-ratio in favour of males was observed (Evans et al. 1985). A number of germ line male chimaeras produced progeny only from the ES cell genome. The authors proposed that such chimaeras resulted from female blastocysts injected with male ES cells; the male ES cells converted the sex of the female blastocysts and XX cells from the latter were excluded from spermatogenesis. The efficiency of germ line colonization by ES cells appears to be partly dependent upon the strain of the host blastocyst. Blastocysts from C57/BL/6 have been shown to permit better incorporation of microinjected ES cell into the germ line as compared to

blastocysts from outbred strains CD1, MF1 (DeChiara et al. 1990; Schwartzberg et al. 1989), and ICR (Pease and Williams 1990). The use of blastocysts from strains harbouring a known genetic lesion affecting fertility may also significantly improve the prospects of germ line transmission of the ES cell component (Evans et al. 1985).

Chimaeras have also been produced by aggregation of ES cells with 8-cell diploid or 4-cell tetraploid embryos (Nagy et al. 1990) and by injection of ES cells under the zona pellucida of 8-cell embryos (Tokunaga and Tsunoda 1992). Although live animals were born from the aggregation of ES cells with 4-cell tetraploid embryos, they failed to survive (Nagy et al. 1990). Injection of ES cells (C57/BL/6 X CBA) into uncompacted 8-cell eggs (CD-1) resulted in significantly higher frequency of chimaerism as compared to injection of ES cells into blastocysts and the degree of chimaerism was higher (Tokunaga and Tsunoda 1992). In this study an exceptionally high proportion of 8-cell egg derived male chimaeras produced progeny exclusively from the ES cell component. It is not clear whether this effect reflects only the stage of the embryo or is the outcome of the interaction of mouse and ES cell strain, and the stage of embryo. However, the conclusion of this study is consistent with that of Lallemand and Brûlet (1990). In this study the extent of chimaerism was measured by staining for expression of a *lacZ* gene carried by the ES cells. Chimaeric embryos were generated by injection of ES cells (129/Sv+/+) either into morulae or blastocysts (F2, BALB/c X SJL/J) and it was found that consistently better chimaerism was obtained from morula injections. Recently, a simple method based on a brief coculture of ES cells and 8-cell embryos has been reported for the production of germ line chimaeras (Wood et al. 1993). 15 pups were born from 40 manipulated embryos and 6 of these pups were chimaeric. Two male

chimaeras transmitted the ES cell component to their progeny. The number of ES cells populating the embryos in culture appeared to be critical for post implantation development; higher contribution from an ES cell line appears to be associated with abnormal development. Replication of these experiments with a number of ES cell lines would be required to draw firm conclusions on this association.

In 1986, two laboratories demonstrated the use of ES cells to generate transgenic mice, using either retroviral infection (Robertson et al. 1986) or the calcium phosphate method (Gossler et al. 1986). ES cells selected for G418-resistance *in vitro* prior to injecting them into blastocysts were demonstrated to be compatible with germ line transmission (Gossler et al. 1986). The use of ES cells to generate mice carrying mutations of a specific gene was shown by two different groups (Hooper et al. 1987; Kuehn et al. 1987). To generate HPRT-deficient mice Hooper et al. (1987) isolated ES cell clones resistant to 6-thioguanine (6-TG). One clone carrying a spontaneous deletion of part of the HPRT gene was used to generate chimaeras, and this clone contributed to the germ line. Kuehn et al. (1987) inactivated the HPRT gene by retroviral insertion and HPRT-deficient clones were selected using 6-TG selection. Two independent ES cell clones were found competent to colonize the germ line; one of these clones was transmitted through both the male and female chimaeras.

A more directed approach to the generation of specific mutations in mice involves gene targeting (see next section). This has spurred the widespread use of ES cells. Initially, ES cells were isolated and grown on feeder layers of either mitotically inactivated fibroblasts (Evans and Kaufman 1981) or primary embryonic fibroblasts (Gossler et al. 1986). The use of feeder layers is cumbersome, and the isolation of

rare clones of ES cells using various selective media has been facilitated by the development of feeder-free culture conditions suitable for proliferation of ES cells. Smith and Hooper (1987) reported that culture medium conditioned by Buffalo rat liver (BRL) cells could inhibit the differentiation of ES cells in the absence of feeder layers. BRL cell-conditioned medium in combination with fibroblast feeder layers was used to isolate a diploid ES cell line (Handyside et al. 1989). The established ES cell line could be maintained on BRL medium alone, and cells of this line can contribute to the germ line (Hooper et al. 1987). Subsequently, a glycoprotein (M_r=43,000) called differentiation inhibiting activity (DIA) from the BRL medium (Smith et al. 1988) or myeloid leukaemic inhibitory factor (LIF) (Williams et al. 1988) was found to be sufficient to inhibit the differentiation of pluripotential ES cells *in vitro* in the absence of feeder layers. ES cell lines isolated in culture medium supplemented with purified LIF and without feeder layers have been shown to be germ line competent (Nichols et al. 1990; Pease and Williams 1990). However, only one of the 15 ES cell lines isolated by Nichols et al. (1990) had a completely normal karyotype. Bradley et al. (1992) have argued that despite the simplicity of feeder-free ES cell culture systems it may be advantageous to use feeders in experiments designed to isolate genetically modified ES cell clones. In the absence of published information on comparison of efficiency of these culture systems, many laboratories prefer to use feeder layers; nevertheless germ line transmission of a number of targeted ES cell clones has been reported in the literature using feeder-free culture systems (Clarke et al. 1992; Selfridge et al. 1992; Thompson et al. 1989).

1.4.2. Gene Targeting in Mammalian Cells

The term gene targeting has been used to describe the directed modification of endogenous genes via homologous recombination between chromosomal DNA sequences and an extraneously introduced DNA fragment. Development of gene targeting strategies in mammalian cells was preceded by the demonstration that extraneously introduced DNA molecules can participate in extrachromosomal recombination (see review by Subramani and Seaton 1988). Although these studies demonstrated homologous recombination in mammalian cells, subsequently it has been shown that extra-chromosomal recombination involves a non-conservative pathway of recombination distinct from that of gene targeting (Lin et al. 1984; Lin et al. 1990; Seidman 1987). In the process of extrachromosomal recombination, one of the substrates is lost, unlike in a gene targeting reaction where both the participating molecules are conserved (Pennington and Wilson 1991). The design of gene targeting plasmids used in mammalian cells has been facilitated by the knowledge gained from similar experiments in yeast. Essentially, a targeting plasmid includes a DNA fragment homologous to the chromosomal region of interest. In cases where the targeted modification does not lead to a selectable phenotype a positive marker gene is placed within homologous sequences. Depending upon the site of a double-strand cut in the targeting plasmid two types of targeting vectors are generated. If the cut is placed within the region of homology a insertion-type gene targeting vector is generated (Orr-Weaver et al. 1981). Homologous recombination of this vector with the target locus leads to the integration of the entire vector sequences resulting in duplication of the target gene sequences included in the vector (Fig. 1.2A). Introduction

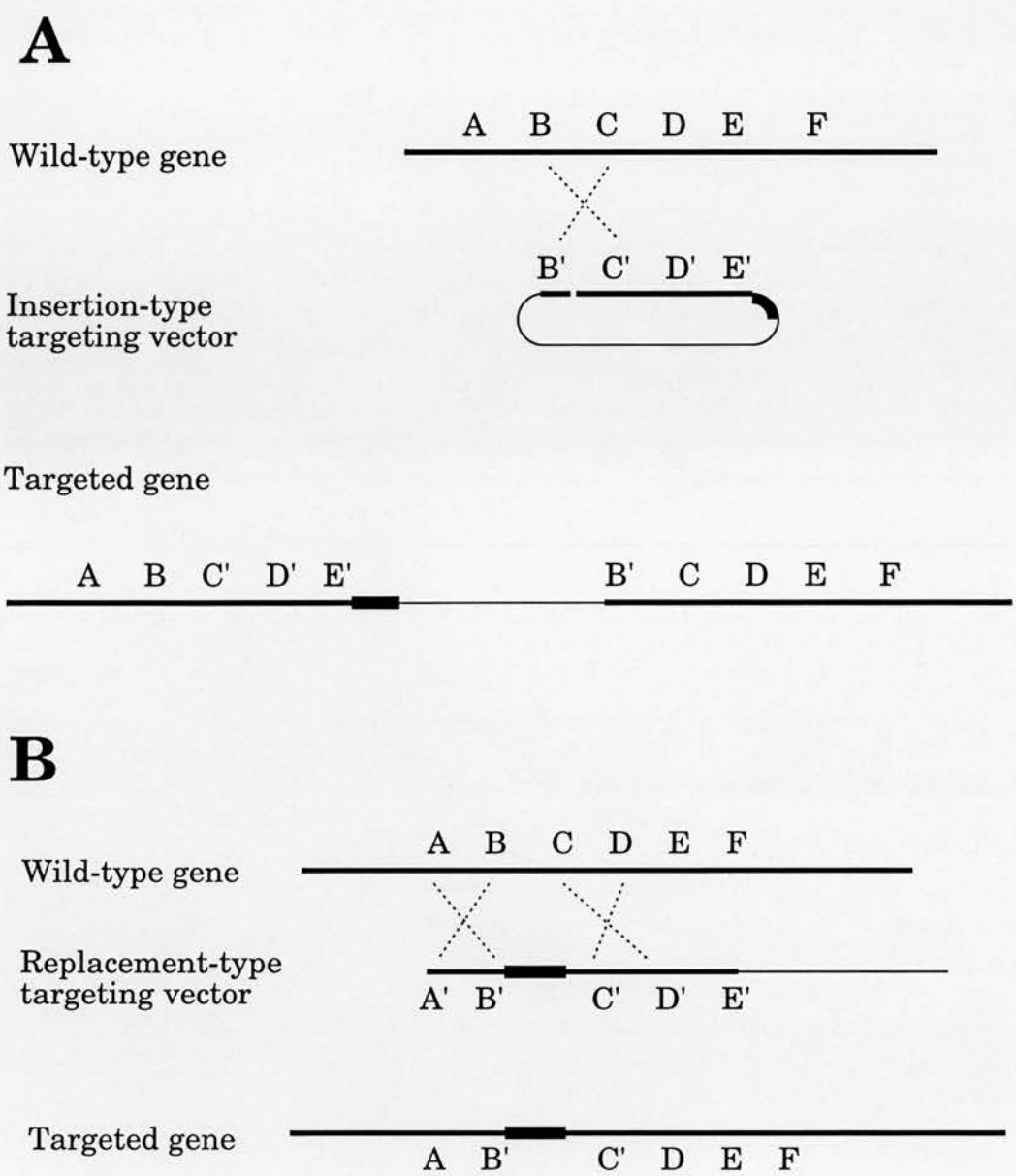


Figure 1.2. Schematic representation of gene targeting.

A) Targeting by an insertion-type vector. The entire targeting vector is integrated into the target gene.

B) Targeting by a replacement-type vector. Sequences in the target gene are replaced by those of targeting vector.

Very thick line: positively selectable marker; Thin line: plasmid backbone.

of a double-strand break at the junction or outside the region of homology in the targeting plasmid results in a gene replacement-type vector (Rothstein 1983). In these vectors the positive marker is flanked by two arms of sequences homologous to the target gene. Homologous recombination yields replacement of endogenous gene sequences with those corresponding to vector sequences, including the positive marker (Fig. 1.2B).

1.4.2.1. Early Gene Targeting Experiments in Mammalian Cells

Initial studies on gene targeting involved correction of mutated tk or neo genes in artificially created loci in cell lines (Lin et al. 1985; Smith and Berg 1984; Song et al. 1987; Thomas et al. 1986). Smith and Berg (1984) stably transfected 3T6 cells with a plasmid carrying a deleted neo gene, and isolated a cell line containing multiple copies of the vector. Calcium phosphate transfection of this cell line with a plasmid carrying another mutant neo gene generated functional neo genes at a frequency of 10^{-3} to 10^{-4} per cell that integrated the transfected DNA. In similar experiments Lin et al. (1985) provided evidence for the creation of functional tk genes through homologous recombination between overlapping 5' and 3' fragments of the tk gene. In this study a non-functional tk gene was corrected at a frequency 10^6-fold lower than that of transformation frequency by a functional tk gene, in only one out of 10 cell lines tested. The lack of generation of the tk^+ phenotype in the remaining nine lines was interpreted as being due to the positions of the loci. An alternative explanation, not considered by the authors, might have been the difference in the precise structures of the reconstructed loci

in these cell lines. Also their scheme of homologous recombination was very inefficient (see below) and it is not possible to rule out that they fortuitously isolated one cell line with higher efficiency of homologous recombination. The low frequency might have been due to the use of supercoiled plasmid and a small region of homology (320 bp) between the integrated and exogenous tk gene fragments. Southern blot analysis of 9 tk^+ cell lines revealed that in only one of these lines the tk gene was corrected by insertion of the incoming plasmid (Lin et al. 1985). In the absence of linearisation of the incoming plasmid it was possible that in some of these lines the tk gene was reconstructed in two steps. In the first step the targeting plasmid might have integrated into the locus via homologous recombination at a point other than tk sequences (the integrated and the incoming plasmids shared the same backbone) and a second intrachromosomal recombination step generated the functional tk gene. The loss of most of the tk gene sequences in one of the clones appears to be consistent with this interpretation of their data.

Thomas et al. (1986) corrected a non-functional neo gene integrated into three chromosomal sites in mouse L cells. These authors microinjected linearised correcting plasmids. In the three cell lines one cell per 1000 cells injected were found to have a functional neo gene. In one experiment, the number of molecules injected was varied to find out the effect on the gene targeting frequency. Thomas et al. (1986) concluded that the number of molecules injected did not influence gene targeting and went on to further suggest that the finding of homologous sequences by the targeting vector in the cell genome was not rate limiting in a gene targeting reaction. The number of targeting events obtained in this experiment was very small and thus, their data were not sufficient to support their conclusions. Similarly their conclusion on the lack of effect

of chromosomal position on gene targeting efficiency requires caution since only three sites were studied. Double reciprocal recombination or gene conversion resulted in the correction of either the integrated neo gene or the neo gene on the incoming plasmid (Thomas et al. 1986). The lack of targeted integration of the incoming plasmids is likely to have been due to the introduction of a double strand-break outside the neo gene sequences in the targeting plasmids in this study rather than an inherent preference of mammalian cells for gene conversion events.

Song et al. (1987) reported correction of integrated non-functional neo genes in hamster and human cells. In various cell lines the ratio of homologous recombination to non-homologous integration varied from 1:76 to 1:500. Southern analysis revealed that functional neo genes were created in three ways: a) targeted integration of the incoming plasmid, b) replacement or gene conversion and c) correction of the incoming plasmid by sequence replacement or gene conversion and its integration elsewhere in the genome. In this study the double-strand break in the incoming plasmid was placed in such a way that a functional neo gene could have been created either by an integration or a replacement-type targeting event.

Smithies et al. (1985) were the first to provide evidence that a cellular gene in mammals can be modified through homologous recombination. They used an insertion-type vector to target the human β-globin gene in EJ carcinoma cells (in which the β-globin gene is silent) and in a hybrid cell line, Hu 11 (in which the β-globin gene is active). The targeting vector included 4.6 kb of DNA from the human β-globin gene, the neo gene to select the cells that incorporated the vector DNA, and a bacterial suppresser gene (*supF*) to facilitate cloning out of the unique fragment generated as result of targeted integration of the

vector in the β-globin gene. Homologous recombination of this vector was predicted to generate a 7.7 kb *Xba*I fragment comprising both *supF* gene and β-globin sequences which were not present in the targeting vector. Smithies et al. (1985) screened genomic DNA from the cells transfected with the targeting vector by rescuing the 7.7 kb *Xba*I fragment into a mutant bacteriophage. The authors showed that approximately one out of every thousand cells transformed by the targeting vector DNA had undergone the predicted modification in the β-globin gene irrespective of the state of transcription of this gene.

These early experiments demonstrated that cellular genes can be modified by gene targeting but that only a minority of the transformed cells undergo homologous recombination and emphasised the need to further refine the techniques. Gene targeting experiments with the HPRT gene have proved very beneficial in understanding the parameters affecting the gene targeting efficiency in mouse ES cells. HPRT is encoded by a single copy gene on the X-chromosome and selection protocols have been available to select for (HAT) and against (6-TG) the activity of this gene. Thus, targeted modification in the HPRT gene in a male cell line leads to a selectable phenotype and the rare targeting events can easily be identified in the background of cell clones generated by random integration of the targeting vectors.

1.4.2.2. Targeting the HPRT Gene in ES cells.

In 1987, two laboratories introduced targeted modification of the HPRT gene. Doetschman et al. (1987) restored HPRT gene function by targeted integration of an insertion-type vector in an ES cell line carrying a deletion in the 5' region of the HPRT gene. Cells

electroporated in the presence of the vector were selected in HAT medium, and HAT-resistant colonies were isolated at a frequency of $1.8x10^{-6}$ per cell treated. Five independent colonies were shown to have undergone the predicted modification at the HPRT locus (Doetschman et al. 1987). In three out of five such clones analysed, a four base pair insertion included near the double-strand break point in one arm of the targeting vector was corrected by the chromosomal sequences.

Complementary experiments were described by Thomas and Capecchi (1987). These authors disrupted the HPRT gene via homologous recombination at this locus using both insertion and replacement-type vectors. Their targeting vectors included a neo gene specially constructed to achieve high level of expression in ES cells and varying amounts of homologous sequences. The targeted clones were isolated after selection in G418 and 6-TG. Thomas and Capecchi (1987) concluded that a 2-fold increase in homology resulted in 20-fold enhancement of the targeting frequency. However, only one experiment was performed with most vectors in this study. Another of their conclusions that insertion and replacement-type vectors targeted with equal frequency needs qualification. First, the neo gene in the vectors lacked a polyadenylation signal; thus, some of the clones targeted by insertion vectors may have been lost by reconstruction of the HPRT functional mRNA after splicing out the sequences corresponding to the targeting vectors. Second, the design of insertion vectors in this study was such that sequences homologous to 5' and 3' region of the HPRT gene were joined together; the total homology was interrupted twice, by a double-strand break and the neo gene. The relevance of this is that these vectors are capable of targeting by both insertion and replacement pathways. Both of the above factors could have resulted in an under-

estimation of targeting frequency of insertion vectors. Withstanding these observations, the absolute targeting efficiency of equivalent replacement and insertion vectors was $0.4x10^{-6}$ and $0.5x10^{-6}$, respectively. Southern analysis revealed that replacement vectors targeted in a predictable manner. It is intriguing to note that although the structure of insertion vectors used in this study was compatible with both insertion and replacement targeting, the majority of the targeted clones (9/12) were the outcome of an insertion-type event. The remaining 3 clones were generated by replacement targeting.

In contrast to the observations made by Thomas and Capecchi (1987), another laboratory reported that an HPRT replacement-type vector targeted at least 5-fold less efficiently in contrast to a corresponding insertion vector (Hasty et al. 1991c; this comparison is based on the absolute targeting efficiency calculated from the data presented by the authors). In this study the targeting efficiency was expressed as the ratio of G418-6-TG-resistant clones to G418-resistant clones while comparing the two types of vectors and it was concluded that insertion vectors were 9-fold more efficient. The above measurement of targeting efficiency may not be an ideal method to compare various vectors since the frequency of G418-resistant clones would be influenced by the design of targeting vectors. Hasty et al. (1991c) showed that the structure of the majority of the clones targeted by their replacement vector did not arise after sequence replacement events at the HPRT locus but these were the product of complex events involving integration of one or more copies of the replacement vector in the HPRT gene. Their conclusion that replacement vectors are prone to aberrant targeting has been challenged by Thomas et al. (1992). These authors targeted the 3' region of the HPRT gene with a replacement vector having 9.1 kb of

homologous sequences distributed as two arms of 6.1 and 3.0 kb. It was observed that in all the 18 targeted clones the structure of the HPRT locus was generated as a result of precise targeting. However, when the short arm of homology was reduced to 0.7 kb, in five out of the 18 clones analysed the targeting events were not the predicted replacement events. Thomas et al. (1992) argued that reduction in the fidelity of targeting by replacement vector may be due to the reduction in the length of the short arm of the replacement vector in their study as well as in that of Hasty et al. (1991c). They have further suggested that a minimum length of homology on each arm may be required for the gene targeting reaction to proceed and concatemerisation of the vector molecule may be advantageous when one arm of the vector is exceptionally short. A trivial alternative explanation may be that the aberrant targeting occurs with low frequency at all times but when one arm of the vector is very short the frequency of precise gene targeting events drops dramatically due to the degradation of the short arm (and consequently loss of the neo gene expression), and that the selection for neo gene expression during the isolation of targeting events apparently enriches for the aberrant targeting events.

The relationship between the gene targeting efficiency and total homology in the targeting vector has been examined by two groups (Deng and Capecchi 1992; Hasty et al. 1991b). Hasty et al. (1991b) observed that the absolute targeting frequency increased by 250-fold with a corresponding increase in homology from 1.3 kb to 6.8 kb in their insertion vector. In this study equimolar vector DNA was not used and the absolute targeting efficiencies were adjusted to take into account differences in the numbers of the vector molecules. The authors compared the targeting efficiency of replacement vectors having a homology of

1.3-6.8 kb. Their data suggested that although the absolute targeting efficiency increased with an increase in the length of homology, the relationship appeared to be linear and not exponential as suggested by Thomas and Capecchi (1987) and subsequently demonstrated by the latter laboratory in an extensive study (Deng and Capecchi 1992). This difference may be reconciled by the fact that Hasty et al. (1991b) probably overestimated the absolute gene targeting efficiencies of their smaller vectors since they used a fixed amount of the vector DNA, and that their replacement vectors did not always target the HPRT gene in a way predicted for these vectors. Further, the conclusion of Hasty et al. (1991b) that targeting efficiency did not respond to an increase in homology beyond 6.0 kb was based on only a single vector. On the other hand Deng and Capecchi (1992) have provided good evidence that gene targeting efficiency increases exponentially with an increase in homology in both replacement and insertion vectors. A plateau effect was observed when the length of homology reached 14 kb for both insertion and replacement vectors (Deng and Capecchi 1992). The targeting efficiency of these two types of vectors appeared to be similar in this study but it may be noted that the 5' and 3' regions of the HPRT gene homology in the majority of the insertion vectors were joined together. As a result these vectors are capable of targeting both by insertion and replacement pathways. Deng and Capecchi (1992) did not analyse the structure of the targeted HPRT gene of the 6-TG-resistant clones.

1.4.2.3. Strategies for Targeting Non-selectable Genes

Targeted mutations in most genes in ES cells do not lead to selectable phenotypes. To identify ES clones transformed by the

incoming DNA either via random integration or targeted integration, a positively selectable marker gene is included in the targeting vector. A neo gene cassette pMC1Neo, constructed by Thomas and Capecchi (1987) for its high level of expression in ES cells has been widely used in gene targeting experiments. pMC1Neo includes the HSV tk gene promoter and an enhancer from the PyF441 polyoma virus in addition to the neo gene coding sequences. The hygromycin gene is another positive selection marker that has been used in gene targeting experiments (Te Riele et al. 1990).

Mammalian cells generally incorporate transfected DNA at various random sites throughout their genome (reviewed by Roth and Wilson 1988). ES cells are no exception and usually, only a small proportion of cells transformed by the targeting vector DNA undergo homologous recombination. Two different approaches have been utilised to enrich for the targeting clones at the expense of clones generated as a result of random integration of the targeting vector.

In the first approach, a promoterless selectable marker gene is included in the targeting vector such that upon homologous recombination the marker gene would come under the transcriptional control of the regulatory sequences of the target gene. The modified gene would allow selection for the marker gene if the target gene is active in the embryonic stem cells. As the marker gene in the targeting vector lacks its control elements random integration of the vector into the ES cell genome would very infrequently lead to its expression. Hence, the relative proportion of the targeted clones would be increased. Doetschman et al. (1988) targeted the HPRT gene with a replacement-type vector using this approach. A translation stop codon and a translation start codon was included upstream of the neo coding

sequences in the vector. It was envisaged that the initiation of translation from the endogenous start codon in the mRNA from the targeted allele would be terminated at the newly introduced in-frame stop codon and the reinitiation of translation would result in the functional neo gene product. Doetschman et al. (1988) were able to isolate targeted clones by this strategy; the comparison of the transformation frequency of their targeting vector and a functional neo gene suggested approximately 100-fold enrichment for the targeting events. Very low level of neo gene expression was achieved from the targeted allele as demonstrated by the failure to isolate targeted clones at the high stringency selection regime. This was probably due to the inefficient reinitiation of translation in the fusion mRNA. A better alternative involves construction of the targeting vector in such a way that the fusion transcript from the targeted allele would direct a fusion protein conferring G418-resistance (Charron et al. 1990; Schwartzberg et al. 1989). The use of a neo gene without its polyadenylation site may further reduce the number of clones generated by random integration (Joyner et al. 1989; Zijlstra et al. 1989) but it may not always be possible to avoid the polyadenylation signal present in the homologous sequences included in the targeting vector. Nevertheless, if the objective of the experiment is to introduce a null mutation it may be desirable to include the polyadenylation signal at the end of the marker gene.

The promoterless selectable marker gene approach has also been used in conjunction with insertion vectors (Jasin and Berg 1988). These authors targeted the SV40 early region in COS1 cells and achieved 100-fold enrichment for the targeting events. It may be noted that the targeted allele in some of the clones was unstable probably

reflecting the weak transcription from the endogenous locus vis-a-vis selection pressure applied to isolate these clones.

An important strategy, the positive-negative selection (PNS) procedure for enrichment for the targeted clones relies on the selective killing of cells that have randomly incorporated a replacement-type targeting vector (Mansour et al. 1988). In this method a negatively selectable marker gene is placed at one or both ends of the vector in addition to the positive marker embedded into the region of homology. Homologous recombination of the replacement vector with the target gene results in the exclusion of the non-homologous terminal DNA of the marker gene, and thus targeted clones are able to survive selection against expression of the negatively selectable marker gene. However, clones generated by random integration of the targeting vector would incorporate the terminally placed marker and negative selection will lead to the killing of cells expressing this marker. Mansour et al. (1988) chose to target the HPRT and *int-2* genes using this approach. They placed the thymidine kinase (tk) gene from Herpes simplex virus at the end of the replacement-type vectors, and after the first round of selection for neo gene expression an additional selection was applied against the cells expressing the HSV tk gene using the drug gancyclovir (GANC). In this study the authors reported approximately 2000-fold enrichment after negative selection. However, the extensive use of the PNS procedures by a large number of laboratories has shown that generally 5-20-fold enrichment is feasible (DeChiara et al. 1990; McMahon and Bradley 1990; Tybulewicz et al. 1991; and many others). The reasons for this discrepancy are not clear. Even with the variable level of enrichment there is a distinct advantage to this procedure. Unlike the promoterless neo approach the PNS procedure does not require the expression of the

target gene in ES cells. The HSV tk gene can be substituted by the diphtheria toxin A (DT-A) gene and approximately 9-30-fold enrichment has been obtained (McCarrick III et al. 1993; Yagi et al. 1990). The obvious advantage of using the DT-A is that no drug is required for negative selection. However, it should be noted that a small proportion of ES cells may be killed by the DT-A gene even in the absence of the stable integration of the targeting vector.

1.4.2.4. Screening for Targeting Events

Despite the various selection and enrichment strategies, a large number of ES cell clones are generally required to be examined to identify the clones that have undergone targeting. The initial screening may be performed either by Southern blotting or by the polymerase chain reaction (PCR). If PCR is used in the initial screening the structure of the targeted gene in the positive clones is confirmed by Southern blotting.

Typically, the targeted allele are discriminated from the wild-type allele by Southern analysis since a number of novel restriction fragments would be created as a result of targeting; these fragments are detected with probes suitable for the target locus. Both the 5' and 3' sites of homologous recombination are examined with probes corresponding to DNA sequences not included in the targeting vector. Additional screening with a probe internal to the targeting vector is necessary to exclude the aberrant targeting events and random integration of the targeting vector in the targeted clones; although the probability of the latter is extremely rare. PCR-based screening is based on the fact that homologous recombination using a replacement-type

vector results in a novel junction in the targeted allele. One primer is chosen from the targeting vector sequences not present in the target gene (usually from the positive marker gene) and the second primer corresponds to target gene sequences not included in the targeting vector. In theory amplification of the predicted fragment is possible only from the targeted allele (Kim and Smithies 1988). The choice of the screening method has implications for the vector design. For the efficient amplification of the predicted PCR product one arm of the targeting vector is kept short (<1 kb). This requirement may be antagonistic to the observation that a very short arm in the targeting vector may reduce the fidelity of gene targeting reaction involving a replacement vector (Thomas et al. 1992). However, by using PCR targeted clones can be selected at an early stage without the need to expand a large number of ES cells clones transformed by targeting vectors.

1.4.2.5. The Effects on Terminal Heterologies on Gene Targeting Efficiency

The PNS procedure (see section 1.4.2.2) is based on the premise that blocking one or both arms of replacement-type vectors does not adversely affect the gene targeting efficiency. The successful use of the PNS procedure in targeting a variety of genes is consistent with this postulation, however, few controlled experiments have been reported in the literature. Hasty et al. (1991c) cut an HPRT gene targeting plasmid in three different ways so that approximately 3 kb heterologous sequences were either excised from the targeting vector or were left on one of the two arms of homology. Gene targeting frequency was expressed as the ratio of 6-TG to G418 clones and the authors concluded that the

presence of heterology did not decrease the targeting frequency. However, this conclusion is not sustained if the data are expressed in terms of absolute targeting efficiency, in fact, the presence of heterology on the short arm appears to increase the efficiency. The results of Hasty et al. (1991c) are inconclusive since most of their replacement vectors targeted the HPRT gene in an imprecise way (see section 1.4.2.1). Mansour et al. (1988) and Deng and Capecchi (1992) have cited unpublished results in support of the neutrality of terminal heterologies on the gene targeting efficiency of replacement-type vectors.

A double-strand cut in the region of homology in the insertion vectors results in a several-fold improvement in gene targeting efficiency in ES cells (Hasty et al. 1992; Valancius and Smithies 1991a). Hasty et al. (1992) have reported that the placement of a 2.3 kb heterology on one of the DNA ends significantly reduced the targeting frequency. The structure of the targeted clones revealed that the terminal heterology had been lost in the majority of the clones. On the other hand, blocking of both the DNA ends of an insertion vector with 13 bp of heterology each did not appreciably affect the gene targeting frequency but in some of the targeted clones the heterology was retained. The authors have suggested that the small terminal heterologies were retained as a result of ligation of the DNA ends during the gene targeting reaction. One prediction of this argument is that the heterologies should be retained in the targeted locus either to the 5' or 3' of the integrated vector sequences with equal probability. However, in five of the six clones analysed the small terminal heterology was retained 5' to the integrated vector sequences. This would be more consistent with a preferential cross-over on the longer of the two arms of homology after recircularisation of the vector. In the sixth clone the altered but identical

heterology was retained on both sides of the vector sequences. This may have been due to the inclusion of heterology on one end in the first instance and subsequent repair via intrachromosomal recombination event might have transferred it to the other end (Hasty et al. 1992). In contrast to the above results Pennington and Wilson (1991) did not find any evidence for the inclusion of small terminal heterologies from an insertion vector in the target locus in any of the 17 targeted clones analysed.

1.4.2.6. Isogenic DNA and Gene Targeting Frequency

The majority of the ES cell lines used in gene targeting experiments have been derived from the mouse strain 129 whereas the targeting vector DNA has been derived from various strains. Te Riele et al. (1992) assayed the effect of the source of vector DNA on the gene targeting efficiency at the retinoblastoma susceptibility (*Rb*) gene. They used either 129- or BALB/c-derived DNA vector to target this gene in 129-derived ES cells. In one experiment with vectors bearing 10.5 kb of *Rb* sequences, the targeting efficiency was 50-fold higher with the 129-derived vector. These authors then performed a more controlled gene targeting experiment at the *Rb* locus. First they converted this locus into a 'selectable locus' by targeting an HPRT minigene into the *Rb* gene in an otherwise HPRT-deficient ES cell line. The modified locus was targeted again by using either isogenic or non-isogenic vectors such that the HPRT minigene was lost as a result of the targeting reaction. This strategy permitted direct isolation of targeted clones using 6-TG selection. In an extensive experiment these authors provided strong evidence that targeting of the modified *Rb* allele was 20-fold more efficient when the

targeting vector was constructed from isogenic DNA. Surprisingly, in another experiment 78% of the colonies transformed by a hygromycin based vector bearing 17 kb of isogenic DNA were generated by homologous recombination. Similar conclusions have been reached by Deng and Capecchi (1992) and Van Deursen and Wieringa (1992) for the HPRT gene and creatine kinase M gene, respectively. In a limited sequence analysis of their non-isogenic DNA Te Riele et al. (1992) and Van Deursen and Wieringa (1992) found that the longest stretch of the perfect match to the target gene was 278 nt and 255 nt, respectively. It is yet to be understood that at what step in homologous recombination the presence of mismatches between the targeting vector and the chromosomal gene becomes rate limiting.

The enhancement of targeting efficiency by the use of isogenic DNA is interesting in that the length of heterologous DNA sequences embedded between two arms of homology in a replacement vector does not appear to affect the targeting efficiency (Mansour et al. 1990). These workers found that the inclusion of variable lengths of heterology (8 bp-12 kb) in the HPRT gene targeting vectors did not affect the targeting efficiency at this locus. These differences can be reconciled by assuming that there is a requirement for a minimum length of perfect homology on both arms of targeting vectors.

1.4.2.7. Targeting of Subtle Mutations

Targeted disruption of mouse genes has been proved to be of immense importance in addressing questions regarding gross functions. Subtle mutations are likely to be important in understanding the fine structure-function relationships of gene products, and the

regulation of gene expression. The ability to introduce point mutations is also vital in the generation of human disease models, and in modifying phenotypes of complex traits of economic importance in farm animals, for example in the properties of milk (see section 1.3.3). In principle a single base pair of an endogenous gene can be modified by gene targeting with a replacement-type gene targeting vector. In practice the isolation of the targeted clone is not straightforward if the targeted mutation does not lead to a selectable phenotype. Steeg et al. (1990) introduced two specific point mutations into the RNA polymerase II gene of mouse ES cells. The point mutations were placed 20 bp apart on a replacement vector having 5.8 kb of homology. One of these mutations conferred resistance to the mushroom toxin α-amanitin and was found to be incorporated into the target locus at a frequency of 1 per 30 cells that incorporated and expressed an equivalent vector DNA. This result demonstrated that gene targeting can proceed in the absence of large stretches of heterology between the two substrates. However, the second point mutation included in the targeting vectors was corrected to the wild-type sequence in the majority of the targeted cells, probably indicating a directional mismatch repair (Steeg et al. 1990). Since only the minority of cells treated in transfection experiments incorporate the DNA and a small fraction of these undergo a gene targeting event, incorporation of positively selectable marker gene into the targeting vector is essential to identify the cell clones that have undergone targeting at non-selectable loci. However, incorporation of a marker gene may also interfere with the expression of the targeted locus. A number of gene targeting strategies have been reported for introducing point mutations into endogenous genes without leaving behind any selectable marker gene sequences.

Zimmer and Gruss (1989) microinjected their vector, designed to target a 20 bp insertion into the *Hox 1.1* gene, and isolated targeted clones by PCR. The vector did not carry a selectable marker gene and one targeting event per 150 cells injected was observed. The authors argued that high transfection efficiency achieved with microinjection and a very small heterology in their targeting vector ensured such an exceptionally high absolute targeting frequency. Contrary to the above argument, Mansour et al. (1990) have provided strong evidence against the effect of the length of internal heterology on the HPRT gene targeting efficiency. The microinjection method of Zimmer and Gruss (1989) has not found general application, possibly due to the cumbrousness of microinjection into ES cells.

Cotransformation of ES cells with a gene targeting vector carrying a subtle mutation and a selectable marker gene (neo-resistance) on a separate DNA fragment has been used to isolate ES cell clones with subtle modifications in the HPRT gene by two laboratories (Davis et al. 1992; Reid et al. 1991). Reid et al. (1991) reported that only 4% of targeted clones incorporated the neo gene elsewhere in the ES cell genome when they identified the targeted clones without the use of selection for neo gene expression. Although some 80-fold enrichment was obtained, cotransformation precluded the isolation of the majority of the targeted clones. Further, the authors could not exclude the possibility of incorporation of the selectable marker gene into the target gene alongside the subtle mutation. On the other hand Davis et al. (1992) obtained subtle mutation of the HPRT gene by cotransformation at an efficiency comparable to that of a vector with a positive marker included in the HPRT gene targeting vector sequences. However, only 8% of the targeted clones had the desired change in the target locus. The remaining clones

had incorporated copies of selectable marker gene inside the target locus in addition to the subtle modification. This method may not be of much practical use for introducing subtle mutations in non-selectable genes.

There was a marked difference in the frequencies of targeting the HPRT gene in the above two studies. One of the explanations for this discrepancy may be that Reid et al. (1991) scored their targeting events after HAT selection which needed correction of a deletion in the HPRT gene and thus, precise homologous recombination was required at least on one arm of their vector. Cointegration of their HPRT gene targeting vector and the coelectroporated neo gene into the HPRT gene could have interfered with the generation of functional HPRT gene. Since Davis et al. (1992) scored their targeting events by observing for the loss of the HPRT gene function they could have isolated all the targeting events irrespective of the precision of the homologous recombination events.

Another approach for targeting subtle mutation in non-selectable loci involves cotransfer of a point mutation alongside a neo gene using either a replacement vector (Deng et al. 1993; Rubinstein et al. 1993; Shesely et al. 1991) or an insertion vector (Hasty et al. 1991a; Valancius and Smithies 1991b). Deng et al. (1993) have suggested that neo gene may be placed at a neutral site of a replacement vector. The suggestion of Deng et al. (1993) that targeting by the replacement vector occurs near its ends implied that the point mutation would be transferred to the target locus irrespective of the distance from the neo marker gene; however, their data did not seem to support this suggestion. On the other hand observations of Rubinstein et al. (1993) did suggest that a subtle mutation placed very close to the neo gene but away from the ends of the vector was transferred to the chromosomal gene in the majority of the

targeted clones. However, it may be almost impossible to choose a suitable neutral site to place neo gene prior to performing the experiment. Further, the cotransfer may be complicated by directional correction of the point mutation after the formation of heteroduplex as suggested by Steeg et al. (1990) or prior to initiation of the targeting reaction if the point mutation is placed near the DNA ends of the vector.

Valancius and Smithies (1991b) and Hasty et al. 1991a) have utilised a two step approach to introduce subtle mutations in non-selectable genes. In the first instance the subtle mutation is cotransferred alongside a positive marker using an insertion-type vector. Such a targeting event results in duplication of sequences at the target locus. In the second step intrachromosomal recombination between the duplicated segments would excise the targeting vector; the isolation of such rare recombination events is facilitated by selection against expression of a negative selection marker included in the backbone of the targeting vector. In some cases depending upon the precise recombination site, the subtle mutation would be left in the target locus. This method is limited by the lack of an enrichment procedure during the first targeting step. Further, the subtle mutation placed near to the DNA ends of the insertion vector might be lost due to degradation of terminal sequences (Deng et al. 1993; Valancius and Smithies 1991a).

Askew et al. (1993) have shown that subtle mutation can be introduced in two successive targeting steps using replacement vectors. In the first step the target locus was disrupted by a neo/HSV tk gene cassette. A second gene targeting step was performed with a replacement vector consisting of a genomic clone carrying a subtle mutation; GANC selection against the HSV tk gene expression was employed to isolate the targeted clones during this step. Although there

was high background of GANC-resistant clones that were generated by inactivation or loss of the tk gene one out every 54 clones analysed had incorporated the subtle mutation. One major advantage of this method is the possibility of generating a variety of mutant alleles from the targeted clones obtained after the first step of targeting (Askew et al. 1993). However, prolonged culture of ES cells involving two successive electroporation and isolation of targeting clones by this method may compromise the germ line potential of ES cells.

The fidelity of gene targeting is crucial in an experiment designed to introduce subtle mutations. Aberrant modifications do occur during gene targeting reactions (Brinster et al. 1989; Doetschman et al. 1988; Hasty et al. 1991c; Thompson et al. 1989), although some of these aberrant modifications might not be associated with the process of gene targeting but might have arisen after gene targeting. An extensive analysis of DNA sequences around the recombination junctions by a method based on chemical cleavage of mismatches has revealed that targeting of the HPRT gene by replacement and insertion vectors can be very precise (Zheng et al. 1991). These workers found only two mismatches in 80 kb of sequences from 44 independent targeted clones. The probability of aberrant targeting events may increase in targeting strategies involving two successive steps (see above).

Mostly gene targeting in ES cells has been used so far to ask questions pertaining to the functions of a number of genes in the context of whole-animal and numerous surprises have come to light in the fields of immunology, oncology and developmental biology. The ability to introduce subtle mutations into the mouse genome should allow more precise definition of these questions. Particularly, regulation of gene

expression is likely to be an area where the application of gene targeting technology should find increasing usefulness. Probably one of the most important contributions of gene targeting in biology includes the realisation that the gene expression data may not always indicate the indispensable *in vivo* function of a gene (see Erickson 1993).

1.5. AIMS OF THE PROJECT

The main aim of the project was to create a mouse strain with a null mutation in the β-casein gene using gene targeting in embryonic stem cells, and to study the phenotype of mutant mice. It was anticipated that this would allow the delineation of any essential biological function of β-casein in lactation and in determining the properties of milk. Further, the β-casein deficient mouse should serve as an animal model to study the dynamics of protein synthesis in the mammary gland.

The second objective was to examine if it would be feasible to enhance the gene targeting efficiency of insertion-type gene targeting vectors by including a negatively selectable marker gene at the double-strand break point. To address this question a series of HPRT gene targeting experiments were performed to determine the effect(s) of terminal heterologies included at the DNA ends of the insertion-type vectors on the gene targeting efficiency of these vectors.

CHAPTER TWO

MATERIALS AND METHODS

2.1. DNA MANIPULATION

2.1.1. Restriction Digestion of Plasmid DNA

Plasmid DNA was digested with various restriction enzymes as per the manufacturers' (Boehringer Mannheim; New England Biolabs) guidelines using the buffers supplied except that enzymes were used in a 3-5 fold excess of the recommended amounts.

2.1.2. Isolation of DNA Fragments from Agarose Gels

Small fragments (< 3kb) of plasmid DNA utilised for the construction of various plasmid vectors or for preparation of radioactive probes were recovered from agarose gels using the Geneclean Kit (Bio 101 Inc.) as per the method provided with the kit by the manufacturer. Generally, 2-5 μg plasmid DNA was digested with appropriate restriction enzymes and electrophoresed on agarose gels (see section 2.8.1.). The DNA fragment was visualised on an agarose gel using long-wave UV light, excised from the gel into an Eppendorf tube and incubated in 3 volumes of NaI solution (6 M) at 50°C for 5 minutes to disolve the agarose. Once the agarose was in solution 5-10 μl of 'Glassmilk' was added and after mixing, the tube was left on ice for 10 minutes. The

'Glassmilk' with bound DNA was pelleted by microcentrifugation for 10 seconds and the pellet was washed thrice with the washing solution provided with the kit (NEW Concentrate, 140 ml ddH_2O, 155 ml of 100% ethanol). Finally, the DNA was eluted in 5-10 µl of ddH_2O at 50°C and microcentrifuged to remove the 'Glassmilk'. 40-50% of the total DNA present in the agarose gel could be recovered by this method.

The above method proved unsatisfactory for isolation of fragments larger than 3 kb. A 'freeze-thaw' procedure was used to isolate DNA fragments of 3-8 kb from agarose gels. A piece of agarose gel containing the fragment of interest was frozen and thawed several times in an Eppendorf tube. DNA was released from the agarose gel into the liquid phase in the tube. The liquid phase was recovered by microcentrifugation, extracted with phenol:chloroform and the DNA was precipitated with 2 volumes of ethanol. Finally, the DNA pellet was washed with 70% ethanol and dissolved in ddH_2O. Typically, 10-15% of the original amount of DNA fragment could be retrieved.

2.1.3. Ligation Reactions

In a typical sticky-end ligation reaction the following mixture was incubated overnight at 14°C :

50-100 ng: Vector and insert DNA (molar ratio 1:3)
2 µl: 10x Ligation buffer (Boehringer Mannheim)
1 µl: T4 DNA ligase (1 unit/ µl: Boehringer Mannheim)
Made up to 20 µl with ddH_2O

In blunt-end ligations ~300 ng of DNA was used. The 3' recessed ends created by restriction enzyme digestion of double strand-

DNA were blunted using Klenow fragment of DNA polymerase I (Boehringer Mannheim) in an end-filling reaction as described by Maniatis et al. (1982).

2.1.4. Transformation of Competent Cells with Plasmid DNA

E. coli DH5α competent cells (Genotype: F^-. endA1, hsdR17(r_k^-, m_k^+), supE44, thi-1, recA1, gyrA96, relA1, Δ(argF-laczya)U169, ϕ80dlacZΔM15, λ–) purchased either from Bethesda Research Laboratories (supplied by Life Technologies, Inc.) or from Clontech were used to subclone various plasmid DNAs. Typically, 50 µl of competent cells were transformed with 1 ng of purified plasmid DNA or 2-5 µl of a ligation mixture (see section 2.1.3) following supplier's instructions.

2.1.5. Preparation of Plasmid DNA

2.1.5.1. Small-Scale Preparation

The alkali lysis method as described by Sambrook et al. (1989) was used omitting the use of lysozyme.

2 ml LB medium (Sambrook et al. 1989), containing 50 $\mu g ml^{-1}$ ampicillin in 15 ml tubes were inoculated with single bacterial colonies and incubated overnight at 37°C in a shaking incubator (Gallenkamp, 200 rpm). 1.5 ml of each culture was decanted into Eppendorf tubes and centrifuged at 12,000g for 10 seconds to pellet the cells. After complete removal of the supernatant the pellets were resuspended in 100 µl of ice-cold TGE solution (50 mM glucose, 25 mM

Tris.Cl pH 8.0, 10 mM EDTA pH 8.0). The cells were lysed and the DNA was denatured by addition of 200 µl of 0.2 N NaOH 1% SDS solution to the cell suspension. The tube was rapidly inverted several times and stored on ice. 150 µl of ice-cold potassium acetate/glacial acetic acid solution (3 M and 2 M, respectively) was added and mixed thoroughly. The tube was allowed to sit on ice for 5-10 minutes and then centrifuged at 12,000g for 5 minutes. The supernatants were removed into fresh tubes and the DNA was precipitated with 2 volumes of ethanol at room temperature. The tubes were centrifuged again for 5-10 minutes to obtain the plasmid DNA pellets. The DNA was washed twice with 70% ethanol. The last traces of supernatant were removed and the pellets were dissolved in TE buffer (10 mM Tris.Cl pH 8.0, 1 mM EDTA) containing DNAase-free pancreatic RNAase (20 $\mu g ml^{-1}$). Extraction of plasmid DNA with phenol:chloroform was not necessary.

