GENE TARGETING AND INSERTIONAL MUTAGENESIS IN EMBRYONIC STEM CELLS

HELEN F. SUTHERLAND

PhD
University of Edinburgh
1991

MRC Human Genetics Unit, Western General Hospital,
Crewe Road, Edinburgh, EH4 2XU.
DECLARATION

I declare:

a. that this thesis is composed by myself and

b. that the work is my own except where stated.
ABSTRACT

Mutations are important in the study of gene function. Until recently mouse mutants available for study were spontaneous mutations and those derived from chemical or X-ray mutagenesis. Transgenic technology brought about the creation of more mouse mutants through insertional mutagenesis by retroviral agent or microinjected recombinant DNA. The insert may be used as a tag to molecularly clone and characterise the mutated gene. With the advent of embryonic stem (ES) cell technology the number of mutations available for study should grow - both specific and random mutations.

ES cells are pluripotent stem cells, that may be manipulated genetically in culture, and yet retain the ability to contribute to normal development and the germ line of a host blastocyst. Thus the phenotypic effects of the introduced mutation may be studied in vivo. Random mutations may be introduced into the ES cell genome by insertional mutagenesis by retroviral agent or by promoter or enhancer trap vectors. Gene targeting by homologous recombination may be used to introduce specific mutations.

*Hox-2.1* is a mouse homeobox-containing gene, belonging to the family of *Hox* genes. It has been cloned and mapped to the *Hox-2* cluster on chromosome 11. *Hox* genes are thought to play important roles in mouse development. They may act as regulatory genes, controlling the expression of structural genes at different positions in the developing embryo. In order to study the function of *Hox-2.1* in development it would be useful to have a mouse mutant for *Hox-2.1*.

It was the aim of this project to produce a mouse mutant for *Hox-2.1*. The first step towards this aim is the targeting/knock-out of the gene in ES cells. During this project various *Hox-2.1* targeting constructs have been designed and built. Replacement vectors, incorporating a promoterless neo or employing the positive-negative selection strategy, were designed to enrich for targeting events. These and two insertion vectors were transfected separately into ES cells. Over 200 selected clones have been screened for disruption of the *Hox-2.1* locus. No homologous recombinants were identified. One clone, known as a "pick-up" clone, was identified. It is thought that the targeting vector has found homology with the target sequence, picked up DNA from the locus (in this case 3' to sequence within the construct) and integrated elsewhere in the genome.

An integration site, which the promoterless *Hox-2.1-neo* targeting vector has integrated into twice, despite non-homology, has been characterised. It is suggested that this locus may be a site of frequent integration. The sequence flanking the construct was amplified by inverse PCR, cloned and sequenced. The properties which may make it highly targetable are discussed. It was shown to have weak promoter function.
ABBREVIATIONS

A  adenine
BAP  bacterial alkaline phosphatase
bp  base pairs
BSA  bovine serum albumin
C  cytosine
CAT  chloramphenical acetyltransferase
dATP  deoxyadenosine triphosphate
dCTP  deoxycytodine triphosphate
dGTP  deoxyguanidine triphosphate
dTTP  deoxythymidine triphosphate
ddNTP  dideoxynucleotide triphosphate
dNTP  deoxynucleotide triphosphate
DMSO  dimethyl sulfoxide
DNA  deoxyribonucleic acid
DTT  dithiothreitol
EDTA  (disodium) ethylenediaminetetra-acetate
ES cells  embryonic stem cells
ganc  gancyclovir
G  guanine
G418  aminoglycoside antibiotic
HAT  hypoxanthine aminopterin thymidine
Hox  homeobox containing
Hprt  hypoxanthine guanine phosphoriblsyl transferase
hrs  hours
HSV-tk  herpes simplex virus thymidine kinase
kb  kilobase pairs
LMT  low melting temperature
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIF</td>
<td>leukaemia inhibitory factor</td>
</tr>
<tr>
<td>mins</td>
<td>minutes</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>neo</td>
<td>neomycin phosphotransferase</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>p.c.</td>
<td>postcoitum</td>
</tr>
<tr>
<td>PNS</td>
<td>positive-negative selection</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>x&lt;sup&gt;r&lt;/sup&gt;</td>
<td>resistance to x</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SOFI</td>
<td>site of frequent integration</td>
</tr>
<tr>
<td>SSC</td>
<td>sodium saline citrate</td>
</tr>
<tr>
<td>6-TG</td>
<td>6-thioguanine</td>
</tr>
<tr>
<td>T</td>
<td>thymine</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethlenediamine</td>
</tr>
<tr>
<td>Tris</td>
<td>tris (hydroxymethyl) aminomethane</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to thank the following people for helping me out over the last three years.

Ian Jackson, my supervisor, for his constant enthusiasm, advice and encouragement.

Martin Hooper and colleagues, for advising on and providing ES cells.

Robin Lovell-Badge and Michaela Riddel, for growing ES cell clones before I could grow them myself.

Sinead, Julia D. and Wendy K., for sharing ES cells and the trials of culturing them.

Donald, Peter, Jane, Doreen, Ruth and Eman, for technical advice.

Sheila C., for ordering enzymes and essentials.

Sandy and Norman, for the photographs.

Paul, for computer assistance.

Nick, for commiserating.

Everyone in the MRC Unit (especially Molecular Genetics Section), for providing a friendly environment in which to work.

I dedicate this thesis to Bob Davis, who won't understand a word of it - but I have to show something for the 3 years spent away in Edinburgh.
CONTENTS

Declaration ii
Abstract iii
Abbreviations iv
Acknowledgements vi
Contents vii

1. INTRODUCTION 1
1.1 The mouse as a genetic model 1
1.2 Pluripotent stem cells 2
1.3 Insertional mutagenesis 7
1.4 Gene targeting by homologous recombination 9
1.5 Mutant mice generated from targeted ES cells 22
1.6 ES cells and new approaches to gene function analysis 24
1.7 Homeobox genes and mouse development 26
   1.7.1 Homeobox genes as transcription factors 27
   1.7.2 Expression of Hox genes during mouse development 28
   1.7.3 Hox gene mutants 29
1.8 Hox-2.1 29
1.9 Summary of thesis 30

2. MATERIALS AND METHODS 33
2.1 Mammalian cell culture 33
   2.1.1 Cell lines 33
   2.1.2 Maintainance of L cells in culture 33
   2.1.3 Maintainance of ES cells 34
   2.1.4 Electroporation 34
   2.1.5 Selection 35
   2.1.6 Picking resistant clones 36
   2.1.7 Expansion and freezing clones 36
2.2 Preparation of nucleic acid from cultured cells 36
   2.2.1 Cell lysate DNA preparation for PCR 36
   2.2.2 Large scale DNA preparation for genomic DNA digests 37
   2.2.3 RNA preparation from cultured cells 37
2.3 Bacterial cell culture 38
   2.3.1 Media and additives 38
2.3.2 Bacterial host strains
2.3.3 Bacterial vectors and/or clones
2.3.4 Preparation of competent JM83/GM161

2.4 Manipulation of plasmid DNA
2.4.1 Transformation of competent JM83/GM161
2.4.2 Small scale plasmid DNA preparation
2.4.3 Large scale plasmid DNA preparation

2.5 Manipulation of DNA by enzymes
2.5.1 Restriction enzyme digestion of DNA
2.5.2 Dephosphorylation of DNA
2.5.3 Conversion of 5' overhang to blunt end
2.5.4 Phosphorylation of double-stranded oligonucleotide linker
2.5.5 Ligation of DNA

2.6 Amplification of DNA by the polymerase chain reaction (PCR)
2.6.1 PCR on plasmid or genomic DNA
2.6.2 Inverse PCR

2.7 Gel electrophoresis of nucleic acid
2.7.1 Solutions
2.7.2 Agarose gel electrophoresis for resolving DNA molecules
2.7.3 Purification of DNA from agarose
2.7.4 Electrophoresis of RNA
2.7.5 Polyacrylamide gel electrophoresis

2.8 Transfer of nucleic acid to nylon membrane
2.8.1 Southern transfer of DNA
2.8.2 Dry-blotting of DNA
2.8.3 Transfer of plasmid DNA from bacterial colonies to nylon membrane
2.8.4 Northern blotting of RNA

2.9 Radiolabelling of DNA
2.9.1 Pandom primed oligo-labelling with Klenow
2.9.2 End labelling of DNA oligonucleotides

2.10 Nucleic acid hybridisation
2.10.1 Hybridisation solutions
2.10.2 Hybridisation methods

2.11 Autoradiography

2.12 DNA sequencing
2.12.1 Double-stranded DNA sequencing of plasmid DNA
2.12.2 Direct sequencing of PCR products
2.13 CAT assays 55
2.14 Gel retardation 57
   2.14.1 Preparation of ES cell nuclear protein 57
   2.14.2 Gel retardation 57
2.15 Oligonucleotides 58

3. TARGETING Hox-2.1 IN ES CELLS 60
3.1 Introduction 60
3.2 Replacement vector with promoterless neo 62
3.3 Replacement vectors based on PNS 69
3.4 Insertion vectors 72
3.5 Frequencies of G418 resistance 74
3.6 Screening of colonies by restriction mapping 78
3.7 Identification of a "pick-up" clone 83
3.8 Conclusions 86

4. IDENTIFICATION OF A SITE OF FREQUENT INTEGRATION IN ES CELLS 89
4.1 Introduction 89
4.2 Restriction site mapping at the integration sites in 5B2 and 7B2 89
4.3 Inverse PCR to obtain flanking sequence 94
4.4 Sequencing of upstream flanking sequence 100
4.5 Sequencing of downstream flanking sequence 100
4.6 Sequencing of the SOFI 103
4.7 Analysis of the SOFI sequence 107
   4.7.1 Search for DNA sequence homologies 107
   4.7.2 Translation of the SOFI sequence 112
   4.7.3 Attempt to map the SOFI 115
4.8 Identification of the SOFI locus promoter function 115
   4.8.1 Promoter elements from the SOFI locus drive neo expression 116
   4.8.2 CAT assays 116
   4.8.3 Gel retardation of sequence from the SOFI locus 120
4.9 Analysis of expression from the SOFI locus 122
4.10 Search for other integration events into the same sequence 122
4.11 Search for integration into the F2 locus 123
4.12 Conclusions 124
5. DISCUSSION

5.1 Targeting Hox-2.1 in ES cells 127
5.2 Future prospects for targeting Hox-2.1 130
5.3 Future prospects for the study of Hox-2.1 function 132
5.4 Identification of a site of frequent integration 132
5.5 Application of inverse PCR 134
5.6 Future prospects for the analysis of the SOFI 135

REFERENCES 137
CHAPTER 1

INTRODUCTION
1. INTRODUCTION

Recent technical advances have widened the scope of knowledge that may be gleaned from the mouse model.

This chapter will introduce the mouse as a genetic model; discuss the power of embryonic stem (ES) cell technology in the manipulation of the mouse genome and outline the potential application of this technology to the study of development, with specific reference to the *Hox-2.1* gene.

1.1 The mouse as a genetic and developmental model.

More than 1300 genetic loci have been described in the mouse (catalogued by Green in Lyon and Searle, 1989). Initially most were discovered due to visible phenotype alteration such as coat colour changes, physical or behavioural effects. Subsequently by experimental determination of strain differences in enzyme activity, antigenic determinants were used and more recently DNA sequence variations were detected. Over a thousand of these loci have been assigned to the mouse chromosome linkage map by the use of biochemical or DNA polymorphisms or chromosomal rearrangements.

Mutations in mouse genetic loci are useful in the study of mammalian development and may also serve as models for human disease. Until recently the mouse mutations available for study arose from natural variation or spontaneous mutation and induction by X-irradiation (Russel, 1951), chemical mutagenesis (Peters, 1985) or retroviral insertion (Jaenisch et al., 1985).

However, large numbers of genetic loci have been cloned and mapped which show no variation between mouse strains and whose functions are unknown. Several mouse genes have been cloned due to homology to known developmental genes of *Drosophila* e.g., the homeobox-containing family of genes (to be described in section 1.7). Function may be surmised from DNA/protein sequence motifs and expression patterns may be determined by *in situ* hybridisation.
To some extent gene function may be determined by over- or ectopic expression. Overexpression may be regarded as a dominant gain of function mutation, similar to those observed in Drosophila melanogaster e.g. Antennapedia (Antp). Genes may be introduced into the genome of transgenic mice by microinjection into pronuclei of fertilised eggs. Transgenic mice, in which Hox-1.4 is overexpressed, exhibit a megacolon phenotype (Wolgemuth et al., 1989). Transgenic mice, in which Hox-1.1 is ectopically expressed, exhibit multiple craniofacial abnormalities (Balling et al., 1989) and variations of cervical vertebra -"transformed" structures and an additional vertebra have more posterior-like characteristics (Kessel et al., 1990).

Function may be determined absolutely by the disruption of the normal process i.e. mutation of the gene. It is difficult to mutate specific genes of interest by the above mentioned methods.

1.2 Pluripotent stem cells.

A more recent technology which allows the mutation of a large number of genes either randomly or specifically has been facilitated by the culturing of mouse pluripotent stem cells. Embryonal carcinoma (EC) cells derived from teratocarcinomas and embryonic stem (ES) cells derived from the normal mouse embryo may differentiate into a number of diverse cell types both in culture and in vivo under the influence of environmental signals.

The technique for the derivation of EC cells is described by Damjanov et al. in Robertson (1987). Teratocarcinomas are established by transplanting 6-7 day old (post implantation egg cylinder) embryos to an ectopic site, such as the kidney capsule of a recipient mouse. Foci of EC cells comprise 20-40% of 8 week old post-transplantation tumours. EC cells (resembling blastocyst cells) may be established as permanent cell lines by explanting onto a layer of feeder cells in tissue culture medium. Under defined culture conditions (removal of feeders) EC cells form a structure known as an embryoid body (EC cells surrounded by endodermal cells)
from which a variety of cell types are observed to proliferate e.g., fibroblasts, cartilage, beating heart muscles, pigmented cells and neural cells (Martin and Evans, 1975). Upon injection of EC cells into a host blastocyst and transfer of that blastocyst into a surrogate mother, EC cells may contribute to normal development (Brinster, 1974 and Mintz and Illmensee, 1975) and may sometimes populate the germ line of the resulting host-EC cell chimaeric mouse (Stewart and Mintz, 1981).

The derivation of ES cells was first described independently in two papers: Evans and Kaufman (1981) and Martin (1981). The technique is described in Hogan (1986) and by Robertson in Robertson (1987). The ES cells are derived from the inner cell mass of 3.5 day blastocysts explanted into culture conditions which prevent differentiation. The cells are maintained under the same culture conditions once established as cell lines.

ES cells were originally derived and cultured on mitomycin-treated STO fibroblast or embryo-derived fibroblast feeders. Subsequently it was discovered by Smith and Hooper (1987) that ES cells cultured on Buffalo rat liver (BRL) cells retain pluripotency. Cells may also be derived in BRL cell conditioned medium in the presence of feeders (Handyside et al., 1989). More recently a glycoprotein of $M_r$ 43000 (generated through extensive glycosylation of a $M_r$=20000 polypeptide), termed differentiation inhibitory factor (DIA) was isolated from BRL cell conditioned medium and found to inhibit ES cell differentiation in vitro (Smith et al., 1988). This factor had been previously partially purified from STO feeder conditioned medium and shown to act as a differentiation retarding factor (Koopman and Cotton, 1984). Purified DIA is identical to the haemopoietic regulatory factors, leukaemia inhibitory factor (LIF) (Williams et al., 1988) and human interleukin for DA cells (HILDA) (Moreau et al., 1988). DIA/LIF may control the differentiation of embryonic ectoderm during normal embryogenesis. ES cell differentiation is inhibited by 0.2 ng/ml (10 pM) DIA (Heath and Smith, 1989). Its action is probably hormonal via interaction with cell surface receptors. ES cells capable of germ-line transmission have been derived in medium supplemented only with DIA (Nichols et
al., 1990). Removal of feeders, conditioned medium or DIA/LIF allows differentiation of ES cells. Culturing ES cells in suspension, without conditioned medium, results in the formation of embryoid bodies which upon attachment give rise to an array of cell phenotypes, e.g., nerve, cartilage and muscle (Robertson in Robertson, 1987). Photographs of ES cells in the undifferentiated state and their subsequent differentiated state upon removal of LIF from the medium are presented in Figure 1.1.

ES cells have a number of advantages over EC cells as an experimental tool - the euploid number of chromosomes is more stable, the chimaerism obtained upon injection into host blastocysts is higher and most importantly they are more able to contribute to the germ line of the chimaeric animal.

Upon microinjection of ES cells into a host blastocyst and replacement of that blastocyst into a surrogate mother, the blastocyst will continue normal development. Well maintained ES cells will contribute to all the tissues of the developing embryo, including the germ line (Bradley et al., 1984) and a chimaeric mouse will result. Differing coat colour markers or glucose phosphate isomerase isozyme (GPI) analysis distinguish the ES cell derivatives from the host cell derivatives. If XY ES cells contribute widely to an XX host the phenomenon of sex conversion may occur, giving rise to a male animal that can transmit only ES cell derived sperm (since XX cells cannot undergo spermatogenesis). Sex is determined by the number of XY cells residing in the developing genital ridge. Sex conversion results in a sex distortion effect such that more that 50% of the chimaeras produced will be male (Robertson, 1986). Extent of chimaerism and efficiency of germ-line transmission is affected by the choice of mouse strain used as the host blastocyst (Schwartzberg et al., 1989). C57B/6 was found to be a good host strain for 129-derived ES cell germ-line transmission when compared to strains CD-1 and MF1.

It has now been shown that ES cells selected on 6-thioguanine (6-TG), HAT (hypoxanthine aminopterin thymidine) medium, G418 and gancyclovir are capable of germ-line transmission. Germ-line chimaeras may be bred to produce offspring
Figure 1.1
ES cells cultured in the presence (2 μM) and absence of LIF.
heterozygous for the genetic alteration. Breeding of two heterozygous mice will result in offspring homozygous for the alteration.

ES cells may be genetically manipulated and yet retain full pluripotency and potential to form germ-line chimaeras. Genetic manipulation in culture may be carried out through the introduction of DNA by microinjection, calcium phosphate precipitation, electroporation or retroviral infection, followed by clonal selection of altered genetic make-up. Large numbers of cells may be screened for the desired genotype before microinjection into the mouse - thus ES cell technology offers a tremendous advantage over microinjection of DNA into fertilised eggs to produce transgenics. Although there is one reported case of the targeted correction of a mutation by DNA microinjection into mouse eggs (Brinster et al., 1989), it is very inefficient. Of 506 transgenics among 1841 mice born, only one had the targeted correction of the MHC class II Eα gene.

One of the first experiments to show the usefulness of ES cells in the generation of new mouse mutations was the generation of hypoxanthine phosphoribosyl transferase (Hprt) deficient mice as a model for the human disease, Lesch-Nyan syndrome. This experiment was carried out independently in two laboratories: Hooper et al. (1987) and Kuehn et al. (1987). Hooper et al. selected for spontaneous Hprt loss of function by selection of ES cells on 6-TG. 6-TG is toxic to cells upon metabolism by the Hprt enzyme. Kuehn et al. retrovirally infected ES cells and selected for Hprt variants.