2.1.5.2. Large-Scale Preparation

Plasmid DNA from large scale bacterial culture was prepared by a modified alkali lysis method (unpublished protocol of R. Treisman described by Sambrook et al. 1989) followed by a caesium chloride (CsCl)-ethidium bromide gradient centrifugation. However, the use of lysozyme to lyse the bacterial cells was not necessary.

0.5 ml of a bacterial culture (grown overnight) was inoculated into a 2 litre flask containing 500 ml of LB medium (containing 50 $\mu g ml^{-1}$ ampicillin) and incubated overnight at 37°C in a shaking incubator (Gallenkamp, 200 rpm).

The bacterial pellet was obtained from the above culture by centrifugation at 5000 rpm at 4°C for 10 minutes in a Sorvall GS3 rotor and resuspended in 20 ml of ice-cold TGE buffer. The cells were

lysed by addition of 40 ml of 0.2 N NaOH, 1% SDS. The resulting viscous solution was mixed thoroughly, but gently, and incubated at room temperature for 10 minutes. To precipitate the proteins and chromosomal DNA, 20 ml of ice-cold potassium acetate/glacial acetic acid solution (3 M and 2 M, respectively) was added, mixed thoroughly, incubated on ice for 10 minutes and the supernatant was recovered after centrifugation at 4°C at 5000 rpm for 15 minutes. The supernatant containing plasmid DNA was filtered through four layers of a nylon gauze. The DNA was precipitated by addition of 0.6 volume of isopropanol. After 10 minutes at room temperature the plasmid DNA was pelleted by centrifugation at 5000 rpm at room temperature for 15 minutes. The pellet was washed twice with 70% ethanol, drained and dissolved in ~2 ml of TE buffer. 5.14 g CsCl and 250 µl of ethidium bromide solution (10 $mgml^{-1}$) were added to the DNA solution and the contents were dissolved and the final volume was made to 6.0 ml with TE buffer.

The bacterial proteins which precipitated in the presence of CsCl and ethidium bromide were removed by centrifugation at 5000 rpm at room temperature for 5 minutes in a Sorvall SS34 rotor. The clear red solution was transferred to a Beckman Quick-Seal tube and centrifuged at 50,000 rpm at 20°C for 16-18 hours. The lower band (distinct from chromosomal and open circular plasmid DNA) was recovered from the tube with a needle and syringe. To avoid contamination by the upper band, only 80% of the lower band was retrieved. The ethidium bromide was removed from the DNA solution by repeated extraction with butan-1-ol saturated with ddH_2O and CsCl. DNA was precipitated by addition of 3 volumes of 70% ethanol. The pellet was drained, washed once with 70% ethanol and dissolved in 450 µl of TE and reprecipitated with 2 volumes of ethanol in the presence of sodium

acetate pH5.0 (50µl 3M). The supernatant was discarded and the DNA pellet was washed twice with 70% ethanol. The liquid was completely drained, the DNA was dissolved in TE buffer and stored at -20°C. The DNA concentration was estimated by measurement of absorbance at 260 nm. The quality of the DNA preparation was checked by running an aliquot of DNA solution on an agarose mini-gel in the presence of ethidium bromide. Irreversibly denatured products of supercoiled pMC1neo (Stratagene) plasmid were frequently observed when this plasmid DNA was prepared by the method described above. This problem was avoided by shortening the incubation time to 5 minutes during the alkali denaturing step.

2.2. CONSTRUCTION OF β-CASEIN GENE TARGETING VECTORS

2.2.1. Correction of Mutation in Plasmids pMC1neo and pMC1neopA

Plasmids pMC1neo and pMC1neopA bought from Stratagene had the mutation reported by Yenofsky et al. (1990). To correct this mutation, *Ecl*XI-*Nco*I fragments in pMC1neo and pMC1neopA were replaced by the corresponding sequences from the plasmid pSV2neo. To confirm these corrections, the transfection efficiency of the two versions of pMC1neo and pMC1neopA were compared in ES cells. The original pMC1neo and pMC1neopA vectors supplied by Stratagene were approximately 10 fold less efficient than the corrected versions [pMC1neo(c) and pMC1neopA(c)] in generating G418 resistant colonies in an ES cell electroporation experiment (data not shown). The corrected plsamids were used for all the DNA constructs reported in this thesis.

2.2.2. Construction of Vector pβ1MC1NEO/TK

A 6.6 kb *Eco*RI fragment of the β-casein gene from C57/BL/6 strain of mouse (Goodman and Rosen 1990) was subcloned into the *Eco*RI site of the vector pPolyIIID (Lathe et al. 1987) and the resulting plasmid was named pPolyIIIDβ (Fig. 2.1A). A HSV tk gene cassette on a 2.1 kb *Xho*I-*Bam*HI fragment from plasmid pSPTK (a gift from Alan Clarke) was subcloned into the corresponding sites of pPolyIIIDβ at the 3' end of the β-casein gene fragment to obtain plasmid pPolyIIIDβ/TK (Fig. 2.1B).

A 1.1 kb *Xho*I-*Bam*HI end-filled fragment containing the neomycin-resistance gene from plasmid pMC1neo(c) was subcloned into the *Asp* 700I site in exon 2 of the β-casein gene fragment in plasmid pPolyIIIDβ/TK such that the β-casein and neo genes were in the same transcriptional orientation. The initiation codon of the β-casein gene is just upstream of this *Asp*700I site. This vector was designated pβMC1NEO/TK (Fig. 2.1C).

To construct the targeting vector pβ1MC1NEO/TK (Fig. 2.1 D), the 7.9 kb *Sca*I- *Bam*HI fragment from the plasmid pβMC1NEO/TK was subcloned into *Pvu*II and *Bam*HI sites of vector pPolyIIID. This targeting vector had 4.7 kb of β-casein gene sequences distributed as two arms of homology; 841 bp and 3.9 kb in length flanking the neo gene.

2.2.3. Construction of Vector pβ2MC1NEO/TK

The targeting vector pβ2MC1NEO/TK (Fig. 2.2D) was similar to vector pβ1MC1NEO/TK except that the former included an additional 1.35 kb of β-casein sequences on the long arm of homology.

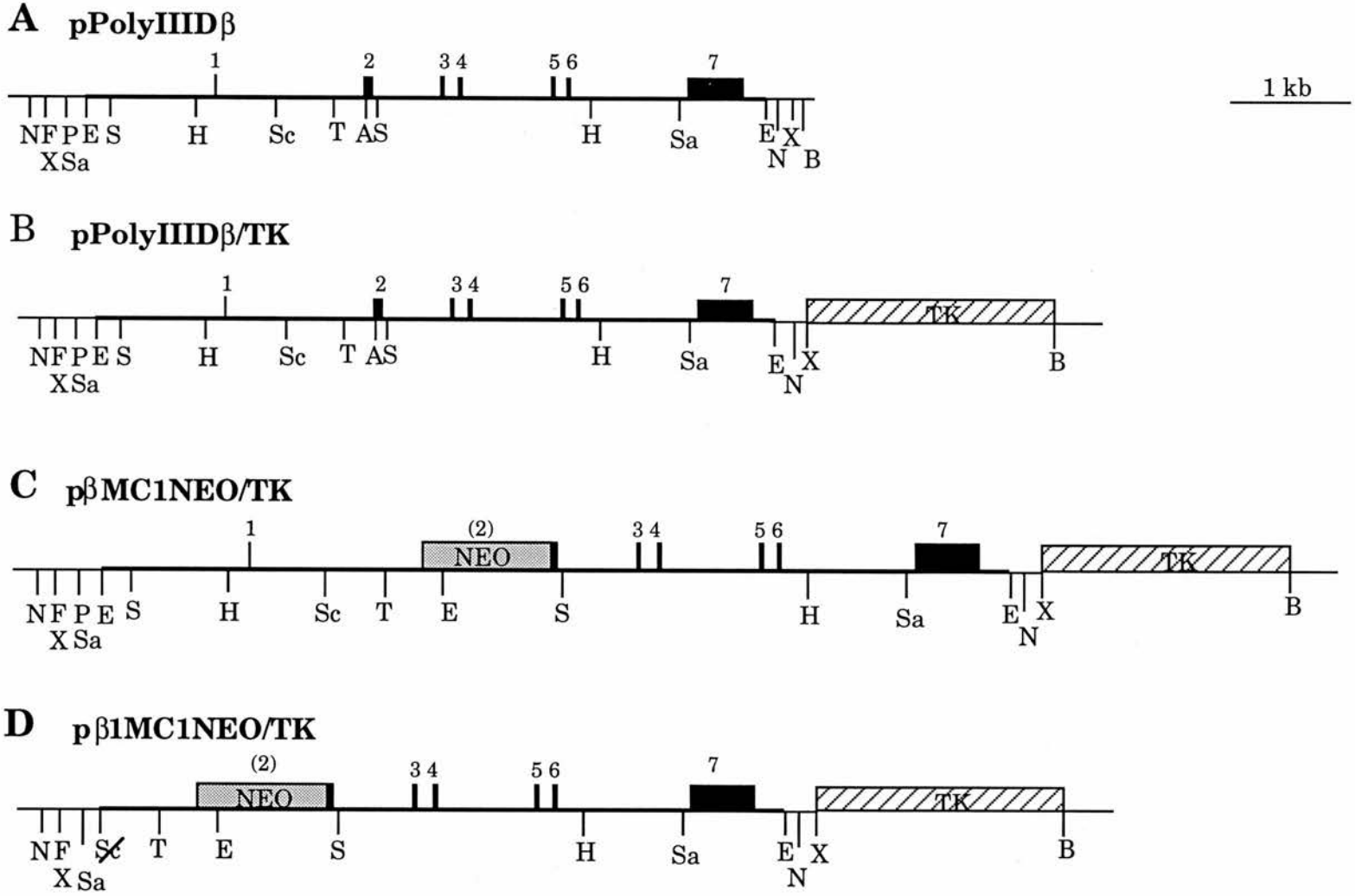

A pPolyIIIDβ
1
2
3 4
5 6
7
1 kb
NF P E S
X Sa
H
Sc
T A S
H
Sa
E X
N B
B pPolyIIIDβ/TK
1
2
3 4
5 6
7
TK
NF P E S
X Sa
H
Sc
T A S
H
Sa
E X
N
B
C pβMC1NEO/TK
1
(2)
NEO
3 4
5 6
7
TK
NF P E S
X Sa
H
Sc
T
E
S
H
Sa
E X
N
B
D pβ1MC1NEO/TK
(2)
NEO
3 4
5 6
7
TK
NF Sc
X Sa
T
E
S
H
Sa
E X
N
B

Figure 2.1. Construction of the mouse β-casein gene targeting vector pβ1MC1NEO/TK.

A) The 6.6 kb *Eco*RI fragment of the mouse β-casein gene was subcloned into the *Eco*RI site of the vector pPolyIIID to obtain plasmid pPolyIIIDβ.

B) The HSV tk gene on *Xho*I-*Bam*HI fragment from pSPTK was subcloned between the *Xho*I and *Bam*HI sites of pPolyIIIDβ to derive the vector pPolyIIIDβ/TK.

C) The neo gene on an *Xho*I-*Bam*HI fragment (end-filled by Klenow DNA polymerase I) from plasmid pMC1neo(c) was subcloned into the *Asp*700I site in exon 2 of β-casein gene fragment in pPolyIIIDβ/TK. The resulting plasmid was named pβMC1NEO/TK.

D) The targeting vector pβ1MC1NEO/TK was derived by subcloning the 7.9 kb *Sca*I-*Bam*HI fragment from plasmid pβMC1NEO/TK between the *Pvu*II and *Bam*HI sites of pPolyIIID.

Thin lines: plasmid polylinker sequences (not to scale); thick line: β-casein gene sequences; boxes numbered 1-7 and (2): β-casein exons 1-7 and exon 2 disrupted by neomycin gene, respectively; shaded box: neomycin gene cassette; hatched box: HSV tk gene cassette. A: *Asp*700I; B: *Bam*HI; E: *Eco*RI; F: *Sfi*I; H: *Hin*dIII; N: *Not*I; P: *Pvu* II; S: *Sac*I; Sa: *Sal*I; Sc: *Sca*I; T: *Bst*EII; X: *Xho*I.

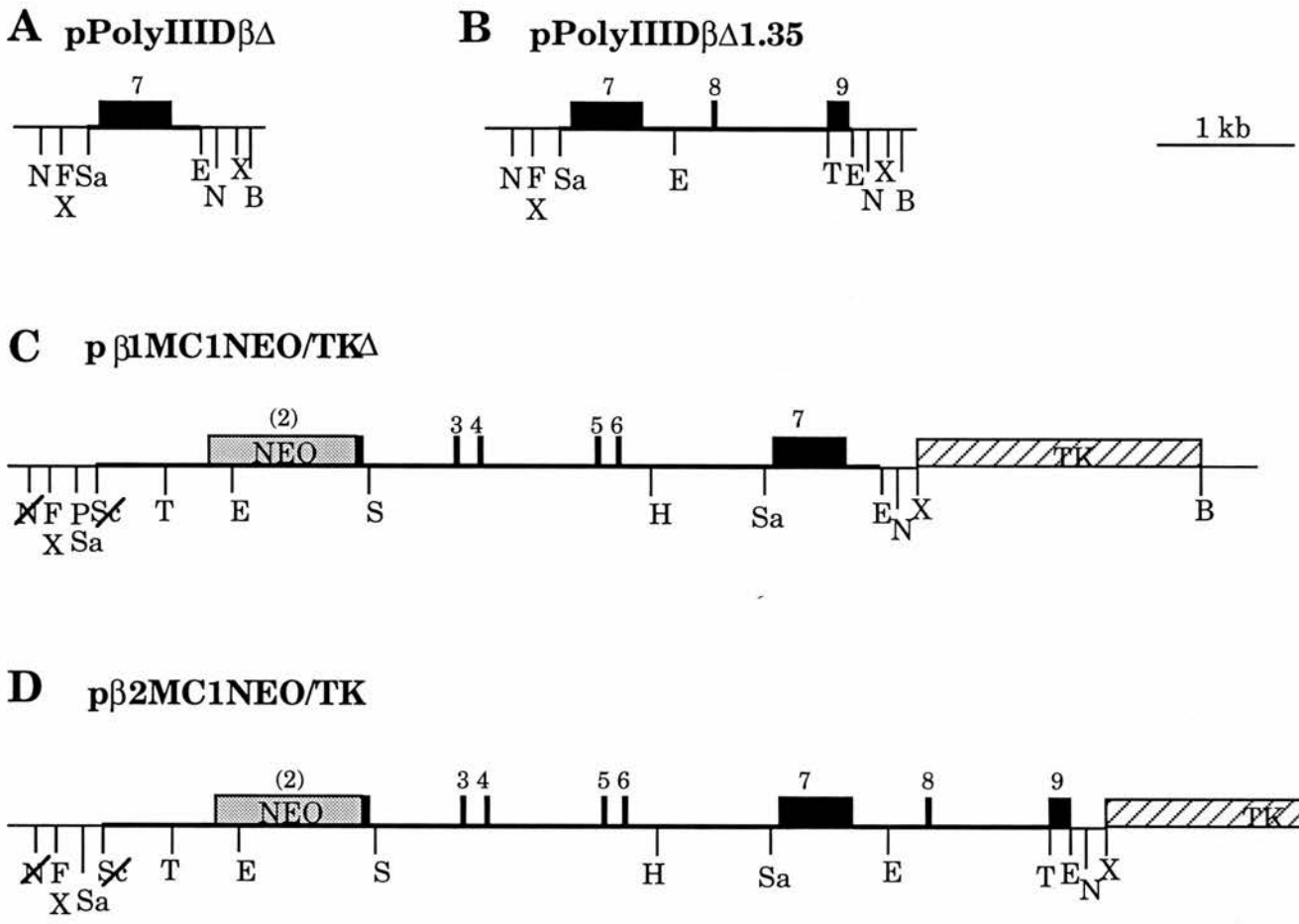

A pPolyIIIDβΔ
7
N F Sa
X
E
N
X
B
B pPolyIIIDβΔ1.35
7
8
9
N F Sa
X
E
T E X
N B
1 kb
C pβ1MC1NEO/TKΔ
(2)
NEO
3 4
5 6
7
TK
F P
X Sa
T
E
S
H
Sa
E
N
X
B
D pβ2MC1NEO/TK
(2)
NEO
3 4
5 6
7
8
9
TK
F
X Sa
T
E
S
H
Sa
E
T E
N
X
B

Figure 2.2. Construction of the mouse β-casein gene targeting vector pβ2MC1NEO/TK.

A) pPolyIIIβΔ was obtained by deletion of 5.2 kb *Sal*I fragment from plasmid pPolyIIIDβ (see Fig. 2.1A).

B) The 1.35 kb *Eco*RI fragment from 3' end of β-casein gene was subcloned into the *Eco*RI site of pPolyIIIβΔ to derive pPolyIIIDβΔ1.35.

C) Plasmid pβ1MC1NEO/TKΔ. was created after destroying a *Not*I site of pβ1MC1NEO/TK (in the polylinker 5' to the neo gene; see Fig. 2.1D) by end-filling with Klenow fragment.

D) The 859 bp *Sal*I-*Not*I fragment from pβ1MC1NEO/TKΔ was replaced with the 2.2 kb *Sal*I-*Not*I fragment from pPolyIIIDβΔ1.35 (Panel B) and the β-casein gene targeting vector pβ2MC1NEO/TK (Panel D) was derived. The vector has 6.1 kb of DNA sequences homologous to β-casein gene.

Thin lines: plasmid polylinker sequences (not to scale); thick line: β-casein gene sequences; boxes numbered 1, 3-9 and (2): β-casein exons 1, 3-9 and exon 2 disrupted by neomycin gene, respectively; shaded box: neomycin gene cassette; hatched box: HSV tk gene cassette. A: *Asp*700I; B: *Bam*HI; E: *Eco*RI; F: *Sfi*I; H: *Hin*dIII; N: *Not*I; P: *Pvu* II; S: *Sac*I; Sa: *Sal*I; Sc: *Sca*I; T: *Bst*EII; X: *Xho*I.

The 5.2 kb *Sal*I fragment was deleted from plasmid pPolyIIIDβ (Fig. 2.1A). The resulting plasmid pPolyIIIDβΔ (Fig. 2.2A) had the 0.75 kb *Sal*I-*Eco*RI fragment consisting of part of intron 6, exon 7 and part of intron 7 of the β-casein gene. A 1.35 kb *Eco*RI fragment comprising the remainder of intron 7, exon 8, intron 8 and part of exon 9 of the β-casein gene (Goodman and Rosen 1990) was inserted into the *Eco*RI site of plasmid pPolyIIIDβΔ. The orientation of this insertion was checked to ascertain that the structure of intron 7 of the β-casein gene was restored in plasmid pPolyIIIDβΔ1.35 (Fig. 2.2B).

A *Not*I site, 5' to the neo gene in plasmid pβ1MC1NEO/TK (Fig. 2.1D) was destroyed by end-filling of linear pβ1MC1NEO/TK after partial digestion with *Not*I and using Klenow polymerase to obtain plasmid pβ1MC1NEO/TKΔ (Fig. 2.2C). The 859 bp *Sal*I-*Not*I fragment in the latter plasmid was substituted with the 2.2 kb *Sal*I-*Not*I fragment from plasmid pPolyIIDβΔ1.35 (Fig. 2.2B) and the resulting vector was designated pβ2MC1NEO/TK. This vector had 6.1 kb DNA sequences homologous to mouse β-casein gene and distributed as two arms of 841 bp and 5.25 kb.

2.3. CONSTRUCTION OF HPRT INSERTION-TYPE GENE TARGETING VECTORS

2.3.1. Construction of Plasmid pIV

Plasmid pMC1neopA(c)ΔB was derived from pMC1neopA(c) by deletion of the unique *Bam*HI site via end-filling with Klenow fragment and recircularisation. A 5.6 kb *Eco*RI (intron I) - *Sau* 3A (exon III) fragment from the HPRT gene (Melton et al. 1984) cloned between

the *Eco*RI and *Bam*HI sites of plasmid pGEM4 (Fig. 2.3A) was available from Dr. J. P. Simons. A neo gene cassette on a *Xho*I-*Sal*I fragment from pMC1neopA(c)ΔB was inserted into the *Sal*I site of pGEM4. The resulting plasmid pIV (Fig. 2.3B) had a unique *Bam*HI site in the HPRT sequences and when linearised with this enzyme this plasmid served as an insertion-type gene targeting vector with two unequal arms of 4.8 kb and 0.8 kb homologous to the HPRT gene.

2.3.2. Construction of HPRT Insertion-Type Targeting Vectors with Terminal Heterologies

A unique *Xho*I site was introduced into the end-filled *Bam*HI site of pIV by cloning in an 8-mer *Xho*I linker (CCTCGAGG). The resulting plasmid, designated pIVX (Fig. 2.3C), when opened at this *Xho*I site had terminal heterologies of 5 nucleotides on one strand and a single base on the second strand of each of its two homologous arms.

A series of insertion-type HPRT gene targeting vectors were generated from the vector pIVX In these vectors the HPRT homology was interrupted by incorporation of heterologous sequences of various lengths (2.1 kb-190 bp) at the *Xho*I site of pIVX. By cloning each one of these heterologous fragments in both orientations, the heterologous DNA could be left on either of the two homologous arms of the HPRT gene targeting vectors after *Xho*I digestion.

Plasmids pIVL2.1 (Fig. 2.3D) and pIVS2.1 (Fig. 2.3E) were obtained by cloning in a HSV tk gene (a 2.1 kb *Xho*I-*Sal*I fragment from pSPTK) into the *Xho*I site of pIVX in both orientations. *Xho*I digestion of plasmids pIVL2.1 and pIVS2.1 leaves 2.1 kb heterologous sequences on the long and the short arms of these vectors, respectively. Plasmid

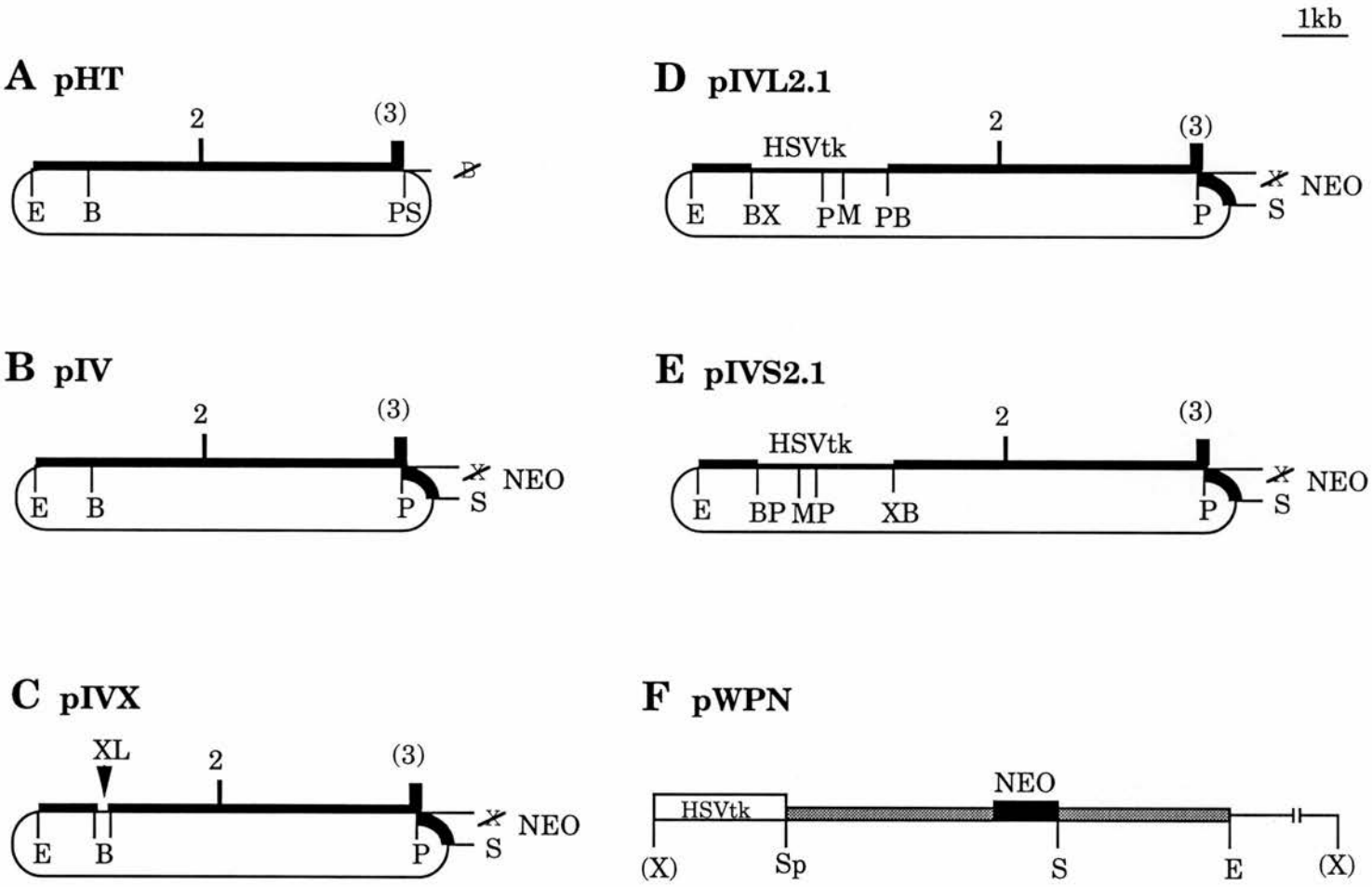

1kb
A pHT
2
(3)
E
B
PS
B pIV
2
(3)
E
B
P
S
NEO
C pIVX
XL
2
(3)
E
B
P
S
NEO
D pIVL2.1
HSVtk
2
(3)
E
BX
PM
PB
P
S
NEO
E pIVS2.1
HSVtk
2
(3)
E
BP
MP
XB
P
S
NEO
F pWPN
HSVtk
NEO
(X)
Sp
S
E
(X)

Figure 2.3. Construction of the HPRT gene targeting vectors and map of the negative control vector pWPN.

A) Plasmid pHT has 5.6 kb fragment of the mouse HPRT gene (part of intron I, exon 2 and part of exon 3) in the backbone of vector pGEM4.

B) Plasmid pIV was derived by subcloning neo gene [*Xho*I-*Sal*I fragment from pMC1neopA(c)] into the *Sal*I site of pHT polylinker. The HPRT and neo gene sequences are in the same transcriptional orientation.

C) An *Xho*I linker was inserted into the unique *Bam*HI site of pIV to obtain pIVX.

D) Plasmid pIVL2.1 was obtained from plasmid pIVX by inserting HSV tk gene (2.1 kb *Xho*I-*Sal*I fragment) into *Xho*I site of the latter. The HPRT and tk gene sequences are in the same transcriptional orientation.

E) Plasmid pIVS2.1 was derived from plasmid pIVX by inserting HSV tk gene (2.1 kb *Xho*I-*Sal*I fragment) into *Xho*I site of the latter. The HPRT and tk gene sequences are in the opposite transcriptional orientation.

Thin line: plasmid vector sequences (not to scale); Thick line: HPRT gene sequences; boxes numbered 2 and (3): HPRT exons 2 and 3 (partial); neo: neomycin gene; HSV tk: HSV thymidine kinase gene from pSPTK; arrow (XL): *Xho*I linker (not to scale). Shaded rectangles: mouse whey acidic protein (WAP) gene sequences. B: *Bam*HI; E: *Eco*RI; M: *Sma*I; P: *Pst*I; S: *Sal*I; X: *Xho*I; Sp: *Sph*I.

pIVL2.1 can also be linearised at the unique *Sma*I site included in the tk gene sequences. The resulting insertion vector had its long and short arms blocked by 0.7 and 1.4 kb heterologous sequences, respectively.

Plasmids pIVL0.7 and pIVS0.7 were constructed from plasmids pIVL2.1 and pIVS2.1, respectively. DNA from the latter vectors were digested with *Sma*I and *Xho*I, end-filled and religated. This step recreated the unique *Xho*I site in pIVL0.7 and pIVS0.7; when opened at this site, these targeting vectors had either the long or the short homologous arms blocked by 0.7 kb tk sequences.

Plasmids pIVL0.33 and pIVS0.33 were made from vector pIVX. A 334 bp *Xho*I-*Sal*I fragment was subcloned into the *Xho*I site of pIVX in both orientations. This 334 bp fragment was obtained from a plasmid pMC1neoΔA. The latter plasmid was derived by deletion of an *Xma*III-*Bam*HI fragment from the plasmid pMC1neopA(c).

A 182 bp *Eco*RI fragment from pMC1neo was end-filled and subcloned into the *Bam*HI site in the polylinker of the vector pGEM4 after end-filling with Klenow fragment of DNA polymerase I. A 194 bp *Xho*I-*Sal*I fragment (containing the 182 bp *Eco*RI fragment) was isolated and inserted into the *Xho*I site of pIVX in both orientations to generate plasmids pIVL0.19 and pIVS0.19.

To act as a negative control for the HPRT gene targeting experiments, pWPN (Fig. 2.3F), a replacement type vector (Mansour et al. 1988) suitable for targeting mouse whey acidic protein (WAP) gene was used. Briefly, this vector had approximately 6.0 kb of WAP gene sequences interrupted by the 1.1 kb *Xho*I-*Sal*I neo cassette from pMC1neopA(c) in the back bone of pSP72 (Promega). The 5' end of the WAP sequences were flanked by 2.1 kb tk gene from pSPTK. For

electroporation experiments the plasmid pWPN was opened at an *Xho*I site just 5' to the tk gene.

2.4. PREPARATION OF GENE TARGETING VECTOR DNA FOR ELECTROPORATION EXPERIMENTS

Gene targeting plasmid DNAs were digested with the appropriate restriction enzyme(s). The completion of digestion was checked on an agarose mini-gel stained with ethidium bromide (0.5 $\mu g ml^{-1}$; see section 2.8.1 for details of electrophoresis of DNA). The digested DNA was extracted once with phenol (equilibrated with 0.5 M Tris.Cl pH 8.0) followed by an extraction with chloroform. The DNA was precipitated from the aqueous phase by addition of 0.1 volume sodium acetate (pH 5.0) and 2 volumes of ethanol at room temperature for 10 minutes and the DNA was pelleted by centrifugation at 12000g for 10 minutes and washed twice with 70% ethanol. The last traces of liquid were removed taking care not to over dry the DNA. The pellet was dissolved in TE buffer by treating at 65°C for 10 minutes. The concentration of DNA was measured by reading the absorbance at 260 nm and the DNA was stored at -20°C until needed.

2.5. CULTURE OF EMBRYONIC STEM CELLS

Gene targeting experiments in mouse embryonic stem (ES) cells were performed at the IAPGR, Roslin and the Cancer Research Campaign (CRC) Laboratories, Department of Pathology, Medical School, University of Edinburgh. The ES cell culture conditions employed at the IAPGR, are described in detail and any significant differences between the two laboratories are highlighted at the end of this section.

E14 embryonic stem cells (Handyside et al. 1989) at passages 13-15 and EFC-1 embryonic stem cells (Nichols et al. 1990) at passage 15 were obtained as frozen aliquots from Dr. Martin Hooper, Department of Pathology, Medical School and Dr. Austin Smith, AFRC Centre for Genome Research, University of Edinburgh, respectively. These cells were expanded for 3-4 passages and refrozen again in several aliquots in liquid nitrogen until needed (see section 2.5.4). The ES cell culture methods described by Robertson (1987) were used with modifications.

ES cells were cultured under strict aseptic conditions including regular wiping of work surfaces with 70% ethanol and Microzid (Sterling Medicare). Mycoplasma infection of cells was monitored by a variety of methods by Moredun Animal Health Ltd, Edinburgh. Mycoplasma infection was not detected during the investigations reported in this thesis.

2.5.1. Tissue Culture Equipment, Glassware and Plasticware

Cell culture work was performed in a vertical laminar flow cabinet (Medical Air Technology Ltd) under aseptic conditions. Culture vessels and medium bottles were handled in the vicinity of Bunsen burner flame inside the flow cabinet. Other important pieces of equipment in the tissue culture facility included three humidified CO_2 incubators set at 5% CO_2 and 37°C, an inverted phase contrast microscope (Nikon Diaphot; magnification: x400, x100 and x40) and a bench-top centrifuge (Jouan).

All the glassware was soaked in water overnight, cleaned with 7X-PF (Flow Laboratories), washed thoroughly several times with water to

remove the detergent, rinsed in deionised distilled water (ddH_2O) and drained. Glass pipettes and Pasteur pipettes were heat sterilised at 180°C for four hours. All other glassware was autoclaved.

2.5.2. Culture Media and Solutions for Tissue Culture

Complete Medium for ES cells (IAPGR, Roslin)

500 ml	Dulbecco's Modification of Eagle's Medium (DMEM; High glucose formulation; Life Technologies)
28 ml	Foetal calf serum*
28 ml	Newborn calf serum*
5.6 ml	Non-essential amino acids (100x; Life Technologies)
5.6 ml	L-Glutamine (200 mM; Life Technologies)
4.4 µl	β-Mercaptoethanol (Sigma)
0.25-1.0 ml	Differentiation inhibiting activity/leukaemia inhibitory factor (DIA/LIF) medium†

All ingredients except the DMEM were mixed and filtered through 0.22 µm filters (Millipore) into the bottle containing DMEM. The complete medium was stored at 4°C and supplemented with fresh L-glutamine if the medium was in use for more than a week. No antibiotics were added to the media used for routine ES cell culture. Penicillin and streptomycin (@ 100 $IUml^{-1}$ and 100 μgml^{-1} of culture media, respectively: Life Tecnologies) were added to culture media used

---

* A number of serum batches were tested for cloning efficiency and their ability to support undifferentiated ES cell proliferation by Ray Ansell (see Robertson 1987 for details). Serum batches giving at least 20 % cloning efficiency were used.

† Medium conditioned by Cos-7 cells transfected (see Chen & Okayama 1987 for transfection method) with pLIF-D, a mouse LIF-D cDNA in an expression vector pMT2 (see Rathjen et al. 1990), was provided by Ray Ansell. The amount of LIF medium used varied depending upon the differentiation inhibitory activity of a given batch of conditioned medium.

in all the HPRT gene targeting experiments and one of the β-casein gene targeting experiments.

CMβ: Complete Medium for ES cells (CRC Laboratories)

45.0 ml	10x Glasgow Modification of Eagle's Medium (Flow Laboratories)
25.0 ml	Foetal calf serum*
25.0 ml	Newborn calf serum*
16.5 ml	$NaHCO_3$ (7.5% Flow Laboratories)
5.0 ml	Non-essential amino acids (100x; Flow Laboratories)
5.0 ml	L-Glutamine (200 mM; Flow Laboratories)
5.0 ml	Sodium Pyruvate (100 mM Flow Laboratories)
4.4 µl	β-Mercaptoethanol (Sigma)
0.25-1.0 ml	DIA/LIF†
375 ml	ddH_2O

TEG (Trypsin EGTA)

100 ml	2.5% Trypsin (Flow)
0.5 g	EGTA (Sigma)
150 mg	PVA (Sigma)
7.0 g	NaCl
0.12 g	$Na_2 HPO_4$
0.24 g	KH_2PO_4
0.37 g	KCl
1.0 g	D-Glucose
3.0 g	Tris
1.0 ml	1% Phenol Red

The above constituents were dissolved in 950 ml of autoclaved AnalaR water (BDH), adjusted to pH 7.6 with HCl, made up to 1,000 ml, filter sterilised through 0.22 µm filter (Millipore) and stored in 20 ml aliquots at -20°C.

---

* Sera already tested for their suitability to support ES cells proliferation without any overt differentiation were available from the CRC Laboratories (see Robertson 1987)

† LIF medium was available from CRC Laboratories. The amount varied depending upon the differentiation inhibitory activity of a given batch of LIF medium.

TVP (Cell Disaggregation Solution)

0.025% (w/v) trypsin (Flow Laboratories 2.5% solution)
1 mM EDTA
1% (v/v) chick serum (Flow Laboratories)
in Ca^{2+}, Mg^{2+}-free PBS (Dulbecco's; Flow Laboratories)

Filter sterilised through 0.22 μm filter (Millipore) and stored in 20 ml aliquots at -20°C.

Gelatin (0.1%)

0.1% gelatin (Sigma) solution was prepared in AnalaR water (BDH), autoclaved twice and stored at room temperature (IAPGR).

At CRC laboratories 1% stock solution of gelatin (Sigma) was prepared, autoclaved twice and stored at 4°C. The stock solution was diluted 10x just prior to use.

HEPES Buffered Saline (HBS)

20.0 mM HEPES (BDH),
137.0 mM NaCl,
5.0 mM KCl,
0.7 mM Na_2HPO_4,
6.0 mM Glucose

Adjusted to pH 7.05, sterilised by autoclaving and stored at RT.

Phosphate Buffered Saline (PBS)

Ready to use Ca^{2+} and Mg^{2+}-free PBS tablets (Oxoid or Flow Laboratories) were dissolved in AnalaR water (BDH) autoclaved and stored at RT.

2.5.3. Maintenance of ES Cell Cultures

Embryonic stem cells were cultured on gelatin (0.1%) coated Nunclon plastic dishes without feeder cells in complete Dulbecco's Modification of Eagle's Medium (see section 2.5.2). The medium was supplemented with recombinant differentiation inhibiting activity/ leukaemia inhibitory factor (DIA/LIF) medium (Smith et al. 1988) to maintain the undifferentiated phenotype of ES cells. In the experiments undertaken to isolate β-casein gene-targeted ES cell clones for production of chimaeric mice no antibiotics were included in the cell culture medium at any stage except G418 in the selective media.

The cells were fed with fresh medium on alternate days and passaged at an interval of 2-3 days. To passage a confluent 25 cm^2 flask the medium was aspirated and the cell surface was washed gently twice with 5 ml of PBS (Section 2.5.2). The cells were covered with 0.5 ml of TEG (trypsin EGTA solution; Section 2.5.2) and monitored using a phase contrast microscope. The flask was tapped after two minutes to dislodge the cells and medium was added. Large clumps of cells were broken by pipetting the contents through a drawn-out Pasteur pipette a number of times. The cells were collected in a 10 ml centrifuge tube (Sterilin) and pelleted by centrifugation at 1000 rpm for 5 minutes. The medium was aspirated and the cell pellet was resuspended in a suitable amount of medium. Normally, the cells were split at 1 in 5 ratio.

2.5.4. Freezing and Thawing of ES Cells

Cells were harvested from a 25 cm^2 flask as described in section 2.5.3 and resuspended in 0.75 ml of complete medium. While

working on ice, 0.75 ml of ice-cold freezing medium (60% complete medium, 20% foetal calf serum, 20% dimethyl sulfoxide; Sigma) was added drop-wise with continuous mixing of the contents. The mixture was dispensed in 3x1.5 ml pre-cooled screw cap tubes (Sarstedt), placed inside a polystyrene box and left in a -70°C freezer to cool slowly. The tubes were transferred to liquid nitrogen after a few days.

To thaw the cells, the vial was quickly transferred from liquid nitrogen to a water bath at 37°C and agitated until the contents had thawed. The contents of the tube were added to ~15 ml of complete medium in a universal tube pre-warmed to 37°C. The cells were pelleted by centrifugation, suspended in 5 ml of fresh medium, plated out in a 25 cm^2 flask and placed in the CO_2 incubator.

2.5.5. Electroporation

Growing cells were harvested by trypsinisation, and resuspended at high density in either HBS or PBS (see section 2.5.2). The cells were counted using an improved Neubauer haematocytometer. Normally, the cell concentration was adjusted to 10^8 ml^{-1}. 0.8 ml cell suspension containing 25 μg to 150 μg of targeting vector DNA was placed in a Bio-Rad Gene Pulser cuvette (width 0.4 cm). The mixture was given a single pulse either at 800V, 3μF or 230V, 500μF. The electroporated cells were allowed to stay at room temperature for 10 minutes and then transferred to the complete medium and plated out in suitable petri dishes. The details of electroporation conditions, plating densities, selection regimes etc. followed in each individual electroporation experiment are described in the relevant sections of the following chapters.

2.5.6. Isolation and Expansion of Individual ES Cell Clones

Following electroporation and selection, individual ES cell colonies with an undifferentiated ES cell phenotype were picked up with a drawn out Pasteur pipette while working under a low magnification microscope (Olympus Zoom Stereo; x7-x40) in a still-air cabinet. Each colony was incubated briefly in a small drop (~50 µl) of trypsin-EGTA solution and was disaggregated using a finer diameter drawn-out Pasteur pipette to achieve an almost single cell suspension. The cell suspension from each colony was plated in a well of a 24-well tissue culture plate containing 1.5 ml of complete culture medium per well and was placed in the CO_2 incubator. The following day the cells were fed with fresh medium. At confluence, generally within a week, the cells were trypsinised and sequentially plated into larger vessels to obtain sufficient cells for preparation of genomic DNA and freezing of cells for future retrieval.

2.5.7. Chromosome Analysis

Chromosome spreads of targeted ES cell clones were analysed by the method described by Robertson (1987) except that the growing ES cells were not treated with colcemid. Cells growing in a sub-confluent 25 cm^2 flask were trypsinised as described in sub-section 2.5.3. Pelleted cells were resuspended in hypotonic solution (0.56% w/v KCl) for 5 minutes at room temperature and were recentrifuged at 500 rpm for 5 minutes to remove the supernatant. The cell pellet was finger flicked to resuspend the cells and the cells were fixed by adding dropwise 1 ml of ice-cold fixative (3 volume of absolute methanol: 1 volume of glacial acetic

acid; freshly prepared). More fixative was added and the tube was left at room temperature for 5 minutes. At least two more washes were performed with the fixative. The metaphase plates were prepared on dry and clean microscopic slides previously stored in 70% ethanol at 4°C. The slides were air dried, stained in 3% (v/v) Gurr's Giemsa stain (BDH) in PBS for 10-15 minutes and the chromosome number was counted by examination of slides under oil immersion objective (x1000 total magnification).

2.5.8. ES Cell Culture at the CRC Laboratories

The ES cells were cultured on Costar (Northumbria Biological Ltd) tissue culture dishes. The cell culture medium composition (CMβ: section 2.5.2) was different from the one in use at the IAPGR, Roslin. The trypsin concentration in cell disaggregating solution (TVP: section 2.5.2) was only one tenth of that in the trypsin solution (TEG) used for trypsinising ES cells at the IAPGR laboratory. The trypsinisation was performed by incubating the PBS washed growing ES cells in a tissue culture dish with TVP for 3-5 minutes at 37°C.

In electroporation experiments each selected colony was expanded to a super-confluent well of a 24-well plate. The cells were trypsinised with 2-3 drops of TVP and trypsinisation was stopped by adding 1 ml of freezing medium (80% CMβ:complete medium, 10%, foetal calf serum, 10% dimethyl sulfoxide; Sigma). One half of the cell suspension was frozen in Nunclon 1 ml freezing ampoules and the other half was processed for the preparation of genomic DNA. The simplicity of this method proved critical in analysing a large number of ES cell clones for the identification of targeted clones.

2.6. ISOLATION OF GENOMIC DNA FROM EMBRYONIC STEM CELLS

2.6.1. Large Scale Preparation

Genomic DNA from ES cells grown to confluence in tissue culture flasks (25 cm^2, 80 cm^2 and 175 cm^2) was prepared by the method of Blin and Stafford (1976) with modifications as described by Sambrook et al. (1989).

Typically, a 25 cm^2 tissue culture flask with ES cells grown to confluence was washed twice with PBS and incubated at 37°C for one hour after addition of 5 ml of extraction buffer (10 mM Tris.Cl pH 8.0, 0.1M EDTA pH 8.0, 20 $\mu g ml^{-1}$ pancreatic RNAase, 0.5% SDS). The flask was incubated for a further 3 hours after addition of proteinase K (100 $\mu g ml^{-1}$; final concentration). The contents of the flask were transferred to a 15 ml centrifuge tube and extracted once with phenol (equilibrated with 0.5 M Tris.Cl pH 8.0) and once with chloroform. The genomic DNA was precipitated from the aqueous phase with 2 volumes of ethanol after the addition of 0.1 volume of 10 M ammonium acetate. The precipitate was recovered on a hook made from a Pasteur pipette and washed twice with 70% ethanol. The last traces of liquid were removed from the DNA without over drying the pellet. Finally, the DNA pellet was suspended in 200 µl of TE buffer in an Eppendorf tube and left at 4°C for several days before measuring the absorbance of the DNA solution.

2.6.2. Small Scale Preparation

Genomic DNA from ES cell clones isolated in the electroporation experiments was prepared by a simplified method described by Laird et al. (1991).

Cells harvested from wells of 24-well tissue culture plates were pelleted by centrifugation in Eppendorf tubes and the medium was removed. 0.5 ml of lysis buffer (100 mM Tris.Cl pH 8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl, 100 $\mu g ml^{-1}$ proteinase K) was added to the cells and the tubes were incubated overnight in a shaking incubator at 37°C. The DNA was precipitated by addition of one volume of isopropanol. The DNA precipitates were recovered by centrifugation for 10 minutes, and washed twice with 70% ethanol. 100 µl of TE was added to each tube. The tubes were incubated at 65°C for 10 minutes and then left at 4°C for 2-3 days for complete dissolution of the DNA. One third of this DNA solution was sufficient for analysis by Southern blotting after digestion with an appropriate restriction enzyme. Restriction digestion reactions included 0.1 mg BSA/ ml of reaction volume (Laird et al. 1991).

2.7. ISOLATION OF MOUSE TAIL DNA

Five week-old mice were anaesthetised (Flecknell 1983) and 1-2 cm tail biopsies (by Frances Thomson or Roberta Wallace) were taken for the preparation of genomic DNA. The tails were incubated in 0.75 ml buffer (10 mM Tris.Cl pH 7.9, 1 mM EDTA pH 8, 0.3 M sodium acetate, 1% SDS and 200 $\mu g ml^{-1}$ proteinase K) overnight in a shaking incubator (37°C). 0.5 ml of this tail digest was extracted with phenol:chloroform once and DNA was precipitated by addition of 1 volume of isopropanol at room

temperature. The DNA was pelleted by centrifugation (12000g for 10 minutes), washed twice with 70% ethanol and dissolved in TE buffer as described in section 2.6. The DNA concentration was obtained by measuring the absorbance of the DNA solution at 260 nm.

2.8. ANALYSIS OF GENOMIC DNA

2.8.1. Electrophoresis of Genomic DNA

DNA samples (10 μg DNA per lane) digested with the appropriate restriction enzymes were electrophoresed on a 0.7-0.8% (w/v) agarose gel either at 1-2 $voltcm^{-1}$ overnight or at 4 $voltcm^{-1}$ for 3-5 hours in a continuous buffer system (40 mM Tris-acetate, 2.5 mM EDTA pH 7.7, with 0.5 μgml^{-1} ethidium bromide). A loading buffer (0.1 volume of a solution containing 40 mM EDTA, 0.1% SDS, 30% ficoll, 1.2 $mgml^{-1}$ bromophenol blue) was added to DNA samples before applying onto the agarose gels.

After completion of electrophoresis the gel was placed on a transilluminator (wavelength 302 nm) and photographed using Polaroid film (667).

2.8.2. Transfer of DNA onto Hybond-N Membranes

The DNA was transferred from the agarose gel onto the Hybond-N membrane (Amersham) using the blotting method described by Southern (1975).

The electrophoresed DNA was depurinated by gently shaking the gel in 0.2N HCl for 7-15 minutes. The latter solution was

replaced with a DNA denaturing solution (1.5 M NaCl, 0.5 M NaOH). After two 15 minutes treatments in this solution the gel was rinsed with ddH_2O and neutralised by shaking in two changes of neutralising solution (1.5 M NaCl, 0.5 M Tris.Cl pH 7.5, 1 mM EDTA) for 15 minutes each.

The gel was placed on two sheets of Whatman 3 MM paper on a glass plate and the paper was connected to a reservoir of 20x SSC solution (3 M NaCl, 0.3 M sodium citrate). A Hybond-N membrane cut to the size of the gel was prewetted in 3x SSC and placed on the top of the gel. Two sheets of wet (soaked in 3x SSC) and one sheet of dry 3 MM paper (all cut to the size of the gel) were placed on the filter. Any air bubbles trapped between gel, membrane and 3 MM papers were carefully removed by rolling with a sterile pipette. Paper towels were placed on top of 3 MM papers followed by a glass plate and a 0.5 -1 kg weight was placed on glass plate to draw liquid through the gel by capillary action into the paper towels. The transfer of DNA from the gel onto the membrane was allowed to proceed for more than 20 hours. The membrane was removed, rinsed in 3x SSC and the DNA was immobilised onto the membrane in an automatic UV linker (Stratagene).

2.8.3. Preparation of Radioactive Probes

Radioactive-labelled DNA probes were synthesised using a Multiprime DNA Labelling Kit (Amersham) based on the methods described by Feinberg and Vogelstein (1983, 1984).

25-50 ng of linear double-strand DNA in 28 µl of ddH2O in an Eppendorf tube was denatured by boiling for 2 minutes. The following

were added in this order and the tube was incubated at room temperature for 3-5 hours:

5 µl Primer solution
10 µl Unlabelled dNTPs except dCTP
5 µl ^{32}P labelled dCTP (Amersham: Sp. Activity 3000 Ci/mM)
2 µl Klenow fragment of DNA polymerase 1 (1unit/1 µl)

The extent of incorporation of radioactivity was measured by the TCA precipitation method described by Sambrook et al. (1989). The probe was extracted with phenol:chloroform, denatured by addition of 0.1 volume of a denaturing solution (1.5 M NaOH, 0.1 M EDTA) and added to the hybridisation solution after 5 minutes.

2.8.4. Hybridisation of Genomic DNA

Homologous probes were used to analyse the structure of the β-casein and HPRT genes in ES cell clones. The hybridisation conditions described by Church and Gilbert (1984) were used with modifications.

Membranes (usually 20x20 cm^2) were prehybridised in 30-40 ml of hybridisation solution (0.5 M phosphate buffer pH 7.2, 7% SDS, 1 mM EDTA) in a sealed perspex box at 65°C for a minimum of 30 minutes. The solution was replaced with the same amount of fresh hybridisation solution containing denatured probe (see above). Hybridisation was continued for at least 12 hours at 65°C. To remove unhybridised probe the membranes were washed twice for 15 minutes each in a washing solution (40 mM phosphate buffer, 1% SDS, 1 mM EDTA) at 65°C. Occasionally, a further more stringent wash (20 mM phosphate buffer,

1% SDS, 1 mM EDTA) was used to remove background signal. Membranes were wrapped in Saran Wrap after draining and exposed to X-ray film (AGFA; CURIX RP1) placed between intensifying screens in an autoradiography cassette at -70°C. Generally, it was possible to obtain sufficient signal from a single copy gene fragment after 24 hours of exposure; otherwise, the exposure duration was extended for up to two weeks. If necessary, the probe was stripped by adding the membrane to boiling 0.1% SDS. The solution was allowed to cool to RT and complete removal of the probe from the membrane was checked by exposing it to film for several days prior to hybridisation with another probe.

2.9. PRODUCTION OF CHIMAERIC MICE

Injections of ES cells into mouse blastocyst (3.5 *d.p.c.*) were carried out by Dr. Martin Hooper's laboratory, and the injected blastocysts were implanted into the uteri of pseudo-pregnant mice by Roberta Wallace.

ES cells carrying targeted mutations were injected into 3.5 day C57/BL/6/Ola x CBA/Ca/Ola F_2 blastocysts as described by Bradley (1987). The manipulated blastocysts were transferred to the uteri of pseudo pregnant MF1 female mice (2.5 *d.p.c.*). The genotypes of the blastocysts and E14 ES cells for their respective coat colour phenotypes are presented in Table 2.1. The parents of the host blastocysts segregated at the *A* locus resulting in agouti and non-agouti mice. Any contribution from ES cells was easily manifest as light yellow patches on an otherwise dark coat colour. Chimaeras were bred with MF1 mice and the germ line contribution from ES cells was identified by the presence of light chinchilla offspring.

Table 2.1. Coat colour genotypes and phenotypes of mice from host blastocysts, embryonic stem cells and tester females.

	Genotype	Coat Colour of Progeny
Host Blastocyst F_2 (C57/BL/6/Ola x CBA/Ca/Ola)	C/C P/P AA, A/a, a/a	Agouti or black
Stem Cells 129/Ola	c^{ch}/c^{ch} p/p A^w/A^w	Chinchilla
Tester Female MF1	c/c a/a	Albino

Modified from Thompson et al. (1989).

2.10. RNA ANALYSIS

2.10.1. Isolation of Total RNA from Mouse Tissues

All solutions and plastics used for RNA analysis were treated with DEPC (Sambrook et al. 1989).

Nursing female mice were sacrificed by cervical dislocation on day-11 of lactation. Approximately 0.5 g of mammary tissue was removed, placed in 2 ml of 'RNazol B' (Biogenesis) and thoroughly homogenised. 0.1 volume of chloroform was added to the homogenate, mixed well and the mixture was kept on ice for 15 minutes. The aqueous phase was recovered by centrifugation (12,000g, 4°C, 15 minutes) and RNA was precipitated from it by addition of 1 volume of isopropanol. After 15 minutes at 4°C the RNA was pelleted (12000g for 15 minutes at 4°C), washed twice with 70% ethanol and dissolved in either 1 mM EDTA

pH 7.0 or formamide (Chomczynski 1992) and stored at -70°C or -20°C, respectively.

2.10.2. Electrophoresis of RNA

Total RNA samples were electrophoresed on a 1% agarose gel prepared in MOPS buffer (0.02 M 3-N-{morpholino} propanesulfonic acid, 5 mM sodium acetate) containing 6.8% formaldehyde, essentially as described by Sambrook et al. (1989).

10 μg of RNA in 10 μl was mixed with 39 μl of sample buffer (500 μl Formamide, 100 μl 10x MOPS, 160 μl of ~40% Formaldehyde), denatured at 65°C for 5 minutes, mixed with 5 μl of loading dye (50% glycerol, 0.1 $mgml^{-1}$ bromophenol blue) and electrophoresed in MOPS buffer.

2.10.3. Northern Blotting and Hybridisation

RNA from the gel was transferred onto Hybond-N (Amersham) membrane according to the methods described in section 2.8, omitting pre treatment of the gel.

2.11. MILKING OF MICE

Milk was obtained from nursing mothers on day 11 of lactation by Frances Thomson and Roberta Wallace as described by Simons et al. (1987). The pups were removed three hours before milking. 0.3 IU of oxytocin (Sigma) was injected intraperitoneally to mothers. After 10 minutes the mice were anaesthetised by the method of Flecknell (1983).

100 µl-200 µl of milk was recovered from each mouse into capillary tubes by gentle massage of mammary glands.

2.12. MILK PROTEIN ANALYSIS

Qualitative analysis of mouse milk samples was performed by SDS polyacrylamide gel electrophoresis according to the method of Laemmli (1970). Milk samples were diluted 1:8 in ddH_2O and defatted by centrifugation for 5 seconds in Eppendorf centrifuge. Defatted milk was further diluted (final dilution 1:200) in sample buffer (62.5 nM Tris.Cl 10% Glycerol, 2% SDS and 5% β-mercaptoethanol, 0.006% bromophenol blue) and a 15 µl aliquot of diluted milk was electrophoresed at a constant current of 32 mA for first hour and 48 mA for a further 5 hours on a 4% stacking gel and 15% resolving gel (gel thickness 1.0 mm). After electrophoresis the gel was fixed and stained overnight in a Coomassie Blue solution (0.04% Coomassie Blue, 2% TCA, 7.5% acetic acid, 50% methanol). Finally, the gel was bathed in destaining solutions (1st wash: 7% acetic acid, 23% ethanol for 30 minutes; second wash: 30% ethanol for 3-4 hours) to reveal individual milk proteins.

2.13. ESTIMATION OF MILK PROTEIN CONCENTRATION

Protein concentration of whole milk and whey was estimated by the micro-Kjeldahl method (Rowland 1938) as described by Davies and Law (1983) except that the N value was multiplied by 6.38 instead of 6.51 to obtain protein concentration.

Briefly, 25 µl of whole milk (or whey from 100 µl of whole milk), was hydrolysed by boiling in 3 ml of H_2SO_4 (nitrogen-free, 98% w/w:

BDH) in the presence of 2.0 g of potassium sulphate-selenium mixture (100 g of potassium sulphate and 1 g of selenium powdered together: BDH) in a graduated micro-Kjeldahl flask. The boiling was continued for 20 minutes after the sulphuric acid had started refluxing in the neck of the flask. When the flask had cooled, the contents were diluted to 70 ml with ddH_2O. The nitrogen concentration in the diluted acid hydrolysate was determined by the colorimetric method of Reardon et al. (1966) as described below.

To 1 ml of hydrolysate, 2.5 ml of reagent A (30.0 g NaOH per litre: BDH), 5.0 ml of reagent B (37.03 g sodium salicylate and 0.26 g sodium nitroprusside per litre: BDH) and 2.5 ml of reagent C (0.54 g sodium dichloroisocyanurate per litre: BDH) were added and mixed thoroughly. After 30 minutes, the absorbance of this solution was measured at a wavelength of 667 nm and compared with that of a standard (0.6604 g of ammonium sulphate per litre; 1 ml is equivalent to 0.14 mg N).

To obtain whey, caseins were removed from whole milk by acid precipitation (reduction of pH to 4.27 by addition of 0.2 M sodium acetate, 0.2 M acetic acid buffer) and followed by centrifugation (2560g for 10 minutes) at room temperature. The supernatant whey was filtered through a 0.22 μm filter prior to acid hydrolysis.

2.14. MEASUREMENT OF CASEIN MICELLE SIZE

The casein micelle size distribution was determined by Drs J. Leaver and D. Horne of Hannah Research Institute, Ayr, using photon correlation spectroscopy (Horne 1984; Horne and Dalgleish 1985).

2.15. STATISTICAL METHODS

Weaning weight data were analysed by ANOVA and the effects of β-casein genotype and sex of the pups were considered. To find out the influence of β-casein genotype on casein micelle size, ANOVA (one-way classification) was performed.

The significance of differences between β-casein genotypes with respect to weaning weight, milk protein concentration, casein micelle size and litter size was tested using Student's *t*-test (Snedecor and Cochran 1967).

To examine the effect of maternal genotype at the β-casein locus on the growth performance of pups, the litters were weighed thrice a week up to eleven days of age. To account for the variation due to the non-genetic effects (litter size and age at weighing), the following linear regression model was fitted by Anthea Springbett for each genotype separately:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

where:

Y = litter weight, X_1 = age at weighing, X_2 = litter size, and α, β_1, β_2 and β_3 were constants, and ε represented the residual effect.

More than 95% of the variation in litter weight was explained by the equation fitted for each genotype. The significance of differences between predicted values of litter weights was tested using Student's *t*-test (Snedecor and Cochran 1967).

CHAPTER THREE

GENERATION AND CHARACTERIZATION OF β-CASEIN-DEFICIENT MICE

3.1. INTRODUCTION

β-casein is a major milk protein and it exists as a component of casein micelles in all species studied so far (Creamer 1991; Davies et al. 1983). The amino acid sequences of β-caseins have been very poorly conserved during evolution; although there are no major deletions and insertions unlike in α-caseins (Bonsing and Mackinlay 1987; Holt and Sawyer 1988). Mouse β-casein is an approximately 26 kD phosphoprotein, and in common with β-caseins of various other species, it has a strongly hydrophobic C-terminal region and a cluster of serine phosphate residues in the hydrophilic N-terminus (Bonsing and Mackinlay 1987; Yoshimura et al. 1986; E. Stevenson and J. Leaver unpublished). Bonsing and Mackinlay (1987) have proposed that the overall architecture of β-casein molecule has been preserved in various species despite the high level of divergence, and that this may reflect functional constraints imposed on the structure of this protein during evolution. Further, the comparison of β-casein sequences from several species has shown conservation of a number of individual Gln, Pro and Lys residues (Holt and Sawyer 1988); it has been suggested that these residues may be important in the determination of the secondary structure and have functional implications. Pearse et al. (1986) reported that addition of partially

dephosphorylated β-casein to artificial casein micelles *in vitro* affected their coagulation time and curd formation reaction. Their interpretation of these data was that β-casein had a role in determining the surface properties of casein micelles.

The mouse β-casein gene, like its counterparts in cattle (Bonsing et al. 1988), rat (Jones et al. 1985), and rabbit (Thépot et al. 1991), is a single copy gene consisting of nine exons (Yoshimura and Oka 1989). In mouse it maps on chromosome 5 and is closely linked with the α-casein gene (Gupta et al. 1982; Geissler et al. 1988; Rosen et al. 1989). Another mouse casein gene, γ-casein (the corresponding protein in mouse milk is yet to be identified) also maps on this chromosome. In cattle α_S1-, α_S2-, β- and κ-casein genes have been found physically linked within 200 kb of DNA on chromosome 6 (Ferretti et al. 1990; Threadgill and Womack 1990). The presence of Ca^{2+}-sensitive casein genes as a close linkage group and conservation of 5' and 3' noncoding regions and sequences corresponding to signal peptides and phoshorylation sites in these genes have led to the suggestion of a casein gene family which arose by successive duplication of an ancient casein gene (Jones et al. 1985; Yu-Lee et al. 1986). By analogy with the β-globin locus (reviewed by Kollias and Grosveld 1992) it has been speculated that Ca^{2+}-sensitive casein genes might be regulated as a complex locus by common control elements (Rosen et al. 1989). However, any such shared regulatory elements are yet to be located in the casein gene complex of any species.

Cloning of major milk protein genes from various species has provided a novel way for production of heterologous proteins through transgenesis in the mammary gland of laboratory and farm animals (reviews: Clark et al. 1987; Hennighausen et al. 1990; Chapter 1: section 1.3). Expression of high levels of heterologous proteins in mammary gland

can be achieved by this approach (Archibald et al. 1990; Simons et al. 1987; Velander et al. 1992; Wall et al. 1991; Wright et al. 1991), but the site of integration and structure of the transgene arrays cannot be controlled *a priori*, and thus, the level of transgene expression is mostly unpredictable (Clark et al. 1990; Palmiter and Brinster 1986; Wall and Seidel Jr. 1992). Every transgene, in principle, represents an insertion mutation; it may not always be possible to breed the transgene to homozygosity if a vital endogenous locus were disrupted (Jaenisch 1988; Smith et al. 1987). These inherent problems of transgenic technology are further accentuated in farm animals due to the lower efficiency of transgenesis and the high cost of transgenic procedures in these species (Clark et al. 1990; Wall and Seidel Jr. 1992). The availability of pluripotential mouse embryonic stem cells (Evans and Kaufman 1981; Martin 1981) has allowed the introduction of targeted modifications into the mouse genome (reviewed by Bradley et al. 1992; Capecchi 1989a, 1989b; Evans 1989; Hooper 1992; Koller and Smithies 1992). Bradley et al. (1992) and Wilmut et al. (1991) have suggested that some limitations of transgenic technology could be overcome by placing the coding sequences of a heterologous protein under the control of the regulatory elements of an appropriate endogenous gene through homologous recombination. It is expected that the pattern of expression of the heterologous protein would be similar to that of the original target locus (Nandi et al. 1988; Mansour et al. 1990; Mouellic et al. 1990). The above technology should be useful to introduce quantitative and qualitative modifications in the endogenous milk proteins to either enhance the nutritional value of milk or to change its physical properties to suit the needs of the dairy industry (see Chapter 1: section 1.3.3).

To date, the mouse is the only vertebrate species in which targeted genetic changes into the endogenous genes have been feasible, and obviously, the availability of ES cells from farm animals may be crucial to effect similar modification in these species. An ideal target locus to achieve mammary gland specific expression of a heterologous protein through gene targeting would have to satisfy at least two important criteria. First, the level of expression of the endogenous protein must be sufficiently high; second, a null mutation in the targeted locus would have to be compatible with the well-being of the animal and normal physiology of mammary gland. β-casein meets the first requirement as it is the most abundant protein in many milks (Davies et al. 1983). However, the biological consequences of a null mutation in this locus will need to be assessed. The information on any essential *in vivo* function of the native protein will be equally important prior to targeting any subtle genetic changes into the endogenous β-casein to modify physical properties of milk of farm animals (Clark 1992; Jimenez-Flores and Richardson 1988).

Casein mRNAs have been detected in mouse cytotoxic T lymphocyte (CTL) cell lines and in thymus (Grusby et al. 1990). These workers have speculated that casein micelles play a role in CTL function and have proposed that the casein micelles may act as a vehicle for the delivery of perforin to the target cells (see review: Yagita et al. 1992 for the role of perforin in CTL function); the absence of free Ca^{2+} in the casein micellar environment would insure against the premature polymerisation of perforin. However, casein micelles or individual caseins are yet to be positively recognised in CTLs, CTL cell lines or thymus and thus, the functional significance of casein gene expression in thymus is not known.

To gain knowledge of the biological function of β-casein, I have generated β-casein-deficient mice through embryonic stem technology (Bradley et al. 1992; Capecchi 1989a; Evans 1989; Hooper 1992). The β-casein-deficient mice may be a suitable *in vivo* system to evaluate the physical properties of genetically engineered bovine β-casein and effects of such modifications on the characteristics of milk (Clark 1992). Expression of sheep β-lactoglobulin in the mouse mammary gland has revealed that the total protein concentration of milk did not increase inspite of the fact that β-lactoglobulin constituted 29% of milk protein (Wilde et al. 1992). Thus, there is a limitation to the level of milk proteins in mouse milk. The β-casein-deficient mice may be a useful animal model to understand the limits and dynamics of protein synthesis in the mammary gland.

This chapter describes the generation and characterization of mice lacking β-casein in their milk. It is shown that β-casein is not essential in either survival or lactation. The effects of this mutation on the milk composition may have implications for the genetic engineering of milk of farm animals.

3.2. RESULTS

3.2.1. Targeted Disruption of the Mouse β-Casein Gene in Embryonic Stem Cells

3.2.1.1. Gene Targeting Vectors and Scheme of Targeting

A β-casein gene targeting plasmid pβ1MC1NEO/TK, based on the positive-negative selection protocol of Mansour et al. (1988), contained the 4.7 kb *Sca*I (intron I) -*EcoR*I (Exon 7) non-isogenic DNA fragment of the mouse β-casein gene. The neomycin gene expression cassette from plasmid pMC1neo(c) was inserted into exon 2 of β-casein gene sequences just 3' of the initiation codon of the β-casein gene, and in the same transcriptional orientation. Plasmid pβ1MC1NEO/TK, when opened at a unique *Sfi*I site, served as a replacement type gene targeting vector with two arms of homology of 841 bp and 3.9 kb flanking the neo gene (Fig. 3.1B). The herpes simplex virus thymidine kinase (HSV tk) gene from pSPTK was included at the 3' end of the β-casein gene sequences in the targeting vector pβ1MC1NEO/TK to select against random integration of the targeting vector into the ES cell genome. A second β-casein gene targeting vector pβ2MC1NEO/TK was derived from the vector pβ1MC1NEO/TK after the addition of a further 1.35 kb of the β-casein gene sequences on the long arm of homology. Construction of the above vectors is described in Chapter 2 (see Figs. 2.1 & 2.2).

Homologous recombination of either of the above two gene targeting vectors with one of the β-casein alleles in the mouse embryonic stem cells would result in insertion of 1.1 kb neo gene into exon 2 of the β-casein gene (Fig. 3.1). This insertion would disrupt β-casein gene sequences coding for the signal peptide and allow

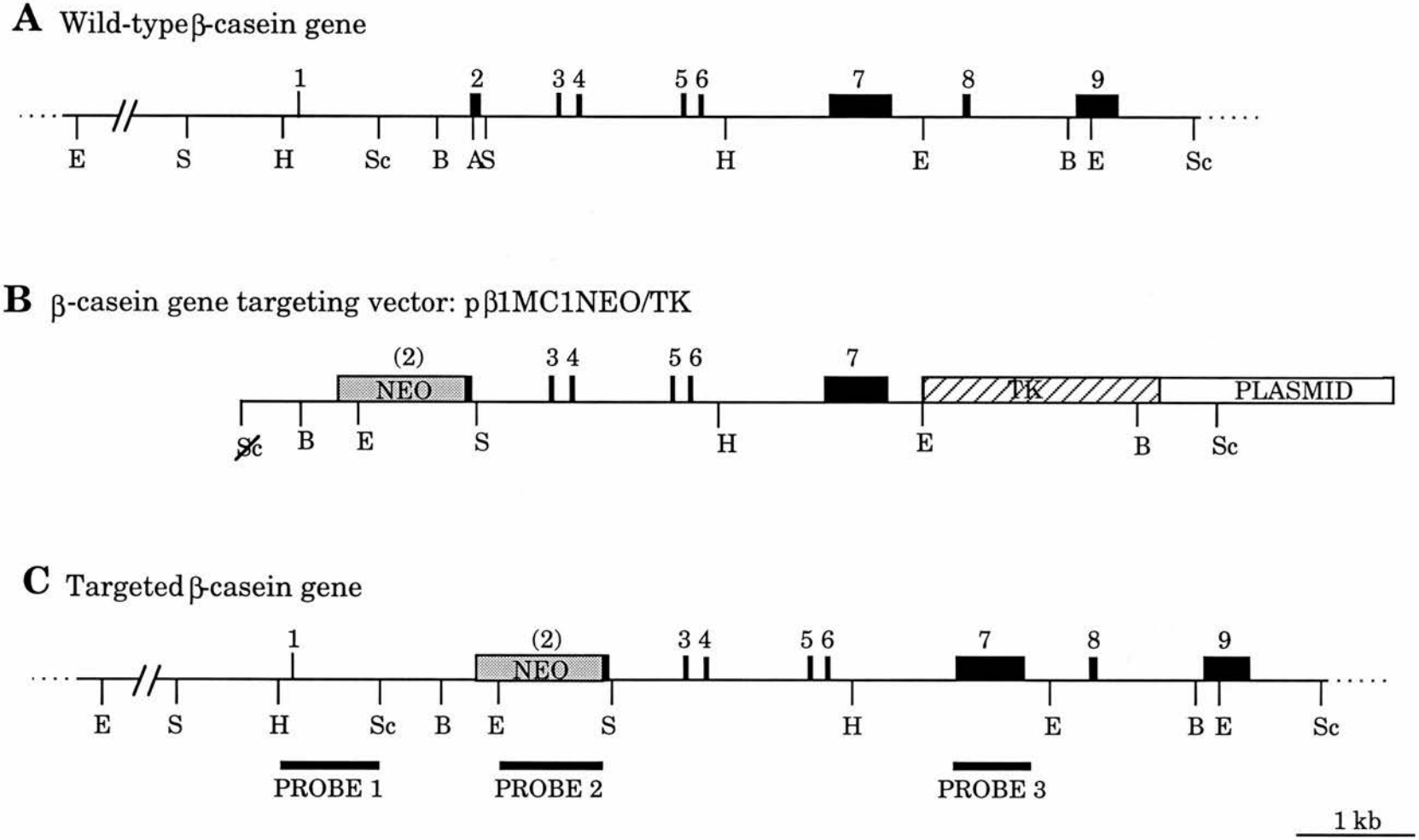

A Wild-type β-casein gene
1
2
3 4
5 6
7
8
9
E
S
H
Sc
B
AS
H
E
B E
Sc
B β-casein gene targeting vector: pβ1MC1NEO/TK
(2)
NEO
3 4
5 6
7
TK
PLASMID
Sc
B
E
S
H
E
B
Sc
C Targeted β-casein gene
1
(2)
NEO
3 4
5 6
7
8
9
E
S
H
Sc
B
E
S
H
E
B E
Sc
PROBE 1
PROBE 2
PROBE 3
1 kb

Figure 3.1. Schematic representation of the mouse β-casein gene targeting in ES cells.

A) Restriction map of wild-type mouse β-casein gene; the translation initiation codon is just 5' of the *Asp*7001 site in exon 2.

B) Restriction map of β-casein gene replacement-type targeting vector pβ1MC1NEO/TK with two unequal arms of homology of 841 bp and 3.9 kb. The HSV tk gene has been included at the 3' end of this vector to allow selection with GANC against random integration of the vector molecules in the ES cell genome.

C) Predicted restriction map of the β-casein gene after targeting by vector pβ1MC1NEO/TK.

Thin line: β-casein gene and flanking sequences; boxes numbered 1-9 and (2): β-casein exons 1-9 and exon 2 disrupted by neo gene, respectively. NEO; neomycin gene cassette from pMC1NEO(c); TK: HSV tk gene cassette from pSPTK; PLASMID: pPolyIIID vector sequences. Thick lines: position of DNA probes used in Southern blotting analysis of β-casein locus. A: *Asp*700I; B: *Bst*EII; E: *Eco*RI; H: *Hin*dIII; S: *Sac*I; and Sc: *Sca*I.

identification of the targeted clones after selection for neomycin gene expression with the drug G418. If transcription from the β-casein gene promoter were not affected by this mutation, a fusion pre-mRNA would be transcribed from the targeted β-casein allele. As the neo sequences in this fusion transcript would lack their own polyadenylation signal, this transcript should be processed into an mRNA species which would be 1.1 kb larger than the β-casein mRNA, assuming that splicing is not affected by insertion of the neo gene into exon 2 of the β-casein gene. Translation from the authentic β-casein AUG codon in this hybrid message would be terminated within the neo gene cassette sequences as there is a stop codon 240 nt downstream in this reading frame. The peptide which could be generated in this way would bear no resemblance to β-casein. There were premature translation termination codons in the other two reading frames as well.

3.2.1.2. Experiment-I: Targeting the β-Casein Gene in High Passage E14 ES Cells

A preliminary experiment was performed to test and characterise the targeting vector pβ1MC1NEO/TK. E14 ES cells (Handyside et al. 1989) at passage 35 were electroporated in the presence of linearised vector DNA (Table 3.1). Double selection with G418 and GANC resulted in approximately 15-fold fewer ES cell colonies as compared with the number of colonies generated after G418 selection alone. Although 15-fold enrichment after GANC selection was much below than the levels of enrichment reported by Mansour et al. (1988), it was still considered reasonable to use this selection regime for further experiments with β-casein gene targeting vectors.

Table 3.1. Targeting the β-casein gene in mouse embryonic stem cells.

Experiment		ES cell line and passage No.	Cells Electro-porated	Vector DNA (μg)	Cells plated	G418-resistantcolonies†	G418-GANC-resistant colonies	Enrichment factor after GANC selection	Targeted /Total clones screened (Targeting efficiency)
I*		E14, P35	$2x10^7$	25	$3.2x10^7$	741	50	14.8	2/19 (1:10)
II**		EFC-I, P21	$8x10^7$	150	$48x10^7$	24864	4848	5.1	0/168 (< 1:168)
III***	a	E14, P20	$8x10^7$	25	$8x10^7$	450	91	5.0	6/83; (1:14)
	b	E14, P19	$8x10^7$	75	$8x10^7$	1232	281	4.4	9/269 (1:29)
	c	E14, P19	$8x10^7$	150	$8x10^7$	672	145	4.6	4/121 (1: 30)

* 0.8 ml of electroporation mixture ($2.0x10^7$ E14 ES cells at passage 35 and 25μg of the linearised β-casein targeting vector pβ1MC1NEO/TK DNA in HBS) was electroporated in duplicate using Bio-Rad Gene Pulser (path length: 0.4 cm) at 750V, 3 μF. The electroporated cells were suspended in complete culture medium, plated in gelatinised 6 cm dishes ($2x10^6$ or $1x10^6$ cells per dish) and incubated at 37°C in 5% CO_2. G418 (Geneticin, Gibco) selection (500 μg/ml Geneticin of complete medium) was applied after 48 hrs. On day-5, GANC ($2x10^{-6}$ M) selection was applied in addition to G418 selection. Individual colonies were picked for clonal expansion between days 12 and 14 inclusive.

** Six aliquots, each consisting of 0.8 ml of electroporation mixture ($8.0x10^7$ EFC-1 ES cells at passage 21 and 150 μg of the linearised β-casein targeting vector pβ2MC1NEO/TK DNA in HBS) were electroporated at 750V, 3 μF. The electroporated cells were plated in 6 cm gelatinised culture dishes ($5x10^6$ cells per dish), grown and selected as described above in experiment-I. Individual colonies were picked for clonal expansion between days 10 and 14 inclusive.

*** Three aliquots of $8.0x10^7$cells each suspended in PBS were electroporated at 800 V, 3μF in the presence of 25 μg, 75 μg or 150 μg of the linearised β-casein targeting vector pβ1MC1NEO/TK DNA. The cells were plated in gelatinised 10 cm dishes ($8x10^6$ cells per dish). The following day G418 (300 μg per ml of medium) selection was applied, and GANC selection ($2x10^{-6}$ M) was applied on day-5. Individual colonies were picked for clonal expansion between days 12 and 14 inclusive.

† 10% of culture dishes in a given experiment were selected with G418 alone and the number of G418 colonies were counted on these plates. This figure was extrapolated to estimate the total number of G418 colonies.

3.2.1.2.1. Screening for targeting at the β-casein locus

Nineteen out of a total of 50 G418-GANC-resistant ES cell clones generated with the targeting vector pβ1MC1NEO/TK were expanded individually. To examine the structure of the β-casein locus in these clones, *Sac*I-digested genomic DNA was analysed by Southern blotting using a β-casein genomic DNA probe (*Hin*dIII-*Sca*I fragment from the 5' region of the β-casein gene: Probe 1; Fig. 3.1). This probe was predicted to hybridise to a 2.7 kb *Sac*I fragment from wild-type allele and to a 3.8 kb fragment from the targeted allele. Randomly integrated molecules of the targeting vector would not be recognised by probe 1 because the corresponding DNA sequences were not included in the targeting vector pβ1MC1NEO/TK.

Probe 1 hybridised to a 2.7 kb *Sac*I DNA fragment from all of the nineteen clones and control E14 cells as was expected from the structure of the non-targeted allele. However, DNA from two of the nineteen clones β1.2 and β1.3 had an extra band of approximately 3.8 kb (Fig. 3.2A). The presence of the 3.8 kb band in addition to the 2.7 kb wild-type band suggested that one of the β-casein alleles in each of these two clones had undergone modification.

To confirm that these two clones were modified as a result of targeting events at the β-casein locus, DNA from these clones was restricted with *Eco*RI or *Hin*dIII and hybridised with probe 1. This probe was expected to hybridise to a 9.8 kb *Eco*RI fragment of the endogenous β-casein gene (Fig. 3.2C). There are two *Eco*RI sites in the neo gene sequences included in the targeting vector pβ1MC1NEO/TK (Fig. 3.1). These two sites would be incorporated into exon 2 of one of the

Figure 3.2. Targeting of the β-casein gene in high passage E14 ES cells - Screening for targeting events and confirmation of the structure of the targeted β-casein gene.

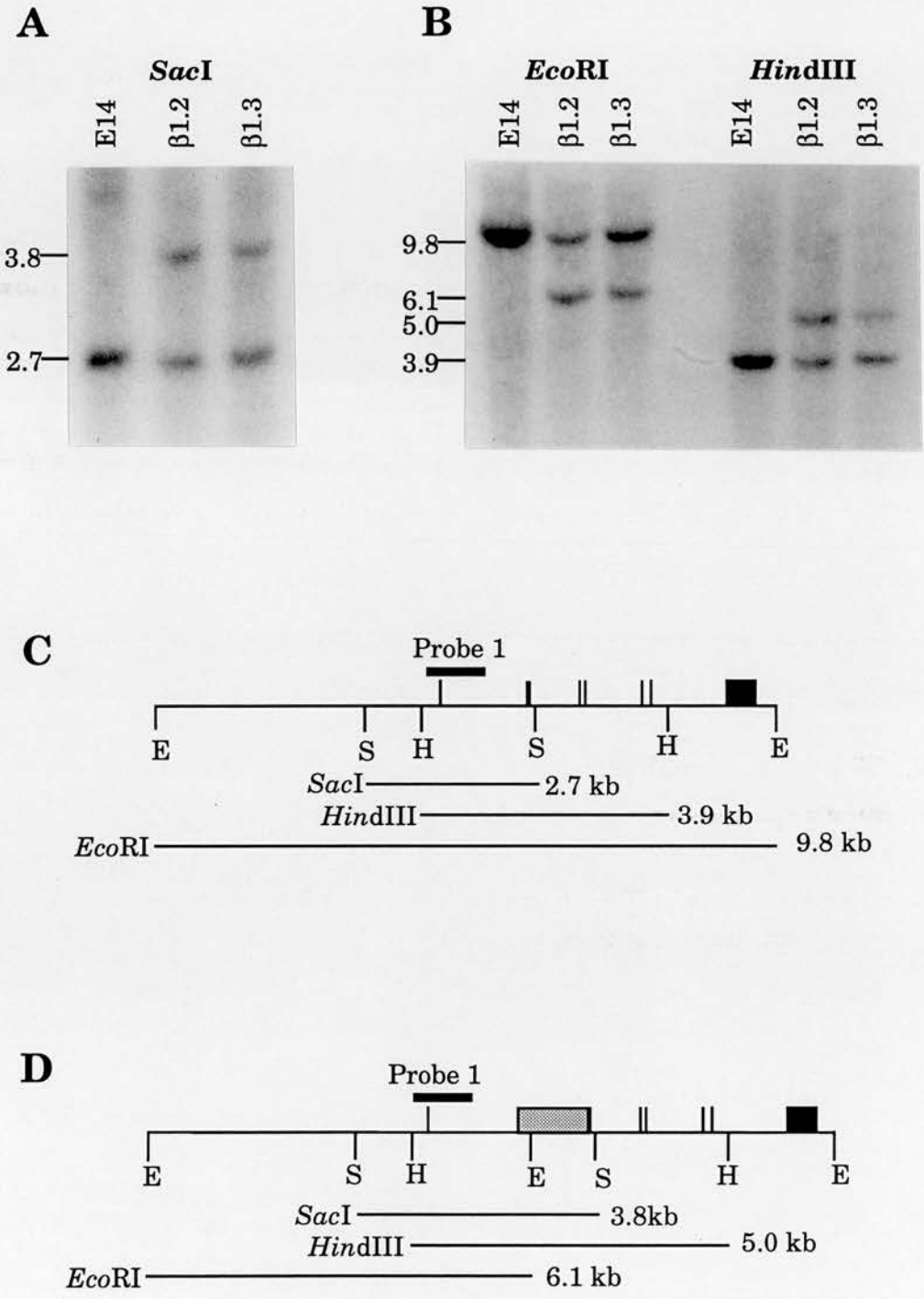
A) *Sac*I-digested genomic DNA from nineteen G418-GANC-resistant ES cell clones were analysed by Southern blotting with probe 1 (see Fig. 3.1). Two clones β1.2 and β1.3 gave a novel 3.8 kb fragment diagnostic of targeting of one of the β-casein alleles. The 2.7 kb *Sac*I fragment in these clones and control E14 cells corresponds to wild-type allele.

B) Confirmation of targeting in clones β1.2 and β1.3 by analysis of *Eco*RI and *Hin*dIII-digested DNA with probe 1; the 6.1 kb *Eco*RI and 5.0 kb *Hin*dIII fragments are diagnostic of targeting.

C) Map of wild-type β-casein gene showing fragments obtained from digests diagnostic of targeting using probe 1.

D) Map of the predicted structure of targeted β-casein gene showing fragments diagnostic of targeting using probe 1.

Boxes: Exons 1 to 7. Thick line: Position of β-casein genomic probe. E: *Eco*RI; H: *Hin*dIII; and S: *Sac*I..


A
SacI
E14
β1.2
β1.3
3.8
2.7
B
EcoRI
HindIII
E14
β1.2
β1.3
E14
β1.2
β1.3
9.8
6.1
5.0
3.9
C
Probe 1
E
S
H
S
H
E
SacI
2.7 kb
HindIII
3.9 kb
EcoRI
9.8 kb
D
Probe 1
E
S
H
E
S
H
E
SacI
3.8kb
HindIII
5.0 kb
EcoRI
6.1 kb

β-casein alleles after homologous recombination and thus, probe 1 should hybridise to a 6.1 kb *Eco*RI fragment from the targeted allele (Fig 3.2D). Southern blot analysis revealed that the clones β1.2 and β1.3 both had an approximately 6.1 kb *Eco*RI fragment in addition to the 9.8 kb *Eco*RI fragment from the wild-type allele (Fig. 3.2B). This observation indicated that clones β1.2 and β1.3 were generated as a result of successful targeting into one of the β-casein alleles in E14 cells by the vector pβ1MC1NEO/TK. Similarly, analysis of the DNA from these two clones with *Hin*dIII gave the novel 5.0 kb fragments in contrast to the 3.9 kb fragments from the non-targeted alleles (Fig. 3.2B), as was expected after targeted modification of the β-casein locus. All of these observations were consistent with the incorporation of the 1.1 kb neo gene cassette into exon 2 of the β-casein gene through homologous recombination between the targeting vector pβ1MC1NEO/TK and one of the β-casein alleles in E14 ES cells.

Finally, the *Sac*I-digested β1.2 and β1.3 DNA was electrophoresed in duplicate on the same gel and transferred onto Hybond-N membrane. One half of the membrane was hybridised with probe 1 and the other half with a 917 bp (*Eco*RI-*Sal*I) DNA probe from the neo gene (Probe 2; see Fig. 3.1). An indistinguishable 3.8. kb *Sac*I fragment was revealed by probes 1 and 2 confirming that the neo gene had been included in the β-casein locus through homologous recombination (data not shown). Further, the absence of any other bands in the DNA of targeted clones upon hybridisation with the neo probe indicated that the targeting vector did not integrate randomly into the genome of these clones.

The above electroporation experiment was planned to check the enrichment level that could be obtained with the

vector pβ1MC1NEO/TK upon negative selection with GANC and to serve as a pilot exercise to aid in the design of further gene targeting experiments. Nevertheless, two ES cell clones out of a total of nineteen clones analysed had undergone a targeting event at the β-casein locus. Taking into account the level of enrichment obtained by GANC-selection, this value represented one targeting event per 140 colonies generated by random integration of the vector (Table 3.1).

3.2.1.2.2. Chromosomal analysis of clones β1.2 and β1.3

Metaphase chromosomal spreads were prepared from the β-casein gene-targeted clones β1.2 and β1.3. Unfortunately, both of these clones had 42 chromosomes. It was thought very unlikely that these targeted clones would be able to populate the germ line of the mouse and form functional spermatozoa (Bradley 1987). The passage number of ES cells used in this experiment was relatively high (passage 35). Probably, a large proportion of these cells had aneuploid chromosomal profiles at the time of the electroporation experiment. As the aim of the project was to introduce the mutation into the mouse germ line and study the resulting phenotype, more experiments were undertaken to target the β-casein gene using low passage ES cells.

3.2.1.3. Experiment-II: Targeting the β-Casein Gene in Low Passage EFC-1 ES Cells

It has been reported that the targeting efficiency at the HPRT locus increases exponentially with a linear increase in the

extent of homology between the target locus and the gene targeting vector within the range of 2-14 kb (Deng and Capecchi 1992; Thomas and Capecchi 1987). The vector pβ1MC1NEO/TK, used in gene targeting experiment-I, had only 4.7 kb of DNA sequences homologous to the endogenous gene. To take advantage of increased homology, another β-casein gene targeting vector pβ2MC1NEO/TK (see Chapter 2: Fig. 2.2D) was used in a gene targeting experiment with EFC-1 ES cells (Nichols et al. 1990). The structure of the vector pβ2MC1NEO/TK was identical to that of the vector pβ1MC1NEO/TK except for the inclusion of an additional 1.35 kb of β-casein gene sequences on the long arm of homology. The vector pβ2MC1NEO/TK, when opened at a unique *Sfi*I site, had a total of 6.1 kb homology distributed as two arms of 841 bp and 5.25 kb flanking the neo gene.

EFC-1 ES cells (Nichols et al. 1990) were grown and electroporated at passage 21 in the presence of linearised pβ2MC1NEO/TK DNA and cells were selected with G418 and GANC (Table 3.1). Double selection gave ~5-fold fewer colonies than G418 selection alone. One hundred and sixty eight individual colonies from 25 independent plates were picked into 24-well tissue culture plates and expanded. From each clone, one confluent 25 cm^2 flask was frozen in three aliquots and another 25 cm^2 flask was processed for the preparation of DNA to screen for the identification of the targeted ES cell clones.

3.2.1.3.1. Screening for targeting at the β-casein locus

To screen for homologous recombination at the β-casein locus, the *Sac*I-digested DNA samples from 168 G418-GANC

colonies were analysed by Southern blotting using probe 1 as described in section 3.2.1.2.1. All of these clones had a 2.7 kb *Sac*I fragment released from the non-targeted endogenous β-casein allele (data not shown). The absence of any additional band with probe 1 was taken as evidence that none of these ES cell clones had undergone targeted modification at the β-casein locus. The level of enrichment in this experiment was only 5-fold in contrast to the 15-fold enrichment observed in experiment-I with the vector pβ1MC1NEO/TK. (Table 3.1). In spite of the increased homologous sequences included in the vector pβ2MC1NEO/TK in contrast to the vector pβ1MC1NEO/TK, and a large number of G418-GANC clones screened in experiment-II (equivalent to 840 G418-resistant clones), it was surprising not to find even a single targeting event in this experiment. However, this result was not directly comparable to that of the experiment-I because both the ES cell lines and targeting vectors used in these two experiments were different.

3.2.1.4. Experiment-III: Targeting the β-Casein Gene in Low Passage E14 ES Cells

A third series of β-casein gene targeting experiments was performed with low passage E14 cells at the CRC Laboratories, Department of Pathology, University of Edinburgh. ES cells grown to passage 19 and 20 were electroporated in the presence of three different concentrations of the vector pβ1MC1NEO/TK DNA (Table 3.1).

Genomic DNA from 473 individual G418-GANC-resistant ES cell clones was analysed by Southern blotting to identify clones carrying the targeted β-casein gene (see section 3.2.1.2.1) *Sac*I-digested DNA samples from these clones were hybridised with probe 1 (for an example see Fig. 3.3). The presence of the novel 3.8 kb *Sac*I

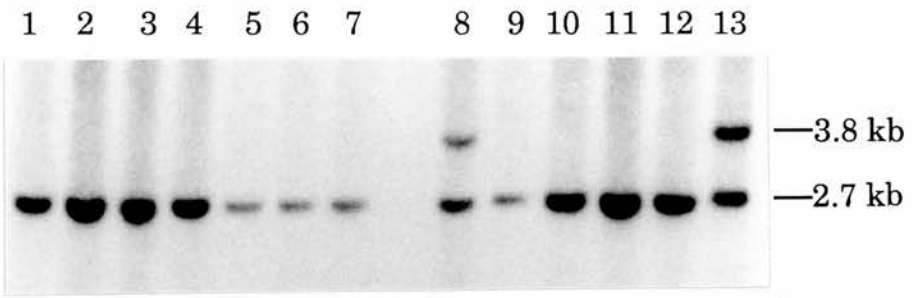


Figure 3.3. Targeting of the β-casein gene in low passage E14 ES cells: Screening for targeting events at the β-casein locus by Southern analysis.

Lanes 1-12: *Sac*I-digested DNA from G418-GANC-resistant E14 clones; lane 13: *Sac*I-digested positive control DNA from β-casein gene targeted clone β1.2 (see Fig. 3.2). The blot was hybridised with probe 1 (see Fig. 3.1). The 3.8 kb fragment was diagnostic of targeting (e.g. lane 8).

fragment in contrast to the 2.7 kb *Sac*I fragment from the endogenous β-casein gene was diagnostic of targeted modification. Nineteen clones were identified as having undergone modification of one of their β-casein alleles.

Overall, one targeting event per 25 G418-GANC-resistant ES cell colonies screened was recorded in experiment-III with the targeting vector pβ1MC1NEO/TK using low passage E14 cells (Table 3.1). After correcting for the level of enrichment, this value represented one targeting event per 125 random integration events and was close to the targeting efficiency observed with this vector using high passage E14 ES cells in Experiment-I (the significance of this comparison is limited by small size of Experiment-I). Thus, with this vector, there is no evidence for an effect of passage number on targeting efficiency, although Koller et al. (1991) have suggested that late passage cells might be easier to target. Three different concentrations of targeting vector DNA were used in the electroporation mixtures in this experiment. The targeting efficiency was ~2-fold higher when 25 μg of vector DNA was used than when 75 μg or 150 μg of DNA was used. The latter two levels of vector DNA gave similar targeting efficiencies. It should be noted that the effect of the vector DNA concentration on gene targeting efficiency was examined only in one experiment and, therefore, while there may be an effect of DNA concentration on targeting efficiency it is not possible to draw any firm conclusions about this.

3.2.1.4.1. Chromosomal analysis

Seven out of the nineteen β-casein gene targeted clones identified in experiment-III were thawed and expanded

for chromosomal analysis and to allow freezing of a number of aliquots of cells per clone. One of these clones differentiated spontaneously and could not be recovered. Metaphase chromosomal spreads were prepared from the six remaining clones (Table 3.2). Two independently isolated clones were selected for confirmation of the structure of their targeted β-casein alleles and production of chimaeric mice. Clone B256 had approximately 80% of cells with 40. Clone A77 had 40 chromosomes in approximately 60% of its cell population.

Table 3.2. Chromosomal analysis of the β-casein gene targeted ES cell clones.

ES cell clone	Chromosome number
A32	40, including one metacentric chromosome (20)*
A77	40 (12); 40? (3); 41 (2); 42 (2); 39(1)
B111	40 (11); 41(9)
B116	40 (13); 41(6); 31(1)
B256	40 (16); 39 (2); 4n† (2)
C128	variable, 38-42 (20)

* The figure in the parenthesis shows the number of metaphase spreads. Twenty metaphase spreads per clone were examined to determine the chromosome number.