Since Hprt is X-linked, females born to a germ-line chimaera and with the ES cell coat colour marker phenotype will be heterozygous for the Hprt defect and all males born wild-type. Upon breeding of the heterozygous females, males hemizygous for the Hprt mutation were born. No defect was observed at the time of publication. However, on recent examination Hprt mice have been observed to show behavioural defects (Hooper, personal communication).
1.3 Insertional mutagenesis.

Mutations resulting from spontaneous loss of function, X-irradiation or chemical mutagenesis are difficult to characterise genetically. Cloning of the genes responsible for the mutant phenotype requires the use of laborious reverse genetic techniques.

Mouse mutations may arise or may be created by the insertion of a DNA element into the genome such that gene sequence or expression is disrupted. The inserting element is useful as a tag for mapping and cloning of the disrupted DNA sequence.

A number of well characterised mouse mutations may have arisen as a result of movement of a naturally occurring transposable element (Jenkins et al., 1981 and Stoye et al., 1988). 500-1000 endogenous retroviral sequences are present in the mouse genome. No two strains have an identical set. Differences in proviral locations may be detected by hybridisation of a viral sequence probe to genomic DNA digests. It was found that an endogenous murine leukaemia virus (MuLV) proviral copy was linked to the dilute (d) coat colour mutation locus (Jenkins et al., 1981). Spontaneous revertants of the d phenotype no longer have the associated proviral sequence. It was found that the hairless (hr) mutation of mice was associated with the presence of a unique proviral copy (Stoye et al., 1988). Sequence cloned from the site was found to be strongly linked to the mutation. Spontaneous excision of the provirus results in reversion of the mutation. The data strongly suggest the hr mutation is caused by the proviral insertion. The insertion is into intronic sequence and proviral sequence may provide a splice acceptor site (Dr. Jonathon Stoye, communication).

Viral insertion has also been used in the generation of new mutations (reviewed by Jaenisch, 1988 and Palmiter and Brinster, 1985). Developing embryos may be exposed to retroviral infection. Stable integration of the provirus into the mouse genome may disrupt gene function. Lines of transgenic mice may be generated. Jaenisch et al. (1983) and Schnieke et al., (1983) have shown that a
recessive embryonic lethal mutation (strain Mov13) is caused by the insertion of the infecting agent Moloney leukemia virus (Mo-MLV) into the α1(1) collagen gene. Mov13 has been useful in studying the role of collagen in development.

Many integration sites from transgenics generated by microinjection of recombinant DNA into fertilised eggs have been cloned (Jaenisch, 1988). However, these sites are often associated with sequence duplication and rearrangement so that determination of the molecular defect causing the mutant phenotype is made difficult. One mutation, Hβ58, resulting from integration of 10-20 copies of human β-globin gene causes a recessive lethal mutation (Radice et al., 1991). Homozygotes for the integration show defective postimplantation development at day 7.5 p.c.. The integration site was mapped to chromosome 10 by in situ hybridisation. A probe, from within a 3 kb deletion of endogenous sequence, detects a transcript from 8.5 days p.c. embryos and represents a strong candidate for the gene responsible for the mutant phenotype.

ES cell technology allows another route for the generation of new mutations by insertional mutagenesis. Random mutations may be generated in the ES cells in culture, characterised at the molecular level and the cells used to generate mutant mice for the study of mutant phenotype.

Robertson et al. (1986) exposed ES cells to infection by the replication defective MPSVneo retroviral vector. Clones derived were found to have an average of 5 integrations. Chimaeric mice were generated from uncloned, infected cells. Chimaeras were bred to generate lines. The viral integration, designated 413.d, was found to cause a recessive lethal mutation (Conlon et al., 1991). Homozygotes for this mutation show disrupted post-implantation development. ES cells derived from the homozygous 413.d mutant embryos were found to have full differentiation potential in vitro and were able to carry out normal development in a chimaeric embryo. It may be concluded that the gene mutated does not code for a cell autonomous product.

Gossler et al. (1989) describe the use of promoter or enhancer trap vectors to
mutate or detect expressing genes in ES cells. It is believed that a number of important developmental genes will be expressed in ES cells. The vectors used by Gossler et al. contain the reporter gene lacZ either attached to a weak promoter (enhancer trap) or promoterless (promoter trap). Upon integration next to an enhancer or active promoter, respectively, the lacZ is expressed and the product, β-galactosidase, easily detected by X-gal staining. Chimaeras were generated from lacZ expressing clones and the pattern of expression during development determined by whole embryo X-gal staining. Several of the cell lines generated were shown to have a time and space specific expression in the developing embryo, indicating that developmental gene enhancer or promoters had been "targeted". Phenotypic affects may be observed upon breeding of the chimaeras and cloning of the disrupted genes facilitated by the presence of known DNA sequence at the site.

A promoter trap vector has been shown by Macleod et al. (1991) to integrate into sites with CpG islands. CpG islands occur at the 5' ends of housekeeping genes and some tissue specific genes. Two integration sites were cloned and shown to detect transcription of endogenous genes.

1.4 Gene targeting by homologous recombination.

Gene targeting by homologous recombination is the process by which introduced DNA integrates into or replaces endogenous sequence due to sequence identity. Specific genes may be targeted to introduce a subtle or null mutation.

Uptake of introduced DNA by homologous recombination into the cell genome was first demonstrated by transformation of the yeast, Saccharomyces cerevisiae (Hinnen et al., 1978). Most transformants (revertants of the Leu2- mutation) were generated as a result of integration of the selectable Leu2 gene-plasmid construct at the Leu2 site due to homology. 20% of the revertants were generated as a result of a double cross-over event at the site of homology. Orr-Weaver et al. (1981) demonstrated that transformation efficiency was increased by 10-1000 fold upon linearisation of the incoming DNA. This may enhance the
recombination by allowing strand invasion or by reducing topological constraints (Mansour et al., 1988). Rothstein (1983) demonstrated, using yeast, how gene disruption may be brought about by homologous recombination. A selectable yeast gene, such as *His3* or *Leu2* is inserted into the target gene sequence in the targeting vector. Upon transformation and selection for presence of the marker gene, the target gene function is lost.

The enzymes necessary to undertake the homologous recombination process are present in mammalian cells. This was shown by a number of experiments involving the correction of defective input DNA molecules such as selectable genes or viral genomes (Wake et al., 1985). Wake et al. showed that a double strand break in the region of homology greatly enhanced the rate of homologous recombination between defective SV40 virus genomes in monkey cells. Lin et al. (1984) showed that homologous recombination rate between two introduced but defective thymidine kinase (*tk*) genes in mouse Ltk- cells was increased by increasing the amount of DNA homology or by linearisation of the DNA.

In contrast to yeast, homologous recombination in mammalian cells is a rare event, occurring against a large background of random integration events. Random integration events have been characterised (Roth et al., 1985 and Macleod et al., 1991) and have been found to be not dependant on large regions of homology and are therefore termed "non-homologous" events. The processes involved in non-homologous recombination are favoured over the processes involved in homologous recombination in mammalian cells.

Different studies show variable frequencies of homologous recombination to random integration: 1/1000 (Smithies et al., 1985 and Thomas and Capecchi, 1987), 1/100 (Thomas et al., 1986) and 1/100000 (Lin et al., 1985). These differences may arise due to differences in mode of DNA introduction or in target sequence.

Gene targeting by homologous recombination into an endogenous gene in mammalian cells was first demonstrated by Smithies et al. (1985). Sequence was integrated into the β-globin gene of human EJ bladder carcinoma cells or human
chromosome 11 hybrid cells. No direct selection method was available to detect homologous integration, therefore a sensitive screening method was used. The targeting vector contained 4.6 kb of homology to the β-globin gene, a neo gene for selection of construct uptake by the cells, supF for the plasmid rescue of integration site sequence and a unique Xba I restriction site to allow the identification of correctly targeted events. Targeting frequency of 1/300-1/1100 selected clones was obtained upon targeting expressing or non-expressing β-globin.

Lin et al. (1985) corrected a defective endogenous tk gene by introduction of a second, differently defective exogenous tk gene. One homologous recombinant occurred for every 100000 cells integrating DNA.

Thomas et al. (1986) corrected mutant copies of integrated neomycin phosphotransferase APH(3′) II enzyme (neo) gene by microinjection of a different mutant copy and selection for homologous recombination or restoration to correct neo function by resistance to G418. Since 1/10 microinjected cells integrate DNA and 1/1000 microinjected cells are G418 resistant (G418r) the overall targeting frequency is 1/100.

Early studies in gene targeting were carried out on the Hprt gene. Hprt has two advantages which make it a suitable gene for the study of factors influencing homologous recombination in mammalian cells. Firstly, loss of Hprt may be selected for by resistance to 6-TG. Secondly, it is located on the X-chromosome and thus in male ES cells only one inactivation event is required to produce the selectable phenotype. Hprt spans 33 kb of genomic DNA and contains 9 exons coded by 1307 kb DNA (Melton et al., 1984).

Thomas and Capecchi (1987) compared the rate of homologous recombination into Hprt obtained using two types of targeting vector: replacement and insertion. These vectors are analogous to the vectors used in yeast (Rothstein, 1983 and Orr-Weaver et al., 1981). The replacement vector contains sequence homologous to the target sequence which is colinear with the endogenous gene. It is designed to replace endogenous sequence (Figure 1.2a) by a double cross-over event at the ends of
homology. The insertion vector contains a double strand break in the region of target homology and is designed to insert into the endogenous sequence (Figure 1.2b) by a single cross-over event. Insertion brings about the duplication of target sequence.

Thomas and Capecchi (1987) included sequence from exon 8 of the Hprt gene in the targeting vectors. The neo gene performs two functions: 1. insertion into the coding sequence to mutate the Hprt gene and 2. acts as a selectable marker to confer G418r on cells stably integrating the construct (only 0.1 % cells integrate DNA after electroporation). Homologous recombinants were identified as 6-TG resistant clones and confirmed by Southern blot hybridisation. Homologous recombinants give rise to a diagnostic restriction band pattern at the target locus.

On varying the length of homology to target sequence the targeting frequency varied. The most efficient targeting frequency obtained was 1 correctly targeted clone per 950 G418r clones upon electroporation of the replacement vector containing 9.1 kb of homology. Comparable frequencies of 1/1100 to 1/1400 was obtained from the insertion vector with 9.3 kb homology. Replacement vectors with 4 and 5.4 kb homology had targeting frequencies of 1/40000, 1/12000 and 1/7000 respectively. The smaller insertion vector, with 3.7 kb homology, gave a targeting frequency of 1/20000. A five-fold increase in target homology (2.9-14.3 kb) results in a 100-fold increase in targeting frequency (Capecchi, 1989). Expressing homologous recombinants as a fraction of the number of electroporated cells, the targeting frequencies vary between 0.02-0.3 x 10^6 (3 x 10^4 for 14.3 kb homology).

However, in a separate study by Doetschman et al. (1988) an overall targeting frequency of 0.4-1.6 x 10^6 per cells electroporated was obtained when 1.3 kb of Hprt exon 3 was targeted using a replacement vector. The comparable frequencies obtained, despite the smaller target sequence than Thomas and Capecchi may be due to the choice of gene sequence to be targeted. Sequences, even from within the same gene, may have different inherent targeting properties.

Doetschman et al. (1987) were able to correct a 10 kb deletion in the Hprt gene in ES cells (created by Hooper et al.,1987) by homologous recombination
Figure 1.2
Schematic representation of disruption of the Hprt gene by gene targeting using a replacement vector (a) and an insertion vector (b). Closed boxes represent exons and hatched boxes represent the neo gene. These diagrams were reproduced from Thomas and Capecchi (1987).
using an insertion vector. Hprt revertants are selectable on HAT medium. The deletion in the ES cells was found to span exons 1 and 2. The targeting vector contained the Hprt promoter and exons 1 and 2 to replace deleted sequence. 2.5-5 kb homology (size not known exactly since 3' deletion end-point had not been mapped) was present in the vector. Upon integration, exon 3 is duplicated. A targeting frequency of 1.4 x 10^{-6} per electroporated cell was obtained. Since the G418^r frequency was 1 x 10^{-5}, the targeting frequency may also be stated as 14/100 G418^r clones.

Thompson et al. (1989) were the first to show that homologous recombination could be used to introduce a targeted alteration in ES cells which could be transmitted to mice, in vivo. Hprt deleted ES cells were corrected as above (Doetshman et al., 1987) and injected into mouse blastocysts to generate germ-line chimaeras. Heterozygous females born as a result of crossing the male chimaera to Hprt^r homozygous females show the normal expression pattern of Hprt from the corrected locus. More of the corrected clones have since been introduced into the mouse via germ-line chimaeras (Koller et al., 1989).

Recently the Hprt system has been used to show that the length of non-homologous DNA sequence included in the replacement targeting vector has no effect on targeting frequency (Mansour et al., 1990). 8 bp, 1 kb or 3.4 kb heterologous DNA sequence inserted into 9.1 kb of Hprt sequence gave similar targeting frequencies; as did 4.3 kb and 12 kb of sequence inserted into Hprt exon 3. Homologous recombinants were selected on 6-TG and checked by Southern blot hybridisation.

As stated earlier, Hprt has the advantage of a selectable phenotype to identify correctly targeted clones. Another gene, which has a selectable phenotype and is X-linked, has been targeted - the adenine phosphoribosyltransferase locus (Adair et al., 1989). Other genes of interest are present in two copies and do not have a selectable phenotype upon disruption. Since the targeting frequency of a given locus is unpredictable and since even with the largest region of homology the most efficient
targeting frequency obtained upon targeting of *Hprt* was 1/300 G418\(^r\) clones (Capecchi, 1989), it is not always feasible to screen for homologous recombinants by Southern blot hybridisation alone.

The polymerase chain reaction (PCR) (Saiki *et al.*, 1988) may be used as a rapid and sensitive screening method to identify homologous recombination events by amplification of the recombinant DNA or novel junction fragment (Kim and Smithies, 1988). PCR amplification between a primer specific for the targeting vector sequence (e.g., *neo*) and a primer specific for the targeted endogenous sequence, within 2 kb of the first primer but not within targeting vector sequence, results in a PCR product of predictable size only if the vector has integrated by homologous recombination. The PCR is able to detect small numbers of homologous recombinant cells in pools of thousands of non-recombinant cells. PCR amplification may be used to identify pools of clones containing homologous recombinants, which may be further analysed by a process of sib-selection to identify single homologous recombinant clones. PCR amplification may be subject to artifact from random integration events if the polymerase fails to amplify *neo* primer through to flanking random DNA. The partially extended product may hybridise to target locus and amplify sequence to the second primer to give a band size similar to that predicted for homologous recombination (Mansour, 1990). Also false positives may arise due to contamination or if the endogenous sequence primer has similarity to and hybridises to sequence near a non-homologous targeting vector insert (Kim *et al.*, 1991).

PCR amplification was used by Joyner *et al.* (1989) to successfully identify targeted events in the *En-2* gene of ES cells. Upon positive identification of a pool of ten clones, cells from individual clones were analysed. DNA from a positive clone was checked by Southern blot hybridisation. A targeting frequency of 1/260 G418\(^r\) clones was obtained.

Zijlstra *et al.* (1989) used PCR to identify homologous recombinants upon targeting *β2*-microglobulin (*β2*-m) with a replacement vector, containing 10 kb of
homology. Thirty independent pools of ES cells were grown to the 8 cell stage and each clone divided in half - one half was grown in a pool for PCR analysis, the second half was grown in an individual well for analysis upon positive identification of a pool. The high targeting frequency obtained (1/25 G418r clones) may be due to a recombination hotspot or because of enrichment for targeted clones by omitting the polyadenylation signal from the neo selectable marker (strategies for enrichment are discussed below). Koller and Smithies (1989) also report the targeting of β2-m (8.4 kb homology included in the vector) and screening for homologous recombinants by PCR. 1/117 G418r clones were correctly targeted.

Zimmer and Gruss (1989) used PCR detection without prior selection after microinjection of ES cells with a Hox-1.1 targeting construct. Non-homologous DNA in the vector was limited to a 20 bp oligonucleotide, which inserts a stop codon inframe to disrupt the Hox-1.1 coding sequence and acts as a priming site for PCR amplification. 719 ES cells were microinjected and analysed for homologous recombination by PCR in 12 pools. At least 5 independent homologous recombination events were identified giving a targeting frequency of 1/150 microinjected cells. Since 20% of cells are estimated to stably integrate DNA (this reduces the need for a selectable marker), a frequency of 1 homologous to 30 non-homologous events is estimated. The authors suggest that microinjection may increase targeting frequency by providing DNA substrate suitable for homologous recombination as opposed to non-homologous recombination. However, microinjection of ES cells is technically difficult and lower frequencies for some genes may mean large numbers would have to be microinjected. A similar strategy was used recently by Kim et al. (1991) in an attempt to correct the human sickle-cell β globin gene in a mouse-human cell hybrid. No homologous recombinants were detected after screening 18000 (1800 to 3600 transformants) microinjected cells. Upon electroporation of a targeting vector containing similar sequence and a neo selectable marker gene a targeting frequency of 1/9700 G418r clones was obtained (Shesely et al., 1991).
The background of non-homologous events may be lowered by addition of dideoxynucleotides to the 3' hydroxyls of the targeting vector (Chang and Wilson, 1987). Dideoxynucleotides prevent end-joining and since end-joining is thought to be involved in non-homologous recombination, these events are reduced. Free DNA ends probably play a different role in homologous recombination events, that of strand invasion, and this process will be unaffected by dideoxynucleotide ends.

Homologous recombination events may also be enriched for by selecting against non-homologous recombination events and the number of clones required for screening thus reduced. There are two enrichment procedures that may be used: 1. the use of cis-acting regulatory elements of the target gene for the expression of the selectable marker or 2. the positive-negative selection (PNS) strategy (Mansour et al., 1988). These strategies have been used to target a number of genes successfully at high targeting frequencies.

Use of a promoterless gene as a selectable marker in the targeting vector is expected to reduce the number of selected clones generated by random integration and enrich for homologous recombination into the expressed target locus gene (Doetschman et al., 1988). Use of this strategy requires target gene expression in the experimental cells. Requirement for enhancer sequence (Jasin and Berg, 1988) or polyadenylation signal (Zijlstra et al., 1989) by the selectable marker may also enrich for integration into the target locus.