† 4n: Tetraploid.

Clones A77 and B256 were selected for production of chimaeras.

3.2.1.4.2. Structure of the 5' end of targeted β-casein allele in clones A77 and B256

Correct targeting by the vector pβ1MC1NEO/TK at the 5' end of the β-casein locus was expected to generate a 3.8 kb *Sac*I and a 5.0 kb *Hin*dIII fragment from the modified

β-casein allele in addition to 2.7 kb and 3.9 kb respective fragments from the non-targeted allele (Fig. 3.4C & D). The novel 3.8 kb *Sac*I and 5.0 kb *Hin*dIII fragments should hybridise to both probe 1 from 5' region of the β-casein gene and probe 2 from the neo gene (Fig. 3.4D). Randomly integrated copies of the targeting vector would not be detectable using probe 1. Following random integration of the targeting vector probe 2 would hybridise to *Sac*I and *Hin*dIII fragments of unpredictable lengths.

DNA from clones A77 and B256 was digested with *Sac*I or *Hin*dIII and hybridised with either probe 1 (Fig. 3.4A) or probe 2 (Fig. 3.4B). Again probe 1 hybridised to a 2.7 kb and a 3.8 kb *Sac*I fragments in both of these clones as opposed to the 2.7 kb *Sac*I fragment from the control E14 cells (Fig. 3.4A). Similarly, this probe highlighted 3.9 kb and 5.0 kb *Hin*dIII fragments from clone A77 and B256 in contrast to the single 3.9 kb *Hin*dIII fragment in the control E14 DNA. The presence of the 3.8 kb *Sac*I and 5.0 kb *Hin*dIII fragments in these two clones suggested that the 5' end of the targeting vector had correctly recombined with the 5' end of one of the β-casein alleles.

The 3.8 kb *Sac*I and 5.0 kb *Hin*dIII fragments from A77 and B256 DNA detected by probe 1 also hybridised to the probe 2 from the neo gene (Fig. 3.4B) confirming that the neo gene had been inserted into the 5' end of one of the β-casein alleles in these clones as was envisaged.

3.2.1.4.3. Structure of the 3' end of the targeted β-casein allele in clones A77 and B256

To analyse the structure of the 3' end of the targeted β-casein allele in clones A77 and B256, DNA was restricted with *Bst*EII or *Sca*I and probed with a 642 bp *Sal*I-*Sac*I β-casein genomic

Figure 3.4. Targeting of the β-casein gene in low passage E14 ES cells: Structure of the 5' end of the targeted allele in clones A77 and B256.

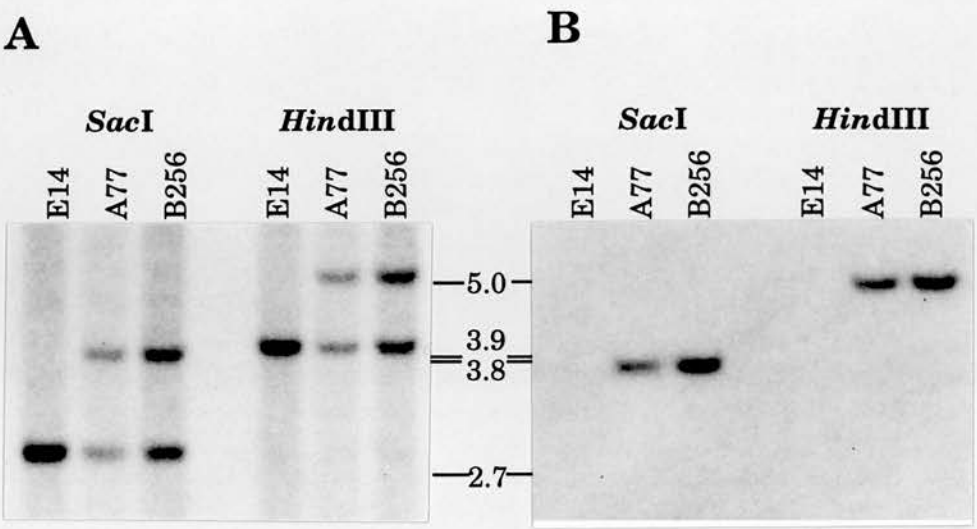
A) *Sac*I and *Hin*dIII-digested genomic DNA was probed with probe 1 (Fig. 3.1). The 3.8 kb *Sac*I and 5.0 kb *Hin*dIII fragments from A77 and B256 DNA are diagnostic of correct targeting at the 5' end of β-casein gene.

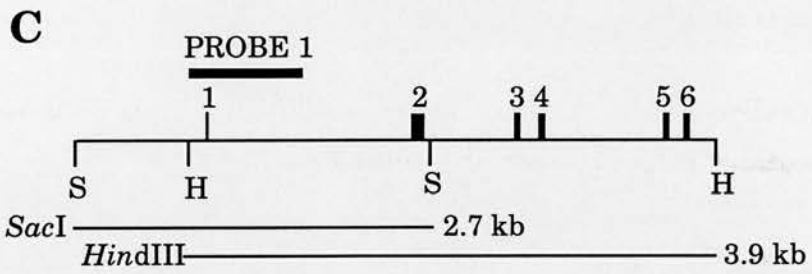
B) *Sac*I and *Hin*dIII-digested genomic DNA was probed with probe 2 (Fig. 3.1). The 3.8 kb *Sac*I and 5.0 kb *Hin*dIII fragments from A77 and B256 DNA are diagnostic of correct targeting at the 5' end of β-casein gene.

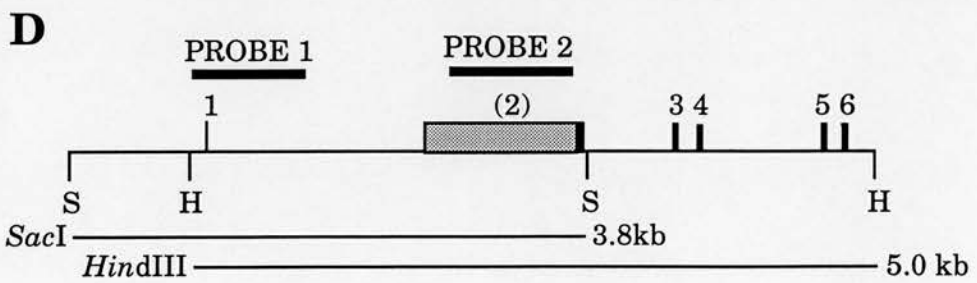
C) Map of wild-type β-casein gene showing fragments obtained from digests diagnostic of targeting using probe 1.

D) Map of the predicted structure of the targeted β-casein gene showing fragments diagnostic of targeting using probes 1 and 2.

Boxes: Exons 1 to 6. Thick lines: Position of genomic DNA probes. H: *Hind*III; S: *Sac*I


A
B
SacI
HindIII
SacI
HindIII
E14
A77
B256
E14
A77
B256
E14
A77
B256
E14
A77
B256
5.0
3.9
3.8
2.7


C
PROBE 1
1
2
3 4
5 6
S
H
S
H
SacI
2.7 kb
HindIII
3.9 kb


D
PROBE 1
PROBE 2
1
(2)
3 4
5 6
S
H
S
H
SacI
3.8kb
HindIII
5.0 kb

probe which contains (most of) exon 7 (probe 3; see Fig. 3.1). These sequences were included in the targeting vector. The results of Southern analysis have been shown in Fig. 3.5A. The probe was stripped off and the filters were reprobed with the neo probe (Fig. 3.5B).

*BstE*II digestion of E14 DNA generated a 5.3 kb fragment with probe 3 (Fig. 3.5C). The targeting vector has two *Bst*EII sites; one in the short arm of homology and the second in the tk gene (Fig. 3.1B). Random integration of the targeting vector into the ES cell genome would result in a 7.1 kb *BstE*II fragment capable of hybridising to both probes 2 and 3. However, correct targeting of the β-casein gene at the 3' end would generate a 6.4 kb *Bst*EII fragment with both of these probes (Fig. 3.5D).

Probing of *Bst*EII-digested DNA from clone A77 and B256 with probe 3 revealed the expected 6.4 kb fragment in addition to the 5.3 kb fragment from the non-targeted allele (Fig. 3.5A). The 6.4 kb *Bst*EII fragment was also recognised by the neo probe as was predicted if the 3' end of the targeting vector recombined with the β-casein gene (Fig. 3.5B). The presence of the 6.4 kb *Bst*EII fragment with these two probes suggested that one of the β-casein allele had undergone a correct targeted modification at its 3' end.

The above interpretation of results was supported by analysis of *Sca*I-digested DNA from the targeted clones (Fig. 3.5). Correct targeting at the 3' end was expected to increase the size of wild-type 7.0 kb *Sca*I fragment to 8.1 kb (Fig. 3.5C & D). Random integration was predicted to generate a *Sca*I fragment of unknown size (>8.2 kb). *Sca*I digestion of A77 and B256 DNA gave 7.0 kb and 8.1 kb fragments hybridising to probe 3 in contrast to a single 7.0 kb fragment from the E14 control (Fig. 3.5A). The novel 8.1 kb *Sca*I fragment from

Figure 3.5. Targeting of the β-casein gene in low passage E14 ES cells: Structure of the 3' end of the targeted allele in clones A77 and B256.

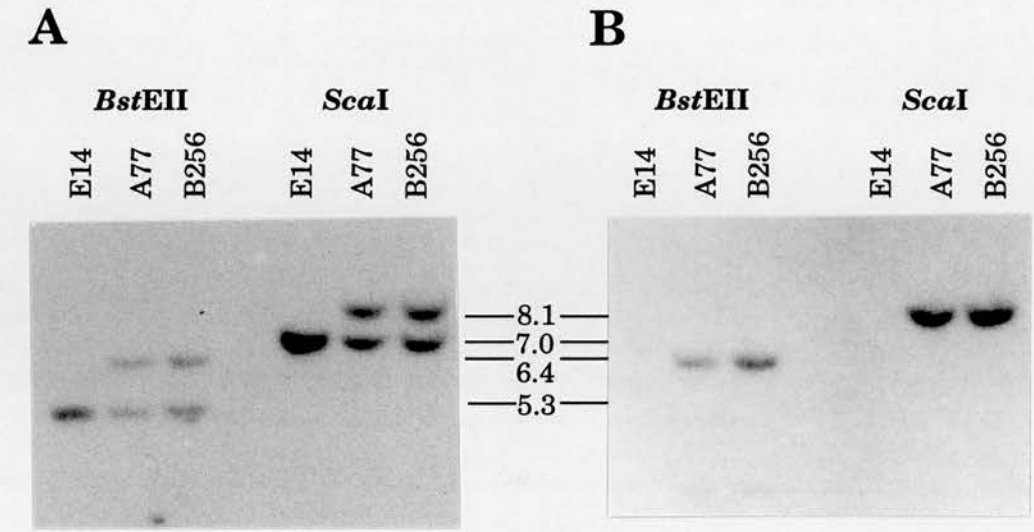
A) *Bst*EII and *Sca*I-digested genomic DNA was probed with probe 3 (Fig. 3.1). The 6.4 kb *Bst*EII and 8.1 kb *Sca*I fragments from A77 and B256 DNA are diagnostic of correct targeting at the 3' end of β-casein gene.

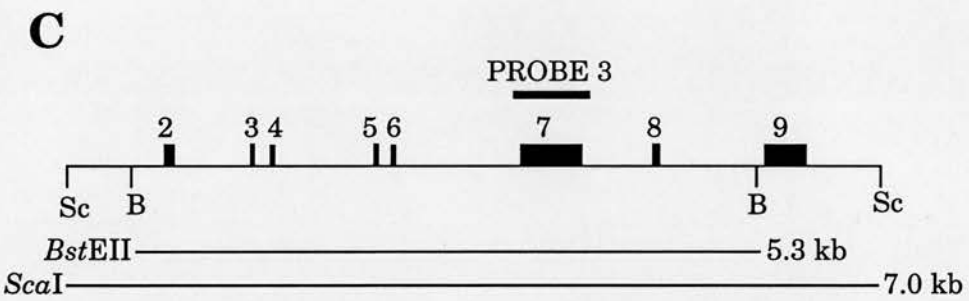
B) *Bst*EII and *Sca*I-digested genomic DNA was probed with probe 2 (Fig. 3.1). The 6.4 kb *Bst*EII and 8.1 kb *Sca*I fragments from A77 and B256 DNA are diagnostic of correct targeting at the 3' end of β-casein gene.

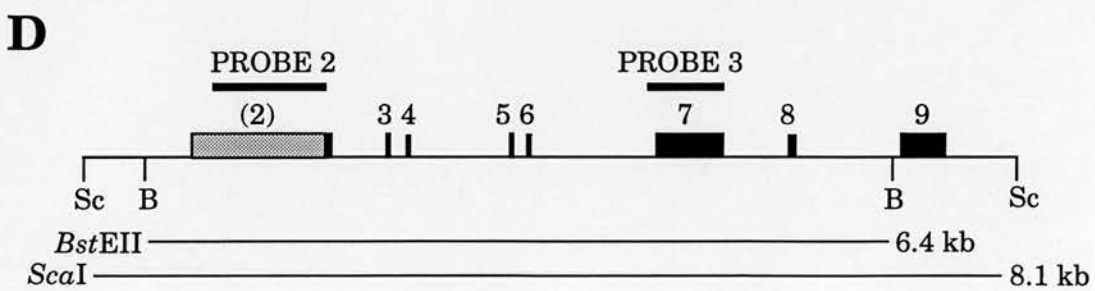
C) Map of the wild-type β-casein gene showing fragments obtained from digests diagnostic of targeting using probe 3.

D) Map of the predicted structure of the targeted β-casein gene showing fragments diagnostic of targeting using probes 2 and 3.

Boxes: Exons 2 to 9. Thick lines: Position of genomic DNA probes. B: *Bst*EII; Sc: *Sca*I.


A
B
BstEII
ScaI
E14
A77
B256
8.1
7.0
6.4
5.3


C
PROBE 3
2
3 4
5 6
7
8
9
Sc
B
B
Sc
BstEII
5.3 kb
ScaI
7.0 kb


D
PROBE 2
PROBE 3
(2)
3 4
5 6
7
8
9
Sc
B
B
Sc
BstEII
6.4 kb
ScaI
8.1 kb

A77 and B256 DNA also hybridised to probe 2 (Fig. 3.5B). These observations were consistent with the precise modification of one of the β-casein alleles of these clones through replacement targeting by the vector pβ1MC1NEO/TK.

Southern blot analysis of A77 and B256 DNA after digestion with a number of restriction enzymes and hybridisation to neo probe as described above, invariably lit up a single band of the size consitent with a homologous recombination event between the endogenous β-casein gene and the targeting vector (Figs. 3.4 & 3.5). Thus, there was no evidence for random integration of the targeting vector molecules in the genome of these clones.

3.2.2. Production of Chimaeras and Germ line Transmission of the Mutation

To produce chimaeric mice from the β-casein gene targeted clones, mouse blastocyst injections were performed by Dr. Martin Hooper's laboratory at the Medical School, University of Edinburgh and the manipulated embryos were transferred to the uteri of psuedopregnant mice by Miss Roberta Wallace at Roslin.

β-casein gene targeted clones A77 and B256 were injected into F_2 C57/BL/6/Ola x CBA/Ca/Ola blastocysts (Thompson et al. 1989). The blastocysts were transferred to psuedopregnant MF1 female mice. The ES cell contribution to mice that developed from injected blastocysts was visually scored by observing for the patches of lighter coat colour in the pups at three weeks of age. The host blastocysts segregated at the *A (agouti)* locus resulting in agouti and non-agouti mice and any contribution from ES cells was easily evident as light patches on a dark coat (Fig. 3.6).

Figure 3.6. Production of chimaeric mice from the β-casein gene targeted ES cell clone B256. A litter of pups derived from F_2C57/BL/6/Ola x CBA/Ca/Ola blastocysts after injection of B256 ES cells. Seven of the nine pups in this litter were chimaeric as demonstrated by the patches of lighter coat colour in the pups born from the manipulated blastocysts.

The efficiency of chimaera production is presented in Table 3.3. Clone A77 was relatively inefficient in generating chimaeras; only 8 out of 30 mice scored were chimaeric. The extent of the ES cell contribution was generally very poor in these chimaeras. Further, there was no distortion of sex-ratio in favour of males as was anticipated after injection of male ES cells into the unsexed blastocysts. However, clone B256 produced extensive chimaeras contributing up to 90% to the coat colour of the pups. More than 50% of the progeny were chimaeric and there was a clear bias in the sex-ratio among these chimaera in favour of males, indicating that some of the males might have resulted from sex conversion of female host blastocysts (Table 3.3).

Table 3.3. Production of chimaeric mice from the β-casein gene targeted E14 ES cell clones.

Clone injected	No. of blastocysts injected	No. of pups at weaning	No. of chimaeras		Extent of coat-colour chimaerism (%)
			Male	Female	
A77	40	30	2	6	1-65
B256	90	57	23	10	1-90

Table 3.4 shows the breeding performance data of the chimaeras mated to MF1 albino mice. The germ line contribution from ES cells was identified by the presence of light chinchilla offspring in the litters of chimaeric males and females. None of the A77 chimaeras produced any light chinchilla mice. Out of a total of 16 fertile male and 6 fertile female B256 chimaeras, nine males and one female produced

Table 3.4. Breeding performance of chimaeras produced from mouse blastocysts injected with the β-casein gene targeted E14 ES cell clones.

Targeted clone	No. of chimaeras tested		No. of fertile chimaeras		No. of germ line chimaeras	
	Male	Female	Male	Female	Male	Female
A77	2	6	2	6	0	0
B256	21	6	16	6	9*	1

* One male chimaera produced progeny only from ES cells.

progeny from the ES cell component (e.g. Fig. 3.7). Five of the B256 male chimaeras tested were sterile; four of these had approximately a 90% contribution from ES cells. Among the germ line chimeras, the extent of the ES cell contribution in their coat colour ranged from 5-90%. Chimera 56 produced only light chinchilla progeny which suggested that this male had a XX-XY chromosomal constitution. The extent of coat colour chimaerism was a poor indicator of germ line transmission potential from a particular chimaera (Table 3.5).

Tail DNA of the light chinchilla offspring from the germ line chimeras was analysed by Southern blotting to confirm that some of these animals had inherited the disrupted β-casein allele. Fig. 3.9 shows results from such an analysis. The *Sac*I-digested DNA samples were analysed with probe 1. Approximately 50% of these mice had the 3.8 kb *Sac*I fragment corresponding to the targeted β-casein allele in addition to the 2.7 kb fragment from wild-type allele. The mutant allele was designated *csnb*$^{-}$.

Figure 3.7. Germ line transmission of the targeted β-casein allele. A litter from one of the germ line male chimaeras B256/30 (shown in the inset) mated to an MF1 female. Agouti and non-agouti pups are from the host blastocysts (F_2 C57/BL/6/Ola x CBA/Ca/Ola). Light chinchilla pups are from the β-casein gene targeted ES cells; 50% of such pups carried the targeted β-casein allele.

Table 3.5. Extent of coat-colour chimaerism in male germline chimaeras and proportion of ES cells derived pups in their progeny.

Male germline chimaera	Coat colour chimaerism (%)*	No. of progeny scored	ES cell-derived progeny (%)
9	75	135	8
18	40	177	27
30	10	100	37
42	5	94	11
43	90	50	6
45	90	30	33
55	50	5	60
56	55	12	100
57	20	112	7

* The ES cell contribution to the coat colour of chimaeric mice was judged by visual appraisal.

3.2.3. Mice Homozygous for the Mutation were Viable and Fertile, and the Females Appeared to Lactate Normally

Mice heterozygous for the targeted mutation in the β-casein gene were mated to produce homozygous mice (*csnb*$^{-}$/*csnb*$^{-}$). Matings amongst full sibs and half sibs were avoided since the mutation was in an outbred genetic background and any inbreeding at this stage could have introduced a bias in the subsequent analysis of phenotypic effect of the mutation.

Tail DNA of pups born from heterozygous intercrosses (Fig. 3.8) was analysed by Southern blotting (Fig. 3.10). Probe 1 revealed a

Figure 3.8. Production of mice homozygous for the mutant β-casein allele. Heterozygous mutant mice were bred to produce homozygous mutant mice. Top: a heterozygous mutant mouse derived from the β-casein gene targeted ES cell clone B256. Bottom row: A litter from *inter-se* mating of heterozygous mutant mice; two of these pups are homozygous mutants.

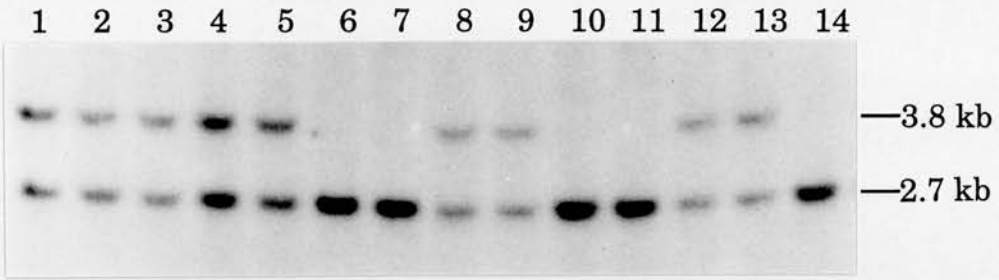


Figure 3.9. Germ line transmission of the targeted β-casein allele. Southern blot analysis of *Sac*I-digested DNA from the tail biopsies of the light chinchilla progeny of one of the male chimaeras mated to MF1 females. Lane 1: Control DNA from the ES cell clone B256 carrying the targeted β-casein allele; lanes 2-14: tail DNA from the light chinchilla offspring of the male chimaera. The filters were hybridised with probe 1 (see Fig. 3.1). The 2.7 kb fragment is released from wild-type allele in all lanes, and the presence of an additional 3.8 kb fragment in lanes 2-5, 8, 9, 12, and 13 is diagnostic of transmission of the targeted β-casein allele originating from ES cell contribution to the germline of the chimaera.

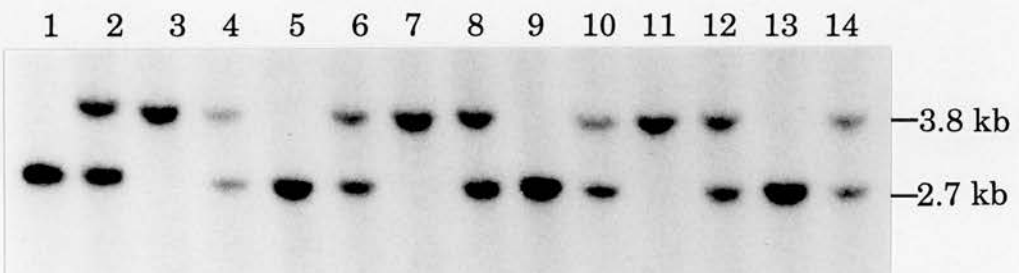


Figure 3.10. Generation of mice homozygous for the targeted β-casein allele. Southern blot analysis of *Sac*I-digested DNA from the tail biopsies of progeny from *inter-se* matings of mice heterozygous for the targeted β-casein allele. Lane 1: Control DNA from the E14 cells; lane 2: control DNA from the targeted clone B256; lanes 3-14: tail DNA from the offspring from heterozygous matings. The filters were hybridised with probe 1 (see Fig. 3.1). The 2.7 kb *Sac*I band is from wild-type allele and the 3.8 kb band is from the targeted allele. Approximately 25% of the progeny screened were homozygous for the mutant allele as in lanes 3, 7 and 11.

single 3.8 kb *Sac*I fragment corresponding to the disrupted β-casein allele in appoximately 25% of the animals in contrast to a 2.7 kb band from wild-type allele. Heterozygous animals (*csnb*$^+$/*csnb*$^-$) were identified by the presence of both the 2.7 kb and 3.8 kb fragments. The relative numbers of wild-type homozygotes (*csnb*$^+$/*csnb*$^+$), heterozygotes (*csnb*$^+$/*csnb*$^-$) and mutant homozygotes (*csnb*$^-$/*csnb*$^-$) were in agreement with the expected Mendelian ratio of 1:2:1, respectively (Table 3.6). Thus, the mutation in the β-casein gene was compatible with normal development. The average weight of the β-casein-deficient pups at weaning was not different from that of wild type pups (Table 3.6) indicating that the mutation in the β-casein gene did not compromise the preweaning growth of the pups. However, the pups heterozygous at the β-casein locus weighed more than the mutant homozygous pups ($p < 0.05$). It was difficult to reconcile how the superiority of the heterozygotes observed in the present study could result from the interaction of the targeted and wild-type β-casein alleles.

The wild-type, heterozygous, and mutant homozygous females were mated to mutant homozygous, heterozygous and wild-type males, respectively. The three groups of females produced litters of similar sizes (Table 3.7). The preweaning growth of the pups feeding β-casein-deficient milk appeared to be slightly slower than that of pups feeding milk from either wild-type or heterozygous mothers (Fig. 3.11A & B), probably indicating the reduced total milk protein in β-casein-deficient milk (see section 3.2.6 below).

Table 3.6. Frequencies of the genotypes of progeny and their average weaning weights from heterozygous (*csnb*$^+$/*csnb*$^-$ x *csnb*$^+$/*csnb*$^-$) intercrosses.

Genotype of Pups*	*csnb*$^+$/*csnb*$^+$	*csnb*$^+$/*csnb*$^-$	*csnb*$^-$/*csnb*$^-$	Total
Number of progeny	57	152	70	279
Weaning weight (g ± SE)	12.1 ± 0.39 (39)**	12.6 ± 0.21 (110)**	11.6 ± 0.32 (48)**	–

* The pups were genotyped at five weeks of age. The observed ratios of the three genotypes were not statistically different from the expected ratios of 1:2:1 (χ^2 - test).

** The average weaning weight of *csnb*$^-$/*csnb*$^-$ pups was statistically different from that of *csnb*$^+$/*csnb*$^-$ pups ($p < 0.01$; *t*-test) but not from that of *csnb*$^+$/*csnb*$^+$ pups ($p > 0.05$; *t*-test). The weaning weights of *csnb*$^+$/*csnb*$^+$ and *csnb*$^+$/*csnb*$^-$ were not significantly different from each other ($p > 0.05$; *t*-test).

Table 3.7. The effect of the *csnb*$^-$ allele on the average litter-size at birth.

Genotype of Dam	*csnb*$^+$/*csnb*$^+$	*csnb*$^+$/*csnb*$^-$	*csnb*$^-$/*csnb*$^-$
Litter-size at birth	8.3 ± 0.60 (26)	9.8 ± 0.76 (26)	8.9 ± 0.63 (23)

* The average litter-size did not statistically differ among the various β-casein genotypes ($p > 0.05$; *t*-test).

Figure 3.11A. A lactating homozygous mutant female mouse with its litter. There was no obvious effect on the fertility. However, the pup growth rate (up to day 11) was reduced.

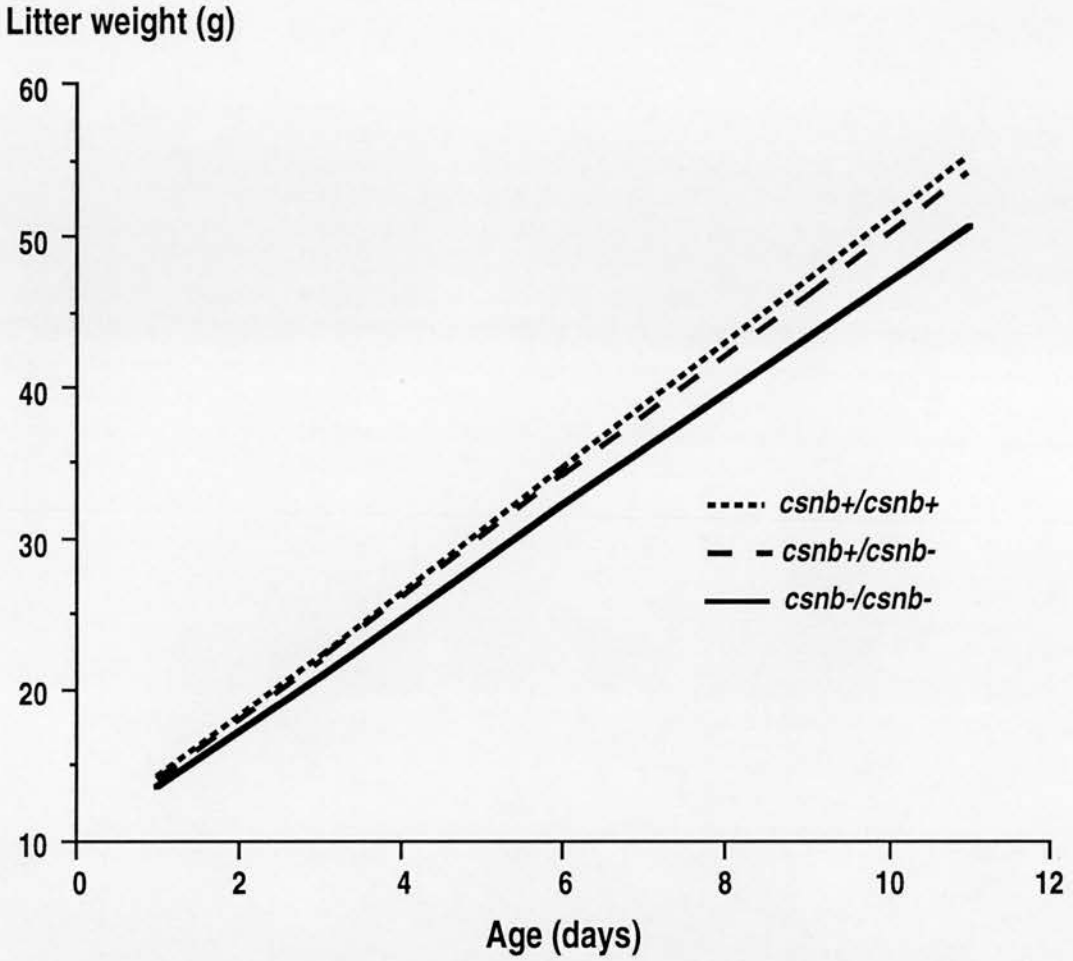


Figure 3.11B. The effect of the mother's genotype at the β-casein locus on the pre-weaning growth of the pups.

For each genotype a linear regression model was fitted to the litter weight data considering the age at weighing and the litter size as independent variables ($R^2 \geq 95\%$; see Chapter 2: section 2.15). The graph has been drawn from the predicted values of litter weights assuming a litter size of eight. The homozygous mutant ($csnb^-/csnb^-$) mothers produced significantly lower litter weight on day 11 of lactation than that of heterozygous mutant ($csnb^+/csnb^-$) and wild-type ($csnb^+/csnb^+$) mothers ($p < 0.001$, t-test). Note that assuming litter sizes of seven, nine and ten also results in statistical significance. The heterozygous mutants were not different from the wild-type except for litter size of 10 ($p < 0.05$).

3.2.4. Transcription of the Mutant β-Casein Gene

The targeted allele *csnb*$^-$ was created by insertion of 1.1 kb neo gene sequences into exon 2 of the β-casein gene in the same transcriptional orientation (see section 3.2.1.1). As the neo gene lacked its own polyadenylation signal, a fusion transcript of 2.3 kb was expected from the targeted allele which should hybridise to the β-casein cDNA probe. Fig. 3.12 shows a Northern analysis of total RNA from the mammary gland of lactating (day 11) females. A cDNA probe (Gupta et al. 1982) homologous to exons 7 and 9 of the mouse β-casein gene failed to detect any corresponding transcript in the mammary gland RNA of homozygous mutant lactating mice (Fig. 3.12A). Consistent with this observation, heterozygous females had approximately half the amount of the β-casein mRNA present in the mammary gland of homozygous wild-type mice. Northern blot analysis of mammary gland RNA from homozygous mutant females with the neo probe gave diffuse hybridisation reminiscent of partially degraded RNA (Fig. 3.12B). Control hybridisation with a rRNA probe gave discrete bands as expected showing that the RNA preparations were not degraded (not shown). Given that the neo hybridisation was specific to mice carrying the mutant allele, these data indicate that the targeted locus was being transcribed in the mammary gland at some level. However, it is not possible to know without further characterization whether the transcription was from the β-casein gene promoter or from the neo promoter included in the modified exon 2 of the targeted β-casein gene. It is also possible that the transcripts detected originate from the other strand. Nevertheless, failure to detect any signal from the mutant allele (*csnb*$^-$) with the β-casein cDNA probe even after exposure of the film for several days (data not

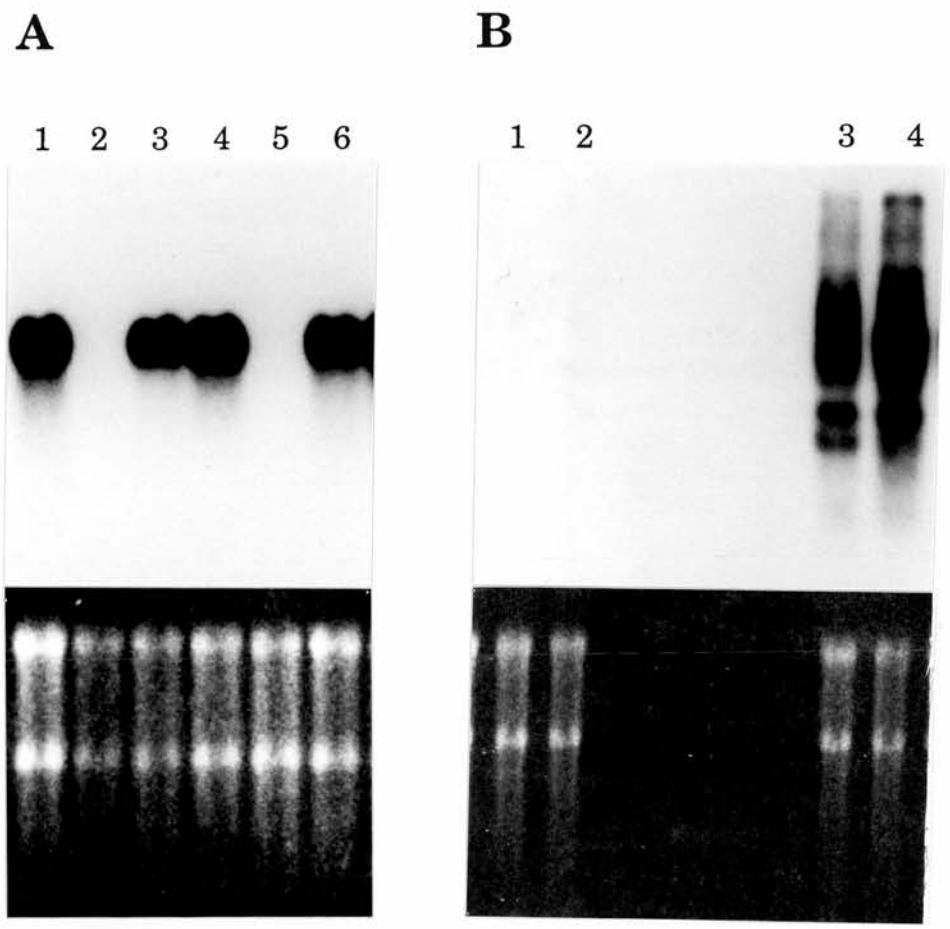


Figure 3.12. Expression of the targeted β-casein gene in the mammary gland.

A) β-casein probed Northern blot. Lanes 1 & 4: wild-type; lanes 2 & 5: mutant homozygote; lanes 3 & 6: mutant heterozygote.

B) Neo probed Northern blot. Lanes 1 & 2: wild-type; lanes 3 & 4: mutant homozygote.

The controls for the amount of RNA loaded are shown at the bottom of panels A & B (visualised by ethidium bromide staining). *N.B.* The intensities of signal in panels A and B are not comparable.

shown) as opposed to the one hour exposure needed to detect β-casein mRNA from wild-type allele indicated that *csnb*$^-$ is effectively a null allele.

3.2.5. Homozygous Mutant Mice Lack β-Casein in their Milk

Fig. 3.13 shows representative milk protein profiles from wild-type homozygote, mutant heterozygote, and mutant homozygote females as revealed by SDS-PAGE and staining with Coomassie Blue. β-casein was not detected in the milk samples from the homozygous mutant mice (lanes 1 and 4) as expected. Milk from heterozygous females had lesser β-casein than that from wild-type females. These observations are consistent with the absence of any β-casein message from the targeted allele *csnb*$^-$ (see section 3.2.4). Thus, the β-casein allele *csnb*$^-$ represents a true null mutation, and the reduction in the amount of β-casein in the heterozygous mice shows that there is a gene dosage-effect at this locus.

3.2.6. Reduced Total Protein Concentration and Partial Compensation for the Loss of β-Casein in the Milk of Homozygous Mutant Females

The micro-Kjeldahl method in conjuction with colorimetry (see Chapter 2: section 2.13) was used to estimate the nitrogen (N) concentration in whole milk and whey samples (day 11 of lactation) from homozygous mutant and homozygous wild-type mice. Protein concentrations were then calculated from the estimated N value. Although the micro-Kjeldahl method overestimates the protein concentration due to inclusion of non-protein nitrogen (NPN), this should

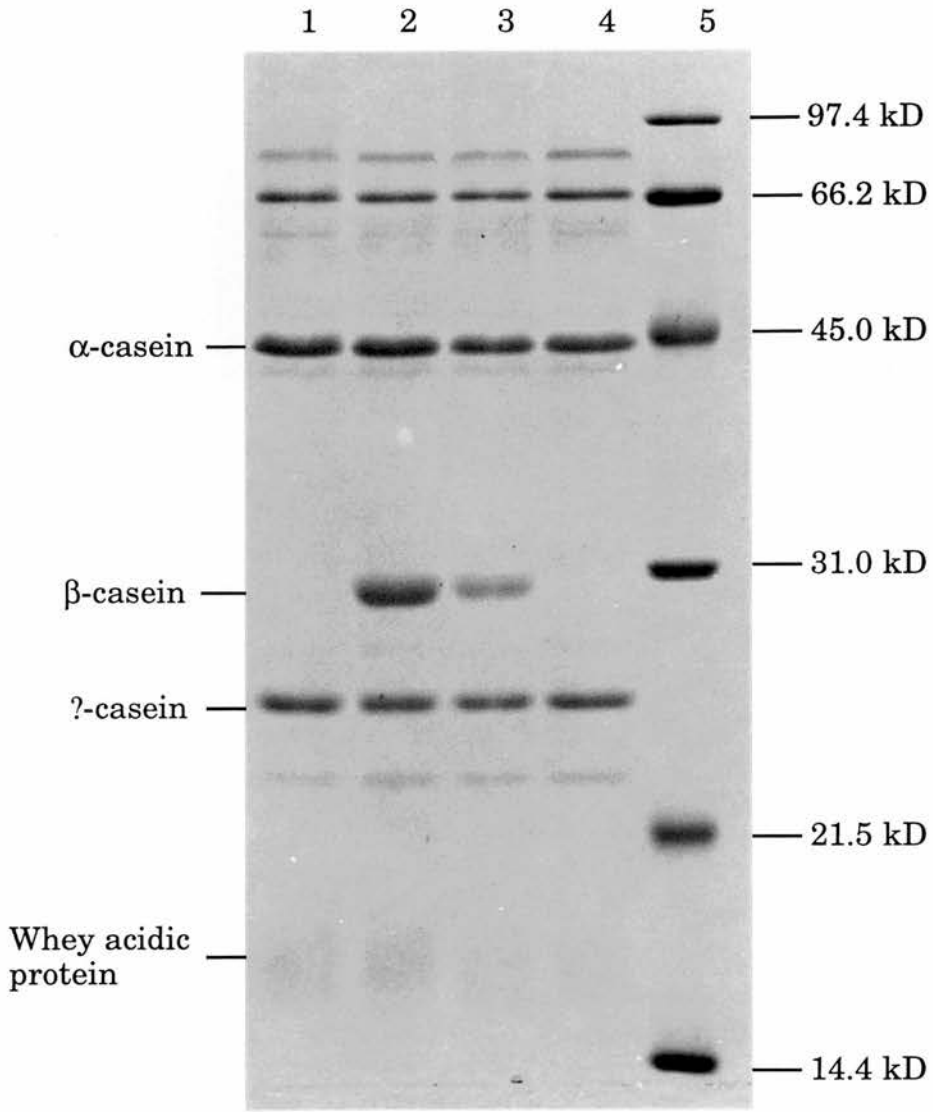


Figure 3.13. Protein composition of milk from β-casein mutant mice. Lanes 1 & 4: milk from homozygous mutant mice; lane 2: milk from wild-type control mouse; lane 3: milk from heterozygous mutant mouse. Lane 5: molecular weight standards.
Diluted and defatted milk (equivalent to 75 nl of the whole milk per lane) was electrophoresed into a 15% polyacrylamide gel under denaturing conditions and the proteins were visualised after staining with Coomassie Blue.

not affect the comparison of milk protein concentration in various genotypes in the present investigations. [In bovine milk NPN constitutes 3-8% of total N; Grappin and Ribadeau-Dumas 1992]. For each genotype, the milk samples from 12 mice were pooled, and four independent estimates of protein concentration were obtained for each of the two pools. In addition, total protein concentrations were also measured in individual mouse milk samples. The average protein concentration in mutant mouse milk was approximately 10% lower than that in wild-type mouse milk (Table 3.8). β-casein represented over 21% of milk proteins in wild-type mice used in the present investigations (E. Stevenson, J. Leaver and A. J. R. Law, unpublished). Thus, reduction in the level of milk protein in the β-casein-deficient mice was much smaller than what was anticipated after the loss of β-casein. These data indicate that the loss of β-casein in the mutant mice was being compensated by up-regulation of synthesis and/ or secretion of milk proteins other than β-casein.

To test whether this compensatory bio-synthesis and/ or secretion of proteins in the mammary gland was spread across caseins as well as whey proteins, the whey was isolated from the pooled β-casein-deficient milk and the pooled control milk after acid precipitation of the caseins. The complete removal of caseins from the whey fractioned was confirmed by SDS-PAGE (data not shown). Whey protein concentrations were measured and the values were expressed $mgml^{-1}$ of whole milk. The casein concentrations were calculated by subtraction of the value for the whey protein from that of the total milk protein (Table 3.8). The mutant mouse milk had approximately 19% more whey protein than that from the control milk (it may be stressed that only one experiment was performed to estimate whey protein concentration and to draw firm conclusions, there is need to replicate this experiment). The casein level

in the milk from the control mice was 78.6 $mgml^{-1}$. As β-casein constituted 27% of this value (see above), the casein concentration in the mutant mice was expected to be approximately 57.6 $mgml^{-1}$. Instead, the concentration of caseins in the milk from the mutant mice was 65.3 $mgml^{-1}$ (Table 3.8). Taken together, the data on total milk protein and

Table 3.8. The effect of *csnb*$^-$ allele on the protein concentration of milk.

Protein concentration ($mgml^{-1} \pm SE$)	Genotype	
	csnb$^+$/*csnb*$^+$	*csnb*$^-$/*csnb*$^-$
1. Total milk protein (pooled sample from 12 mice in each group; estimated 4 times)	97.0 ± 0.97	89.1 ± 0.12
2. Total milk protein (milk samples from 6 individual mice in each group)	97.3 ± 2.96	83.0 ± 1.55
3. Total milk protein (weighted average of 1 and 2)	97.1 ± 1.71	87.3 ± 0.89
4. Whey protein (whey prepared from pooled milk from 18 mice; estimated 3 times)	18.5 ± 0.25	22.0 ± 0.21
5. Total caseins (Total milk protein concentration minus whey protein concentration)	78.6	65.3

The total milk protein and whey protein concentrations were affected by the β-casein genotypes ($p < 0.001$; *t*-test).

whey protein levels in mutant mice were consistent with the hypothesis that the loss of β-casein was being compensated by enhanced synthesis and/ or secretion of other proteins in the mammary gland.

3.2.7. β-Casein is not Essential for Micelle Assembly and Secretion but the Average Micelle Size is Reduced in the β-Casein-Deficient Milk

Casein micelle sizes were determined by dynamic light scattering spectroscopy and the results are presented in Table 3.9 (measurements performed by J. Leaver and D. Horne of Hannah Institute, Ayr). The reduction of the β-casein level in the milk of the heterozygous animals or the absence of β-casein in the milk of the mutant mice resulted in significantly smaller casein micelles. Also, casein micelles of the mutant mice were significantly smaller than those of the heterozygous animals. Caseins other than β-casein were being organised as micelles, and thus, β-casein is not essential in either formation or secretion of these micelles.

Table 3.9. The effect of *csnb*$^-$ allele on the average casein micelle size.

Genotype	*csnb*$^+$/*csnb*$^+$	*csnb*$^+$/*csnb*$^-$	*csnb*$^-$/*csnb*$^-$
Average micelle size (nm) ± SE	280.5 ± 3.7 (12)	268.6 ± 3.6 (11)	255.8 ± 2.3 (14)

* The casein micelle size of each genotype was statistically different from those of the remaining two genotypes ($p < 0.05$; *t*-test).

3.3 DISCUSSION

3.3.1 Targeting β-Casein Gene in ES Cells and Germ line Transmission of the Mutation

The mouse β-casein gene was disrupted in E14 ES cells through homologous recombination using a replacement-type gene targeting vector pβ1MC1NEO/TK. This vector was based on the positive-negative selection protocol and included approximately 4.7 kb of non-isogenic DNA of the β-casein gene. GANC selection gave 4-15 fold enrichment in various experiments. This level of enrichment is far below the 2000-fold enrichment originally reported by Mansour et al. (1988) but is in agreement with the results of a number of other workers (DeChiara et al. 1990; McMahon and Bradley 1990; Tybulewicz et al. 1991; and many others). Twenty one out of 492 GANC-G418-resistant ES cell clones (one targeting event per ~24 colonies screened) had one of their β-casein alleles targeted (Table 3.1). It is well established that the gene targeting efficiency increases exponentially with the linear increase in the length of homology included in the targeting vector (Deng and Capecchi 1992), and isogenic DNA targets more efficiently than non-isogenic DNA (Deng and Capecchi 1992; Te Riele et al. 1992; and Van Deursen and Wieringa 1992). Despite the comparatively small and non-isogenic β-casein DNA fragment in the vector pβ1MC1NEO/TK the targeting efficiency of this vector in the present study was relatively high. The absence of a polyadenylation signal from the neo gene in a targeting vector has been shown to indirectly enrich for the targeting events by suppressing the frequency of random integration (Joyner et al. 1989; Zijlstra et al. 1989). The neo gene in pβ1MC1NEO/TK lacked its own polyadenylation site but

the presence of a polyadenylation signal in the HSV tk gene placed at the 3' end of this targeting vector could have enhanced the efficiency of the neo gene expression.

To take advantage of increased homology, a β-casein gene targeting experiment was performed in EFC-1 ES cells using another targeting vector pβ2MC1NEO/TK. This vector had 1.35 kb more β-casein gene sequences on its long arm of homology in contrast to the vector pβ1MC1NEO/TK. Unexpectedly, no targeting events were detected in 168 G418-GANC colonies obtained after electroporation of EFC-1 ES cells with the vector pβ2MC1NEO/TK (Table 3.1). The additional 1.35 kb β-casein gene sequences in this vector were upstream of the HSV tk gene and it was possible that these sequences could have adversely affected the tk gene expression and thus lowered the level of enrichment. The similar level of enrichment achieved with both the above targeting vectors would discount this possibility (Table 3.1). The comparison of these two vectors is limited by the fact that the experiments were done at different times using different cell lines; although there is no evidence to suggest that EFC-1 cells are refractory to targeting (my unpublished results). A relatively high level of G418 selection regime (500 $mgml^{-1}$) was used in the experiment with pβ2MC1NEO/TK. It may be noted that in an experiment with E14 cells, the β-casein gene targeted clones were isolated using the same concentration of G418 (see section 3.2.1.2). On the other hand, if the sensitivity of E14 and EFC-1 cells to G418 differed it is possible that the β-casein gene targeted EFC-1 clones could not survive the above level of G418 selection.

The failure to detect any targeting events with pβ2MC1NEO/TK may also be reconciled assuming that 1.35 kb additional β-casein gene sequences in this vector had a number of polymorphic sites

with respect to the β-casein locus in ES cells. Thus, regional polymorphism may be responsible for very poor targeting efficiency of this vector (Deng and Capecchi 1992). There are likely to be still uncharacterised factors that influence the gene targeting efficiency of a vector since the repeatability of targeting efficiency of a given vector in experiments performed under similar conditions but on different days is very poor (Hasty et al. 1992; Koller et al. 1991; Piedrahita et al. 1992; Chapter 4: Table 4.1A). If in further controlled experiments the comparative gene targeting efficiencies of β-casein gene targeting vectors pβ1MC1NEO/TK and pβ2MC1NEO/TK remain similar to that observed in the present experiments it is tempting to suggest that such data may shed some light on the mechanism(s) of gene targeting with replacement vectors.

E14 cells used in experiment-I were obtained originally from Dr Martin Hooper's laboratory at an early passage and were frozen after these were grown for about 12-15 passages at the IAPGR Roslin. At the time of electroporation the cells were at passage 35. It is very likely that these cells may have drifted to aneuploidy while these were grown at the IAPGR because both targeted clones had 42 chromosomes.

A number of β-casein gene targeted clones identified in the subsequent experiments using low passage E14 ES cells were also found to be aneuploid (Table 3.2) The consistency of the chromosomal abnormality in clone A32 would suggest that the abnormality might have existed in the cells prior to electroporation. The E14 ES cell line has been characterised as a euploid cell line with a good efficiency of chimaera formation and of germ line transmission from selected clones (Clarke et al. 1992; Handyside et al. 1989; Thompson et al. 1989). However, it appears that the low passage cells used in the above experiments had

accumulated chromosomal abnormalities at some stage. Two independent β-casein gene targeted clones were microinjected into blastocysts to produce chimaeric mice. One of these clones A77 had 60% of metaphases with a modal number of 40 chromosomes. Upon microinjection into blastocysts, this clone showed limited capacity for differentiation and failed to contribute to the germ line. The modal chromosomal count may be an important prerequisite for germ line transmission but it may not be sufficient. The other clone B256 had 80% of cells with 40 chromosomes. It proved to be very efficient in chimaera formation and several chimaeras produced progeny from the ES cell component (Table 3.3 & 3.4). Germ line transmission of the ES component was observed from one female chimaera in the present experiments in common with the observations of other workers (A. Clarke personal communication; Selfridge et al. 1992; Stewart et al. 1992).

A number of ES cell clones modified at different loci through gene targeting have contributed to the mouse germ line (reviewed by Hooper 1992). However, the factors rendering some of the ES clones incompatible with germ line transmission are not well understood. It has been suggested that the use of the feeder layers is critical for the growth of ES cell, and LIF alone may not be sufficient (Bradley et al. 1992). However, no controlled experiments have been reported in the literature comparing the various culture conditions. Chromosomal instability is noticed even when ES cells are grown on feeder layers (my unpublished observations). A better understanding of the factors responsible for diversion of karyotype of ES cells in culture and for the loss of their ability to populate the mouse germ line would obviously improve the efficiency of introducing a planned change in the mouse genome.

3.3.2. Null Mutation in the Mouse β-Casein Gene

The absence of β-casein from mutant mouse milk as revealed by SDS-PAGE (Fig. 3.13) and the failure to detect any corresponding mRNA from the mammary gland of lactating mutant mice (Fig. 3.12) are two strong lines of evidence in support of the conclusion that the *csnb*$^-$ represents a null mutation.

The lack of any signal with the β-casein cDNA probe in Northern blot analysis was (naively) unexpected, and instead a fusion message was anticipated since the neo gene lacked its own polyadenylation signal. It appears that either transcription or stability of RNA or both are reduced. Formally, it is possible that transcription may have been terminated prematurely within the β-casein transcription unit or the fusion transcript may include cryptic polyadenylation sites such that the processed transcript would contain only the neo message. Skipping of β-casein exons during processing of the hybrid transcript could be another explanation. In the targeted β-casein allele there is an in frame translation stop codon in exon 2 upstream of the neo gene coding sequences. Premature nonsense mutations have been documented to cause reduction in the amount of nuclear and cytoplasmic mRNA without affecting transcription (Cheng and Maquat 1993). Nonsense mutations have also been associated with skipping of exons containing these mutations (Dietz et al. 1993).

3.3.3. Structure of Casein Micelles

In cow's milk, micelles comprise α_{s1}-, α_{s2}-, β- and κ-caseins in addition to calcium phosphate and small amounts of magnesium,

sodium, potassium and citrate. Most studies concerning the structure of the casein micelles have been performed on cow's milk. However, a similar electron microscopic structure of casein micelles has been documented in a number of species (see review by Rollema 1992). As in cow's milk (Dalgleish et al. 1989; Davies and Law 1983; Donnelly et al. 1984; McGann et al. 1980) the inverse relationship observed between micelle size and κ-casein concentration in human (Azuma et al. 1985) and goat (Ono and Creamer 1986) milk would tend to support the argument that similar broad principles of micelle organisation exist in various milks. However, it should be kept in mind that the caseins are one of the most poorly conserved families of proteins (Bonsing and Mackinlay 1987). There is also wide variation in the relative concentrations of different caseins across the species (see Chapter 1: Table 1.2). No information exists in the literature on the structure of mouse casein micelles, and therefore, for the purpose of the following discussion it has been assumed that the casein micelle in this species has an organisation similar to that in the cow.

The organisation of at least some kind of micelles in β-casein-deficient mouse milk immediately demonstrates that this protein is not essential in the organisation or secretion of these micelles into mouse milk. The currently accepted models of bovine micelle structure (Slattery and Evard 1973; Schmidt 1982) neither predict nor exclude any specific or essential role of α- or β-caseins in the organisation of sub-micelles or in the association of sub-micelles into micelles. However, there are a number of explicit and implicit suggestions in the literature indicating some special role of β-casein in the organisation of micelles. 1) In spite of the lack of conservation of the primary amino acids sequences of caseins the overall architecture of β-casein has been conserved, and

unlike α-caseins, no rearrangements have occurred (Bonsing and Mackinlay 1987; Holt and Sawyer 1988). 2) β-casein is present in all milks examined; on the other hand goat and human milk have a very small proportion of α-caseins (see Creamer 1991) 3). Addition of dephosphorylated β-casein to artificial milk micelles increased the rennet coagulation time and decreased the extent of syneresis (Pearse et al. 1986). The latter authors have argued that these results reflected the importance of β-casein in curd formation in the stomach. Bonsing and Mackinlay (1987) have further suggested that β-casein may have a role in determining the surface properties of the casein micelles. However, in the present investigations no overt health problems were noticed in the pups feeding on β-casein-deficient milk. The slightly slower growth rate of these pups (Fig. 3.11B) could suggest such a role of β-casein, but this is more likely due to the reduction in the total protein concentration of β-casein-deficient milk (Table 3.8). Without further characterization of the properties of β-casein-deficient micelles it is not possible to know the role of β-casein either in the organisation or the properties of the native micelles but it is very likely that the composition of sub-micelles reflects the available proteins in a particular milk rather than any specific organisation involving α- and β-caseins as suggested by Ono and Obata (1989). The above argument is consistent with the fact that the altered ratio of α- : β-casein in the heterozygous mutant mouse milk (Fig. 3.13) does not affect micelle formation *per se*.

The total casein concentration in the homozygous mutant mouse milk is approximately 17% lower than that of wild-type milk despite the compensation for the loss of β-casein (Table 3.8). Although the concentrations of individual caseins were not measured in the present experiments, the loss of β-casein concurrent with the decreased casein

concentration in the mutant mouse would mean that the ratio of κ-casein to the remaining caseins has increased in the homozygous mutant mice (Table 3.8). The observation that in the β-casein-deficient mouse milk there is a compensatory increase in the level of other proteins (including whey proteins) would suggest that even the absolute concentration of κ-casein may have increased. The ratio of the κ-casein to the remaining caseins has an inverse relationship with the micelle size (Dalgleish et al. 1989; Davies and Law 1983; Donnelly et al. 1984; McGann et al. 1980) and these results have been the basis for localising most of the κ-casein on the surface of casein micelles in various models of micelle structure. These studies were performed after physical separation of casein micelles differing in sizes and subsequent determination of protein composition of various fractions. The present investigations provide the first *in vivo* evidence that the size of casein micelles is dependent upon the relative concentration of κ-casein in milk of a species (Table 3.9). An alternative interpretation of these results may be that reduction of micelle size in the homozygous mutant mouse milk reflects merely the absence of β-casein; there is suggestion that the outer layer of the micelle may include a small amount of β-casein (Dalgleish et al. 1989). However, the reduction in micelle size in the heterozygous mutant milk as compared with that of wild-type milk would not support this explanation (Table 3.9). Thus, the present results provide sustenance to the casein micelle models of Slattery and Evard (1973), and Schmidt (1982) that most of the κ-casein is present in or on the surface of the micelle. However, these results would argue against any essential role of β-casein in the protein-protein interactions between sub-micelle as envisaged in the model of Slattery and Evard (1973).

3.3.4. Limit to Milk Protein Biosynthetic Capacity of the Mammary Gland

High level expression of a heterologous milk protein gene in the mouse mammary gland does not lead to the increased total protein levels in milk (Wilde et al. 1992). Although the heterologous protein (ovine β-lactoglobulin) constituted 29% of the total milk protein, the total protein concentration of transgenic mouse milk was not different from that of control milk indicating some physiological limit to milk protein level (Wilde et al. 1992; M. McClenaghan personal communication). The corollary of the above results would be that the removal of an endogenous protein may enhance the level of the remaining milk proteins.

In the present studies, the significantly lower milk protein concentration of β-casein-deficient milk (Table 3.8), and the lack of β-casein mRNA in the mammary gland of the mutant mouse (Fig. 3.12) would suggest that the steady-state level of mRNA is limiting in the synthesis of milk proteins in the mammary gland of these mice. It is not known if the loss of β-casein mRNA has been compensated by the enhanced transcription or stability of other milk protein mRNAs. The reduction in β-casein concentration (Fig. 3.13) and in the corresponding mRNA level in the heterozygous mutant mouse (Fig. 3.12) also shows that the β-casein mRNA is limiting in the mammary gland of the heterozygous mutant mouse. The partial compensation of the loss of β-casein by enhanced synthesis and/or secretion of the remaining milk proteins (see section 3.2.6) is consistent with the *in vivo* results of Wilde et al. (1992) that there is an physiological limit to the bio-synthetic capacity of the mammary gland of the wild-type mouse. Disruption of β-casein gene probably releases that limit but now mRNA has become

limiting. Further characterization of the β-casein-deficient mouse in terms of steady-state mRNA levels of other milk proteins, and transcription rates of all milk protein genes including that of the mutated β-casein would be required to address the nature of limit seen in the wild-type mice.

The β-casein-deficient mice are not suitable to address questions concerning the co-ordinated expression of the casein locus as hypothesised by Rosen et al. (1989) because the targeted β-casein locus is transcriptionally active as evident from the Northern analysis of mammary gland RNA using the neo gene probe (Fig. 3.12B). Targeted mutations in the promoters of various casein genes may be informative in this respect.

The β-casein-deficient animals born to heterozygous mothers did not show any overt abnormality. Given the fact that casein micelles can be organised in the absence of β-casein, it is not possible to exclude the functional significance of casein gene expression in the thymus as proposed by Grusby et al. (1990). Targeted deletion of other milk proteins would be necessary to examine the proposed role of casein micelles in the delivery of perforin, one of the effector molecules in T-cell mediated cytotoxicity.

In a number of species β-caseins have been shown to contain peptide sequences with opioid activities (reviewed by Teschemacher and Koch 1991). During the present studies the pups feeding on β-casein-deficient milk appeared to grow slower than the wild-type control pups. This difference is likely to be due to the reduction in the total protein in β-casein-deficient milk. To address the non-nutritional significance of β-casein it would be necessary to substitute for the loss of this protein in milk of these animals (see below).

The fact that the loss of β-casein has been compensated by the increased level of the remaining milk proteins (including whey proteins) may be useful to enhance the level of heterologous/pharmaceutical protein in the milk of transgenic farm animals (Velander et al. 1992; Wright et al. 1991; I have crossed β-casein-deficient mice with transgenic mice secreting BLG in milk to examine whether BLG expression is up-regulated to compensate for the loss of β-casein). Little information is available on the effect of expression of a foreign protein on total milk protein level in farm animals. There is some genetic evidence in cattle to suggest that protein bio-synthetic capacity may be limiting (Hill 1993). The A allele at the β-lactoglobulin locus in cattle is associated with a significantly high level of this protein in milk but this increase is at the expense of α-lactalbumin and casein concentrations (Hill 1993). Ideally, if most of the endogenous milk proteins can be deleted from the 'milk' of animals secreting pharmaceutical protein(s) such an approach may even be useful in the purification procedures needed to harvest these proteins. However, this may very well be an oversimplification as there is some evidence to suggest that 3' UTR of differentiation specific RNAs may be required in the maintenance/ promotion of differentiation (Rastinejad and Blau 1993) and thus, deletion of some of the milk protein genes may be incompatible with lactation.

To conclude, I have described the production and characterization of the phenotype of β-casein-deficient mice in this chapter. The mouse β-casein gene in embryonic stem cells was disrupted through homologous recombination, using a replacement-type gene targeting vector based on the positive and negative selection protocol. Germ line transmission of the mutation was obtained. The intercrossing of heterozygous animals resulted in viable and fertile homozygous mutant offspring. The females

appeared to lactate normally and suckled their pups successfully. Analysis of the β-casein-deficient milk revealed that β-casein is not required in the formation and secretion of casein micelles; however, the reduction (in mutant heterozygote) and absence of β-casein (in mutant homozygotes) result in significantly smaller casein micelles. These data provide the first *in vivo* evidence in support of the surface localisation of κ-casein in casein micelles. Finally, the synthesis and or secretion of milk proteins other than β-casein were up-regulated to partially compensate for the loss of β-casein in milk of these mice; this may be useful in enhancing the level of heterologous protein in the mammary gland of farm animals.

CHAPTER FOUR

THE EFFECTS OF TERMINAL HETEROLOGIES ON GENE TARGETING BY INSERTION VECTORS: TOWARDS 'POSITIVE-NEGATIVE' SELECTION

4.1. INTRODUCTION

If bovine embryonic stem cells were to become available, removal of a major protein from milk by targeted disruption of the corresponding gene may be beneficial to modify nutritional properties and the processing behaviour of milk, β-lactoglobulin being an obvious example (Bremel et al. 1989; Clark 1992). However, subtle changes in the structure of milk caseins have been suggested to be more important for modifying industrial properties of milk (Clark 1992; Jimenez-Flores and Richardson 1988).

A two step approach involving targeted deletion of an endogenous milk protein gene (see Chapter 3) and reintroduction of the subtly modified gene into the germ line by pronuclear microinjection (see Chapter 1: section 1.3) may be useful in the manipulation of milk structure. However, such an approach is likely to be hampered by the lack of control of transgene expression levels and other limitations of transgenic technology (reviewed by Clark et al. 1990; Wall and Seidel Jr. 1992). Direct targeting of subtle mutations into most endogenous genes do not lead to a biochemically selectable phenotype in ES cells, and therefore, it is necessary to include some positively selectable marker gene into the targeting vector. Incorporation of a marker gene alongside a

subtle mutation might interfere with the subsequent expression of the targeted locus. Recently, a number of strategies have been reported which either obviate introduction of extraneous sequences into the targeted locus (Davis et al. 1992; Reid et al. 1991; Zimmer and Gruss 1989) or allow excision of the marker gene in a second homologous recombination step (Askew et al. 1993; Hasty et al. 1991a; Valancius and Smithies 1991b) [reviewed in Chapter 1: section 1.4.2.7]. A two step 'hit and run' (Hasty et al. 1991a) or 'in-out' (Valancius and Smithies 1991b) method seems to be the most promising approach for targeting subtle mutation into non-selectable loci. One desirable refinement of the 'hit and run' strategy would be an enrichment procedure for targeting events during the first step which involves targeted integration of an insertion-type vector.

Targeting with an insertion vector is stimulated by a double-strand break or a gap (Orr-Weaver et al. 1981) in the region of homology in yeast, and this has also been found to be the case in mouse ES cells (Hasty et al. 1992; Valancius and Smithies 1991a). A positive-negative selection enrichment procedure in conjunction with insertion-type vectors, similar to that available for replacement-type gene targeting vectors (Mansour et al. 1988), would require blocking of the DNA ends at the double-strand break point with some negatively selectable marker gene. This would be an apparent contradiction with the observation that recombination is stimulated by double-strand break in the region of homology. However, no evidence existed in the literature on the effect and fate of large heterologous sequences included in an insertion vector at the double-strand break point. Therefore, I examined the targeting efficiency of a number of HPRT insertion-type gene targeting vectors with

variable lengths of heterologous sequences -from 0.19 kb to 2.1 kb- included at the point of double-strand break in these vectors.

In this chapter, I provide evidence for differential effects of heterology on the targeting efficiency of insertion-type vectors depending upon the precise design of such vectors. While large heterology on the longer of the two arms of homology of the targeting vectors significantly reduced the efficiency the effect of identical heterology included on the shorter arm was less pronounced. Mostly, these heterologous sequences were excluded from the targeted locus as revealed by Southern analysis. Surprisingly, an insertion vector with both homologous arms blocked by large heterologies targeted with high efficiency and fidelity. The data provide some information about the way in which targeting vectors are processed. Thus, it may be possible to include a negatively selectable marker gene at the double-strand break point in insertion vectors to enrich for targeted clones. These results have also been reported elsewhere (Kumar and Simons 1993: see appendix).

4.2. RESULTS

4.2.1. HPRT Gene Targeting Vectors and Strategy

Plasmid pIV (Fig. 4.1B; see Chapter 2: section 2.3 for construction) comprises 5.6 kb of the HPRT sequences, neo resistance selectable gene and a pGEM4 plasmid backbone. When opened at the unique *Bam*HI site in the HPRT sequences this plasmid yields an insertion-type gene targeting vector with arms of 4.8 kb and 0.8 kb. Homologous recombination of this vector with the HPRT gene is predicted to result in integration of all the vector sequences into the locus

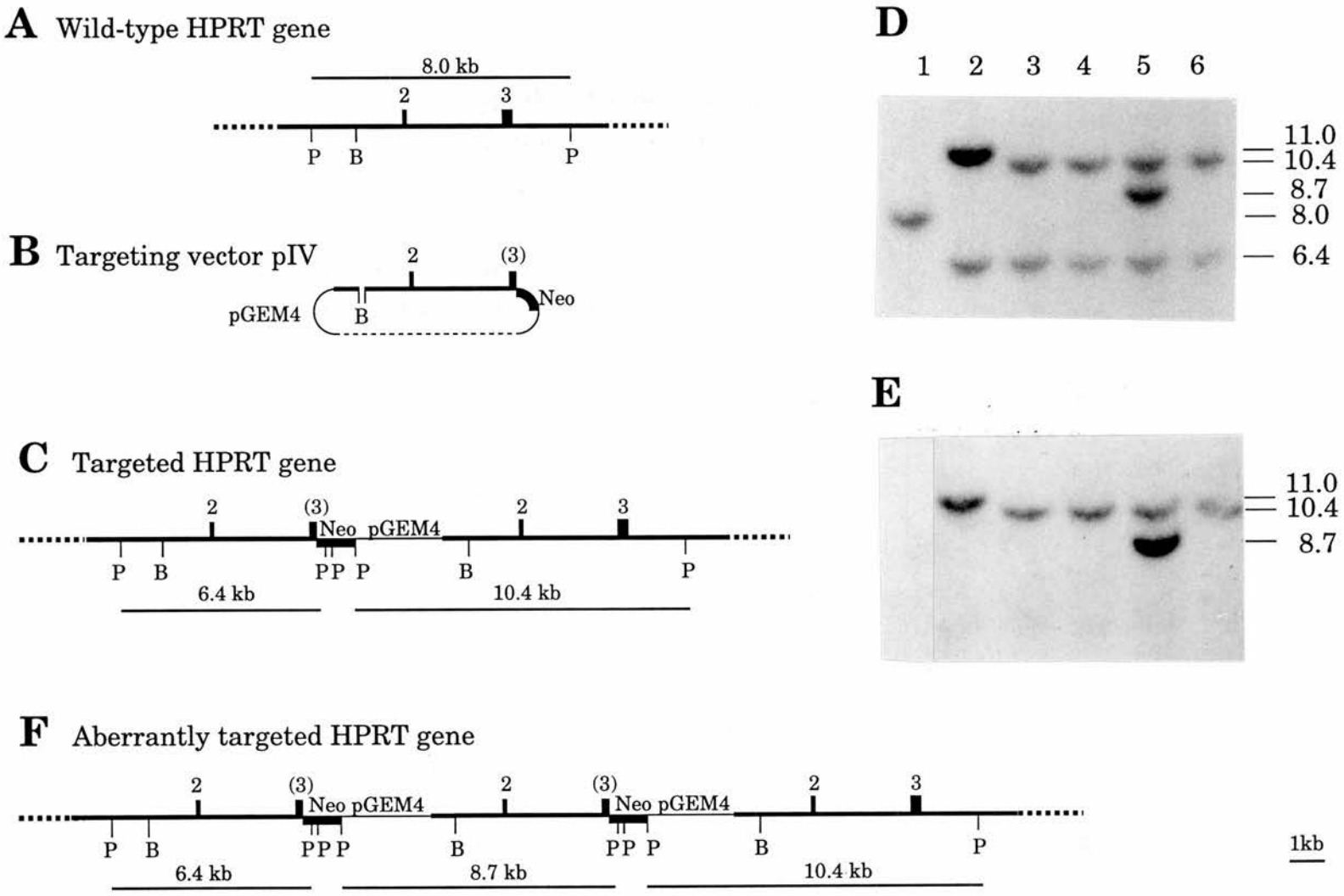

A Wild-type HPRT gene
8.0 kb
2
3
P
B
P
B Targeting vector pIV
2
(3)
pGEM4
B
Neo
C Targeted HPRT gene
2
(3)
Neo
pGEM4
2
3
P
B
P P P
B
P
6.4 kb
10.4 kb
D
1 2 3 4 5 6
11.0
10.4
8.7
8.0
6.4
E
11.0
10.4
8.7
F Aberrantly targeted HPRT gene
2
(3)
Neo pGEM4
2
(3)
Neo pGEM4
2
3
P
B
P P P
B
P P P
B
P
6.4 kb
8.7 kb
10.4 kb
1kb

Figure 4.1. Scheme of targeted integration of an insertion vector pIV into the HPRT gene and Southern blot analysis of targeted ES cell clones.

A) A restriction map of wild-type HPRT gene focusing on exons 2 and 3.

B) Structure of the HPRT gene targeting vector pIV.

C) Predicted restriction map of the targeted HPRT gene.

D) Southern blot analysis of *Pst*I-digested DNA from five pIV clones (lanes 2-6) with the HPRT cDNA probe. The 10.4 kb and 6.4 kb fragments in lanes 3, 4 and 6 are diagnostic of correct targeting and eight further clones gave similar fragments. The 8.7 kb fragment in lane 5 is probably from integration of a head-to-tail multimer of the vector (see Panel F).

E) Southern blot analysis of *Pst*I-digested DNA of five pIV clones (as in lanes 2-6 in Panel D) with pGEM4neo probe. The 10.4 kb fragment in lanes 3, 4 and 6 confirms correct targeting in these clones and eight further clones gave a similar pattern. While the 10.4 kb fragment in lane 5 is consistent with targeting on the short arm of vector an additional intense 8.7 kb band is probably due to integration of a head -to tail multimer of the vector.

F) Proposed structure of an aberrantly targeted pIV clone (based on Southern blot data; see Panels D & E: lane 5). A head-to-tail vector multimer integrated into the HPRT locus. *N.B.* Only two copies of pIV are included in this diagram, more may be involved.

Thick lines: HPRT sequences; box numbered 2, 3 and (3): HPRT exons 2, 3 and incomplete exon 3, respectively. Boxes labelled Neo: Neo gene; thin broken line: pGEM4 vector sequences (not to scale). B: *Bam*HI; P: *Pst*I.

as depicted in Fig. 4.1C. As a consequence the 5.6 kb HPRT gene sequences included in the targeting vector would be duplicated. The duplicated sequences would be separated by the pGEM4 backbone and neo sequences disrupting the endogenous HPRT gene. The polyadenylation signal in the neo gene would further ascertain against the possibility of regeneration of functional HPRT mRNA from the modified locus, which could in principle occur by splicing of exon 2 from 5' of the neo gene to down-stream exon 3 of the HPRT gene. An ES cell clone modified by targeted integration of pIV DNA should be selectable with G418 (for integration) and 6-TG (for the loss of HPRT gene function).

Plasmid pIV served as the basis for construction of targeting vectors which carried terminal heterology, and as a positive control targeting vector. The predicted modification by other HPRT gene targeting vectors having terminal heterologous sequences (see Chapter 2: section 2.3.2 for construction) should be similar to that shown for pIV in Fig. 4.1C, assuming that the heterologous sequences were removed from these vectors before their targeted integration. Alternatively, the targeted integration of the latter vectors without removal of heterologous sequences would also be predicted to result in the loss of the HPRT gene function.

4.2.2. G418, 6-TG-Resistant Colonies Were Targeted for the HPRT Gene

E14 ES cells electroporated in the presence of vector DNAs were selected for G418 resistance from 1 or 2 days until 5 days after electroporation. G418, 6-TG selection was applied for the next nine days to assess the number of the HPRT gene targeted colonies (Table 4.1A).

Table 4.1A. Targeting frequencies of the HPRT gene targeting vectors.

Relative targeting efficiency; No. of colonies resistant to G418 and 6-TG.

Experiment	A	B	C	D	E	F	G	H	I	J	
Vector											Remark
pIV	1.0; 65	1.0; 50	1.0; 14*	1.0; 43*	1.0; 21	1.0; 154	1.0; 112	1.0; 272	1.0; 179	1.0; 108	Positive control: no terminal heterologies
pIVX	0.98; 64									0.88; 95	*Xho*I linker
pIVL2.1	0.28; 18	0.36; 18	0.29; 04*	0.28; 12*				0.33; 89	0.36; 65		2.1 kb on long arm
pIVS2.1		0.64; 32	0.71; 10*	0.72; 31*							2.1 kb on short arm
pIVL0.7		0.40; 20						0.59; 160	0.57; 102	0.55; 59	0.7 kb on long arm
pIVS0.7		0.70; 35								0.89; 90	0.7 kb on short arm
pIVL0.33				0.57; 21						0.30; 32	0.33 kb on long arm
pIVS0.33				1.03; 38						0.53; 57	0.33 kb on short arm
pIVL0.19					0.57; 12		1.02; 114				0.19 kb on long arm
pIVS0.19					0.76; 16		1.16; 130				0.19 kb on short arm
pIVL0.7S1.4				1.03; 38	1.10; 23	0.85;131		0.59; 161	0.54; 97		0.7 kb on long arm, 1.4 kb on short arm
pWPN	1	0	0	0	0	0	1	1	0	1	Negative control

In all experiments, 0.8 ml of E14 ES cells (at passage 24-40) were suspended at 10^8 ml^{-1} in PBS and electroporated at 500 μF, 230V (experiments A-F) or 3 μF, 800V (experiments G-J) in the presence of 5 nM of targeting vector DNA. Electroporated cells were diluted in complete medium, and $5x10^6$ cells were plated per 6 cm petri dish (except experiment C, $3.2x10^6$ cells were plated per dish). G418 (Geneticin; Gibco) selection was applied after 48 hrs (500 μg ml^{-1}; experiments A-F) or after 24 hours (300 μg ml^{-1}; experiments G-J). On day 6 after electroporation 80% of the dishes received medium with $2x10^{-6}$M 6-TG and G418. On day 14, the number of 6-TG-G418-resistant colonies was counted after fixing and staining the dishes. Independent 6-TG-G418-resistant colonies were picked on day 14 prior to fixation for clonal expansion and DNA preparation.

The relative targeting efficiency was calculated as the ratio of number of 6-TG-G418 colonies (on 12 petri dishes per vector except *: 14 dishes) from a targeting vector to the number of colonies from a positive control vector pIV within an experiment. The values corresponding to the negative control vector are the number of spontaneous 6-TG-resistant colonies.

N.B.: For experiments H and I HPRT gene targeting vectors were all obtained from plasmid pIVL2.1; pIV was pIVL2.1 digested with *Bam*HI, pIVL2.1 was pIVL2.1x*Xho*I, pIVL0.7 was pIVL2.1x*Xho*I+*Sma*I and pIVL.7S.1.4 was pIVL2.1x*Sma*I

Table 4.1B. Transfection efficiency of the HPRT gene targeting and negative control vectors.

	Transfection efficiency						
Experiment	A	B	C	D	E	F	
Vector							Remark
pIV	3.37×10^{-4}	1.39×10^{-4}	0.50×10^{-4}	1.70×10^{-4}	1.54×10^{-4}	1.14×10^{-4}	Positive control: no terminal heterologies
pIVX	3.60×10^{-4}						*Xho*I linker
pIVL2.1	1.91×10^{-4}	0.94×10^{-4}	0.53×10^{-4}	1.02×10^{-4}			2.1 kb on long arm
pIVS2.1		1.24×10^{-4}	0.61×10^{-4}	1.36×10^{-4}			2.1 kb on short arm
pIVL0.7		1.30×10^{-4}					0.7 kb on long arm
pIVS0.7		1.40×10^{-4}					0.7 kb on short arm
pIVL0.33				2.12×10^{-4}			0.33 kb on long arm
pIVS0.33				1.82×10^{-4}			0.33 kb on short arm
pIVL0.19					1.40×10^{-4}		0.19 kb on long arm
pIVS0.19					1.23×10^{-4}		0.19 kb on short arm
pIVL0.7S1.4				1.42×10^{-4}	1.03×10^{-4}	0.96×10^{-4}	0.7 kb on long arm, 1.4 kb on short arm
pWPN	2.59×10^{-4}	1.79×10^{-4}	0.48×10^{-4}	1.86×10^{-4}	1.56×10^{-4}	1.71×10^{-4}	Negative control

Transfection efficiency was calculated as the ratio of G418-resistant colonies to the number of cells plated. 20% of the dishes in an given experiment were kept on G418 selection alone and the number of colonies was counted on day 10 after electroporation (see legend to Table 4.1A for details of electroporation and selection protocols).

The positive control HPRT gene targeting vector pIV and a negative control vector pWPN (see Chapter 2: section 2.3) gave G418-resistant colonies at a comparable frequency (Table 4.1B). However, in ten electroporation experiments the vector pIV generated 1018 6-TG-resistant colonies in contrast to only four such colonies detected with pWPN (Table 4.1A). Thus, it was concluded that most 6-TG-resistant colonies in these experiments were generated by the targeted integration of insertion vector pIV into the HPRT locus rather than by random inactivation or deletion of the HPRT gene. To confirm this conclusion, Southern analysis of 6-TG-resistant colonies generated by vector pIV was performed.

A 340 bp *Hpa*II-*Hin*cII fragment from the mouse HPRT cDNA (Konecki et el. 1982), essentially corresponding to exons 2 and 3 (Melton et al. 1984) should hybridise to an 8.0 kb *Pst*I fragment from wild-type HPRT gene in E14 ES cells (Fig. 4.1A). Targeted integration of vector pIV into the HPRT gene was predicted to yield two *Pst*I fragments of approximate sizes of 10.4 kb and 6.4 kb (Fig. 4.1C). The 10.4 kb *Pst*I fragment should also hybridise with the pGEMneo probe (neo gene in the backbone of the plasmid pGEM4).

Genomic DNA from thirteen pIV-generated G418, 6-TG-resistant ES cell clones was analysed to assess the structure of the HPRT locus (Fig. 4.1D). The HPRT cDNA probe hybridised to an 8.0 kb *Pst*I fragment from control E14 ES cells. In all thirteen clones, the 8.0 kb *Pst*I band was absent as was expected after successful targeting of the HPRT gene by pIV. Eleven clones gave two fragments of 10.4 and 6.4 kb. These sizes are consistent with modification of the HPRT locus by integration of pIV as shown in Fig. 4.1C. To confirm this structure, the *Pst*I-digested DNA was probed with pGEMneo DNA (Fig. 4.1E). The presence of a

single 10.4 kb fragment with this probe in these eleven clones was taken as a further evidence of correct targeting by pIV.

One clone gave the above described 10.4 and 6.4 kb fragments with the cDNA probe, and an additional 8.7 kb signal with both the cDNA and pGEMneo probes (Fig. 4.1D & E: lane 5). The intensity of the 8.7 kb band was higher than that of 10.4 kb band indicating integration of multiple copies of the targeting vector. The probable modification in this clone was the targeted integration of a head-to-tail tandem array of pIV into the HPRT gene (Fig. 4.1F). Another clone gave an intense band of approximately 11 kb and a 6.4 kb band with the cDNA probe (Fig. 4.1D: lane 2). The larger band was also recognised by the pGEMneo probe (Fig. 4.1E: lane 2). These observations taken together with the absence of the 8.0 kb wild-type signal indicated that this clone had undergone precise modification on the long arm only. The nature of the recombination event on the short arm of the vector cannot be determined from these data.

Thus, all the thirteen 6-TG-resistant pIV clones analysed were generated as a result of targeted integration of pIV into the HPRT gene. The above results demonstrated that frequency of 6-TG-resistant colonies was a valid measurement of targeting efficiency of pIV.

4.2.3. Comparison of Targeting Efficiency of Different HPRT Gene Targeting Vectors

Ten gene targeting experiments (Table 4.1A) were performed with HPRT vectors to determine the effects of terminal heterologies on gene targeting efficiency. Each experiment included the positive control vector pIV and the negative control vector pWPN in addition to various HPRT vectors carrying terminal heterologies. In every

case, an approximately same number of ES cells was electroporated in the presence of 5 nM of the respective vector DNA (see legend to Table 4.1A). To allow comparison between experiments, the targeting frequency (number of 6-TG-resistant colonies per cell electroporated) of each HPRT vector in a given experiment was expressed relative to the targeting frequency of vector pIV (Table 4.1A) [The relative targeting efficiency of pIV = 1.0]. The absolute targeting frequency of a given vector varied between experiments. However, the relative targeting efficiency of a vector was fairly consistent across the experiments. Variable targeting frequency of a particular targeting vector has also been observed by other workers (Hasty et al. 1992; Koller et al. 1991; Piedrahita et al. 1992). The targeting frequency in the present experiments did not bear any relationship with passage number of cells at electroporation (data not shown) as suggested by Koller et al. (1991).

The frequency of random integration, measured by the number of G418-resistant colonies per cell electroporated varied from experiment to experiment for a given vector (Table 4.1B). This value also varied among different targeting vectors in a particular experiment. However, the latter variability was not unexpected because there were major structural differences between the various HPRT targeting vectors (see Chapter 2: section 2.3) which may influence the expression of the neo gene. In some studies (Hasty et al. 1991c; Hasty et al. 1992) the comparison of the gene targeting efficiency of a number of HPRT vectors has been made after expressing targeting frequency as the number of targeting events (number of 6-TG-resistant colonies) per number of random integration events (number of G418-resistant colonies). Hasty et al. (1991b) compared the absolute targeting efficiency of various vectors but did not use equimolar vector DNA. Such comparisons across targeting

vectors could be misleading. Therefore, in the present study the number of 6-TG-resistant colonies per cell electroporated in the presence of equimolar vector DNA was used as an estimate of the absolute targeting frequency.

4.2.4. Targeting Efficiency Was Not Appreciably Affected by Small Heterologies at the Double-Strand Break Point

Plasmid pIVX was derived from plasmid pIV by insertion of *Xho*I linker in a unique *Bam*HI site of the latter (see Chapter 2: Fig. 2.3C). When opened at this *Xho*I site, pIVX was similar in structure to the positive control vector pIV except for the presence of five terminal heterologous nucleotides on one strand and a single nucleotide on the second strand of each of the two homologous arms. The relative targeting efficiency of pIVX in two experiments was 0.98 and 0.88 (Table 4.1A). Thus, small heterologies on both homologous arms did not appreciably affect the targeting efficiency.

4.2.5. Targeting Efficiency Was Differentially Affected by Large Heterologies Included on One Arm

Targeting efficiencies of a series of HPRT vectors (see Chapter 2: section 2.3) carrying different lengths of heterologous sequences (2.1 kb - 0.19 kb) on one of their arms and a small heterology (5 nucleotides & 1 nucleotide) on the other arm were compared with that of the positive control vector pIV (Table 4.1A). The heterologies reduced the targeting efficiencies of these vectors and the effect was dependent on the position and length of such heterologous sequences. When 2.1 kb

heterology was placed on the long arm, pIVL2.1 targeted the HPRT gene with relative efficiencies ranging from 0.28 to 0.36 (Table 4.1A). When the same 2.1 kb sequences were included on the short arm (vector pIVS2.1) the targeting efficiencies were 0.64, 0.71 and 0.72 in three different experiments. This positional effect was consistent for each pair of targeting vectors with the terminal heterologies of 0.7 kb, 0.33 kb and 0.19 kb (Table 4.1A); in each case the relative targeting efficiency was higher when the heterologous sequences were placed on the short arm.

The targeting efficiency was also sensitive to the length of heterology on a given homologous arm. As the length of heterology on the long arm decreased from the 2.1 kb in vector pIVL2.1 to 0.19 kb in pIVL0.19 the relative targeting efficiencies increased (Table 4.1A). This trend was also noticeable when a similar comparison was made within the vectors carrying heterologies on the short arm but the corresponding changes were small. This was not unexpected as the largest heterology on the short arm led to a relatively low decrease in targeting efficiency.

4.2.6. Terminal Heterologous Sequences Were Excluded from Most Clones Targeted by pIVL2.1 and pIVS2.1

Targeting of the HPRT gene by pIVL2.1 (Fig. 4.2.A) or pIVS2.1 (Fig. 4.3A) with removal of the 2.1 kb terminal heterology is predicted to result in a structure identical to that obtained after targeting by pIV (Fig. 4.1C). *Pst*I digested DNA from such a targeted clone should give the 10.4 kb and 6.4 kb fragments with the HPRT cDNA probe, and a 10.4 kb fragment with the pGEMneo probe. On the other hand, if targeting occurs without removal of the 2.1 kb heterology from a recircularised pIVL2.1 or pIVS2.1 molecule the 2.1 kb heterology would

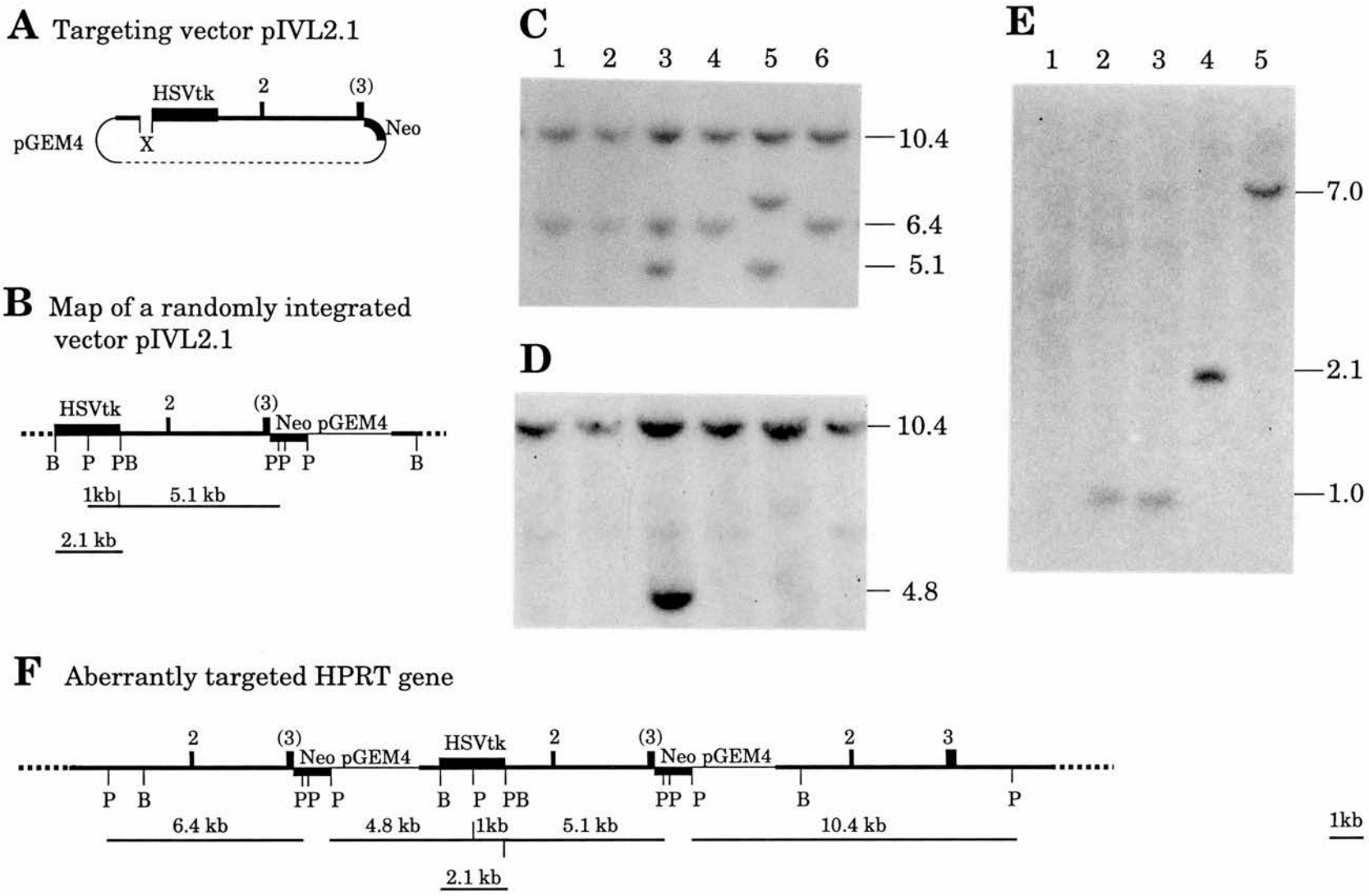

A Targeting vector pIVL2.1
HSVtk
2
(3)
pGEM4
X
Neo
B Map of a randomly integrated vector pIVL2.1
HSVtk
2
(3)
Neo pGEM4
B P PB
PP P
B
1kb
5.1 kb
2.1 kb
C
1 2 3 4 5 6
10.4
6.4
5.1
D
10.4
4.8
E
1 2 3 4 5
7.0
2.1
1.0
F Aberrantly targeted HPRT gene
2
(3)
Neo pGEM4
HSVtk
2
(3)
Neo pGEM4
2
3
P B
PP P
B P PB
PP P
B
P
6.4 kb
4.8 kb
1kb
5.1 kb
10.4 kb
2.1 kb
1kb

Figure 4.2. Structure of the HPRT gene targeting vector pIVL2.1 and Southern blot analysis of targeted ES cell clones.

A) Structure of the HPRT gene targeting vector pIVL2.1.

B) Map of a randomly integrated molecule of pIVL2.1.

C) Southern blot analysis of *Pst*I-digested DNA of six pIVL2.1 clones with the HPRT cDNA probe. The 10.4 kb and 6.4 kb *Pst*I fragments originate from correct targeting after removal of the 2.1 kb terminal heterology. A 5.1 kb fragment in lanes 3 and 5 indicates retention of the 2.1 kb heterology. Analysis of four further clones was consistent with correct targeting after removal of terminal heterology. See Fig. 4.1D for controls

D) Confirmation of the structure of six pIVL2.1 clones (as in Panel C) using pGEM4neo probe. The 10.4 kb *Pst*I fragment is consistent with correct targeting on the short arm. The 4.8 kb fragment in lane 3 indicates retention of terminal heterology in a head-to-tail dimerised vector (see Panel F).

E) Confirmation of retention of terminal heterology in two pIVL2.1 targeted clones (same as clones 3 & 5 in Panels C & D). *Pst*I (lanes 2 & 3) and *Bam*HI (lanes 4 & 5) digested DNA was hybridised with HSV tk 3' sequences. Lanes 2 & 4: clone 3; lanes 3 & 5: clone 5. Lane 1 contains *Pst*I-digested negative control DNA.

F) Structure of the HPRT locus modified as a result of aberrant targeting by pIVL2.1 (see lane 3 in Panels C & D; lane 2 and 4 in Panel E). A head-to-tail vector dimer is integrated into the HPRT locus; one copy of 2.1 kb terminal heterology is retained between two copies of the vector.

Thick lines: HPRT sequences; box numbered 2, 3 and (3): HPRT exons 2, 3 and incomplete exon 3, respectively. Boxes labelled Neo: Neo gene; thin broken line: pGEM4 vector sequences (not to scale); HSVtk: tk gene from pSPTK. B: *Bam*HI; P: *Pst*I; X: *Xho*I.