Dorin et al. (1989) report a targeting frequency of 3/11 G418r clones upon targeting of SV40 sequence in human-mouse cell hybrid Cl21 cells using a replacement vector in which a promoterless neo gene is fused inframe with the SV40 sequence. Sedivy and Sharpe (1989) claim an 100-fold enrichment for disruption of the polyoma middle T antigen in NIH 3T3 cells upon targeting with a replacement vector in which a promoterless neo is fused inframe to chloroamphenical acetyltransferase gene amino-terminal portion sequence (to ensure correct neo function) and T antigen target sequence. A targeting frequency of ~ 1/100 G418r clones was obtained. A similar construct design was used to target c-abl in ES cells.
(Schwartzberg et al., 1989 and 1990) at a targeting frequency of 1/34 G418\textsuperscript{r} clones. Clones carrying the \textit{c-abl} base substitution mutation were subsequently introduced into blastocysts to derive germ-line chimaeric mice. A promoterless \textit{neo} vector was used to target \textit{N-myc} in ES cells at a frequency of 1/5 (Charron et al., 1990) and a promoterless/AUG\textsuperscript{r} \textit{neo} vector used to target \textit{pim} in ES cells at a frequency of 34/40 G418\textsuperscript{r} clones (te Riele et al., 1990).

As an alternative to the generation of a target gene-\textit{neo} fusion protein, the termination-reinitiation strategy (Peabody and Berg, 1986) has been suggested. A stop site is inserted immediately 5' to the \textit{neo} AUG start site. However, Schwarzberg et al. (1990) found the latter mechanism to be relatively ineffective. Doetschman et al. (1988) found the \textit{neo} gene was not expressed sufficiently to protect targeted cells from G418 at 400 \textmu g/ml.

Jasin and Berg (1988) showed an ~ 100-fold enrichment upon deletion of the transcriptional enhancer from a SV40 sequence targeting vector, to obtain a targeting frequency of 25/57 selected clones. Use of a \textit{neo} gene lacking the polyadenylation signal reduced the number of G418\textsuperscript{r} clones 4-fold upon targeting of \textit{\beta2-microglobulin} (Zijlstra et al., 1990). A targeting frequency of 1/25 G418\textsuperscript{r} clones was obtained.

The second strategy for the enrichment of homologous recombinants is that of PNS (Mansour et al., 1988). PNS uses a positive selectable marker, such as \textit{neo}, to select for integration events, and a negative selectable marker gene, such as herpes simplex virus thymidine kinase (HSV-tk), to select against random integration events. Metabolism of base analogues, gancyclovir (ganc), acyclovir or FIAU, recognised by HSV-tk but not cellular \textit{tk}, leads to cytotoxicity and cell death. The PNS vector design is illustrated in Figure 1.3. \textit{Neo} is inserted into target sequence and HSV-tk is placed adjacent to target sequence. Upon replacement of endogenous sequence in an homologous recombination event, HSV-tk does not integrate, and cell lines derived will be doubly G418\textsuperscript{r} and ganc resistant (ganc\textsuperscript{r}). Upon random
Positive – Negative Selection.

a. Gene Targeting

Gene X

\[ x^{-}\text{neo}^{\prime} \text{HSV-tk}^{-}\langle G418^{r}, \text{GANC}^{r}\rangle \]

b. Random Integration

\[ x^{+}\text{neo}^{\prime} \text{HSV-tk}^{+}\langle G418^{r}, \text{GANC}^{r}\rangle \]

Figure 1.3

Schematic representation of the disruption of Gene X using the positive-negative selection strategy.

a. Homologous recombination between the targeting vector and genomic X DNA results in the disruption of the Gene X and loss of HSV-tk sequence. Such cells will be X\(^{-}\), G418\(^{r}\) and ganc\(^{r}\).

b. Random integration of the targeting vector results in insertion of both neo and HSV-tk genes. Such cells will be X\(^{+}\), G418\(^{r}\) and ganc sensitive.

Closed boxes represent exons and cross-hatch boxes represent neo or HSV-tk genes.

This diagram is reproduced from Mansour et al. (1988).
integration via the ends of the targeting vector, HSV-tk sequence is integrated, and non-homologous recombinant cell lines derived will be G418r but ganc sensitive. Random integration events, in which HSV-tk is impaired upon integration, are G418r and gancr.

Mansour et al. (1988) report a 2000-fold enrichment factor upon application of ganc, such that 1 in 2000 G418r clones is also gancr. Targeting of Hprt resulted in a targeting frequency of 19/24 double resistant clones and targeting of int-2 gave a targeting frequency of 4/81. These frequencies may be expressed as a fraction of G418r clones: 1/2500 for Hprt and 1/40000 for int-2. It is suggested that lower accessibility of target locus lowers targeting frequency for some genes.

Johnson et al. (1989) report a 300-2000-fold enrichment with ganc selection and targeting frequencies of 1/16 for c-fos, 2/22 for adipsin and 12/142 for aP2, when the frequency is expressed as a fraction of doubly resistant clones. Adipsin and aP2 are not detectably expressed on Northern blots of ES cell RNA. This demonstrates that the open chromatin structure of expressing genes is not essential for recombination. This was also shown for the targeting of insulin-like growth factor II (DeChiara et al., 1990) and β-globin in EJ bladder carcinoma cells (Smithies et al., 1985). However, elevated recombination rates may be observed in transcriptionally active DNA. This was shown in the experiments of Thomas and Rothstein (1989), in which recombination at the GAL10 locus of S. cerevisiae was found to be 15 X higher when constitutively expressed in gal80 mutants than in the repressed state of gal4 mutants.

Many other genes have been targeted successfully using the PNS strategy. However, reports vary on the enrichment factor obtained through negative selection with ganc. Enrichment varies from 8-fold for the targeting of the immunoglobulin μ gene (Kitamura et al., 1991); 13-fold for the insulin-like growth factor II (DeChiara et al., 1990); 23-fold for the transcription factor GATA-1 (Pevny et al., 1991) to 70-fold for the insulin receptor gene (Accili et al., 1991). Targeting frequencies were 1/38, 3/100, 6/79 and 2/24 double resistant clones, respectively. These frequencies vary.
may be expressed as 1/304, 1/429, 1/302 and 1/804 G418r clones.

Use of two tk genes in the targeting vector - one at either end of the target sequence - may be expected to enrich for homologous recombinants still further. Different tk gene sequences are used to reduce the potential for intraplasmid recombination. Mombaerts et al. (1991) report an enrichment of only 2-fold upon targeting T-cell antigen receptor and McMahon and Bradley (1990) report enrichment of 26-fold upon targeting of Wnt-1. Targeting frequencies were 1/500 and 1/275 double resistant clones, respectively or 1/1000 and 1/7500 G418r clones.

These reported enrichment factors may not be "true" enrichment factors. Tybulewicz et al. (1991) report an apparent 5.2-19.6 fold enrichment upon ganc selection. However, upon transfection of cells with a neo vector and subsequent selection on G418 or G418+ganc, two times as many G418r as G418r+ganc clones were generated. 50% of apparent enrichment is therefore due to non-specific cytotoxicity of ganc rather than the presence of HSV-tk.

ES cells selected on ganc or FIAU (McMahon and Bradley, 1990) are able to populate the germ line of chimaeric mice. However, ganc has the potential to cause random mutations. A negative selectable marker gene which does not require drug selection may be used to decrease the likelihood of adverse effect on the cells. One such gene is that encoding the cytotoxic Diphtheria toxin A-fragment (DT-A). When used in a PNS vector to target c-fyn, DT-A gave an enrichment factor of 10-fold and a targeting frequency of 8/16 selected clones (Yagi et al., 1990). Germ-line chimaeras were obtained.

A novel method for enrichment and selection, epitope addition, is used by Jasin et al. (1990). The murine cell surface marker Thy-1 is used as a positive selectable marker. Since the Thy-1 marker used requires promoter activity, AUG and a leader sequence for proper expression on the cell surface, homologous recombination events are highly enriched for. Flow cytometry was used to detect Thy-1 marker on the cell surface. 70% of expressing clones were correctly targeted at the human CD4 locus in the human T-cell line, JM. A similar strategy is used by
Itzhaki et al. (1991) to target an interferon inducible gene, 6-16 in HeLa cells. A promoterless human growth hormone (hGH) is used as positive selectable marker. An antibody assay was used to detect hGH in the growth medium of pooled clones. 9/10 interferon inducible hGH secreting cell lines were homologous recombinants.

The factors affecting homologous recombination in mammalian cells are not yet fully understood. Different genes or experiments show tremendous variation in targeting frequency. Vector design may be optimised by incorporating large regions of target sequence homology and a strongly expressed neo. An enrichment strategy may be employed to reduce non-homologous recombination events. It has been suggested that use of isogenic DNA in vector and cells enhances targeting frequency by ensuring complete homology (Hein te Riele and Anton Berns, personal communication). It is also apparent that the DNA sequences vary in their inherent targetability. DNA sequence, chromatin structure, enhancers or inhibitors of neo expression may affect frequency of homologous recombination or detection.

It must be noted that comparison of gene targeting frequencies is complicated. Frequencies expressed as a fraction of the number of cells electroporated may be complicated by cell death, electroporation conditions or plating efficiencies. Frequencies expressed as a fraction of the number of G418r colonies is complicated by the different neo genes used in the targeting construct - weak promoters or enrichment strategies may reduce numbers.

1.5 Mutant mice generated from targeted ES cells.

In the last two years mice heterozygous and homozygous for targeted mutations have been generated from ES cells. Phenotypes have been studied in the hope of gaining some insight into the function of the mutated genes.

Heterozygous mice, born to male chimaeras of ES cells, carrying a null mutation in the insulin-like growth factor II (IGF-II) gene were found to be 60% normal size (DeChiara et al., 1990). Homozygotes were also 60% normal size, while heterozygotes born to female heterozygotes were normal size (DeChiara et al.,
This study demonstrates that IGF-II is indispensable for normal embryonic growth and is subject to parental imprinting.