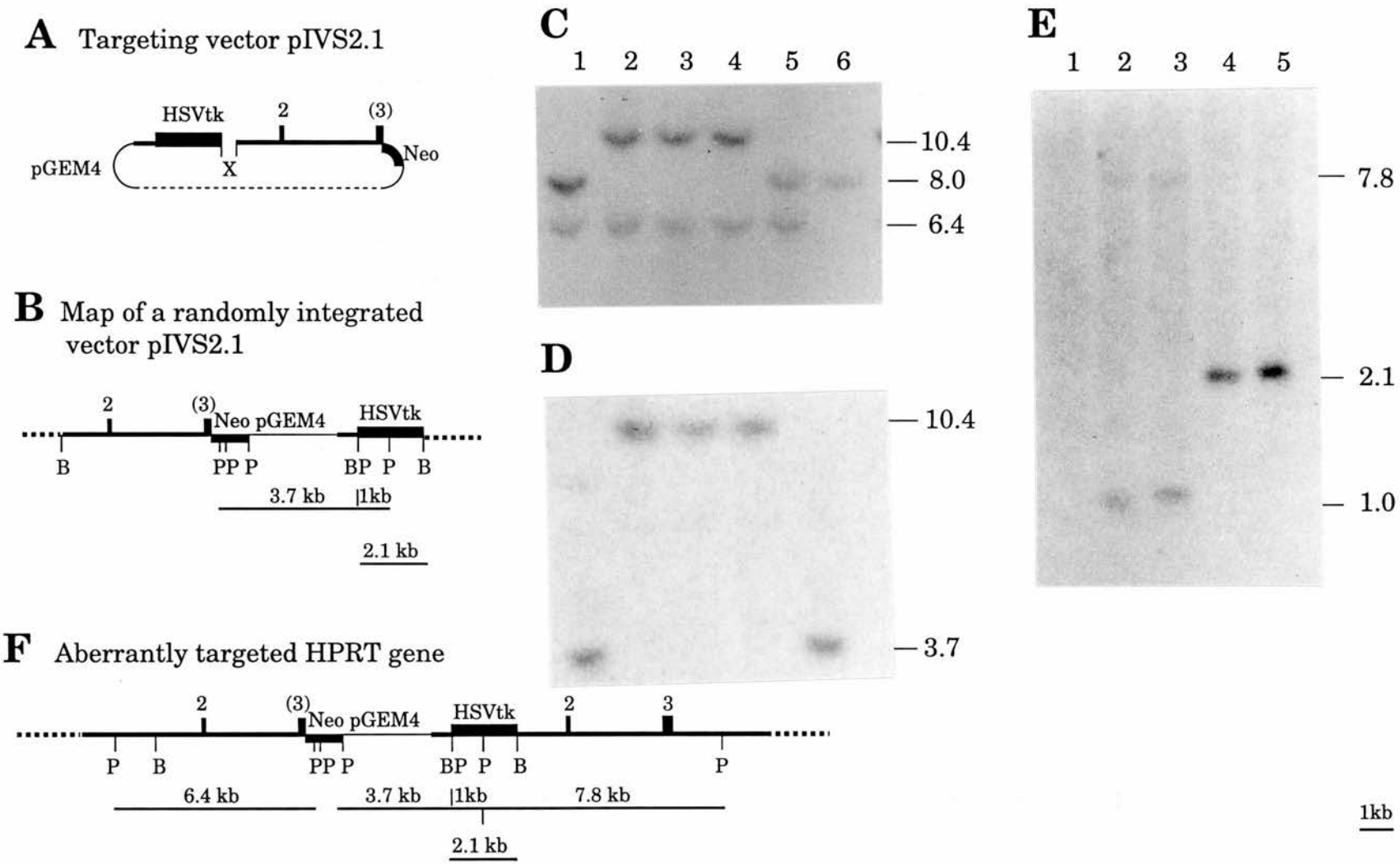

A Targeting vector pIVS2.1
HSVtk
2
(3)
pGEM4
X
Neo
B Map of a randomly integrated vector pIVS2.1
2
(3)
Neo pGEM4
HSVtk
B
PP P
BP P B
3.7 kb
1kb
2.1 kb
C
1 2 3 4 5 6
10.4
8.0
6.4
D
10.4
3.7
E
1 2 3 4 5
7.8
2.1
1.0
1kb
F Aberrantly targeted HPRT gene
2
(3)
Neo pGEM4
HSVtk
2
3
P B
PP P
BP P B
P
6.4 kb
3.7 kb
1kb
7.8 kb
2.1 kb

Figure 4.3. Structure of the HPRT gene targeting vector pIVS2.1 and Southern blot analysis of targeted ES cell clones.

A) Structure of the HPRT gene targeting vector pIVS2.1.

B) Map of a randomly integrated pIVS2.1 molecule.

C) Southern blot analysis of *Pst*I-digested DNA of six pIVS2.1 clones (lanes 1-5) with the HPRT cDNA probe. The 10.4 kb and 6.4 kb *Pst*I fragments (lanes 2, 3 & 4) originate from correct targeting after removal of the 2.1 kb terminal heterology from the short arm of pIVS2.1. One further clone gave a similar pattern. The 8 kb fragment in lanes 1 and 5 indicates inclusion of terminal heterology into the HPRT locus (see Panel F). Lane 6 has *Pst*I-digested control E14 DNA.

D) Confirmation of the structure of five pIVS2.1 clones (the same as in Panel C) using pGEM4neo probe. The 10.4 kb *Pst*I fragment shows correct targeting on the short arm. The 3.7 kb fragment in lanes 1 and 5 indicates retention of terminal heterology. (see Panel F).

E) Confirmation of retention of terminal heterology in two pIVS2.1 clones (the same as the clones 1 & 5 in Panels C & D). *Pst*I (lanes 2 & 3) and *Bam*HI (lanes 4 & 5) digested DNA was hybridised with HSV tk 3' sequences. Lanes 2 & 4: clone 1.; lanes 3 & 5: clone 5. Lane 1 contains *Pst*I-digested negative control DNA.

F) Structure of an aberrantly targeted pIVS2.1 clone (see lane 5 in Panels C & D). A single copy of pIVS2.1 has integrated into the HPRT locus without removal of 2.1 kb terminal heterology.

Thick lines: HPRT sequences; box numbered 2, 3 and (3): HPRT exons 2, 3 and incomplete exon 3, respectively. Boxes labelled Neo: Neo gene; thin broken line: pGEM4 vector; HSVtk: tk gene from pSPTK. B: *Bam*HI; P: *Pst*I; X: *Xho*I.

be incorporated into the HPRT gene either 5' or 3' of the vector sequences depending upon the point of cross-over in the recircularised vector. As there are two *Pst*I sites in the tk gene, the incorporation of these sequences into the targeted HPRT locus would result in *Pst*I fragments which would be much smaller than 10.4 kb or 6.4 kb (for example see Fig 4.3F)

Ten G418, 6-TG-resistant ES cell clones generated using pIVL2.1 were analysed for the structure of their HPRT locus. Southern analysis revealed that in eight clones, pIVL2.1 targeted the locus resulting in the same structure that was generally obtained following targeting with pIV. The *Pst*I-digested genomic DNA from these clones gave two fragments of 10.4 kb and 6.4 kb which hybridised to the HPRT cDNA probe (Fig. 4.2C). The 10.4 kb fragment was also recognised by the pGEMneo probe (Fig. 4.2D). There was no evidence of retention of the 2.1 kb heterologous sequences (tk) in these clones (data not shown).

One clone (Fig. 4.2C: lane 3) gave a third 5.1 kb *Pst*I fragment with the cDNA probe in addition to the 10.4 kb and 6.4 kb fragments. The 10.4 kb and 6.4 kb fragments are consistent with targeting via both long and short arms of the vector after removal of terminal heterology. A 5.1 kb *Pst*I fragment can be explained by either targeting by a dimer of pIVL2.1 or random integration of a second copy pIV. Targeting by a head-to-tail dimer would be predicted to give a *Pst*I fragment of 4.8 kb which should hybridise with pGEMneo (Fig. 4.2F) Random integration would yield a fragment of >3.8 kb, the size of which would depend on the structure of the integration site. Probing of *Pst*I-digested DNA with pGEMneo sequences gave a 10.4 kb band as expected after correct targeting by the short arm, and an additional 4.8 kb fragment. The latter fragment indicated that targeting had probably

occurred by a vector dimer as shown in Fig. 4.2F, and a randomly integrated vector molecule was not the likely source of the 5.1 kb *Pst*I band detected by the cDNA probe. The presence of the 5.1 kb *Pst*I fragment predicted retention of at least some of the tk sequences. Probing of the *Pst*I and *Bam*HI-digested DNA with 3' HSV tk gene sequences (1.5 kb *Eco*RV fragment from pSPTK) revealed a 1.0 kb and a 2.1 kb fragment, respectively (Fig. 4.2E: lanes 2 & 4). The above observations suggested the participation of more than one vector molecule in the targeting reaction.

Another clone (Fig. 4 2C: lane 5,) revealed three *Pst*I fragments of 10.4 kb, 7.4 kb and 5.1 kb when hybridised with the HPRT cDNA probe. The 10.4 kb *Pst*I fragment suggests precise targeting by the short arm of the vector. The absence of the 8.0 kb wild-type fragment is also consistent with targeting of the HPRT gene. However, the absence of a 6.4 kb fragment indicated the lack of correct recombination by the long arm of the vector. The presence of the 5.1 kb fragment is evidence for retention of terminal heterology in this clone. The 7.4 kb *Pst*I fragment detected by cDNA probe could not be explained and thus from these data, it is not possible to deduce the precise nature of the modification of the HPRT locus in this clone. The pGEMneo probe hybridised to a 10.4 kb *Pst*I fragment consistent with correct targeting by the short arm (Fig. 4.2D: lane 5). A 1.0 kb *Pst*I fragment hybridising with HSV tk probe confirmed retention of terminal heterology but a 7.0 kb *Bam*HI signal suggested rearrangement/ loss of tk sequences in this clone (Fig. 4.2E: lanes 3 & 5).

The structure of the HPRT gene of six pIVS2.1-derived clones was examined (Fig. 4.3). Detection of the 10.4 kb and 6.4 kb *Pst*I fragments with the HPRT cDNA probe in four of these clones indicated

targeted recombination by the short and the long arms of the vector, respectively, with removal of terminal heterology (Fig. 4.3C). The wild-type 8.0 kb fragment was absent. Hybridisation with the pGEMneo probe also revealed the expected 10.4 kb fragment in these clones (Fig. 4.3D).

In two other clones a 6.4 kb *Pst*I fragment was detected with the HPRT cDNA probe, suggesting targeting by the long arm of the vector. However, an approximately 8 kb fragment instead of the expected 10.4 kb fragment was inconsistent with targeted recombination by the short arm and removal of the terminal heterology. It may be noted that wild type HPRT gene gives the 8.0 kb *Pst*I fragment with this probe. Probing of the *Pst*I-digested DNA with the pGEMneo probe is predicted to give a 3.7 kb fragment if the terminal heterology has been retained and indeed a 3.7 kb band was detected (Fig. 4.3D). However, the 3.7 kb band could result either from targeted or random integration of pIVS2.1 without removal of terminal heterology (Fig. 4.3B). Similarly, 1.0 kb *Pst*I and 2.1 kb *Bam*HI fragments with tk probe were consistent with either of these two possibilities (Fig. 4.3E). However, an additional 7.8 kb *Pst*I fragment with the tk probe (Fig. 4.3E: lanes 2 & 3) could be explained by targeted integration of pIVS2.1 into the HPRT gene without removal of the terminal heterology (Fig. 4.3F). The structure depicted in Fig. 4.3F is also consistent with the 8 kb *Pst*I fragment observed with the cDNA probe. Taken together, the data suggested the targeted integration of the vector pIVS2.1, including the terminal heterology, into the HPRT gene of the above two clones. The higher intensity of the 8 kb fragment in one of these clones (Fig. 4.3C: lane 1) could be due to the presence of multiple copies of pIVS2.1.

In twelve of the sixteen pIVL2.1 and pIVS2.1 clones, the terminal heterology present on one arm of these vectors was excluded

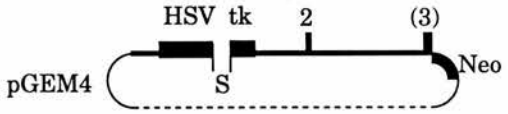
from the targeted HPRT locus. At least, in one of the remaining four clones there was evidence to suggest that terminal heterology was removed from one of the two copies of the vector participating in the targeting reaction.

4.2.7. Targeting by an Insertion Vector with Both of its Arms Blocked by Large Heterologies

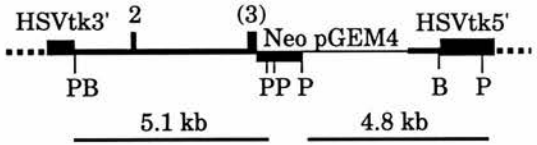
Plasmid pIVL2.1 was opened at a unique *Sma*I site included in the 2.1 kb heterologous sequences. The resulting targeting vector pIVL0.7S1.4 had 0.7 kb heterology on the long arm and 1.4 kb on the short arm (Fig. 4.4A). Surprisingly, this vector targeted the HPRT gene with a reasonable efficiency. In five experiments the relative targeting efficiency varied between 0.54 to 1.1 (Table 4.1A). It may be noted that the total length of heterology in vectors pIVL2.1, pIVS2.1 and pIVL0.7S1.4 was the same. However, the targeting efficiency of pIVL0.7S1.4 was always greater than that of pIVL2.1 (0.28 - 0.36) and comparable to that of pIVS2.1 (0.64 - 0.72).

The above unexpected observation was reinforced by comparison of the relative targeting efficiency of pIVL0.7 and pIV0.7S1.4. pIVL0.7 had 0.7 kb heterology on the long arm while pIVL0.7S1.4 had 1.4 kb on the short arm in addition to 0.7 kb on the long arm. Having observed the reduction in targeting efficiency by blocking either the long or short arm with large heterologies, it was expected that pIVL0.7S1.4 should target less efficiently than pIVL0.7. However, in five experiments, the relative targeting efficiency of pIVL0.7S1.4 ranged from 0.54 to 1.10; when compared with targeting efficiency of pIVL0.7 (from 0.40 to 0.59). It is clear that when the long arm carried 0.7 kb heterology, the presence of

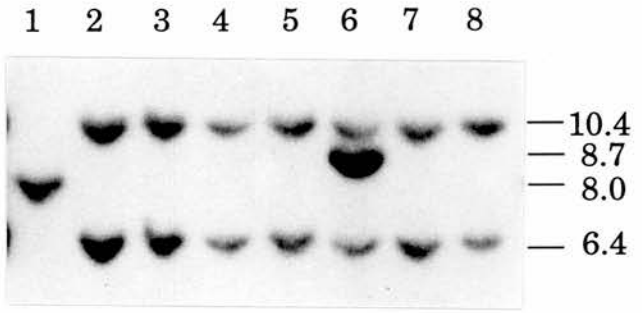
A Targeting vector pIVL0.7S1.4



B Map of a randomly integrated vector pIVL0.7S1.4



C



D

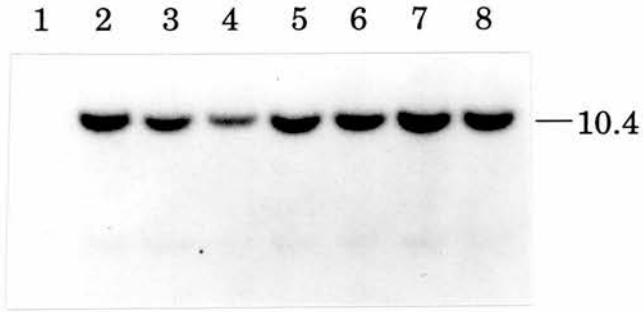


Figure 4.4. Structure of the HPRT gene targeting vector pIVL0.7S1.4 and Southern blot analysis of targeted ES cell clones.

A) Structure of the HPRT gene targeting vector pIVL0.7S1.4.
B) Map of a randomly integrated pIVL0.7S1.4 molecule.
C) Southern blot analysis of *Pst*I-digested DNA of seven pIVL0.7S1.4 clones (lanes 2-8) with the HPRT cDNA probe. The 10.4 kb and 6.4 kb fragments result from correct targeting after removal of the terminal heterologies from both arms of the vector. An additional 8.7 kb intense fragment (lane 6) is probably the outcome of amplification of the vector sequences after targeting or targeting by a vector multimer. Sixteen further clones gave signals consistent with correct targeting after removal of terminal heterologies as in lanes 2-5, 7 and 8. Lane 1 contains *Pst*I-digested control E14 DNA.
D) Confirmation of the structure of pIVL0.7S1.4 clones (lanes 3-8; different clones from those shown in Panel D) with pGEM4neo probe. The 10.4 kb *Pst*I fragment confirms correct targeting on the short arm. Sixteen further clones gave the similar signal. One clone (Panel C: lane 6) gave an additional 8.7 kb fragment (not shown). Lane 1: *Pst*I-digested negative control DNA; lane 2: *Pst*I-digested positive control DNA (pIV, see Fig. 4.1E).

Thick lines: HPRT sequences; box numbered 2, 3 and (3): HPRT exons 2, 3 and incomplete exon 3, respectively. Boxes labelled Neo: Neo gene; thin broken line: pGEM4 vector sequences (not to scale); HSVtk5' and HSVtk3': 5' and 3' fragments of tk gene, respectively from pSPTK. B: *Bam*HI; P: *Pst*I; S: *Sma*I

1.4 kb of heterology on the short arm had little effect or may even have enhanced targeting. These results show that the effects on targeting efficiency of blocking two arms of homology with large heterologies were not cumulative.

4.2.8. Terminal Heterologies Were Excluded from the Clones Targeted by pIVL0.7S1.4

pIVL0.7S1.4 could, in principle, integrate into the HPRT gene in four different possible ways with respect to the terminal heterologies: 1) Integration of the recircularised vector without the loss of terminal heterologies. 2) The loss of terminal heterology from the short arm and retention of heterology on the long arm. 3) The loss of terminal heterology from the long arm and retention of heterology on the short arm. 4). Removal of terminal heterologies from both arms. The HSV tk gene fragments on both the short and long arms of pIV0.7S1.4 have one *Pst*I site each (Fig 4.4B). Probing of the *Pst*I-digested DNA of clones targeted by this vector with the HPRT cDNA would readily distinguish the fourth possibility from the remaining three. Removal of heterologies from both arms prior to targeted integration of pIVL0.7S1.4 would lead to a 10. 4 kb and 6.4 kb *Pst*I fragments, similar to those obtained by pIV targeting. A 5.1 kb *Pst*I fragment is predicted from either of the first two possibilities (as in the case of random integration of the vector, see Fig. 4.4B). Retention of heterology on the short arm (possibilities 2 and 3) would lead to a 7.0 kb *Pst*I fragment detectable with the cDNA probe (not shown).

Twenty three G418, 6-TG-resistant ES cell clones generated by pIVL0.7S1.4 were analysed for their HPRT locus structure. 22 clones

gave the 6.4 kb and 10.4 kb *Pst*I fragments with the HPRT cDNA probe (Fig. 4.4C). These sizes are characteristic of correct targeting by this vector with removal of terminal heterologies from both arms. Probing with pGEMneo highlighted the 10.4 kb fragment (Fig. 4.4D) confirming the results obtained from the cDNA probe.

One remaining clone had in addition to the 6.4 kb and 10.4 kb bands, an 8.7 kb band of higher intensity which hybridised with both the cDNA probe (Fig. 4.4C: lane 6) and the pGEMneo probe (data not shown). The absence of any signal upon hybridisation with the tk probe confirmed that terminal heterologies had been removed from the vector before targeting (data not shown). It may be recalled that head-to-tail concatemerisation of pIV generates 8.7 kb *Pst*I fragment with the cDNA and pGEMneo probes (see Fig. 4.1F). The 8.7 kb band might have resulted from amplification of the duplicated HPRT sequences after targeted integration of the vector. The alternate explanation of targeting by an array of vector molecules seems less likely because there was no evidence of retention of terminal heterologies which would have been expected at least in some of the internal copies of pIVL0.7S1.4 array.

4.2.9. Insertion Vectors and Positive-Negative Selection?

The above results provided ample evidence to conclude that the presence of terminal heterologies does not exclude the insertion vector from the gene targeting pathway. Moreover, the fidelity of the targeting reaction was not compromised. Targeting vectors pIVL2.1 and pIVS2.1 had HSV tk gene on the long and short arm, respectively. The HSV tk gene would permit GANC selection against ES cells transformed by random integration of these two vectors (Mansour et al. 1988). Two

experiments were undertaken to determine the level of enrichment with GANC selection. ES cells were electroporated in the presence of either pIVL2.1 or pIVS2.1 DNA. Electroporated cells were selected for G418, GANC-resistance and G418-resistance. Both of these vectors gave similar but very low (~2-fold) enrichment after GANC selection.

4.3. DISCUSSION

Inclusion of a negative selection marker gene at the double-strand break point of an insertion-type gene targeting vector would be useful to select against the random integration of the vector provided that blocking of the DNA ends does not interfere with the efficiency and fidelity of homologous recombination. Data presented in this chapter show that terminal heterologies included on the two arms reduced the targeting efficiency of the vector but the extent of reduction depended upon the length and position of the heterology. Longer heterology tended to decrease the targeting efficiency to a greater degree. Heterology limited to a few base pairs did not appreciably influence targeting efficiency. Surprisingly, while a 2.1 kb heterology on the long arm of the vector was considerably detrimental, the same piece of DNA either on the short arm or distributed on two arms resulted in a much smaller reduction in the targeting efficiency.

Hasty et al. (1992) found that heterologies of 13 bp on each of the two arms of an HPRT insertion vector did not appreciably affect its targeting efficiency. However, 2.3 kb heterology on the long arm significantly reduced the targeting efficiency. In another study a 37 bp heterology was compatible with gene targeting by an insertion vector in Chinese hamster ovary cells (Pennington and Wilson 1991). These

observations are consistent with the present results obtained for cognate vectors.

Out of 39 clones generated by HPRT targeting vectors carrying large terminal heterologies, 35 clones did not carry the heterologous sequences into the target locus. This is a strong evidence against recircularisation of these vectors before their participation in homologous recombination. These results also exclude a non-conservative gene targeting pathway via a multimerised vector and support the conclusions of Pennington and Wilson (1991) that, in mammalian cells, gene targeting by insertion vectors proceeds by a conservative pathway. In principal, the structure of two pIVS2.1 clones (Fig. 4.3C: lanes 1 & 5) could have arisen by homologous recombination of a recircularised vector molecule with the chromosomal gene. However, the relatively high targeting efficiency of pIVS2.1 (Table 4.1A) would suggest against this being the major path of recombination as it is well established that circular vectors target very inefficiently (Hasty et al. 1992; Orr-Weaver et al. 1981; Valancius and Smithies 1991a). The majority of randomly integrated vector molecules during the present investigations retained the terminal heterologies (data not shown). A number of previous studies (Folger et al. 1982; Macleod et al. 1991; Roth et al. 1985; Thomas et al. 1986; Toneguzzo et al. 1988;) have also shown that randomly integrated single copy DNA molecules more or less preserve their terminal sequences. Thus, the targeting efficiency and Southern analysis data reported in this chapter could only be explained by postulating removal of terminal heterologies from targeting vectors in conjunction with homologous recombination.

Double-strand-break repair (DSBR) model (Fig. 4.5) is the currently accepted model of meiotic homologous recombination in yeast

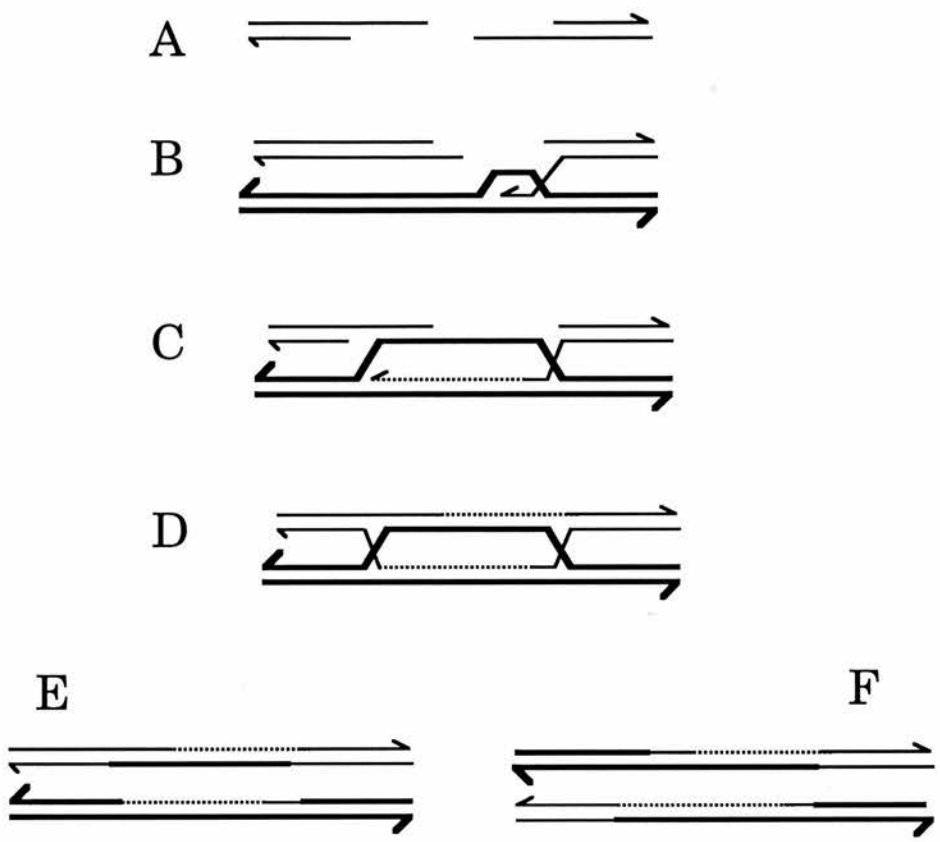


Figure 4.5. Schematic representation of the Double-Strand Break Repair model of homologous recombination (Redrawn from Szostak et al. 1983).
A) A double-strand break is enlarged into a double-strand gap by exonucleases and 3' single-strands are exposed on both sides of the gap.
B) One of the 3' single-strands invades the homologous double-strand DNA and one strand from the latter is looped out.
C) The invading 3' single-strand is extended by repair-synthesis and the loop is enlarged. The second 3' end now anneals to the enlarged loop.
D) Repair-synthesis from the second 3' end converts the looped single-strand into double-strand and two Holliday junctions are formed.
E) Resolution of the two Holliday junctions by two cuts in either the outer or the inner strands results in non-crossover products.
F) Crossover products result if one of the Holliday junctions in resolved by a cut in the outer strand and a second cut in the inner strand.

(Sun et al. 1989, 1991; Szostak et al. 1983; Zenvirth et al. 1992). This model proposes that a double-strand break in the targeting vector is extended by a 5'—›3' exonuclease concurrently from both the DNA ends. One of the resulting 3' ends of the vector invades the target duplex and displaces the corresponding strand from it. The displaced strand is primed by the second 3' tail of the invading molecule. The 3' ends of the vector are extended by copying information from the target duplex resulting in two Holliday junctions. Resolution of these junctions can lead to either cross-over or non-cross-over products. The double-strand break repair model (Szostak et al. 1983) predicted 3' tails at the double-strand break point on both the ends of the invading molecule and indeed, indirect (Lichten et al. 1990) and direct (Sun et al. 1991) evidence exists for association of 3' tails with a meiosis specific double-strand break at the ARG4 locus in yeast. It has been suggested that these tails are generated by 5'—›3' exonuclease digestion. A purified strand exchange protein (SepI) from yeast mitotic cells has been demonstrated to have this activity (Johnson and Kolodner 1991).

Gene targeting by insertion-type vectors in the mouse ES cells appears to be consistent with the DSBR model (Valancius and Smithies 1991a). These authors observed that small heterologies and deletions embedded in the homologous sequences of insertion vectors were co-converted to the chromosomal sequences, particularly if such heterologies were within 600 bp from the point of the double-strand break in the vector. In these experiments the ends of various targeting vectors were always in the region of homology. To explain the directional correction, the authors argued against the inclusion of heterologies into the heteroduplex DNA and instead suggested extension of double-strand

break by exonucleases. The proposed exonucleases would have to degrade both strands of the targeting vector.

4.3.1. Mechanism(s) of Removal of Terminal Heterologies

It is very unlikely that double-strand endonuclease(s) removed the terminal heterologies from HPRT insertion vectors in the present study. Such a proposition would have to assume at least one endonuclease site in each of the two arms of homology. There is no direct evidence for the presence or absence of these sites. However, double-strand endonuclease(s) cannot account for differential effect of the length and position of heterology on targeting efficiency of the various HPRT vectors (Table 4.1A).

Two alternative mechanisms could be envisaged: 1) digestion of both strands by 5'—›3' and 3'—›5' exonucleases, and 2) exonuclease digestion of one strand followed by non-random removal of second strand by a single-strand endonuclease.

Removal of both strands of the terminal heterologies by exonucleases is compatible with differential effects of heterologies on targeting efficiency. The structure of pIVL0.7S1.4 targeted clones is a proof that terminal heterologies are removed from both arms (Fig. 4.4C & D). If exonuclease reactions proceed concurrently in both directions from the double-strand break point removal of 2.1 kb heterologous sequences from the long arm of vector pIVL2.1 would be accompanied by complete degradation of the 0.8 kb short arm. The vector processed this way would not be compatible with targeting in the manner observed. In some cases if the reactions on both arms are not completely synchronous, partial homology may be left on the short arm when the removal of heterology

from the long arm is complete. Such a vector would still be able to participate in a targeting reaction assuming repair of the gap (Valancius and Smithies 1991a). The above arguments can explain reduced targeting efficiency of pIVL2.1. On the other hand removal of 2.1 kb heterology from the short arm of vector pIVS2.1 is not likely to be accompanied by complete removal of 4.8 kb homology from the long arm. Thus, there would be a smaller effect on the targeting efficiency of pIVS2.1 and the same has been observed experimentally (Table 4.1A). A similar explanation could easily account for targeting efficiency data obtained with vector pIVL0.7S1.4. Heterology on the small arm in this vector would protect the homologous sequences on this arm, and thus, the targeting by this vector would be apparently less sensitive to the presence of terminal heterologies.

The above interpretation of the experimental data is based on the premise that the reduction in targeting efficiency of vectors carrying heterologies on the long arm is not merely the consequence of blockage of homology on the long arm of these vectors. This assumption is validated by the observation that reduction in efficiency of vectors carrying heterologies on the long arm is responsive to change in the length of heterology (Table 4.1A).

A 5'—› 3' exonuclease, central to the DSBR model, has been copurified with a multiprotein recombinational complex (RC-1) preparation from calf thymus nuclei (Jessberger et al. 1993). This protein complex can catalyse repair of double-strand gaps and deletions via recombination *in vitro* (Jessberger et al. 1993). Low level of 3'—›5' exonuclease potency has also been detected in association with RC-1 by these workers. Assuming that similar activities exist in mouse ES cells, the removal of terminal heterologies (the present study) and internal

heterologies (Valancius and Smithies 1991a) could be effected by such exonucleases.

A single-strand exonuclease degradation followed by the action of a single-strand non-random endonuclease is another possible mechanism for the removal of terminal heterologies. According to the DSBR model, the polarity of the exonuclease activity would have to be 5'—›3'. The action of such an exonuclease on different targeting vectors would determine their efficiencies in exactly the same way as has been explained above for degradation of both strands by exonucleases. Exonuclease could completely remove one strand of homology from the short arm of the vector pIVL2.1 before one of the strands of 2.1 kb heterology is removed from the long arm of the vector. Single-strand homologous sequences would still be left on the short arm. However, double-strand homology on both arms would be necessary for the formation of two Holliday junctions (Fig. 4.5). If after the action of single-strand exonuclease the vector retains some double-strand homology on both arms and the second strand of heterology is removed by a single-strand endonuclease activity, the intermediate vector molecule could easily be assimilated in homologous recombination reaction after gap repair. Recently, Fishman-Lobell and Haber (1992) have invoked a single-strand endonuclease to explain removal of non-homologous ends during extra-chromosomal recombination in yeast.

A mechanism for enlargement of double-strand break by exonuclease digestion of both strands will be consistent with homology search by the vector either before or after exposure of homologous ends. It may be noted that a recombinase activity from human cell extract needed homologous ends for homology search (Hsieh et al. 1986). On the other hand single-strand exonuclease and endonuclease based mechanism

would require homology search at least prior to the removal of the second strand of terminal heterology (Fishman-Lobell and Haber 1992). The latter workers have suggested that paranemic base pairing (Schutte and Cox 1987) could achieve the homology search in the absence of homologous ends on the invading DNA molecule. However, paranemic base pairing as a mechanism of homology search is yet to be demonstrated in eukaryotic cells.

In conclusion, the results presented in this chapter show that it may be possible to include a negative selection marker gene at the double-strand break point. Negative selection against HSV tk gene included in replacement-type gene targeting vectors can provide a significant level of enrichment against random integration of targeting vectors (Mansour et al. 1988). In preliminary experiments, poor enrichment was obtained with HPRT insertion vectors pIVL2.1 and pIVS2.1. Poor enrichment is not uncommon (Itohara et al. 1993; Mombaerts et al. 1991; Tybulewicz et al. 1991). Some important points emerged from the present investigations; these should be helpful in the design of insertion-type gene targeting vectors to be used in tandem with negative selection against random integration. The length of marker gene should be kept as short as possible to minimise the reduction of targeting efficiency. Equal distribution of total homology on two arms of the vectors may be advantageous. Finally, it may be desirable to block each arm of homology with a different negative marker gene. Although the absolute targeting efficiency of such a gene targeting vector may be slightly reduced, the probability of finding a targeting event can significantly increase depending upon the level of enrichment obtained after selection against the negative marker gene.

CHAPTER FIVE

OVERVIEW

In principle, any endogenous gene can be manipulated through gene targeting in embryonic stem cells. It is not yet known whether there are regions of the genome that are refractory to targeting, although successful targeting of a wide variety of mouse genes so far would suggest that most of the genome may be amenable to targeting.

The availability of mouse ES cells has provided impetus to studies intended to understand the process of gene targeting in mammalian cells and as a result of these studies better design of gene targeting vectors is now possible. These methods will find immediate application in farm animals provided germ line-competent ES cells are available from these species.

ES-like cell lines have been isolated from day 7-9 post-oestrus blastocysts of pigs (Evans et al. 1990; Notarianni et al. 1990, 1991, and Piedrahita et al. 1990a, 1990b). Morphologically these cells appeared as flattened epithelial cells. When grown to high density or in suspension culture, cells from one cell line differentiated into a variety of cell types indicating pluripotency (Evans et al. 1990). One ES-like cell line from a day 8 post-oestrus sheep blastocyst has been reported (Notarianni et al. 1991); spontaneous differentiation of this cell line on reaching confluence into various cell-types reflected pluripotency. To date, there are no published reports of chimaeras generated with ES cells from livestock (reviewed by Anderson 1992). In part this is due to the difficulties of

working with livestock species and may reflect species differences in early development (Evans et al. 1990).

Recently, the establishment of ES-like cell lines from mouse primordial germ cells (PGC) has been demonstrated (Matsui et al. 1992; Resnick et al. 1992). While these EG (embryonic germ) cells may not be strictly identical to ES cells, their isolation has opened up the intriguing possibility of similar experiments in livestock species as an alternate source of pluripotential cell lines (Resnick et al. 1992; reviewed by Rossant 1993). The use of basic fibroblast growth factor (bFGF) in combination with the transmembrane form of steel factor and LIF is critical in the establishment of cell lines from PGC. Matsui et al. (1992) have demonstrated that once established these EG cell lines can be grown on STO feeders in the absence of any additional factors, and that they are pluripotent. Two independent lines contributed to chimaerism when injected into blastocysts (Matsui et al. 1992). However, it is yet to be shown whether EG cell lines are germ line-competent, a prerequisite for use in genetic manipulation experiments; the potential of EG cell lines may depend on the state of genomic imprinting in these lines (see Rossant 1993).

As the technology stands at present considerable time must elapse between the creation of a mutation in ES cells and having a mutant animal because the manipulated ES cells can only be used to produce chimaeric animals in the first generation. Therefore, the progress in using gene targeting in the improvement of farm animals is likely to be slow. A possible solution may involve nuclear transfer from the genetically manipulated ES cells to enucleated oocytes. Nuclear transfer from inner cell mass (ICM) cells to enucleated oocytes has been used to produce a live lamb (Smith and Wilmut 1989). Although ES cells are

derived from the ICM, the nature of modification that allows the *in vitro* propagation of these cells is not known. If we assume that ES cell nuclei would be able to support development to term, nuclear transfer would facilitate generation of live animals from the genetically manipulated ES cells without the need to produce chimaeric intermediates. Consecutive targeting of both alleles of a given gene in ES cells before nuclear transfer would allow production of animals homozygous for the mutation in the first generation.

Another possible approach for the manipulation of endogenous genes that requires consideration involves microinjection of targeting vectors into eggs. Successful manipulation of an endogenous gene in mouse eggs in one study has suggested that eggs may have cellular machinery capable of supporting gene targeting (Brinster et al. 1989). Extrachromosomal homologous recombination has also been demonstrated in mouse eggs (Pieper et al. 1992). Although Brinster et al. (1989) observed only one targeting event out of 500 transgenic mice screened and a number of new mutations were introduced in the target locus in addition to the intended modification, it is very likely that this frequency may represent an underestimation because there were numerous polymorphisms between the target gene and the targeting vector DNA. It is yet to be established whether additional mutations introduced in the target locus represented the inherent imprecision of gene targeting machinery in eggs in contrast to that in ES cells or whether this discrepancy was the result of the extensive mismatches between the target gene and vector DNA. Nevertheless, the error-prone targeting may still be suitable for generating null mutations.

The use of isogenic DNA in the construction of gene targeting vectors has been shown to significantly improve gene targeting efficiency

in ES cells (Van Deursen and Wieringa 1992; Te Riele et al. 1992) Interestingly, in one experiment 78% of the cells that were transformed by the targeting vector DNA underwent a targeting event (Te Riele et al. 1992). A targeting efficiency of this magnitude, if routinely attainable in eggs, should be sufficient to modify endogenous genes without the use of ES cells. One potentially important advantage of such a system would be the possibility to diagnose the targeted embryos using PCR before embarking upon the production of live animals. Such an approach, if feasible, could tremendously reduce the cost of genetic manipulation of farm animals. However, it should be noted that the use of isogenic DNA in livestock species would be very difficult as no inbred strains exist in these species. Cloning of the gene of interest directly from the putative ES cell lines of farm animals would assure isogenicity of targeting vector DNA to one of the two alleles in such cells.

A better understanding of the mechanisms of homologous recombination and random integration in mammalian cells may lead to improvements in gene targeting efficiency. It is not clear at what stage a DNA molecule is committed to either random integration or the homologous recombination pathway. Do these two apparently different processes overlap? Do random integration events represent the failure of homologous recombination or were the randomly integrated molecules never available for long enough to be able to participate in the homologous recombination reaction? The results presented in Chapter 4 of this thesis suggest that the fate of large terminal heterologies placed at the DNA ends of an insertion vector depends on the type of integration. While in targeted integration these heterologies are lost the integrity of the DNA ends is more or less maintained in the random integration events. Although both random and targeted integrations are stimulated

by linearisation of the vector DNA, the underlying mechanisms of this stimulation may very well be quite different in these processes. Mammalian cells are very efficient in the ligation of DNA ends (see review by Roth and Wilson 1988) and concatemerisation of DNA molecules can be detected within 5-10 minutes after microinjection into the mouse eggs (Burdon and Wall 1992). In COS-1 cells, the ratio of intramolecular ligation of DNA ends to extrachromosomal homologous events involving the same ends has been shown to be reduced by addition of dideoxynucleotides at the 3' ends (Chang and Wilson 1987). Although it is not possible to conclude from these results that addition of dideoxynucleotides suppressed the ligation reaction (the alternate explanation, which is unlikely but cannot be excluded, may be that intramolecular recombination was enhanced) it may be worth examining the effect of this modification on the frequency of random integration. In view of the removal of terminal heterologies from insertion-type gene targeting vectors as shown in the present study, it will be interesting to study the effects on gene targeting of placing dideoxynucleotides at the DNA ends.

The mechanisms of targeting by replacement vectors is not understood at all except that it is known that linear molecules target more efficiently than circular DNA and that the placement of terminal heterologies is compatible with a gene targeting reaction. However, it is not clear whether terminal heterologies reduce the targeting efficiency of these vectors, and how the terminal heterologies are excluded from the target gene (see Chapter 1: section 1.4.2.5). Both the single-strand annealing model (Lin et al. 1984, 1990) and the double-strand break repair model (Orr-Weaver et al. 1981; Szostak et al. 1983) would require a double-strand break or gap in the target gene in the chromosome. Such double-strand breaks may be produced either actively by the targeting

vector after the homology search or may arise randomly. If a chromosomal double-strand break is rate limiting in gene targeting by replacement vectors, one prediction would be that the introduction of sequence specific breaks in the endogenous genes should enhance targeting. At present no methods are available to introduce *in vivo* site-specific double-strand breaks that would enable the above hypothesis to be tested.

The efficiency of gene targeting may need further improvements in order to extend this technology to farm animals at a reasonable cost. The available methodologies in mouse are useful to model the genetic modifications intended in livestock species. Extrapolation of the results from the β-casein-deficient mice to farm animals would suggest that to achieve high level and mammary gland specific expression, foreign genes may be targeted into this locus without any adverse effects on lactation, and thus it may be possible to avoid position-effect, generally seen in the context of transgenes. However, it may be noted that the optimum design of a reconstructed hybrid gene cannot always be predicted. Poor expression of a number of heterologous cDNA sequences inserted into the otherwise efficiently expressed sheep β-lactoglobulin gene in transgenic mice (McClenaghan et al. 1991; Whitelaw et al. 1991) is a reminder that much remains to be learnt about the design of transgenes. Although extensive characterization of the properties of wild-type mouse milk would be necessary prior to undertaking such projects, β-casein-deficient mice may be a useful animal model to test the effect of genetically engineered bovine β-caseins on physical properties of milk.

The genetic control of most traits of economic importance in livestock is very complex. Current efforts in genome mapping have begun to locate major genes controlling economic traits of farm animals.

Progressive degenerative myeloencephalopathy (weaver syndrome) in Brown-Swiss cattle is associated with higher milk and milk-fat yield (Hoeschele and Meinert 1990). Recently, Georges et al. (1993) have identified a linkage between a microsatellite and the weaver syndrome. Gene targeting may allow positive identification of the candidate gene for this syndrome. Also it may be possible to address the nature of association between the weaver and favourable dairy phenotypes. More importantly, if it turns out that these two traits are controlled by two different but closely linked genes it may be feasible to break the unfavourable linkage through gene targeting. Porcine stress syndrome (PSS) is another such example (reviewed by MacLennan and Phillips 1992). This is a recessive trait and the homozygous animals are prone to death as a result of stress but this trait is associated in heterozygotes with leaner pork. Meat derived from homozygous affected animals is pale, soft and exudes fluid. Strong linkage has been established between this phenotype and a missense mutation in the skeletal muscle ryanodine receptor gene (Fujii et al. 1991); efforts are underway to mutate this gene in mouse ES cells to create a small animal model to understand the molecular basis of PSS and its relationship to leanness (Alan Archibald personal communication).

Once embryonic stem cells from farm animals become available, gene targeting is likely to contribute to positive identification of major genes of economic importance, and it would be feasible to modify endogenous genes to manipulate the physiology of livestock species.

REFERENCES

Al-Shawi, R., Kinnaird, J., Burke, J. and Bishop, J. O. (1990). Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Mol. Cell. Biol. 10*:1192-1198.

Aleandri, R., Buttazzoni, L. G., Schneider, J. C., Caroli, A. and Davoli, R. (1990). The effects of milk protein polymorphisms on milk components and cheese-producing ability. *J. Dairy Sci. 73*:241-255.

Anderson, G. B. (1992). Isolation and use of embryonic stem cells from livestock species. *Anim. Biotechnol. 3*:165-175.

Archibald, A. L., McClenaghan, M., Hornsey, V., Simons, J. P. and Clark A. J. (1990). High-level expression of biologically active human α_1-antitrypsin in the milk of transgenic mice. *Proc. Natl. Acad. Sci. USA 87*:5178-5182.

Askew, G. R., Doetschman, T. and Lingrel, J. B. (1993). Site-directed point mutations in embryonic stem cells: a gene-targeting tag-and-exchange strategy. *Mol. Cell. Biol. 13*:4115-4124.

Axelrod, H. R. (1984). Embryonic stem cell lines derived from blastocysts by a simplified technique. *Develop. Biol. 101*:225-228.

Azuma, N., Kaminogava, S. and Yamauchi, K. (1985). Reconstitution of human casein micelle and its properties. *Agric. Biol. Chem. 49*:2655-2660.

Baribault, H. and Kemler, R. (1989). Embryonic stem cell culture and gene targeting in transgenic mice. *Mol. Bio. Med. 6*:481-492.

Bayna, E. M. and Rosen, J. M. (1990). Tissue-specific, high level expression of the rat whey acidic protein gene in transgenic mice. *Nucl. Acids Res. 18*:2977-2985.

Blin, N. and Stafford, D. W. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res. 3*:2303.

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

Bonsing, J. and Mackinlay, A. G. (1987). Recent studies on nucleotide sequences encoding the caseins. *J. Dairy Res. 54*:447-461.

Bonsing, J., Ring, J. M., Stewart, A. F. and Mackinlay, A. G. (1988). Complete nucleotide sequence of the bovine β-casein gene. *Aust. J. Biol. Sci. 41*: 527-537.

Bradley, A. (1987) Production and analysis of chimaeric mice. In: *Robertson, E. (ed), Teratocarcinomas and embryonic stem cells: a practical approach, pp. 113-151.* IRL Press, Oxford.

Bradley, A., Evans, M., Kaufman, M. H. and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature 309*:255-256.

Bradley, A., Hasty, P., Davis, A. and Ramirez-Solis, R. (1992). Modifying the mouse: Design and desire. *Bio/Technology 10*:534-539.

Bremel, R. D., Yom, H-C. and Bleck, G. T. (1989). Alteration of milk composition using Molecular Genetics. *J. Dairy Sci. 72*:2826-2833.

Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E. and Palmiter, R. D. (1988). Introns increase transcriptional efficiency in transgenic mice. Proc. *Natl. Acad. Sci. USA 85*:836-840.

Brinster, R. L., Braun, R. E., Lo, D., Avarbock, M. R., Oram, F. and Palmiter, R. D. (1989). Targeted correction of a major histocompatibility class II *Eα* gene by DNA microinjected into mouse eggs. *Proc. Natl. Acad. Sci. USA 86*:7087-7091.

Bühler, Th. A., Bruyère, Th., Went, D. F., Stranzinger, G. and Bürki, K. (1990). Rabbit β-casein promoter directs secretion of human interleukin-2 into the milk of transgenic rabbits. *Bio/Technology 8*:140-143.

Burdon, T. G. and Wall R. J. (1992). Fate of microinjected genes in preimplantation mouse embryos. *Mol. Reprod. Develop. 33*:436-442.

Burdon, T., Wall, R. J., Shamay, A., Smith, G. H. and Hennighausen, L. (1991). Over-expression of an endogenous milk protein gene in transgenic mice is associated with impaired mammary alveolar development and a milchlos phenotype. *Mech. Develop. 36*:67-74.

Capecchi, M. R. (1989a). Altering the genome by homologous recombination. *Science 244*:1288-1292.