Mutation in the immunoglobulin \( \mu \) gene leads to abnormal B-cell development (Kitamura et al., 1991) and mutation in the \( \beta 2\)-m gene leads to absence of certain T-cell populations due to lack of MHC I presentation (Zijlstra et al., 1990). The importance of the transcription factor GATA-1 in erythrocyte development was demonstrated in ES cell derived chimaeric mice. Erythrocytes failed to develop from the ES cells, in which the X-linked GATA-1 had been mutated by homologous recombination (Pevny et al., 1991).

Homozygotes for a null mutation in \( c-abl \) (Tybulewicz et al., 1991) and for a deletion of the c-terminus of \( c-abl \) (Schwartzberg et al., 1991) have been generated. Both mutations result in similar phenotypes demonstrating the c-terminus (involved in subcellular localisation) is very important for \( c-abl \) function. Homozygotes die 2-3 weeks after birth. The immune system, thymus and B-cell population are severely affected.

Homozygotes for the disrupted \( Hox-1.5 \) gene die neonatally due to pulmonary defects (Chisaka and Capecchi, 1991). Absence of thymus and parathyroid, coronary defects and throat region abnormalities may be due to the lack of \( Hox-1.5 \) in branchial arches and pharyngeal pouches prior to migration or mixing of cell derivatives.

Homozygotes for \( Wnt-1 \) (\( int-1 \)) mutations have been generated independently by Thomas and Capecchi (1990) and McMahon and Bradley (1990). The majority of homozygotes die neonatally, though one mouse surviving to adulthood suffered severe ataxia (Thomas and Capecchi, 1990). Degrees of penetrance may vary due to mouse strain background. The dorsal and ventral caudal midbrain were found to be absent at 9.5 days p.c. (McMahon and Bradley, 1990). This correlates with a broad band of \( Wnt-1 \) expression observed on \textit{in situ} hybridisation. CNS development, especially of the cerebellum, is defective.

The \( Hox-1.5 \) and \( Wnt-1 \) mutants show severe phenotypes in regions that
would be predicted to be affected because of the previously observed pattern of gene expression. However, these genes are also expressed in regions which appear unaffected by the mutation. The gene product may have no role in these areas despite its expression. A more likely explanation is that of redundancy of function - the gene function may be replaced by similar or related proteins.

Functional redundancy between En-1 and En-2 in shared regions of expression is the probable explanation for the normal development of En-2 null mutation homozygotes (Joyner et al., 1991). A subtle phenotype is observed in the cerebellum - the only region in the developing mouse embryo where En-2 is expressed and En-1 is not. En-2 may have developed a special function in this region. If redundancy is operating, it may be predicted that upon breeding of En-1 mutant with En-2 mutant mice to generate mice homozygous for both mutations, that a more severe phenotype than either mutation alone would be observed. The En-1 gene has not yet been targeted.

Overlap in function between other protein kinases may also explain why a null mutation in c-src does not lead to a more extensive phenotype (Soriano et al., 1991). c-src is expressed widely in the mouse and yet the mutant phenotype is limited to bone. Other widely expressed protein kinases are being targeted.

1.6 ES cells and new approaches to gene function analysis.

Cell lines carrying null alleles of normally expressed genes may be instrumental in assessing the function of that gene in vitro or in vivo. Cell lines may be particularly useful in the study of gene mutations causing early developmental lethality. Cell lines may be derived from mutant embryos or by the consecutive inactivation of both alleles by homologous recombination in cultured cells. te Riele et al. (1990) describe the consecutive inactivation of both pim-1 alleles in ES cells. The first disruption was achieved by integration of a neo gene using a pim-1 replacement vector and the second disruption made by integration of a different selectable marker gene, hygromycin (hyg). The bacterial hyg gene confers resistance
to hygromycin B. Both selectable markers were promoterless. 21/34 hygromycin B resistant clones, from the second targeting, were correct homologous recombination events. In six of these the hyg had replaced neo sequence and in the remaining 15 cell lines the hyg had integrated into the second pim-1 allele. Double knock-out cells show no difference from wild-type cells in morphology, growth or differentiation. Double knock-out cells may also be introduced into blastocysts to form chimaeras. Otherwise developmentally lethal mutant cells may thus be rescued by normal cells and the ES cell contribution to developing tissues monitored. Modulation of gene activity by consecutive gene targeting of the creatine kinase M allele in ES cells is reported by van Deursen et al. (1991). One allele was knocked out and then was modulated by an insertion vector to give altered expression of the gene.

Subtle mutations in a gene locus may be introduced by the hit and run procedure (Hasty et al., 1991). This is achieved by a two-step process and exploits the locus duplication produced by integration of an insertion vector. A subtle mutation, such as a missense mutation, is introduced into a duplicated copy of the gene upon insertion of the targeting vector. The plasmid backbone of the targeting vector carries a neo marker for positive selection and a tk marker for selection of a reversion event. Upon reversion, duplicated DNA sequence resolves by intrachromosomal recombination and the plasmid sequence, along with the tk gene, is lost. Reversion events are selectable in FIAU. Hasty et al. demonstrated the efficacy of this procedure on the Hprt and Hox-2.6 loci of ES cells. Duplication reversion rates were 4.3 x 10^{-6} and 3.8 x 10^{-3} per cell generation, respectively. A proportion of the revertants were found to carry the introduced mutation.

Replacement of target gene in ES cells with reporter gene, lacZ, allows the fate of the mutated cells to be followed in the developing chimaeric embryo (Le Moullieec et al., 1990 and Mansour et al., 1990). Cell ablation studies, similar to those carried out by Palmiter et al. (1987), would be possible if the target gene in ES cells was replaced with reporter genes encoding toxin. Timed ablation, in mouse chimaeras, of the cells expressing the reporter gene would be useful in the study of
genes functional through various stages of development.

Pavan et al. (1990) and Craig Barton et al. (1990) demonstrate the feasibility of targeting specific mammalian loci using the recombination abilities of S. cerivisiae. Large DNA fragments cloned into yeast artificial chromosomes (YAC) may be targeted efficiently. Homologous recombination in yeast could be used to introduce mutations rapidly into a cloned locus which then could be used to make conventional targeting vectors. It has been shown that YACs could be transfected and integrated into EC cells (Pavan et al., 1990). If ES cells could be transformed in a similar manner, YAC DNA could be introduced into mice. Correction of mutations in ES cells and eventually in vivo, by an integrating YAC, may be possible. YAC integration into ES cells would offer an advantage over microinjection transgenic technology, because of the DNA locus size that may be introduced. It may also be possible to target genes in ES cells by homologous recombination of YACs - this may prove to be highly efficient because of the large region of homology.

1.7 Homeobox genes and mouse development.

The genetics of Drosophila melanogaster development has been extensively studied through mutant analysis (Nusslein-Volhard et al., 1987; reviewed by Akam, 1987 and Ingham, 1988). At least three groups of genes have been identified which form a hierarchical network of control throughout the different stages of the fly's development - the maternal effect genes, the segmentation genes and the homeotic genes. Cloning and characterisation of the Drosophila developmental genes has had a spin-off in the study of mammalian development. Drosophila genes contain functional domains which are highly conserved within gene families across a number of species, including mouse and man (reviewed by Kessel and Gruss, 1990). One such domain is the homeobox: a region of 183 bp DNA coding for 61 amino acids (McGinnis et al., 1984b and reviewed by Gehring, 1987). The homeobox is present in all three Drosophila gene classes, as well as genes of invertebrate and vertebrate species (Holland and Hogan, 1986).
More than 30 homeobox-containing (Hox) genes have been located in the mouse genome, cloned and sequenced (Kessel and Gruss, 1990). Genes were cloned through cross-hybridisation of Drosophila homeobox sequence. Mo 10 (now named Hox-1.5) was the first mouse Hox gene to be cloned (McGinnis et al., 1984a). Many of the Hox genes show greatest similarity to the Drosophila Antp homeobox. Most of the Antp-like Hox genes are present in 4 clusters (Hox-1,-2,-3 and -5) in the mouse genome on chromosomes 6, 11, 15 and 2, respectively. It is possible that the ancestral Hox gene was duplicated successively to form a gene cluster. During vertebrate evolution the cluster appears to have duplicated and diverged (Graham et al., 1989, Duboule and Dolle, 1989 and Schugart et al., 1989). Paralogs between complexes, have greater sequence similarity (in homeobox and other regions) than neighbouring genes in the same complex and appear to share a common ancestor with a related Drosophila gene. Relative position in the complex is maintained and is found to correlate with the domain of expression along the anteroposterior axis in both Drosophila and mouse (Graham et al., 1989 and Duboule and Dolle, 1989). Clustered homeobox genes, present before evolutionary divergence of insects and vertebrates, may function as important systems for controlling development.

There are three further lines of evidence to suggest a developmental role for murine Hox genes: their likely role as transcription factors, their expression pattern in the developing embryo and analyses of Hox mutants. It is likely that they determine the expression of other genes in a compartmental-specific and temporal-specific manner within the embryo.

1.7.1 Homeobox genes as transcription factors.

The DNA binding helix-turn-helix motif of the homeobox may bind to promoter or enhancer regions of genes to allow the regulation or autoregulation of transcription. The homeobox shows homology to the yeast transcription factors Matα1 and Matα2 (Shepherd et al., 1984). These proteins control cellular differentiation into mating-type a or α cells, or into spores, by controlling the expression of other genes (reviewed by Nasmyth, 1982). α2 protein binds to DNA
sequence upstream of one target gene (Johnson and Herskowitz, 1985) and it has been shown that the region encoding DNA binding corresponds to the homeodomain (Porter and Smith, 1986). The homeodomain shows similarity to prokaryotic gene regulatory proteins, cro and repressor protein of bacteriophage λ (Laughon and Scott, 1984). The DNA-binding domain is determined to be a helix-turn-helix structure (Matthews et al., 1982). In vitro analyses of putative homeodomain DNA recognition sites have been carried out. These studies identify a 10-12 bp sequence with a central TAAT motif (reviewed by Hayashi and Scott, 1990; Budd and Jackson, 1991).

Many homeobox proteins are located within the nucleus as detected by antibody binding (White and Wilcox, 1984 and see review by Gehring, 1987).

Direct evidence of homeodomain transcriptional regulation comes from cell culture assays, in which activity of a promoter of interest is monitored by expression of a reporter gene, upon cotransfection with a homeobox protein expression construct (reviewed by Hayashi and Scott, 1990).

1.7.2 Expression of Hox genes during mouse development.

The expression patterns throughout development are demonstrated by in situ hybridisation to mRNA (review Holland and Hogan, 1988a; Kessel and Gruss, 1990). Although it was known for some time that Antp-like genes showed spatially and temporally restricted expression patterns, from gastrulation (7.5 days p.c.) to mid- to late-organogenesis (12.5 days p.c.-15 days p.c.), the link between gene expression and body structure was discovered only upon analysis of Hox gene expression in 9.5 days p.c. embryos (Murphy et al., 1989 and Wilkinson et al., 1989). These studies showed that members of the Hox-2 cluster had expression patterns correlating with segmental structures, termed rhombomeres (Lumsden and Keynes, 1989), present in the developing hindbrain. Hox-2.9 expression is confined to a single rhombomere (rhombomere 4) (Murphy et al., 1989 and Wilkinson et al., 1989), while more 5' members of the Hox-2 cluster show expression throughout the neural tube with anterior boundaries of expression at two rhombomere intervals.
(Wilkinson et al., 1989). It is suggested that overlapping domains of expression provide positional signals in the embryo and that Hox genes play a role in determining segment phenotype. Clear boundaries of Hox-2 gene expression are also observed in neural crest cells in the neural plate at 8.5 days p.c. (Hunt et al., 1991). Paralogs from the other gene clusters may also show similar domains of expression along the anteroposterior axis (Gaunt et al., 1991; Murphy and Hill, 1991) but tissue-specific variations suggest these related genes perform specialised functions in specifying parts in the complex body plan (reviewed by Gaunt and Singh, 1990).

Expression pattern of Hox genes during development may also be studied in transgenic mice. Puschel et al. (1990 and 1991) attached the promoter region of Hox-1.1 to the reporter gene, lacZ, and monitored lacZ expression in transgenic embryos. LacZ offers a more sensitive detection system than in situ hybridisation. However, these studies are only reliable if all the promoter function is retained in the construct.

1.7.3 Hox gene mutants.

Developmental abnormalities are observed in Hox gene gain-of-function or loss-of-function mutants. Transgenic mice over-expressing the gene Hox 1.4 exhibit abnormal gut development (Wolgemuth et al. 1989) and those expressing Hox 1.1 ectopically show multiple developmental defects (Balling et al., 1989). Hox-1.5 null mutants, derived from targeted knock-outs in ES cells (Chisaka and Capecchi, 1991), display defective development in the regions normally expressing the gene (see section 1.5).

1.8 Hox-2.1

Hox-2.1 is located within the Hox-2 cluster on chromosome 11. It was isolated by hybridisation of the Drosophila Antp and Ultrabithorax homeobox probes to a mouse genomic library (Jackson et al., 1985 and Hart et al., 1985) and separately cloned by its hybridisation to a human homeobox sequence from the Hu 2 gene, which had itself been isolated by hybridisation to the Antp homeobox probe.
(Joyner et al., 1985 and Levine et al., 1984). The Hox-2.1 cDNA sequence is published (Krumlauf et al., 1987). It is paralogous with Hox-1.3 and Hox-3.4 and is homologous to the Drosophila Sex chromosome reduced (Scr) of the Antennapedia complex (Graham et al., 1989).

The expression of Hox 2.1 has been studied (Jackson et al., 1985, Krumlauf et al., 1987, Holland and Hogan, 1988b and Wilkinson et al., 1989). mRNA is shown to be present in embryonal tissues 7.5-16.5 days p.c. by RNase protection studies and Northern blot hybridisation. The highest expression is in the spinal cord at 12.5 days p.c. and adult kidney. *In situ* hybridisation to embryo sections shows spatial restriction of the mRNA within the central nervous system at 9.5 days p.c. There is a clear anterior boundary of expression at the hindbrain-spinal cord boundary, which lies caudal to the hindbrain rhombomere segments; but no clear posterior boundary. Segmentation is not observed in the spinal cord. There is a temporal shift in pattern with the anterior boundary remaining but posterior expression decreasing after 12.5 days p.c. There is also expression in the peripheral nervous system - dorsal root ganglia (neural crest derived cells) - and mesodermal derivatives in the lung and stomach. This expression pattern might indicate that Hox 2.1 has a role in position determination in "collaboration" with other homeobox or developmental genes (e.g. transition between hindbrain segmentation to spinal cord CNS organisation) and may be involved in organogenesis of the lung and stomach.

1.9 Summary of Thesis.

The study of Hox-2.1 function during mouse development would be greatly facilitated by the availability of a mouse mutant for Hox-2.1. Alteration in or lack of structures would indicate the necessity of Hox-2.1 for normal development. Redundancy of function may also become apparent. Figure 1.4 illustrates the series of steps involved in the generation of a mouse mutant for Hox-2.1.

The first step in the series is the introduction of a null mutation into Hox-2.1 in ES cells. Chapter 3 describes the attempts at targeting, using different Hox-2.1
targeting vectors, designed along the principles discussed in Section 1.4. No correctly targeted cell line was obtained. Thus the \textit{in vivo} analysis of an \textit{Hox-2.1} mutant was not made possible.

Chapter 4 describes the identification and analysis of an integration site, which a promoterless \textit{Hox-2.1-neo} replacement targeting vector has integrated into twice, despite non-homology. It is suggested that properties of this sequence might make it a highly targetable site.
**Figure 1.4**

Schematic drawing of steps involved in the generation of a mouse mutant for *Hox-2.1*. ES cells are transformed to G418-resistant by electroporation with a *Hox-2.1* targeting vector. A targeted clone is identified by screening and cells introduced into host blastocysts by microinjection. ES cell germ-line chimaeric mice may be bred to produce mice heterozygous for the *Hox-2.1* mutation. Phenotypes of *Hox-2.1* mutant homozygotes born to the heterozygotes may be studied.
CHAPTER 2

MATERIALS AND METHODS
2. MATERIALS AND METHODS.

2.1 Mammalian cell culture.

2.1.1 Cell lines.

The L cell L-M (tk-) cell line is a subline of the L-M mouse fibroblast cell line, resistant to 5-bromo-2-deoxyuridine (BUdR) and lacking the enzyme thymidine kinase (Kit et al., 1963).

ES cells cultured throughout the project were of the E14 cell line (Handyside et al., 1989), kindly provided at an early passage number (14 through to 19) by Dr. Martin Hooper's laboratory. E14 cells are derived from strain 129/Ola (homozygous for the ch chinchilla coat colour marker) mouse blastocysts. They have an XY karyotype.

ES cells of the line CCE (Robertson, 1986; Evans and Kaufman, 1981) were grown in Dr. Robin Lovell-Badge's laboratory. CCE cells were derived from 129/Sv/Ev (homozygous for the BB AA black agouti coat colour marker) mouse blastocysts. They have an XY karyotype.

2.1.2 Maintenance of L cells in culture.

L cells were maintained in Dulbecco's modified Eagles medium (DMEM) (Flow Laboratories) supplemented with heat inactivated 10% foetal calf serum (FCS) (Gibco). Penicillin and streptomycin (Gibco) were added to cultures at 100 units/ml and 100 μg/ml respectively. All cultures were grown at 37°C in 25, 75 cm² (Nunclon) and 175 cm² (Falcon) tissue culture flasks in a 10% CO₂ atmosphere.

Cells were washed in Dulbecco's phosphate buffered saline (PBS) before trypsinisation with 10% trypsin in versene.

Cells were resuspended at a concentration of 10⁶/ml in FCS containing 7-10% DMSO and aliquotted into ampoules (Nalgene cryotubes) for freezing. The ampoules were frozen slowly overnight in a polystyrene box at -70°C and transferred to liquid nitrogen. Cells were recovered from frozen stock by rapidly
thawing at 37°C and transferring to 10 ml medium. The cells were pelleted at 1000 rpm for 5 mins before plating in fresh medium.

2.1.3 Maintenance of ES cells.

ES cells were cultured in the presence of DIA/LIF (differentiation inhibitory activity/leukaemia inhibitory factor) (Smith et al., 1988 and Williams et al., 1988) or on fibroblast feeder layers (Robertson, 1987) to prevent their differentiation.

ES cells which were cultured on feeder layers were seeded onto a monolayer of fibroblast feeder (STO) cells which had been mitotically inactivated by mitomycin C. The cells were cultured in DMEM (Flow Labs), with the addition of 20% foetal calf serum, 0.1 mM β-mercaptoethanol and 50 units/ml penicillin and streptomycin.

ES cells which were cultured with LIF were grown in Glasgow's modified Eagles medium (GMEM) (Flow Labs) supplemented with 1 X non-essential amino acids (Gibco), 1 mM sodium pyruvate (Flow Labs), 0.3% glutamine, 0.1 mM β-mercaptoethanol and 10% heat inactivated FCS (Globepharm Ltd., tested for toxicity by Dr. Sinead Jones as described by Robertson, 1987)). LIF was harvested by Wendy Kimber from Cos-7 cells containing the pC10-6R expression plasmid (Moreau et al., 1988) (obtained from Dr. Austin Smith). The amount of LIF necessary for culture was determined by the differentiation of ES cells in a dilution series. Flasks (Costar 25 cm² and 225 cm² with vented lids) for the culture of ES cells were coated with 0.1% gelatin (Sigma, cell culture tested).