Capecchi, M. R. (1989b). The new mouse genetics: Altering the genome by gene targeting. *Trends Genet. 5*:70-76.

Chang, X-B. and Wilson, J. H. (1987). Modification of DNA ends can decrease end joining relative to homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci. USA 84*:4959-4963.

Charron, J., Malynn, B. A., Robertson, E. J., Goff, S. P. and Alt, F. W. (1990). High-frequency disruption of the N-*myc* in embryonic stem and

pre-B cell lines by homologous recombination. *Mol. Cell. Biol. 10*:1799-1804.

Chen, C. and Okayama, H. (1987). High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Bio. 7*:2745-2752.

Cheng, J. and Maquat, L. E. (1993). Nonsense codons can reduce the abundance of nuclear mRNA without affecting the abundance of pre-mRNA or the half-life of cytoplasmic mRNA. *Mol. Cell. Biol. 13*:1892-1902.

Choi, T., Haung, M., Gorman, C. and Jaenisch, R. (1991). A generic intron increases gene expression in transgenic mice. *Mol. Cell. Biol. 11*:3070-3074.

Chomczynski, P. (1992). Solubilization in formamide protects RNA from degradation. *Nucl. Acids Res. 20*:3791-3792.

Church, G. M. and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. 81*:1991-1995.

Clark, A. J. (1992). Prospects for genetic engineering of milk. *J. Cell. Biochem. 49*:121-127.

Clark, A. J., Archibald, A. L., McClenaghan, M., Simons J. P., Wallace, R. and Whitelaw, C. B. A. (1993). Enhancing the efficiency of transgene expression. *Phil. Trans. R. Soc. Lond. B. 339*:225-232.

Clark, A. J., Archibald, A. L., McClenaghan, M., Simons, J. P., Whitelaw, C. B. A. and Wilmut, I. (1990). The germline manipulation of livestock: progress during the past five years. *Proc. N. Z. Soc. Anim. Prod. 50*:167-179.

Clark, A. J., Bessos, H., Bishop, J. O., Brown, P., Harris, S., Lathe, R., McClenaghan, M., Prowse, C., Simons, J. P., Whitelaw, C. B. A. and Wilmut, I. (1989). Expression of human anti-haemophilic factor IX in the milk of transgenic sheep. *Bio/Technology 7*:487-492.

Clark, A. J., Cowper, A. , Wallace, R., Wright, G. and Simons, J. P. (1992). Rescuing transgene expression by co-integration. *Bio/Technology 10*:1450-1454.

Clark, A. J., Simons, J. P., Wilmut, I. and Lathe, R. (1987). Pharmaceuticals from transgenic livestock. *Trends Biotechnol. 5*:20-24.

Clarke, A. R., Maandag, E. R., Van Roon, M., Van der Lugt, N. M. T., Van der Valk, M., Hooper, M. L., Berns, A. and Te Riele, H. (1992). Requirement for a functional *Rb-1* gene in murine development. *Nature 359*:328-330.

Creamer, L. K. (1991). Some aspects of casein micelle structure. In: *Parris, N. and Barford, R. (eds), Interaction of food proteins, pp. 148-163.* American Chemical Society, Washington, DC.

Dalgleish, D. G. (1993). Bovine milk protein properties and the manufacturing quality of milk. *Live. Prod. Sci. 35*:75-93.

Dalgleish, D. G., Horne, D. S. and Law, A. J. R. (1989). Size-related differences in bovine casein micelles. *Biochim. Biophys. Acta 991*:383-387.

Damiani, G., Ferretti, L., Rognoni, G. Sgaramella, V. (1990). Restriction fragment length polymorphism analysis of the κ-casein locus in cattle. *Anim. Genet. 21*:107-114.

Davies, D. T., Holt, C. and Christie, W. W. (1983). The composition of milk. In: *Mepham, T. B. (ed), Biochemistry of lactation, pp. 71-117.* Elsevier Science Publication. B. V. Amsterdam.

Davies, T. D. and Law, A. J. R. (1983). Variation in the protein composition of bovine casein micelles and serum casein in relation to micellar size and milk temperature. *J. Dairy Res. 50*:67-75.

Davis, A. C., Wims, M. and Bradley, A. (1992). Investigation of cotransformation as a method for introducing small mutations into embryonic stem cells. *Mol. Cell. Biol. 12*:2769-2776.

DeChiara, T. M., Efstratiadis, A. and Robertson, E. J. (1990). A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature 345*:78-80.

Deng, C. and Capecchi, M. R. (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.

Deng, C., Thomas, K. R. and Capecchi, M. R. (1993). Location of crossovers during gene targeting with insertion and replacement vectors. *Mol. Cell. Biol. 13*:2134-2140.

Denicourt, D., Sabour, M. P. McAllister, A. J. (1990). Detection of bovine κ-casein genomic variants by the polymerase chain reaction method. *Anim. Genet. 21*:215-216.

Dietz, H. C., Valle, D., Francomano, C. A., Kendzior Jr, R. J., Pyeritz, R. E. and Cutting, G. R. (1993). The skipping of constitutive exons *in vivo* induced by nonsense mutations. *Science 259*:680-683.

Dils, R. R. (1986). Comparative aspects of milk fat synthesis. *J. Dairy Sci. 69*:904-910.

DiTullio, P., Cheng, S. H., Marshall, J., Gregory, R. J., Ebert, K. M., Meade, H. M., Smith, A. E. (1992). Production of cystic fibrosis

transmembrane conductance regulator in the milk of transgenic mice. *Bio/Technology 10*:74-77.

Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S. and Smithies, O. (1987). Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature 330*:576-578.

Doetschman, T., Maeda, N. and Smithies, O. (1988). Targeted mutation of the *Hprt* gene in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA 85*:8583-8587.

Donnelly, W. J., McNeill, G. P., Buchheim, W. and McGann, T. C. A. (1984). A comprehensive study of the relationship between size and protein composition in natural bovine casein micelles. *Biochim. Biophys. Acta 789*:136-143.

Ebert, K. M., Selgrath, J. P., DiTullio, P., Denman, J., Smith, T. E., Memon, M. A., Schindler, J. E., Monastersky, G. M., Vitale, J. A. and Gordon, K. (1991). Transgenic production of a variant of human tissue-type plasminogen activator in goat milk: generation of transgenic goats and analysis of expression. *Bio/Technology 9*:835-838.

Eistetter, H. R. (1989). Pluripotent embryonal stem cell lines can be established from disaggregated mouse morulae. *Develop. Growth Differ. 31*:275-282.

Erickson, H. P. (1993). Gene knockouts of c-*src*, transforming growth factor β1, and tenascin suggest superfluous, non-functional expression of proteins. *J. Cell Biol. 120*:1079-1081.

Evans, M. J. (1989). Potential for genetic manipulation of mammals. *Mol. Biol. Med. 6*:557-565.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature 292*:154-156.

Evans, M. J., Notarianni, E., Laurie, S. and Moor, R. M. (1990). Derivation and preliminary characterization of pluripotent cell lines from porcine and bovine blastocysts. *Theriogenology 33*:125-128.

Evans, M., Bradley, A. and Robertson, E. (1985). EK cell contribution to chimaeric mice: From tissue culture to sperm. In *Jaenisch, R. and Costantini, F. (eds), Banbury Report 20. Genetic Manipulation of the early Mammalian Embryo, pp. 93-102.* Cold Spring Harbor Laboratory Press. New York.

Falconer, D. S. (1989). *Introduction to quantitative genetics (3rd ed).* John Wiley and Sons, Inc., New York.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabellling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem. 132*:6-13.

Feinberg, A. P. and Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Addendum. Anal. Biochem. 137*:266-267.

Ferretti, L., Leone, P. and Sgaramella, V. (1990). Long range restriction analysis of the bovine casein genes. *Nucl. Acids Res. 18*:6829-6833.

Fishman-Lobell, J. and Haber, J. E. (1992). Removal of nonhomologous DNA ends in double-strand break recombination: The role of yeast ultraviolet repair gene *RADI. Sci. 258*:480-484.

Flecknell, P. A. (1983). Anaesthesia in rodents and rabbits. *Vet. Rec. 113*:574.

Folger, K. R., Wong, E. A., Wahl, G. and Capecchi, M. R. (1982). Pattern of integration of DNA microinjected into cultured mammalian cells: Evidence for homologous recombination between injected plasmid DNA molecules. *Mol. Cell. Biol. 2*:1372-1387.

Fox, P. F. (1982). Heat-induced coagulation of milk. *In: Fox, P. F. (ed), Developments in dairy chemistry (vol. 1), p. 189.* Applied Science Publications, New York.

Fujii, J., Otsu, K., Zorzato, F., De Leon, S., Khanna, V. K., Weiler, J. E., O'Brien, P. J. and MacLennan, D. H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science 253*:448-451.

Geissler, E. N., Cheng, S. V., Gusella, J. F. and Housman, D. E. (1988). Genetic analysis of the dominant white-spotting (*W*) region on mouse chromosome 5: Identification of cloned DNA markers near *W*. *Proc. Natl. Acad. Sci. USA 85*:9635-9639.

Georges, M., Dietz, A. B., Mishra, A., Nielsen, D., Sargeant, L. S., Sorensen, A., Steele, M. R., Zhao, X., Leipold, H., Womack, J. E. and Lathrop, M. (1993). Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative trait locus. *Proc. Natl. Acad. Sci. USA 90*:1058-1062.

Gibson, J. P. (1987). The option and prospects for genetically altering milk composition in dairy cattle. *Anim. Breed. Abstr. 55*:231-243.

Gibson, J. P. (1989). Altering milk composition through genetic selection. *J. Dairy Sci. 72*:2815-2825.

Goodman, H. S. and Rosen, J. M. (1990). Transcriptional analysis of the mouse β-casein gene. *Mol. Endo. 4*:1661-1670.

Gossler, A., Doetschman, T., Korn, R., Serfling, E. and Kemler, R. (1986). Transgenesis by means of blastocysts-derived embryonic stem cell lines. *Proc. Natl. Acad. Sci. USA 83*:9065-9069.

Grappin, R. and Ribadeau-Dumas, B. (1992). Analytical methods for milk proteins. In: *Fox, P. F. (ed), Advanced dairy chemistry-1: Proteins, p 6.* Elsevier Applied Science, London and New York.

Grosclaude, F. and Martin, P. (1992). The role of casein polymorphisms-An example of QTL: the α_{s1}-casein of the goat. *Anim. Genet. 23 (Suppl.)*:1--2.

Grosclaude, F., Mahé, M. F., Brignon, G., Stasio, L. D. and Jeunet, R. (1987). A Mendelian polymorphism underlying quantitative variations of goat α_{s1}-casein. *Genet. Sel. Evol. 19*:399-412.

Grusby, M. J., Mitchell, S. C., Nabavi, N. and Glimcher, L. H. (1990). Casein expression in cytotoxic T lymphocytes. *Proc. Natl. Acad. Sci. USA 87*:6897-6901.

Gupta, P., Rosen, J. M., D'eustachio, P. and Ruddle, F. H. (1982). Localisation of the casein gene family to a single mouse chromosome. *J. Cell Biol. 93*:199-204.

Handyside, A. H., O'Neill, G. T., Jones, M. and Hooper, M. L. (1989). Use of BRL-conditioned medium in combination with feeder layers to isolate a diploid embryonal stem cell line. *Roux's Arch. Develop. Bio. 198*:48-55.

Hasty, P., Ramírez-Solis, R., Krumlauf, R. and Bradley, A. (1991a). Introduction of a subtle mutation into the *Hox-2.6* locus in embryonic stem cells. *Nature 350*:243-246.

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

Hasty, P., Rivera-Pérez, J. and Bradley, A. (1992). The role and fate of DNA ends for homologous recombination in embryonic stem cells. *Mol. Cell. Biol. 12*:2464-2474.

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

Hennighausen, L. G. and Sippel, A. E. (1982). Mouse whey acidic protein is a novel member of the family of 'four-sulphide core' proteins. *Nucl. Acids. Res. 10*:2677-2684.

Hennighausen, L., Ruiz, L. and Wall, R. (1990). Transgenic animals-production of foreign proteins in milk. *Curr. Opinion Biotechnol. 1*:74-78.

Herskowitz, I. (1987). Functional inactivation of genes by dominant-negative mutations. *Nature 329*:219-222.

Hettinga, D. H. (1989). Why alter milk composition? *J. Dairy Sci. 72*:2790-2800.

Hill, J. P. (1993). The relationship between β-lactoglobulin phenotypes and milk composition in New Zealand dairy cattle. *J. Dairy Sci. 76*:281-286.

Hoeschele, I. and Meinert, T. R. (1990). Association of genetic defects with yield and type traits: The weaver locus effect on yield. *J. Dairy Sci. 73*:2503-2515.

Holt, C. and Sawyer, L. (1988). Primary and predicted secondary structures of the caseins in relation to their biological functions. *Protein Engin. 2*:251-259.

Hooper, M. L. (1992). *Embryonal stem cells-Introducing planned changes into the animal germline. Vol. 1 of the book series Modern Genetics, (ed) Evans, H. J.* Harwood Academic Publishers, Switzerland.

Hooper, M., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonisation by cultured cells. *Nature 326*:292-295.

Horne, D. S. (1984). Determination of the size distribution of bovine casein micelles using photon correlation spectroscopy. *J. Colloid Interface Sci. 98*:537-548.

Horne, D. S. and Dalgleish, D. G. (1985). A photon correlation spectroscopy study of size distributions of casein micelle suspensions. *Eur. Biophys. J. 11*:249-258.

Hsieh, P., Meyn, M. S. and Camerini-Otero, R. D. (1986). Partial purification and characterization of a recombinase from human cells. *Cell 44*:885-894.

Itohara, S., Mombaerts, P., Lafaille, J., Iacomini, J., Nelson, A., Clarke, A. R., Hooper, M. L., Farr, A. and Tonegawa, S. (1993). T cell receptor δ gene mutant mice: Independent generation of αβ T cells and programmed rearrangement of γδ TCR genes. *Cell 72*:337-348.

Jaenisch, R. (1988). Transgenic animals. *Science 240*:1468-1474.

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

Jenness, R. (1986). Lactational performance of various mammalian species. *J. Dairy Sci. 69*:869-885.

Jessberger, R., Podust, V. Hübscher, U. and Berg, P. (1993). A mammalian protein complex that repairs double-strand break and deletions by recombination. *J. Biol. Chem. 268*:15070-15079.

Jimenez-Flores, R. and Richardson, T. (1988). Genetic engineering of the caseins to modify the behaviour of milk during processing: A review. *J. Dairy Sci. 71*:2640-2654.

Johnson, A. W. and Kolodner, R. D. (1991). Strand exchange protein from *Saccharomyces cerevisiae. J. Biol. Chem. 266*:14046-14054.

Jones, W. K., Yu-Lee, L., Clift, S. M., Brown, T. L. and Rosen, J. M. (1985). The rat casein multigene family: Fine structure and evolution of the β-casein gene. *J. Biol. Chem. 260*:7042-7050.

Joyner, A. L., Skarnes, W. C. and Rossant, J. (1989). Production of a mutation in mouse *En-2* gene by homologous recombination in embryonic stem cells. *Nature 338*:153-156.

Kang, Y. and Richardson, T. (1985). Genetic engineering of caseins. *Food Technol. Oct. 1985*:89-94.

Katsuki, M., Sato, M., Kimura, M., Yokoyama, M., Kobayashi, K. and Nomura, T. (1988). Conversion of normal behaviour to shiverer by myelin basic protein antisense cDNA in transgenic mice. *Science 241*:593-595.

Kim, H-S, Smithies, O. (1988). Recombinant fragment assay for gene targeting based on the polymerase chain reaction. *Nucl. Acids Res. 16*:8887-8903.

Koller, B. H. and Smithies, O. (1992). Altering genes in animals by gene targeting. *Annu. Rev. Immunol. 10*:705-730.

Koller, B. H., Kim, H-S., Latour, A. M., Brigman, K., Boucher Jr., R. C., Scambler, P., Wainwright, B. and Smithies, O. (1991). Toward an animal model of cystic fibrosis: Targeted interruption of exon 10 of the cystic fibrosis transmembrane regulator gene in embryonic stem cells. *Proc. Natl. Acad. Sci. USA 88*:10730-10734.

Kollias, G. and Grosveld, F. (1992). The study of gene regulation in transgenic mice. In: *Grosveld, F. and Kollias, G. (eds), Transgenic animals, pp. 79-98.* Academic Press Limited, London.

Konecki, D. S., Brennand, J., Fuscoe, J. C., Caskey, C. T. and Chinault, A. C. (1982). Hypoxanthine-guanine phosphoribosyltransferase genes of mouse and Chinese hamster: construction and sequence analysis of cDNA recombinants. *Nucl. Acids Res. 10*:6763-6775.

Krimpenfort, P., Rademakers, A., Eyestone, W., Van der Schans, A., Van den Broek, S., Kooiman, P., Kootwijk, E., Platenburg, G., Pieper, F.,

Strijker, R. and De Boer, H. (1991). Generation of transgenic dairy cattle using *'in vitro'* embryo production. *Bio/Technology 9*:844-847.

Kuehn, M. R., Bradley, A., Robertson, E. J. and Evans, M. J. (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. *Nature 326*:295-298.

Kuhn, M. J. (1983). The biosynthesis of lactose. In: *Mepham, T. B. (ed), Biochemistry of lactation, pp. 159-176.* Elsevier Science Publication. B. V. Amsterdam.

Kumar, S. and Simons, J. P. (1993). The effects of terminal heterologies on gene targeting by insertion vectors in embryonic stem cells. *Nucl. Acids Res. 21*:1541-1548.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature. 227*:680-685.

Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M. A., Jaenisch, R. and Berns, A. (1991). Simplified mammalian DNA isolation procedure. *Nucl. Acids Res. 19*:4293.

Lallemand, Y. and Brûlet, P. (1990). An *in situ* assessment of the routes and extents of colonisation of the mouse embryo by embryonic stem cells and their descendants. *Development 110*:1241-1248.

Lathe, R., Clark, A. J.. Archibald, A. L., Bishop, J. O., Simons, P. and Wilmut, I. (1986). Novel products from livestock. In *Smith, C., King, J. W. B. and McKay, J. C. (eds), Exploiting new technologies in animal breeding: Genetic Developments, pp. 91-102.* Oxford University Press, Oxford.

Lathe, R., Vilotte, J. L. and Clark, A. J. (1987). Plasmid and bacteriophage vectors for excision of intact inserts. *Gene 57*:193-201.

Lee, K-F., DeMayo, F. J., Atiee, S. H. and Rosen, J. M. (1988). Tissue-specific expression of the rat β-casein gene in transgenic mice. *Nucl. Acids Res. 16*: 1027-1041.

Levin, S. D., Anderson, S. J., Forbush, K. A. and Perlmutter, R. M. (1993). A dominant-negative transgene defines a role for $p56^{lck}$ in thymopoiesis. *EMBO J. 12*:1671-1680.

Lichten, M., Goyon, C., Schultes, N. P., Treco, D., Szostak, J. W., Haber, J. E. and Nicolas, A. (1990). Detection of heteroduplex DNA molecules among the products of *Saccharomyces cerevisiae* meiosis. *Proc. Natl. Acad. Sci. USA 87*:7653-7657.

Lin, C. Y., Sabour, M. P. and Lee, A. J. (1992). Direct typing of milk proteins as an aid for genetic improvement of dairy bulls and cows: a review. *Anim. Breed. Abstr. 60*:1-10.

Lin, F-L, Sperle, K. and Sternberg, N. (1985). Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. *Proc. Natl. Acad. Sci. USA 82*:1391-1395.

Lin, F-L. M., Sperle, K. and Sternberg, N. (1990). Repair of double-stranded DNA breaks by homologous DNA fragments during transfer of DNA into mouse L cells. *Mol. Cell. Biol. 10*:113-119.

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

MacLennan, D. H. and Phillips, M. S. (1992). Malignant hyperthermia. *Science 256*:789-794.

Macleod. D., Lovell-Badge, R., Jones, S. and Jackson, I. (1991). A promoter trap in embryonic stem (ES) cells selects for integration of DNA into CpG islands. *Nucl. Acids Res. 19*:17-23.

Mahé, M. F. and Grosclaude, F. (1989). α_{s1}-Cn^{D}, another allele associated with a decreased synthesis rate at the caprine α_{s1}-casein locus. *Genet. Sel. Evol. 21*:127-129.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, New York.

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

Mansour, S. L., Thomas, K. R., Deng, C. and Capecchi, M. R. (1990). Introduction of *lacZ* reporter gene into the mouse *int-2* locus by homologous recombination. *Proc. Natl. Acad. Sci. USA 87*:7688-7692.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA 78*:7634-7638.

Matsui, Y., Zsebo, K. and Hogan, B. L. M. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell 70*:841-847.

McCarrick III, J. W., Parnes, J. R., Seong, R. H., Solter, D. and Knowles, B. B. (1993). Positive-negative selection gene targeting with the diphtheria toxin A-chain gene in mouse embryonic stem cells. *Transg. Res. 2*:183-190.

McClenaghan, M., Archibald, A. L., Harris, S., Simons, J. P., Whitelaw, C. B. A., Wilmut, I. and Clark, A. J. (1991). Production of human α_1-antitrypsin in the milk of transgenic sheep and mice: targeting

expression of cDNA sequences to the mammary gland. *Anim. Biotech. 2*:161-176.

McGann, T. C. A., Donnelly, W. J., Kearney, R. D. and Buchheim, W. (1980). Composition and size distribution of bovine casein micelles. *Biochim. Biopys. Acta 630*:261-270.

McKnight, R. A., Shamay, A., Sankaran, L., Wall, R. J. and Hennighausen, L. (1992). Matrix-attachment regions can impart position-independent regulation of a tissue specific gene in transgenic mice. *Proc. Natl. Acad. Sci. USA 89*:6943-6947.

McLean, D. M., Graham, E. R. B., Ponzoni, R. W. and McKenzie, H. A. (1984). Effects of milk protein genetic variants on milk yield and composition. *J. Dairy Res. 51*:531-546.

McMahon, A. P. and Bradley, A. (1990). The *Wnt*-1 (*int*-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell 62*:1073-1085.

Meade, H., Gates, L., Lacy, E. and Lonberg, N. (1990). Bovine $alpha_{s1}$-casein gene sequences direct high level of expression of active human urokinase in mouse milk. *Bio/Technology 8*:443-446.

Medrano, J. F. and Aguilar-Cordova, E. (1990). Genotyping of bovine kappa-casein loci following DNA sequence amplification. *Bio/Technology 8*:144-146.

Melton, D. W., Konecki, D. S., Brennand, J. and Caskey, C. T. (1984). Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proc. Natl. Acad. Sci. USA. 81*:2147-2151.

Mepham, T. B. (1987). *Physiology of lactation.* Open University Press, Milton Keynes and Philadelphia.

Mercier, J. C. (1986). Genetic engineering applied to milk producing animals: some expectations. In *Smith, C., King, J. W. B. and McKay, J. C. (eds), Exploiting new technologies in animal breeding: Genetic Developments, pp. 122-131.* Oxford University Press, Oxford.

Mombaerts, P., Clarke, A. R., Hooper, M. L. and Tonegawa, S. (1991). Creation of a large deletion at the T-cell antigen receptor β-subunit locus in mouse embryonic stem cells by gene targeting. *Proc. Natl. Acad. Sci. USA 88*:3084-3087.

Mouellic, H. L., Lallemand, Y. and Brûlet, P. (1990). Targeted replacement of the homeobox gene *Hox-3.1* by the *Escherichia coli lacZ* in mouse chimaeric embryos. *Proc. Natl. Acad. Sci. USA 87*:4712-4716.

Muysson, D. J. and Gibbins, A. M. V. (1989). The alteration of milk content by genetic engineering and recombinant DNA-mediated selection techniques. *Can. J. Anim. Sci. 69*:517-527.

Nagy, A., Gócza, E., Diaz, E. M., Prideaux, V. R., Iványi, E. Markkula, M. and Rossant, J. (1990). Embryonic stem cells alone are able to support fetal development in the mouse. *Development 110*:815-821.

Nandi, A. K., Roginski, R. S., Gregg, R. G., Smithies, O. and Skoultchi, A. I. (1988). Regulated expression of genes inserted at the human chromosomal β-globin locus by homologous recombination. *Proc. Natl. Acad. Sci. USA 85*:3845-3849.

Ng-Kwai-Hang, K. F., Monardes, H. G. and Hayes, J. F. (1990). Association between genetic polymorphism of milk proteins and production traits during three lactation. *J. Dairy Sci. 73*:3414-3420.

Nichols, J., Evans. E. P. and Smith, A. G. (1990). Establishment of germ-line-competent embryonic stem cells (ES) cells using differentiation inhibiting activity. *Development 110*:1341-1348.

Notarianni, E., Galli, C., Laurie, S., Moor, R. M. and Evans, M. J. (1991). Derivation of pluripotent, embryonic cell lines from the pig and sheep. *J. Reprod. Fert., Suppl. 43*:255-260.

Notarianni, E., Laurie, S., Moor, R. M. and Evans, M. J. (1990). Maintenance and differentiation in culture of pluripotential embryonic cell lines from pig blastocysts. *J. Reprod. Fert., Suppl. 41*:51-56.

Ono, T. and Creamer, L. K. (1986). Structure of goat casein micelles. *N. Z. J. Dairy Sci. and Technol. 21*:57-64.

Ono, T. and Obata, T. (1989). A model for the assembly of bovine casein micelles from F2 and F3 subunits. *J. Dairy Res. 56*:453-461.

Orban, P. C., Chui, D. and Marth, J. D. (1992). Tissue- and site-specific DNA recombination in transgenic mice. *Proc. Natl Acad. Sci. USA 89*:6861-6865.

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

Palmiter, R. D. and Brinster, R. L. (1986). Germ-line transformation of mice. *Annu. Rev. Genet. 20*:465-499.

Palmiter, R. D., Brinster, R. L., Hammer, R. E., Trumbauer, M. E., Rosenfeld, M. G., Birnberg, N. C. and Evans, R. M. (1982). Dramatic growth of mice that develop from eggs micrinjected with metallothionein-growth fusion genes. *Nature 300*:611-615.

Palmiter, R. D., Sandgren, E. P., Avarbock, M. R., Allen, D. D. and Brinster, R. L. (1991). Heterologous introns can enhance expression of transgenes in mice. Proc. *Natl. Acad. Sci. USA 88*:478-482.

Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C. T., Findlay, J. B. C., Sivaprasadarao, R., Jones, T. A., Newcomer, M. E. and Kraulis, P. J. (1986). The structure of β-lactoglobulin and its similarity to retinol binding protein. *Nature 324*:383-385.

Pearse, M. J., Linklater, P. M., Hall R. J. and Mackinlay, A. G. (1986). The effect of casein composition and casein dephosphorylation on the coagulation and syneresis. *J. Dairy Res. 53*:381-390.

Pease, S. and Williams, R. L. (1990). Formation of germline chimaeras from embryonic stem cells maintained with recombinant leukaemia inhibitory factor. *Exp. Cell Res. 190*:209-211.

Pennington, S. L. and Wilson, J. H. (1991). Gene targeting in Chinese hamster ovary cells is conservative. *Proc. Natl. Acad. Sci. USA 88*:9498-9502.

Pepin, M-C., Pothier, F. and Barden, N. (1992). Impaired type II glucocorticoid-receptor function in mice bearing antisense RNA transgene. *Nature 355*:725-728.

Persuy, M-A., Stinnakre, M-G., Printz, C., Mahe, M-F. and Mercier, J-C. (1992). High expression of the caprine β-casein gene in transgenic mice. *Eur. J. Biochem. 205*:887-893.

Piedrahita, J. A., Anderson, G. B. and BonDurant, R. H. (1990a). Influence of feeder layer type on the efficiency of isolation of porcine embryo-derived cell lines. *Theriogenology 34*:865-877.

Piedrahita, J. A., Anderson, G. B. and BonDurant, R. H. (1990b). On the isolation of embryonic stem cells: Comparative behaviour of murine, porcine and ovine embryos. *Theriogenology 34*:879-901.

Piedrahita, J. A., Zhang, S. H., Hagaman, J. R., Oliver, P. M. and Maeda, N. (1992). Generation of mice carrying a mutant apolipoprotein E gene inactivated by gene targeting in embryonic stem cells. *Proc. Natl. Acad. Sci. USA 89*:4471-4475.

Pieper, F. R., De Wit, I. C. M., Pronk, A. C. J., Kooiman, P. M., Strijker, R., Krimpenfort, P. J. A., Nuyens, J. H. and De Boer, H. A. (1992). Efficient generation of functional transgenes by homologous recombination in murine zygotes. *Nucl. Acids Res. 20*:1259-1264.

Pittius, C. W., Hennighausen, L., Lee, E., Westphal, H., Nicols, E., Vitale, J. and Gordon, K. (1988). A milk protein gene promoter directs the expression of human tissue plasminogen activator cDNA to the mammary gland in transgenic mice. *Proc. Natl. Acad. Sci. USA 85*:5874-5878.

Rastinejad, F. and Blau, H. M. (1993). Genetic complementation reveals a novel regulatory role for 3' untranslated regions in growth and differrentiation. *Cell 72*:903-917.

Rathjen, P. D., Toth, S., Willis, A., Heath, J. K. and Smith, A. G. (1990). Differentiation inhibiting activity is produced in matrix-associated and diffusible forms that are generated by alternate promoter usage. *Cell 62*:1105-1114.

Reardon, J., Foreman, J. A. and Searcy, R. L. (1966). New reactants for the colorimetric determination of ammonia. *Clinica Chimica Acta 14*:403-405.

Reid, L. H., Shesely, E. G., Kim, H-S. and Smithies, O. (1991). Cotransformation and gene targeting in mouse embryonic stem cells. *Mol. Cell. Biol. 11*:2769-2777.

Resnick, J. L., Bixler, L. S., Cheng, L. and Donovan, P. J. (1992). Long-term proliferation of mouse primordial germ cells in culture. *Nature 359*:550-551.

Rijnkels, M., Kooiman, P., Pieper, F., Nuijens, J., Krimpenfort, P. and De Boer, H. A. (1993). Expression of the bovine α_{s1}-casein gene in transgenic mice. *Live. Prod. Sci. 35*:196-197 (Abstract).

Robertson, E. (1987). Embryo derived stem cell lines. *In: Robertson, E. (ed), Teratocarcinomas and embryonic stem cells: a practical approach, pp. 71-112.* IRL Press, Oxford.

Robertson, E. J. (1991). Using embryonic stem cells to introduce mutations into the mouse germ line. *Biol. Reprod. 44*:238-245.

Robertson, E., Bradley, A., Kuehn, M., Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature 323*:445-448.

Rollema, H. S. (1992). Casein association and micelle formation. In: *Fox, P. F. (ed), Advanced dairy chemistry-1: Proteins, pp. 111-140.* Elsevier Applied. Science, London and New York.

Rosen, J. M., Bayna, E. and Lee, K-F. (1989). Analysis of milk protein gene expression in transgenic mice. *Mol. Biol. Med. 6*:501-509.

Rossant, J. (1993). Immortal germ cells? *Curr. Biol. 3*:47-49.

Roth, D. and Wilson, J. (1988). Illegitimate recombination in mammalian cells. In: *Kucherlapati, R. and Smith, G. R. (eds), Genetic recombination, pp. 621-653.* ASM, Washington, DC.

Roth, D. B., Porter, T. N. and Wilson, J. H. (1985). Mechanisms of nonhomologous recombination in mammalian cells. *Mol. Cell. Biol. 5*:2599-2607.

Rothstein, R. J. (1983). One-step gene disruption in yeast. *Meth. Enzymol. 101*:202-211.

Rowland, S. J. (1938). The determination of the nitrogen distribution in milk. *J. Dairy Res. 9*:42-46.

Rubinstein, M., Japón, M. A. and Low, M. J. (1993). Introduction of a point mutation into the mouse genome by homologous recombination in embryonic stem cells using a replacement type vector with a selectable marker. *Nucl. Acids Res. 21*:2613-2617.

Ruettimann, K. W. and Ladisch, M. R. (1987). Casein micelle: Structure, properties and enzymatic coagulation. *Enzyme Microb. Technol. 9*:578-589.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular cloning: A laboratory manual (2nd ed).* Cold Spring Harbor Laboratory Press, New York.

Schaar, J., Hansson, B. and Pettersson, H-E. (1985). Effects of genetic variants κ-casein and β-lactoglobulin on cheesemaking. *J. Dairy Res. 52*:429-437.

Schmidt, D. G. (1982). Association of caseins and casein micelle structure. In: *Fox, P. F. (ed), Developments in dairy chemistry (Vol. 1), pp. 61-86.* Applied Science Publication, New York.

Schutte, B. C. and Cox, M. M. (1987). Homology-dependent changes in adenosine 5'-triphosphate hydrolysis during *recA* protein promoted DNA strand exchange: Evidence for long paranemic complexes. *Bichemistry 26*:5616-5625.

Schwartzberg, P. L., Goff, S. P. and Robertson, E. J. (1989). Germ-line transmission of a c-*abl* mutation produced by targeted gene disruption in ES cells. *Science 246*:799-803.

Seidman, M. M. (1987). Intermolecular homologous recombination between transfected sequences in mammalian cells is primarily non-conservative. *Mol. Cell. Biol. 7*:3561-3565.

Selfridge, J., Pow, A. M., McWhir, J., Magin, T. M., Melton, D. W. (1992). Gene targeting using a mouse HPRT minigene/HPRT-deficient embryonic stem cell system: Inactivation of the mouse *ERCC-I* gene. *Som. Cell Mol. Genet. 18*:325-336.

Shamay, A., Pursel, V. G., Wilkinson, E., Wall, R. J. and Hennighausen, L. (1992). Expression of the whey acidic protein in transgenic pigs impairs mammary development. *Transgen. Res. 1*:124-132.

Shesely, E. G., Kim, H-S, Shehee, W. R., Papayannopoulou, T., Smithies, O. and Popovich, B. W. (1991). Correction of a human β^s-globin gene by gene targeting. *Proc. Natl. Acad. Sci. USA 88*:4294-4298.

Simons, J. P., McClenaghan, M and Clark, A. J. (1987). Alteration of the quality of milk by expression of sheep β-lactoglobulin in transgenic mouse. *Nature 328*:530-532.

Sippel, A. E., Saueressig, H., Winter, D., Grewal, T., Faust, N., Hecht, A. and Bonifer, C. (1992). The regulatory domain organization of eukaryotic genome: implication for stable gene transfer. In: *Grosveld, F. and Kollias, G. (eds), Transgenic animals, pp. 1-26.* Academic Press Ltd, London.

Slattery, C. W. and Evard, R. (1973). A model for the formation and structure of casein micelles from subunits of variable composition. *Biochim. Biophys. Acta 317*:529-538.

Smith, A. G. and Hooper, M. L. (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Develop. Biol. 121*:1-9.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J. Stahl, M. and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature 336*:688-690.

Smith, A. J. H. and Berg, P. (1984). Homologous recombination between defective *neo* genes in mouse 3T6 cells. *Cold Spring Harbor Symp. Quant. Biol. 49*:171-181.

Smith, C., Meuwissen, T. H. E. and Gibson, J. P. (1987). On the use of transgenes in livestock improvement. *Anim. Breed. Abstr. 55*:1-10.

Smith, L. C. and Wilmut, I. (1989). Influence of nuclear and cytoplasmic activity on the development *in vivo* of sheep embryos after nuclear transplantation. *Biol. Reprod. 40*:1027-1035.

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

Snedecor, G. W. and Cochran, W. G. (1967). *Statistical methods (6th ed).* Iowa State University Press,Iowa.

Song, K-Y, Schwartz, F., Maeda, N., Smithies, O. and Kucherlapati, R. (1987). Accurate modification of a chromosomal plasmid by homologous recombination in human cells. *Proc. Natl. Acad. Sci. USA 84*:6820-6824.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol. 98*:503-517.

Steeg, C. M., Ellis, J. and Bernstein, A. (1990). Introduction of specific point mutations into polymerase II by gene targeting in mouse embryonic stem cells: Evidence for a DNA mismatch repair mechanism. *Proc. Natl. Acad. Sci. USA. 87*:4680-4684.

Stewart, C. L., Kaspar, P., Brunet, L. J., Bhatt, H., Gadi, I., Köntgen, F., Abbondanzo, S. J. (1992). Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature 359*:76-79.

Subramani, S. and Seaton, B. L. (1988). Homologous recombination in mitotically dividing mammalian cells. In: *Kucherlapati, R. and Smith, G. R. (eds), Genetic recombination, pp. 549-573*. ASM, Washington, DC.

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

Sun, H., Treco, D., Schultes, N. P. and Szostak, J. W. (1989). Double-strand breaks at an initiation site for meiotic gene conversion. *Nature 338*:87-90.

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

Te Riele, H., Maandag, E. R. and Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl. Acad. Sci. USA 89*:5128-5132.

Te Riele, H., Maandag, E. R., Clarke, A., Hooper, M. and Berns, A. (1990). Consecutive inactivation of both alleles of the *pim-1* proto-oncogene by homologous recombination in embryonic stem cells. *Nature 348*:649-651.

Teschemacher, H. and Koch, G. (1991). Opioids in milk. *Endoc. Regul. 25*:147-150.

Thépot, D., Devinoy, E., Fontaine, M. L. and Houdebine, L. M. (1991). Structure of the gene encoding rabbit β-casein. *Gene 97*:301-306.

Thomas, K. R. and Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell 51*:503-512.

Thomas, K. R., Deng, C. and Capecchi, M. R. (1992). High-fidelity gene targeting in embryonic stem cells by using sequence replacement vectors. *Mol. Cell. Biol. 12*:2919-2923.

Thomas, K. R., Folger, K. R. and Capecchi M. R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell 44*:419-428.

Thompson, M. D., Dave, J. R. and Nakhasi, H. L. (1985). Molecular cloning of mouse mammary gland κ-casein: comparison with rat κ-casein and rat and human γ-fibrinogen. *DNA 4*:263-271.

Thompson, S., Clarke, A. R, Pow, A. M., Hooper, M. L. and Melton, D. W. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell 56*:313-321.

Threadgill, D. W. and Womack, J. E. (1990). Genomic analysis of the major milk protein genes. *Nucl. Acids Res. 18*:6935-6942.

Tokunaga, T. and Tsunoda, Y. (1992). Efficacious production of viable germ-line chimaeras between embryonic stem (ES) cells and 8-cell stage embryos. *Develop. Growth Differ. 34*:561-566.

Toneguzzo, F., Keating, A., Glynn, S. and McDonald, K. (1988). Electric field-mediated gene transfer: characterization of DNA transfer and pattern of integration in lymphoid cells. *Nucl. Acids Res. 16*:5515-5532.

Tybulewicz, V. L. J., Crawford, C. E., Jackson, P. K., Bronson, R. T. and Mulligan, R. C. (1991). Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-*abl* proto-oncogene. *Cell 65*:1153-1163.

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

Valancius, V. and Smithies, O. (1991b). Testing an 'In-Out' targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. *Mol. Cell. Biol. 11*:1402-1408.

Van den Berg, G., Escher, J. T. M., De Koning, P. J. and Bovenhuis, H. (1992). Genetic polymorphism of κ-casein and β-lactoglobulin in relation to milk composition and processing properties. *Neth. Milk Dairy J. 46*:145-168.

Van Deursen, J. and Wieringa, B. (1992). Targeting of the creatine kinase M gene in embryonic stem cells using isogenic and nonisogenic vectors. *Nucl. Acids Res. 20*:3815-3820.

Velander, W. H., Johnson, J. L., Page, R. L., Russel, C. G., Subramanian, A., Wilkins, T. D., Gwazdauskas, F. C., Pittius, C., Drohan, W. N. (1992). High-level expression of a heterologous protein in the milk of transgenic swine using cDNA encoding human protein C. *Proc. Natl. Acad. Sci. USA 89*:12003-12007.

Vilotte, J-L., Soulier, S., Stinnakre, M-G., Massoud, M. and Mercier, J-C. (1989). Efficient tissue-specific expression of bovine α-lactalbumin in transgenic mice. *Eur. J. Biochem. 186*:43-48.

Vonderhaar, B. K. and Ziska, S. E. (1989). Hormonal regulation of milk protein gene expression. *Annu. Rev. Physiol. 51*:641-652.

Wall, R. J. and Seidel Jr., G. E. (1992). Transgenic farm animal - A critical analysis. *Theriogenology 38*:337-355.

Wall, R. J., Pursel, V. G., Shamay, A., McKnight. R. A., Pittius, C. W. and Hennighausen, L. (1991). High-level synthesis of a heterologous milk protein in the mammary glands of transgenic swine. *Proc. Natl. Acad. Sci. USA 88*:1696-1700.

Whitelaw, C. B. A., Archibald, A. L., Harris, S., McClenaghan, M., Simons, J. P. and Clark, A. J. (1991). Targeting expression to the mammary gland: intronic sequences can enhance the efficiency of gene expression in transgenic mice. *Transg. Res. 1*:3-13.

Whitelaw, C. B. A., Harris, S., McClenaghan, M., Simons, J. P. and Clark, A. J. (1992). Position-independent expression of the ovine β-lactoglobulin gene in transgenic mice. *Biochem. J. 286*:31-39.

Wilde, C. J., Clark, A. J., Kerr, M. A., Knight, C. H., McClenaghan, M. and Simons, J. P. (1992). Mammary development and milk secretion in transgenic mice expressing the sheep β-lactoglobulin gene. *Biochem. J. 284*:717-720.

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature 336*:684-687.

Wilmut, I., Archibald, A. L., Harris, S., McClenaghan, M., Simons, J. P., Whitelaw, C. B. A. and Clark, A. J. (1990). Modification of milk composition. *J. Reprod. Fert., Suppl. 41*:135-146.

Wilmut. I., Archibald, A. L., McClenaghan, M., Simons, J. P., Whitelaw, C. B. A. and Clark, A. J. (1991). Production of pharmaceutical proteins in milk. *Experientia 47*:905-912.

Wood, S. A., Pascoe, W. S., Schmidt, C., Kemler, R., Evans, M. J. and Allen, N. D. (1993). Simple and efficient production of embryonic stem cell-embryo chimeras by coculture. *Proc. Natl. Acad. Sci. USA 90*:4582-4585.

Wright, G., Carver, A., Cottom, D., Reeves, D., Scott, A., Simons, P., Wilmut, I., Garner, I. and Colman, A. (1991). High level expression of active human alpha-1-antitrypsin in the milk of transgenic sheep. *Bio/Technology 9*:830-834.

Yagi, T., Ikawa, Y., Yoshida, K., Shigetani, Y., Takeda, N., Mabuchi, I., Yamamoto, T. and Aizawa, S. (1990). Homologous recombination at c-*fyn* locus of mouse embryonic stem cells with use of diphtheria toxin A-fragment gene in negative selection. *Proc. Natl. Acad. Sci. USA 87*:9918-9922.

Yagita, H., Nakata, M., Kawasaki, A., Shinkai, Y. and Okumura, K. (1992). Role of perforin in lymphocyte-mediated cytolysis. *Adv. Immunol. 51*:215-242.

Yenofsky, R. L., Fine, M. and Pellow, J. W. (1990). A mutant neomycin phosphotransferase II gene reduces the resistance of transformants to antibiotic selection pressure. *Proc. Natl. Acad. Sci. USA 87*:3435-3439.

Yoshimura, M. and Oka, T. (1989). Isolation and structural analysis of the mouse β-casein gene. *Gene 78*: 267-275.

Yoshimura, M., Banerjee, M. R. and Oka, T. (1986). Nucleotide sequence of a cDNA encoding mouse beta casein. *Nucl. Acids Res. 20*:8224.

Yu-Lee, L., Richter-Mann, L., Couch, C. H., Stewart, A. F., Mackinlay, A. G. and Rosen, J. M. (1986). Evolution of the casein multigene family: conserved sequences in the 5' flanking and exon regions. *Nucl. Acids Res. 14*:1883-1902.

Zenvirth, D., Arbel, T., Sherman, A., Goldway, M., Klein, S. and Simchen, G., (1992). Multiple sites for double-strand breaks in whole meiotic chromosomes of *Saccharomyces cerevisiae*. *EMBO J. 11*:3441-3447.

Zheng, H., Hasty, P., Brenneman, M. A., Grompe, M., Gibbs, R. A., Wilson, J. H. and Bradley, A. (1991). Fidelity of targeted recombination in human fibroblasts and murine embryonic stem cells. *Proc. Natl. Acad. Sci. USA 88*:8067-8071.

Zijlstra, M., Li, E., Sajjadi, F., Subramani, S. and Jaenisch, R. (1989). Germ line transmission of a disrupted β_2-microglobulin gene produced by homologous recombination in embryonic stem cells. *Nature 342*:435-438.

Zimmer, A. and Gruss, P. (1989). Production of chimaeric mice containing embryonic stem (ES) cells carrying a homeobox *Hox 1.1* allele mutated by homologous recombination. *Nature 338*:150-153.

APPENDIX

Nucleic Acids Research, 1993, Vol. 21, No. 7 1541–1548

The effects of terminal heterologies on gene targeting by insertion vectors in embryonic stem cells

Satish Kumar and J.Paul Simons*
Department of Molecular Genetics, AFRC Institute of Animal Physiology and Genetics Research, Edinburgh Research Station, Roslin, Midlothian EH25 9PS, UK

Received January 29, 1992; Revised and Accepted March 3, 1993

ABSTRACT

We have examined the effects of placing non-homologous DNA on the ends of an insertion-type gene targeting vector. The presence of terminal heterologies was found to be compatible with insertion targeting, and the terminal heterologies were efficiently removed. Terminal heterologies reduced the frequency of gene targeting to variable extents. The degree of inhibition of targeting was dependent on the length and the position of the heterology: 2.1kb heterologous sequences were more inhibitory than shorter regions of heterology, and heterology placed on the end of the long (4.8kb) arm of homology was more inhibitory than heterology positioned on the end of the short (0.8kb) arm. When heterology was placed on both arms of the targeting vector the targeting efficiencies were similar to or higher than when heterology was present on the long arm only. These results suggest that terminal sequences are removed simultaneously from both ends of targeting vectors. The removal of terminal sequences probably occurs by exonucleolytic degradation of both strands at each end, and removal of at least one of the strands is intimately coupled with the process of homologous recombination. These findings have implications for the design of gene targeting vectors.

INTRODUCTION

DNA introduced into cells can undergo recombination with homologous chromosomal sequences. The modification of chromosomal sequences by homologous recombination is termed gene targeting and is being widely used in embryonic stem (ES) cells for the generation of mice carrying defined mutations (1–3). There are broadly two types of gene targeting: replacement targeting and insertion targeting (1–3). Replacement targeting is used much more widely than insertion targeting, and results in the substitution of chromosomal sequences by sequences from the targeting vector. Insertion targeting on the other hand results in the integration of the targeting vector with no loss of chromosomal sequences, and a consequence of this is the presence of a sequence duplication in the targeted locus. Mutations generated by replacement targeting cannot revert to wild-type; mutations generated by insertion targeting can excise the targeting vector by homologous recombination between the duplicated sequences. Targeting with insertion vectors which carry a mutation within the homologous sequences, followed by selection for excision of the vector results in either complete reversion to wild-type or to reversion with the mutation retained in the chromosome. The usual products of gene targeting experiments are gross disruptions; the two-step insertion-excision method (known as hit-and-run targeting [4], and as in-out targeting [5]) has been shown to be useful for the targeting of subtle mutations, and is likely to be of increasing importance.

There have been many studies of the mechanism of homologous recombination in mammalian cells (see ref. 6 for review). In the majority of experiments the two substrates for the recombination were introduced together; in individual studies, the experiments are usually designed to assess either extrachromosomal recombination (prior to integration) or intrachromosomal recombination (between co-integrated homologous sequences). A comparison of extrachromosomal and intrachromosomal recombination has clearly demonstrated that fundamental differences exist in the mechanisms of recombination in these two experimental systems (7). Extrachromosomal recombination proceeds predominantly by a non-conservative pathway (8–10). Intrachromosomal recombination and gene targeting occur by a conservative pathway (11,12). The currently most favoured model for the mechanism of conservative recombination is double-strand-break (gap) repair (DSBR) (13–16).

Valancius and Smithies (16) found that predictions of the double-strand-break repair model hold for gene targeting in ES cells. Adjacent homologous ends were found not to be required for insertion vector gene targeting in ES cells, and that terminal gaps were always repaired during integration. The presence of small gaps in the targeting vector did not reduce the efficiency of gene targeting, arguing that the same recombination pathway is operating whether gapped or non-gapped targeting vectors are used. Further, it was shown that internal sequence heterologies within either of the arms of the targeting vector were repaired during gene targeting. The efficiency of such repair was very high for heterologies a short distance (0.2kb) from the double strand break, approaching 100%, while heterologies 0.8kb from

* To whom correspondence should be addressed

the double strand break were repaired, but at a significantly lower frequency. The repair of heterologies was unidirectional: the vector-borne heterologies were always corrected by the chromosomal target. Together, the data suggested that the normal pathway of insertion vector gene targeting includes an exonucleolytic mechanism for the removal of terminal sequences from a targeting vector, and for the use of chromosomal target sequences as template to repair gaps introduced by this process.

Chromosomal integration of exogenous DNA in mammalian cells occurs predominantly by non-homologous recombination. In order to minimise the amount of screening required in gene targeting experiments, Mansour et al (17) devised the positive-negative selection (PNS) method to select against cells in which the gene targeting vector had been integrated at random. This method is based on the finding that random integration occurs predominantly via the ends of the exogenous DNA molecule, whereas targeting of a replacement vector results from recombination within the two arms of homology. To adopt the PNS approach, a negatively selectable gene is placed on the end of one or both arms of homology; selection is applied for the positively selectable marker included within the targeting vector and against the negatively selectable gene. Very significant enrichment for gene targeting can result from the use of this method.

In principle, it would be possible to use a PNS approach to the enrichment of gene targeting with insertion vectors. A prerequisite for the application of PNS to insertion vectors is that the presence of a long heterologous DNA tail on one or both ends of the vector should not interfere with gene targeting. In the experiments described here, we have used the properties of the HPRT gene (negatively selectable and hemizygous in male cells) to allow the facile determination of targeting frequencies and isolation of targeted clones. We find that the presence of terminal heterologies does have effects on the frequency of insertion vector targeting of the HPRT gene, and that the effects depend on the position and length of the heterologous sequences. Gene targeting with vectors carrying terminal heterologies is similarly accurate as targeting with the control vector, the terminal heterologies are removed in the majority of targeted clones. We discuss these results in particular with respect to the mechanism of removal of the terminal heterologous sequences.

MATERIALS AND METHODS

Targeting vectors

The mutation in the neomycin resistance gene (18) in the pMC1neo and pMC1neopola (19, Stratagene) cassettes was corrected by exchange of an *Xma*III−*Nco*I fragment with the equivalent fragment form pSV2neo; the corrected plamids, pMC1neo(C) and pMC1neopola(C), gave significantly higher transformation frequencies than the original plasmids. The *Bam*HI site in pMC1neopola(C) was destroyed by filling in, to yield pMC1neopola(C)ΔB.

pIV was constructed by sequential insertion into pGEM4 of a 254bp *Eco*RI−*Sau*3A PCR-derived fragment from intron 2 and exon 3 of the HPRT gene (inserted into *Eco*RI and *Bam*HI-digested vector), the 5.4kb *Eco*RI fragment (intron 1, exon 2, intron 2) of the mouse HPRT gene (20), inserted into the *Eco*RI site, and the *Sal*I−*Xho*I cassette from pMC1neopola(C)DB, inserted into the *Sal*I site. Digestion of pIV with *Bam*HI yielded vector IV.

pIV+ was constructed by linearisation of IV with *Bam*HI, filling in and insertion of an *Xho*I linker (cctcgagg). Insertion of the 2.1kb *Sal*I−*Xho*I HSVtk cassette from pSPTK into the XhoI site of pIV+ yielded pIV+2.1L and pIV+2.1S, which differed in the orientation of the insert. pIV+0.7L and pIV+0.7S were derived from pIV+2.1L and pIV+2.1S respectively by digestion with *Xho*I and *Sma*I, filling in and religation. pIV+0.33L and pIV+0.33S were derived by insertion into the XhoI site of pIV+ in either orientation of a 334bp *Sal*I−*Xho*I fragment from pMC1neoD, itself derived from pMC1neopola(C) by deletion of the *Xma*III−*Bam*HI fragment. pIV+0.19L and pIV+0.19S were obtained by insertion into the XhoI site of pIV+ in either orientation of a 194bp *Sal*I−*Xho*I fragment which contains the enhancer from pMC1neo. Digestion of these plasmids with *Xho*I gave the corresponding targeting vector (e.g. pIV+2.1S×*Xho*I = vector IV2.1S); pIV+2.1L when digested with *Sma*I, yielded targeting vector IV+0.7L1.4S. For experiments H and I (table 1) the targeting vectors were all derived from pIV+2.1L: Vector IV was pIV+2.1L×*Bam*HI, IV+2.1L was pIV+2.1L×*Xho*I; IV+0.7L was pIV+2.1L×*Xho*I + *Sma*I and IV+0.7L1.4S was pIV+2.1L×*Sma*I.

The control vector (WAPneoTK) is a replacement vector designed for targeting of the whey acidic protein (WAP) gene which contains an unique *Xho*I site, and comprises the pSP72 vector, 6kb of WAP sequences, the neo cassette from pMC1neopola(C), the HSVtk cassette from pSPTK.

Cell culture and electroporation

E14 embryonic stem cells (21) were cultured in gelatin-treated flasks or petri dishes in the absence of feeder cells in DMEM (high glucose formulation, Life technologies), supplemented with NEAA, 0.1mM β-mercaptoethanol, 5% FCS and 5% NBCS, and recombinant DIA/LIF (22). For electroporation, plasmids were digested with the appropriate restriction enzymes and purified by phenol extraction, ethanol precipitation and washing with ethanol. Growing cells (passage 24 to 40) were harvested by trypsinisation and resuspended at 10^8 ml^{-1} in PBS with DNA at a concentration of 5nM. Each electroporation consisted of 0.8 ml cells which were electroporated using a Bio-Rad Gene Pulser at 500μF, 230V (575 Vcm^{-1}) (experiments A−F) or 3μF, 800V (2000 Vcm^{-1}) (experiments G−J). After 10 minutes at room temperature, the cells were diluted to 10^6 ml^{-1} in complete medium and 5×10^6 cells were plated per 6cm diameter petri dish (in experiment C, 3.2×10^6 cells were plated in each dish). G418 selection was either applied after 2 days (0.5mg Geneticin ml^{-1} [~0.25mg/ml active G418], experiments A−F) or after 24 hours (0.3mg Geneticin ml^{-1} [~0.15mg/ml active G418], experiments G−J). Six days after electroporation, 2.0mM 6-TG selection was applied to 80% of the plates, plates were fixed and stained for colony counts after a further four days (G418-selected) or after a further eight days (G418- and 6-TG-selected). When required, colonies were picked immediately prior to fixation from independent plates, and expanded for preparation of DNA.

Southern blot analysis

DNA was isolated, digested and electrophoresed according to standard methods and blotted onto HybondN (Amersham). Probes were labelled by random priming, and hybridised at 65°C by the method of Church and Gilbert (23); post-hybridisation washes were performed with 40mM sodium phosphate pH7.2, 1% sds,

1mM EDTA at 65°C. The probes used were a 339bp *Hinc*II−*Hpa*II fragment of mouse HPRT cDNA clone pHPT5 (24) which is essentially specific for exons 2 and 3 (20), linearised pGEM4-neo which contains the neo gene from pMC1neo(C) in pGEM4, and a 1.5kb *Eco*RV HSV-tk fragment from pSPTK.

RESULTS

Insertion vector targeting of the mouse HPRT gene

An insertion type targeting vector, designated IV, was constructed for targeting of the mouse HPRT gene. This vector contains approximately 5.6kb of (non-isogenic) HPRT sequences, the neo^r gene from pMC1neoPolA(C) and the plasmid pGEM4 (see Figure 1). There is an unique *Bam*HI site 0.8kb from the 5′ end of the HPRT sequences; linearisation at this site yields an insertion vector with asymmetric homologous arms: a long arm of 4.8kb and a short arm of 0.8kb. The structure predicted following targeted integration of IV is shown in figure 1. The region of homology in the unmodified endogenous HPRT gene lies within an 8.0kb *Pst*I fragment. Following targeted integration of IV, the absence of the 8.0kb wild-type *Pst*I fragment and new 10.4kb and 6.4kb *Pst*I fragments diagnostic of targeting are predicted. The 6.4kb fragment contains the 5′ recombination junction and the 10.4kb fragment contains the 3′ recombination junction; these fragments are derived from recombination with the target locus of the long and the short arms (respectively) of the targeting vector.

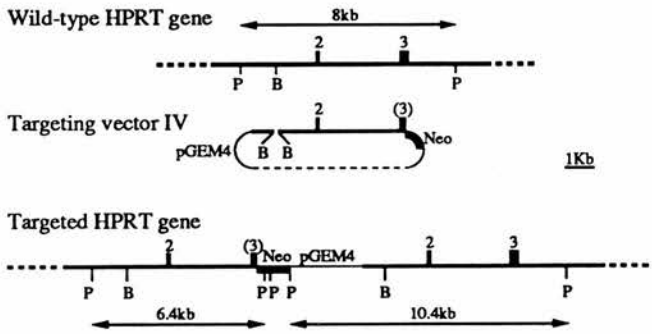


Figure 1. Targeting of the HPRT gene with insertion vector IV. Integration of IV into the HPRT gene by homologous recombination results in the insertion of the entire vector, with the loss of the wild-type 8.0Kb *Pst*I fragment, and new fragments of 6.4kb and 10.4kb. Thick lines: HPRT sequences; boxes numbered 2, 3 and (3): HPRT exons 2, 3 and incomplete exon 3. Boxes labelled Neo: Neo^r gene; thin line, pGEM4 plasmid vector sequences. P: *Pst*I sites; B: *Bam*HI sites.

Cells of the male ES cell line E14 were electroporated with IV, and with the control vector WAPneoTK, which contains no HPRT sequences and which yields G418-resistant colonies at a similar frequency to IV. G418 and 6-thioguanine (6-TG) selection were applied to select transformed cells which did not express HPRT. In each experiment, IV yielded numerous 6-TG-resistant colonies whereas WAPneoTK yielded at most one colony per experiment (Table 1), strongly suggestive of HPRT targeting by IV. To confirm that the 6-TG resistant colonies obtained following electroporation with IV were HPRT-deficient due to targeting, representative colonies were picked and expanded for Southern blot analysis. None of the thirteen clones analysed gave the 8.0kb *Pst*I fragment as found in the wild-type HPRT gene after probing with HPRT cDNA; twelve gave the 10.4 and 6.4kb *Pst*I fragments (see Figure 2), as predicted if targeted integration had occurred. One of the clones yielded, in addition to the predicted fragments, an extra ~8.7kb fragment (Figure 2, lane 5), consistent with targeted integration of multiple copies of the targeting vector in a head-to-tail tandem array (Figure 3A). The

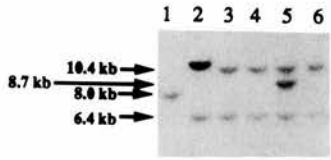


Figure 2. Southern blot analysis of clones targeted with IV. *Pst*I-digested DNA from representative clones hybridised with an HPRT cDNA probe. Lane 1: wild-type E14 DNA; lanes 2−6: DNA from G418- and 6-TG-resistant clones. Eight further clones gave patterns indistinguishable from those in lanes 3, 4 and 6. The 10.4kb and 6.4kb fragments are diagnostic of gene targeting; the 8.7kb fragment indicates the presence of a head-to-tail repeat of the targeting vector (see Figure 3A).

Table 1. Targeting frequencies

Experiment	Relative targeting efficiency (No. of colonies) A	B	C	D	E	F	G	H	I	J	
Vector											Comments
IV	1.0 (65)	1.0 (50)	1.0 (14¶)	1.0 (43¶)	1.0 (21)	1.0 (154)	1.0 (112)	1.0 (272)	1.0 (179)	1.0 (108)	No heterology: positive control
IV+	0.98 (64)									0.88(95)	*Xho*I linker
IV+2.1L	0.28 (18)	0.36 (18)	0.29 (4¶)	0.28 (12¶)				0.33 (89)	0.36 (65)		2.1kb on long arm
IV+2.1S		0.64 (32)	0.71 (10¶)	0.72 (31¶)							2.1kb on short arm
IV+0.7L		0.40 (20)						0.59 (160)	0.57 (102)	0.55 (59)	0.7kb on long arm
IV+0.7S		0.70 (35)								0.83 (90)	0.7kb on short arm
IV+0.33L				0.57 (21)						0.30 (32)	0.33kb on long arm
IV+0.33S				1.03 (38)						0.53(57)	0.33kb on short arm
IV+0.19L					0.57 (12)		1.02 (114)				0.19kb on long arm
IV+0.19S					0.76 (16)		1.16 (130)				0.19kb on short arm
IV+0.7L1.4S				1.03 (38)	1.10 (23)	0.85 (131)		0.59 (161)	0.54 (97)		0.7kb on long arm, 1.4kb on short arm
WAPneoTK	0.02 (1)	<0.02 (0)	<0.072 (0)	<0.03 (0)	<0.05 (0)	<0.007 (0)	0.009 (1)	0.004 (1)	<0.006 (0)	0.009 (1)	Negative control

Summary of the relative targeting efficiencies with each vector. For each experiment (A−J), the targeting frequencies were normalised to the positive control vector IV (relative targeting efficiency = 1.0), and were calculated from the total number of colonies resistant to G418 and 6-TG on 12 petri dishes (except ¶: 14 petri dishes). Each petri dish received 5×10^6 electroporated cells, except in experiment C in which 3.2×10^6 cells were plated per dish. The figures for the negative control vector represent frequencies of spontaneous HPRT$^-$ mutations, and are not targeting frequencies.

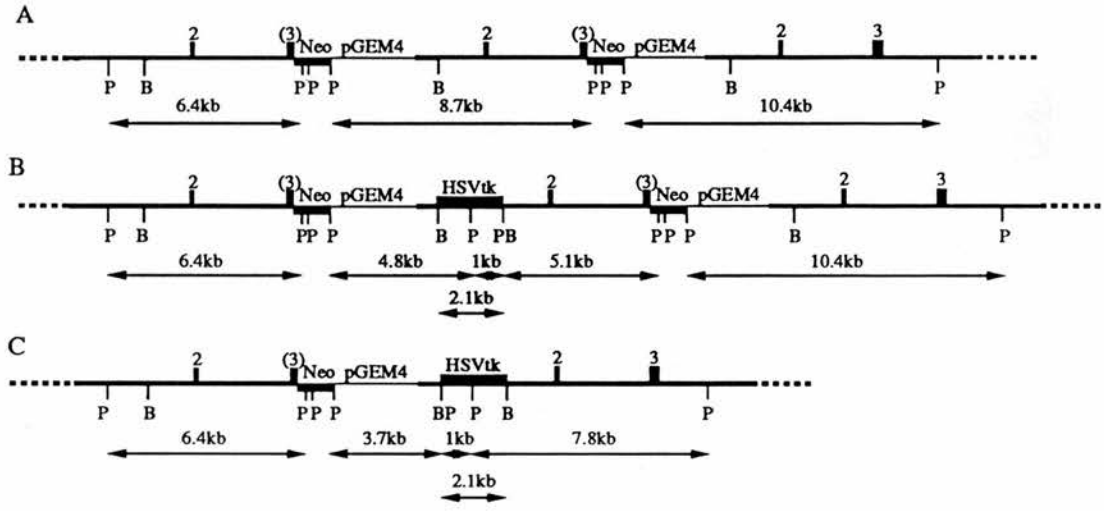


Figure 3. Structures of targeted loci in abnormally targeted clones. **A)** Structure of the targeted HPRT gene in one clone, abnormally targeted with vector IV (see Figure 2, lane 5). Tandemly repeated (Head-to-tail) copies of the vector were integrated (two copies are shown here). **B)** Structure of the targeted HPRT gene in one clone, abnormally targeted with vector IV+2.1L (see Figure 5A, lane3 and 5B, lanes 1 and 3). Tandemly repeated (Head-to-tail) copies of the vector were integrated; the terminal heterology (HSVtk) was retained internally, and removed from the end copy of the targeting vector. **C)** Structure of the targeted HPRT gene in two clones abnormally targeted with vector IV+2.1S (see Figure 5C, lanes 1 and 5; 5D, lanes 1−4). Targeting ocurred by homologous recombination within the long arm, and the HSVtk terminal heterology was retained. A single copy of the targeting vector was integrated in one clone, tandem head-to-tail repeats (not shown) of the targeting vector were integrated in the other.

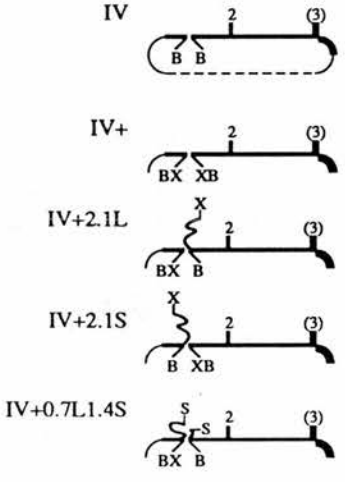


Figure 4. Targeting vectors with terminal heterology. The vectors illustrated carry: no heterology (IV); *Xho*I linker heterology on both ends (IV+); 2.1kb of heterology on the long arm and *Xho*I linker heterology on the short arm (IV+2.1L); 2.1kb of heterology on the short arm and *Xho*I linker heterology on the long arm (IV+2.1S); and 0.7kb of heterology on the long arm and 1.4kb of heterology on the long arm (IV+0.7L1.4S). The vectors which are not shown here are similar to IV+2.1L and IV+2.1S, with shorter regions of terminal heterology. Thick straight lines: HPRT sequences; boxes labelled 2 and (3): HPRT exon 2 and incomplete exon 3; thin lines: pGEM4 sequences; unlabelled boxes neo^r gene; wavy lines: terminal heterologies; B: *Bam*HI; X: *Xho*I; S: *Sma*I.

intensity of the 8.7kb fragment is higher than either the 6.4 or 10.4kb fragments, suggesting that at least three copies of the targeting vector were integrated into the HPRT gene in this clone. The remaining clone gave a fragment of 6.4kb, indicative of homologous recombination of the long arm of the vector; the 10.4kb band was absent, and a fragment of approximately 11kb was observed (Figure 2, lane 2); the intensity of this fragment indicates the integration of multiple copies of the targeting vector. Although this clone is the product of a complex integration event and the structure cannot be precisely determined, the absence of the 8.0kb fragment and the presence of the 6.4kb fragment make it likely that the HPRT gene in this clone underwent one-sided homologous recombination, being targeted by the long arm of the vector.

To confirm this interpretation of the results, the blots were reprobed with pGEM4neo, a plasmid containing all of the sequences in IV other than those derived from the HPRT gene (data not shown). As expected, in the twelve clones which gave the 10.4kb fragment when probed with HPRT cDNA, this fragment also hybridised strongly using pGEM4neo as probe. In the two clones which gave fragments which were not predicted, these fragments also hybridised to pGEM4neo, as expected if the interpretation is correct. Although only eleven out of the thirteen clones were targeted by simple insertion, the two other clones were both targeted.

These results demonstrate that the targeting vector IV integrates into the HPRT gene by homologous recombination, and show that the frequency of 6-TG-resistance is a good approximation to the targeting frequency.

The effects of terminal heterologies on insertion vector targeting efficiencies

To examine the effects of terminal heterologies on the efficiency of insertion vector targeting, a series of derivatives of IV were made. In addition to the sequences present in IV, these vectors contain inserts at the *Bam*HI site; varying from 12nt (*Xho*I linker) to 2.1kb (HSV1-tk). By digestion with *Xho*I, these vectors can be opened to leave very short heterology on both arms (vector IV+), or short heterology on one arm and long heterology on the other (see Figure 4). The designation of these vectors indicates

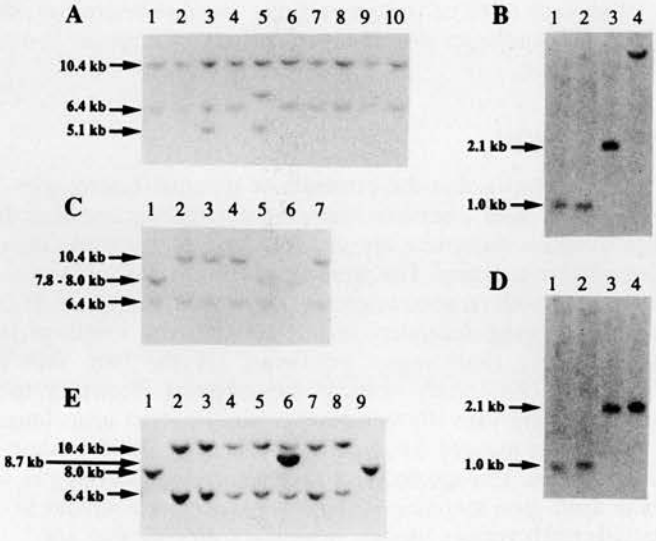


Figure 5. Southern blot analysis of clones targeted with vectors carrying terminal heterology. **A)** Clones derived from electroporation with IV+2.1L. Lanes 1−10: *Pst*I-digested DNA from representative clones hybridised with an HPRT cDNA probe. The 10.4kb and 6.4kb fragments are diagnostic of gene targeting; the 5.1kb fragment indicates retention of the terminal heterology (see Figure 3B). **B)** Confirmation of the presence of the terminal heterology in two clones derived from electroporation with IV+2.1L. Lanes 1 and 3: DNA from the clone shown in Figure 5A, lane 3; lanes 2 and 4: DNA from the clone shown in Figure 5A, lane 5. Lanes 1 and 2: *Pst*I digests; lanes 2 and 4: *Bam*HI digests. Probed for HSVtk sequences. **C)** Clones derived from electroporation with IV+2.1S. *Pst*I-digested DNA from representative clones hybridised with an HPRT cDNA probe. Lane 6: wild-type E14 cell DNA; lanes 1−5 and 7: G418- and 6-TG-resistant clones. The 10.4kb and 6.4kb fragments are diagnostic of gene targeting; the wild-type gene yields a fragment of 8.0kb; head-to-tail integration of the targeting vector without removal of the terminal heterology gives a 7.8kb fragment (see Figure 3C). **D)** Confirmation of the presence of the terminal heterology in two clones derived from electroporation with IV+2.1S. Lanes 1 and 3: DNA from the clone shown in lane 1 of Figure 5C; lanes 2 and 4: DNA from the clone shown in Figure 5C, lane 5. Lanes 1 and 2: *Pst*I digests; lanes 2 and 4: *Bam*HI digests. Probed for HSVtk sequences. **E)** Clones derived from electroporation with IV+0.7L1.4S. *Pst*I-digested DNA from representative clones hybridised with an HPRT cDNA probe. Lanes 1 and 9: wild-type E14 DNA; lanes 2−8: DNA from G418- and 6-TG-resistant clones. Sixteen further clones gave patterns indistinguishable from those in lanes 2−5, 7 and 8. The 10.4kb and 6.4kb fragments are diagnostic of gene targeting with removal of terminal heterology from both arms; the 8.7kb fragment in lane 6 indicates the presence of a head-to-tail repeat of the targeting vector (see Figure 3A).

the length and position of the heterology (e.g. IV+2.1L carries 2.1kb of heterology on the long arm).

The frequency of targeting into HPRT with these vectors was measured and the data are summarised in table 1. In common with others (25−27) we observed significant variability between experiments of absolute targeting efficiency with any given vector. Targeting efficiencies are consistent within experiments and therefore we have normalised the results relative to a positive control (IV) included in each experiment. Targeting efficiencies are expressed as a ratio of targeting frequency obtained with each vector to that obtained with the IV control.

Two experiments were performed with vector IV+; the short (one base pair and four nucleotides) linker heterologies which were present on both arms of the vector had little or no effect on the targeting frequency (Table 1, experiments A & J). In contrast, long terminal heterology consistently reduced the efficiency of gene targeting (Table 1). This effect was most pronounced with the vector carrying 2.1kb of heterologous sequences on the long arm (IV+2.1L): in six experiments, the relative targeting efficiencies ranged between 0.28 and 0.36. Suppression of gene targeting by 2.1kb of heterology on the short arm (vector IV+2.1S) was also observed: in each of three experiments the relative targeting efficiencies ranged from 0.64 to 0.72. Similarly, in each of six experiments with shorter regions of terminal heterology (0.7kb, 0.33kb or 0.19kb), the relative targeting efficiency was higher when the heterology was on the short arm than when it was on the long arm. The data (Table 1) also suggest a possible effect of the length of the heterologous sequences on the degree of suppression of gene targeting: the shorter heterologies suppressing targeting to a lesser degree.

The effect on targeting efficiency of long heterologies on both arms

Given that long terminal heterology on either arm of the targeting vector reduces the targeting efficiency, it may be expected that this effect would be compounded by the presence of heterology on both arms. To examine this, experiments were performed using a vector carrying 0.7kb of heterologous sequences on the long arm and 1.4kb on the short arm (IV+0.7L1.4S). The relative targeting efficiencies in five experiments ranged from 0.54 to 1.10 (Table 1). This contrasts with the relative targeting efficiencies with 0.7kb of heterology on the long arm, which ranged between 0.40 and 0.59 (four experiments). In four of the experiments, other vectors were assessed in parallel. Without exception, vectors which carried 0.7kb or less heterology on the long arm gave targeting efficiencies which were similar to or lower than IV+0.7L1.4S (Table 1). Surprisingly, these data show that when there is 0.7kb of terminal heterology on the long arm, targeting is insensitive to, or is enhanced by the presence of 1.4kb of heterology on the short arm.

Terminal heterologies and the fidelity of gene targeting

Southern blot analysis was performed to determine the structures of the HPRT gene in representative clones targeted with the vectors IV+2.1L, IV+2.1S and IV+0.7L1.4S.

Of the ten clones analysed from targeting with IV+2.1L none gave the wild-type 8.0kb *Pst*I fragment, and nine gave 10.4 and 6.4kb *Pst*I fragments when probed with HPRT cDNA (Figure 5A), as predicted if targeting was accompanied by removal of the terminal heterologous DNA. One of the nine clones gave an additional band of approximately 5.1kb (Figure 5A, lane 3); the presence of three hybridising fragments indicates the integration of at least two copies of the targeting vector. A *Pst*I fragment of 5.1kb is expected if the heterologous HSV-tk sequences were retained. *Pst*I- and *Bam*HI-digested DNAs were probed for 3′ HSV-tk sequences; respectively, 1kb and 2.1kb fragments were detected (Figure 5B, lanes 1 and 3), confirming the incorporation of the heterology. A 2.1kb *Bam*HI fragment is predicted on recircularisation or head-to-tail ligation of the targeting vector. A further prediction of recircularisation or head-to-tail ligation is a 4.8kb *Pst*I fragment containing pGEM4 sequences; a band of this size was obtained following probing with pGEM4-neo (data not shown). Together, these results suggest that recombination has occurred either by insertion of a head-to-tail vector multimer with removal of the heterology from the end copy, or by replacement recombination of a head-to-tail multimer. The probable structure of the HPRT locus in this clone is shown in Figure 3B.

The tenth clone, when probed with HPRT cDNA (Figure 5A, lane 5) gave the predicted 10.4kb *Pst*I fragment, indicative of homologous recombination of the short arm of the targeting vector, and two other fragments of 7.4 and 5.1kb. These data again indicate the involvement of more than one copy of the targeting vector and incorporation of HSV-tk sequences. Probing of *Pst*I- and *Bam*HI-digested DNAs confirmed the presence of HSV-tk sequences. The internal 1kb *Pst*I fragment was detected (Figure 5B, lane 2); with *Bam*HI, a large fragment was detected (Figure 5B, lane 4), the origin of which cannot be accounted for by a simple vector ligation event. The short-arm origin of the 10.4kb *Pst*I fragment was confirmed by hybridisation with pGEM4-neo (data not shown). This clone is thus the product of an imprecise targeting event, and the precise structure of the modified HPRT locus is not clear.

Six clones from targeting with IV+2.1S were analysed (Figure 5C). Four of the clones gave both the predicted 6.4kb and 10.4kb *Pst*I fragments when probed with the HPRT cDNA probe, demonstrating that these clones were targeted and that the terminal heterology was removed. The remaining two clones gave the 6.4kb fragment indicative of targeting via the long arm and a fragment of approximately 8.0kb (Figure 5C, lanes 1 and 5). The presence of the wild-type 8.0kb *Pst*I fragment is incompatible with gene targeting which is indicated by the presence of the 6.4kb fragment. This inconsistency can be reconciled if recircularisation or head-to-tail ligation of this targeting vector occurred. This would be predicted to give a fragment of 7.8kb; targeting via the long arm of such an intermediate would yield *Pst*I fragments of 6.4kb and ~8kb, with integration of the HSVtk terminal heterology (Figure 3C). Probing for HSV-tk sequences of *Pst*I- and *Bam*HI-digested DNAs gave fragments of 1kb and 2.1kb (Figure 5D), as predicted by this explanation. This explanation is further supported by the presence of a 3.7kb *Pst*I fragment which hybridises to pGEM4-neo (data not shown); a fragment of this size is predicted by the above explanation, but because this size of fragment would also be obtained following random integration of IV+2.1S, this support is weak. The intensity of the 7.8kb *Pst*I band in one clone is higher (Figure 5C, lane 1), indicating that this band is derived from multiple copies of the vector. Of the six IV+2.1S clones analysed, four clones were targeted with removal of the terminal heterology. The other two clones were also targeted, but with retention of the terminal heterology. In these two clones recombination occurred within the long arm of the vector, and in one it appears that multiple copies of the targeting vector were integrated in a head-to-tail tandem array.

Twenty three G418- and 6-TG-resistant clones obtained with vector IV+0.7L1.4S were analysed. None gave the 8.0kb wild-type *Pst*I fragment when probed with the HPRT cDNA probe. Remarkably, all of these clones gave the 6.4 and 10.4kb *Pst*I fragments (see Figure 5E); one clone gave a single intense additional band of approximately 8.7kb (Figure 5E, lane 6). When DNA from this clone was analysed further, no hybridisation was detected in *Pst*I- and *Bam*HI-digested DNA probed for HSV-tk (data not shown), showing that the terminal heterology had been removed. Taken together, these data suggest that the HPRT gene in this clone carries multiple copies of the targeting vector from which the terminal heterology has been removed (see Figure 3A). This interpretation of the data was confirmed by hybridisation with pGEM4neo (data not shown). All of the IV+0.7L1.4S clones were targeted with removal of the terminal heterology: twenty two carry a single copy of the targeting vector and one carries multiple copies.

With each of three vectors carrying terminal heterology, the terminal heterologies were removed during gene targeting in most or all clones.

DISCUSSION

We have shown that the presence of terminal heterologies is compatible with insertion vector gene targeting and that the heterologous sequences are usually removed from the vector during gene targeting. The presence of terminal heterologies was associated with reduced targeting frequencies, the extent of the reduction being dependent on the position and length of the heterologies. Heterologies positioned on the long arm of homology consistently reduced the targeting frequency to a greater extent than did heterologies on the short arm; longer heterologies reduced the targeting frequency more than shorter heterologies. Unexpectedly, a vector carrying heterologies on both arms gave targeting frequencies which were similar to or greater than vectors bearing heterology on only one arm.

Hasty et al (25) targeted the HPRT gene in mouse ES cells with insertion vectors possessing a 5.6kb long arm of homology and a 1.2kb short arm; one of the vectors carried 2.3kb of heterologous sequences on the long arm of homology. The heterology reduced targeting efficiency in each of four experiments, giving relative targeting efficiencies from 0.13 to 0.46. The terminal heterology was removed in most clones. The data of Hasty et al (25) are consistent with the results we present here.

Gene targeting by a single targeting vector can occur by different pathways, and this can result in different structures of the targeted gene. For example, one targeting vector designed for replacement type gene targeting was found to undergo the predicted simple replacement targeting in only a minority of clones (28). The structures of most of the targeted genes suggested either replacement targeting by a head-to-tail multimer of the original vector, or insertion type recombination of a recircularised vector. With an insertion vector such as IV, it is not possible to distinguish between simple insertion targeting, recircularisation and insertion targeting or replacement recombination between the target gene and adjacent copies of IV in a multimer, because all three pathways would yield the same structure. Targeting by vectors carrying terminal heterology by either of the latter two possible pathways would however be detectable by the absence of one of the junction fragments and the retention of the terminal heterologous sequences. In our experiments, the majority of clones had undergone a simple insertion whether the vector carried heterology on the long arm, the short arm or both arms, showing that simple insertion targeting with removal of terminal heterologies is more efficient than targeting by either of the above alternative pathways.

Although we have shown that long terminal heterologies reduced the frequencies of gene targeting, the degree of suppression was not very great: at most 3.6-fold. This, together with the efficient removal of the terminal heterologies argues that targeting proceeds by essentially the same pathway of recombination whether or not the vector ends carry terminal heterologies. A prediction of this argument is that the pathway of insertion vector targeting must include the means for removal of terminal heterologies.

How are the terminal heterologies removed? A trivial explanation would be that random physical or endonuclease-induced double strand DNA breaks separate the terminal heterologous sequences from the homology. This is unlikely for

the following reasons. Firstly, frequent random double strand breaks would lead to a low frequency of incorporation of full length molecules; randomly integrated DNA is usually integrated without suffering deletions (29−33). Secondly, targeting efficiencies are significantly increased following the introduction of double strand breaks into the targeting vector (16,25); if random double strand breaks were common, pretreatment of targeting vectors would have little or no effect on targeting frequencies. Because both strands of the terminal heterologies must be removed, there are three alternative explanations for the removal of heterologies: exonuclease digestion of both strands, non-random double-strand endonuclease cleavage, and single-strand exonuclease digestion with non-random single-strand endonuclease cleavage of the other strand.

Are these alternative mechanisms compatible with the effects we observed on targeting frequency? Non-random endonuclease digestion could act before the homology search at preferred sites within the targeting vector, or alternatively endonuclease could act after the homology search, recognising the presence of terminal heterology. For the first explanation to be valid, there would have to be at least one preferred recognition site for the endonuclease within each arm of the targeting vector, and cleavage of both sites would have to be efficient; this is unlikely. Further, a prediction of this mechanism is that with the vectors used, the heterology would be removed efficiently in non-targeted clones. We have found that the majority of randomly integrated copies retain the terminal heterologies (data not shown), arguing strongly against this mechanism. This argument applies for both double-strand endonuclease, and single-strand exonuclease with single-strand endonuclease, and effectively rules out the possibility of non-random endonuclease cleavage prior to the homology search. There is no simple way to reconcile double strand endonuclease recognition and removal of terminal heterologies after the homology search with the observed differential effects of heterologies on targeting frequencies.

Removal of both strands of terminal heterologies by exonuclease(s) is compatible with the effects we observed of heterologies on targeting efficiency if exonuclease digestion occurs concurrently from both ends of targeting vectors. Removal of terminal heterology from one arm of the vector would be accompanied by removal of sequences from the other arm. If for example we take the case of vector IV+2.1L, which carries 2.1kb of terminal heterology on the long arm, then if exonuclease digestion from the two ends is precisely synchronous, the 0.8kb short arm of homology would be completely removed before complete removal of the 2.1kb heterologous sequences, and this would have profound effects on targeting. Complete removal of one arm of homology is not compatible with the structures of targeted loci that were predominantly obtained. Partial removal of homologous sequences would lead to a gap within the region of homology: it is known that such gaps are repaired efficiently during gene targeting (16). Concurrent but not precisely synchronous exonuclease digestion from both ends could explain the reduced targeting frequency observed with this vector. If we consider the case of vectors carrying terminal heterology on the short arm, removal of the heterology would be accompanied in general by incomplete removal of homologous sequences from the long arm, and therefore a smaller reduction in targeting frequency. Again, this accords with the results we obtained. The effect of terminal heterology on both arms of a targeting vector would be to reduce the extent of degradation of sequences from the arms of homology when compared with the same amount of total heterology positioned on a single arm. Another way of looking at this is that, given terminal heterology on one arm of a targeting vector, heterology on the other arm would be protective against removal of the homologous sequences. Again, this is entirely consistent with the effects we observed on targeting frequencies. All of the observed effects of length and position of terminal heterologies on targeting frequencies can be accounted for by removal of both strands of the heterologous sequences by exonuclease(s).

The exonuclease digestion would have to be coupled with recombination for this mechanism to be compatible with the efficient removal of terminal heterologies during gene targeting and with the usual retention of terminal sequences after random integration. This coupling could be by exonuclease digestion occurring after the homology search. Alternatively, exonuclease digestion could proceed until a homologous target is found; failure to find homology would result in complete degradation of the vector. A requirement of the homology search preceding exonuclease digestion is that free homologous ends are not required for the search. Recombinase activity in human cell extracts has been found to require free homologous ends (34) arguing against this possibility although it remains possible that other recombinase activities exist which do not require free ends. A second activity which promotes DNA strand exchange has been identified and purified (homologous pairing protein-1 or HPP-1 [35]), which possesses 3′→5′ exonuclease activity in addition to its pairing activity (36). An activity like HPP-1 could couple the removal of heterology with the homology search.

Let us now consider the third possible mechanism for removal of terminal heterologies: single-strand exonuclease digestion followed by non-random single-strand endonuclease cleavage. Random removal of one strand by single strand exonuclease activity is compatible with randomly integrated DNA retaining terminal sequences, because ligation of single-stranded vector ends to chromosomal DNA could occur followed by filling in of the single strand gap(s). The argument used above, of simultaneous exonuclease digestion from both ends of targeting vectors could apply, but this would not explain the effects on targeting efficiencies with the different vectors, because homologous sequences would remain on the second strand. If however, for efficient targeting there is a requirement for a region of double stranded DNA within both arms of homology, this mechanism would fit the data. Complete removal of one strand from the one arm of homology would preclude the formation of a Holliday junction by that arm, as predicted by the DSBR model of recombination (13); the absence of a Holliday junction however need not be incompatible with recombination. Removal of a single strand by 3′→5′ exonuclease would lead to a single stranded 5′ end. Although according to the Meselson-Radding model of recombination, the target sequence would be invaded by a 5′ end, this is incompatible with gap repair, which we invoke and which has been shown to occur (16). Thus, if one strand is removed by exonuclease and the other by endonuclease, the polarity of the exonuclease would have to be 5′→3′.

The findings of Valancius and Smithies (16) that internal heterologies are frequently lost during gene targeting and that the heterologies are always corrected to the chromosomal sequence suggested that the terminal sequences are subjected to exonucleolytic degradation during gene targeting. The loss of terminal homologous (and internal heterologous) sequences is likely to occur by the same mechanism as the removal of terminal heterologies. The unidirectional correction of internal heterologies (16) argues against the presence of the heterologies within regions of heteroduplex DNA, because repair or replication of heterology-

containing heteroduplexes would give bidirectional correction. This argument would be invalidated if repair of heteroduplexes is highly efficient and intrinsically directional. The removal of terminal (homologous or heterologous) sequences from insertion-type targeting vectors is thus probably due to exonucleolytic digestion of both DNA strands.

The results of our experiments suggest that PNS could be of use for insertion vector gene targeting. Even if gene targeting is suppressed by terminal heterologies, a level of enrichment significantly greater than the degree of suppression would suffice to increase the proportion of targeted clones amongst those analysed. Although we have used only one target locus, the results suggest guidelines which could be of use in the design of insertion vectors intended for PNS: the heterology should be as short as possible, and if the arms are of different sizes, the heterology should be positioned on either the shorter arm or on both arms. In preliminary experiments however, we obtained only poor enrichment with vectors IV+2.1L and IV+2.1S when ganciclovir selection was applied. We do not understand the reason for the low level of enrichment, but this result is consistent with the results of others (37–39) and with our experience with replacement vectors.

Finally, the hit-and-run/in-out method (4,5) for the generation of subtle mutations may be complicated by the unwanted correction of the subtle mutation during the insertion step (16). If this proves to be a problem, positioning terminal heterology on the arm of the vector which carries the mutation could be useful to reduce the frequency of the correction.

ACKNOWLEDGEMENTS

We would like to thank John Clark for support, Martin Hooper for E14 cells and for comments on the manuscript, David Melton for HPRT clones, Austin Smith for DIA/LIF and a DIA/LIF expression vector; Ray Ansell for technical assistance, Anthea Springbett for advice and Norrie Russell for photography. This work was supported in part by AFRC grant TAP2B. SK was supported by an Indian National Scholarship and an ORS award from the CVCP.

REFERENCES

1. Bradley,A., Hasty,P., Davis,A. and Ramirez-Solis,R. (1992) *Bio/Technology*, **10**, 534–539.
2. Capecchi,M.R. (1989) *Science*, **244**, 1288–1292.
3. Koller,B.H. and Smithies,O. (1992) *Ann. Rev. Immunol.*, 10, 705–730.
4. Hasty,P., Ramírez-Solis,R., Krumlauf,R. and Bradley,A. (1991) *Nature*, **350**, 243–246.
5. Valancius,V. and Smithies,O. (1991) *Mol. Cell. Biol.*, 11, 1402–1408.
6. Bollag,R.J., Waldman,A.S. and Liskay,R.M. (1989) *Ann. Rev. Genet.*, **23**, 199–225.
7. Waldman,A.S. and Liskay,R.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5340–5342.
8. Lin, F.-L.M., Sperle,K. and Sternberg,N. (1990) *Mol. Cell. Biol.*, **10**, 103–112.
9. Seidman,M.M. (1987) *Mol. Cell. Biol.*, **7**, 3561–3565.
10. Lin, F.-L., Sperle,K. and Sternberg,N. (1984) *Mol. Cell. Biol.*, **4**, 1020–1034.
11. Bollag,R.J. and Liskay,R.M. (1988) *Genetics*, **119**, 161–169.
12. Pennington,S.L. and Wilson,J.H. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9498–9502.
13. Szostak,J.W., Orr-Weaver,T.L., Rothstein,R.J. and Stahl,F.W. (1983) *Cell*, **33**, 25–35.
14. Sun,H., Treco,D., and Szostak,J.W. (1991) *Cell*, **64**, 1155–1161.
15. Zenvirth,D., Arbel,T., Sherman,A., Goldway,M., Klein,S. and Simchen,G. (1992) *EMBO J.*, **11**, 3441–3447.
16. Valancius,V. and Smithies,O. (1991) *Mol. Cell. Biol.*, 11, 4389–4397.
17. Mansour,S.L., Thomas,K.R. and Capecchi, M.R. (1988) *Nature*, 336, 348–352.
18. Yenofsky,R.L., Fine,M. and Pellow,J.W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3435–3439.
19. Thomas,K.R. and Capecchi,M.R. (1987) *Cell*, **51**, 503–512.
20. Melton,D.W., Konecki,D.S., Brennand,J. and Caskey, C.T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2147–2151.
21. Handyside,A.H., O'Neill,G.T., Jones,M. and Hooper, M.L. (1989) *Roux's Arch. Dev. Biol.*, **198**, 48–55.
22. Smith,A.G., Heath, J.K., Donaldson, D.D., Wong, G.G. Moreau, J., Stahl,M. and Rogers, D. (1988) *Nature*, **336**, 688–690.
23. Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
24. Konecki,D.S., Brennand,J., Fuscoe,J.C., Caskey,C.T. and Chinault,A.C. (1982) *Nucl. Acids Res.*, **10**, 6763–6775.
25. Hasty,P., Ramírez-Solis,R. and Bradley,A. (1992) *Mol. Cell. Biol.*, **12**, 2464–2474.
26. Piedrahita,J.A., Zhang, S.H., Hagaman,J.R., Oliver,P.M. and Maeda,N. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4471–4475.
27. Koller,B.H., Kim,H.-S., Latour, A.M., Brigman,K., Boucher,R.C.,Jr., Scambler,P., Wainwright,B. and Smithies,O. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10730–10734.
28. Hasty,P., Ramírez-Solis,R., Chang,C. and Bradley,A. (1991) *Mol. Cell. Biol.*, **11**, 4509–4517.
29. Macleod,D., Lovell-Badge,R., Jones,S. and Jackson,I. (1991) *Nucl. Acids Res.*, **19**, 17–23.
30. Folger,K.R., Wong, E.A., Wahl,G. and Capecchi,M.R. (1982) *Mol. Cell. Biol.*, 2, 1372–1387.
31. Roth,D.B., Porter,T.N. and Wilson,J.H. (1985) *Mol. Cell. Biol.*, 5, 2599–2607.
32. Thomas,K.R., Folger,K.R. and Capecchi,M.R. (1986) *Cell*, 44, 419–428.
33. Toneguzzo,F., Keating,A., Glynn,S. and McDonald,K. (1988) *Nucleic Acids Res.*, **16**, 5515–5532.
34. Hsieh,P., Meyn,M.S. and Camerini-Otero,R.D. (1986) *Cell*, **44**, 885–894.
35. Moore,S.P. and Fishel,R. (1991) *J. Biol. Chem.*, **265**, 11108–11117.
36. Fishel,R., Derbyshire,M.K., Moore,S.P. and Young, C.S.H. (1991) *Biochimie*, 73, 257–267.
37. Le Mouellic,H., Lallemand,Y. and Brûlet,P. (1992) *Cell*, **69**, 251–264.
38. Mombaerts,P., Clarke,A.R., Hooper,M.L. and Tonegawa,S. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3084–3087.
39. Tybulewicz,V.L.J, Crawford,C.E., Jackson, P.K., Bronson,R.T. and Mulligan, R.C. (1991) *Cell*, **65**, 1153–1163.