Disposable containers and pipettes were used at all times to ensure that the ES cells did not come into contact with residual detergent.

1 X trypsin/EDTA was bought from Flow Labs for the trypsinisation of the cells. ES cells were split usually at a factor of 1 in 8. The passage number of the cells was noted on each flask after splitting.

2.1.4 Electroporation.

Construct DNA was linearised (or whole inserts released) by the appropriate
restriction enzyme, prior to electroporation. The DNA was extracted once with phenol, once with chloroform and ethanol precipitated before resuspending in H₂O at a concentration of 1 mg/ml.

L cells were electroporated on a workshop made electroporator. Cells were suspended in 1 ml PBS at a concentration of 1 x 10⁶ per ml with 20 μg linearised DNA.

ES cells were electroporated using a Bio Rad Gene Pulser. Cells were suspended at a concentration of 5 X10⁶ to 1 X10⁷ per ml in PBS with 20-100 μg linearised DNA. For targeting experiments, 0.8 ml cells were pulsed with capacitance of 3 μF and voltage of 800 V at path length 0.4 cm (Thompson et al., 1989). For transfection of CAT (chloramphenicol acetyltransferase) constructs, 0.8 ml cells were pulsed at 960 μF (with capacitance extender) and 200 V at path length 0.4 cm.

Cells were incubated at room temperature before plating in 10 cm diameter culture dishes (Costar).

2.1.5 Selection.

Cells containing the neomycin phosphotransferase APH(3') II enzyme gene (neo) are resistant to the G418 aminoglycoside antibiotic.

G418 (Gibco) was added to the medium 24 hours after electroporation at a dose of 600 μg/ml to 1 mg/ml for L cells or 150 μg/ml to 400 μg/ml for ES cells, depending on batches of G418. Kill curves for batches of G418 were determined on untransfected cells. G418 selection was maintained for 10-14 days before the picking of clones.

Cells transfected with the positive-negative selection vectors were selected with 2 μM gancyclovir (Syntex Pharmaceuticals Ltd.), a base analogue toxic to cells upon metabolism by the HSV thymidine kinase (HSV-tk) gene, for 5 days on the fourth or fifth day after electroporation.
2.1.6 Picking resistant clones.

Medium was aspirated off and replaced by PBS. A colony was observed under the microscope and scraped from the surface using a drawn out pasteur pipette (diameter equal to size of colony). The colony was drawn into the tip of the pasteur by attachment to a mouth pipetter and transferred to a drop of 1 X trypsin/EDTA. 12 foci were picked in this manner at one time. The trypsin drops were incubated at 37°C for 5 minutes. Using a fine drawn out pasteur (diameter half or quarter size of the colony) a trypsin drop was sucked up and down until the clump of cells was observed to break up into single cells. The cells were drawn into the pasteur and transferred to a medium-containing well of a 24 well plate (Flow Labs). All remaining colonies were broken up and transferred in a similar manner.

2.1.7 Expansion and freezing of clones.

Cells were maintained on G418. Clones were either expanded to 75 cm² flasks or frozen in a 2 cm² well of a 24 well plate (Chan and Evans, 1991).

Cells from 75 cm² flasks were divided into two pellets. One pellet was resuspended in 2 ml freezing mix (7% DMSO in FCS) and transferred to two freezing vials for storage in liquid nitrogen. The second pellet was stored at -20°C until the preparation of DNA.

Alternatively a clone upon reaching confluence in a 2 cm² well was used to seed a second 2 cm² well at a density of 5 X 10⁴ - 1 X 10⁵ cells per well. Any remaining cells were used to seed a 25 cm² flask. 24 hours later the medium in the plate was replaced with 250 µl freezing mix and the plate stored at -70°C in a polystyrene box. Upon reaching confluence the cells from the 25 cm² flask were pelleted and stored as before for the preparation of DNA.

2.2 Preparation of nucleic acid from cultured cells.

2.2.1 Cell lysate DNA preparation for PCR.

This method was used to prepare crude cell lysate DNA from small quantities
of cells for PCR analysis (Joyner et al., 1989).

Cells were pelleted in an eppendorf and resuspended in 100 µl H₂O. The suspension was incubated on dry ice for up to 2 hrs, then 94°C for 8 mins. 2 µg Proteinase K was added and the suspension incubated at 94°C for 6 mins to destroy the proteinase K. The suspension was stored on ice for up to 30 mins before dividing between two PCR reaction.

2.2.2 Large scale DNA preparation.

Adapted from R. Lovell-Badge's method in Robertson (1987).

Cells were harvested from a 25 cm² flask, washed in PBS and pelleted at 1000 rpm for 5 minutes. Pellets could be stored at -20°C at this stage. The pellets were resuspended in 2 ml ice-cold Tris EDTA (50 mM Tris, pH 8 and 100 mM EDTA). 112 µl 4 M NaCl and 30 µl 20 mg/ml Proteinase K was added. The contents were mixed gently. 0.6 ml 10% SDS was added and the tubes rocked gently. The tubes were incubated at 50°C for 2 hours or at 37°C overnight. The contents were transferred to polypropylene tubes for one phenol (equilibrated in TE (10 mM Tris, pH 8 and 1 mM EDTA, pH 8)), one phenol/chloroform (1:1) and one chloroform extraction. 2 volumes of ethanol were added to the final aqueous layer and DNA precipitated by rocking the tube gently. The DNA was looped out using a sealed glass pasteur pipette and resuspended in TE buffer (pH 8).

The DNA was left to resuspend overnight, shaking at 4°C.

The OD readings at wavelength 260 nm were made to determine DNA concentration. An OD₂₆₀ absorbance reading of 1 corresponds to 50 µg/ml double-stranded DNA.

2.2.3 RNA preparation.

Cells were harvested, washed in PBS and pelleted, as above.

1 ml solution D (4 M guanidium thiocyanate, 25 mM sodium citrate (pH 7), 0.5% sarcosyl and 0.1 ml 2-mercaptoethanol) was added to the pellet, which was
then homogenised in a glass-teflon homogeniser. The suspension was transferred to a polypropylene tube and 0.1 ml 2 M sodium acetate (pH 4), 1 ml water saturated phenol and 0.2 ml chloroform-isoamyl alcohol (49:1) were added sequentially. The final suspension was shaken vigorously for 10 seconds and cooled on ice for 15 mins. After centrifugation at 10000 rpm for 20 mins at 4°C, the RNA-aqueous phase was transferred to a fresh tube. The RNA was precipitated by the addition of 1 ml isopropanol and placing at -20°C for at least 1 hr. The RNA was pelleted by centrifugation at 10000 rpm for 20 mins. The RNA pellet was dissolved in 0.3 ml solution D, transferred to an eppendorf and reprecipitated with 0.3 ml isopropanol at -20°C for 1 hr. After centrifugation for 10 mins at 4°C, the RNA pellet was washed in 75% ethanol, vacuum dried and resuspended in 0.2 ml DEPC H₂O (see Section 2.7.1).

The OD readings at 260 nM were made to determine RNA concentration. An OD₂₆₀ absorbance reading of corresponds to 40 μg/ml.

2.3 Bacterial cell culture.

2.3.1 Media and additives

**L-Broth and Agar** 10 g tryptone (Difco), 5 g yeast extract (Difco), 10 g NaCl, 2.46 g MgSO₄ per litre of distilled water.

15 g agar (Oxoid Ltd.) was added per litre of broth for L-agar.

Media was sterilized by autoclaving.

**Terrific Broth** 12 g tryptone, 24 g yeast extract, 4 ml glycerol per 900 ml distilled H₂O, autoclaved. To this was added 100 ml autoclaved phosphate solution (0.1 M KH₂PO₄ and 0.72 M K₂HPO₄)

**Ampicillin** Stock solution of ampicillin (Sigma) was made up at a concentration of 25 mg/ml, filter sterilized and stored at -20°C. Ampicillin was added to media and agar at a concentration of 50 μg/ml.

**X-gal** Stock solution of X-gal (Boehringer Mannheim) was made up at a concentration of 20 mg/ml in DMSO and stored, protected from light at -20°C.
2.3.2 Bacterial host strains.

**JM83** ara, Δ(lac-proAB), rspL, FSO, LacZΔM15, (rk+mk+)

(Vieira and Messing, 1982). This host strain was used for Bluescribe or pUC plasmids.

**GM161** F-, thr1, leu6, thi1, supE44, lacY1, tonA21, dam4, hss2(r-m-)

The bacterial dam gene encodes an enzyme that methylates the N6 position of the sequence 5′GATC3′ (Hattman et al., 1978) The restriction enzyme Bcl I will not cut DNA when this sequence is methylated. Plasmid to be cut at the Bcl I site was grown in dam- GM161.

The stock of GM161 was obtained from Dr. Robin Allshire who had obtained the strain from Dr. J. Bishop, Department of Genetics, University of Edinburgh.

2.3.3 Vectors and/or clones available for cloning constructs in bacteria.

Bluescribe (Stratagene) is one of the pUC series of plasmid vectors based on pBR322 (Bolivar et al., 1977). The bacterial origin of replication allows replication in bacteria to a high copy number and the ampicillin resistance (AmpT) gene (β-lactamase) (Vieira and Messing, 1982) allows for selection of plasmid transformation. The polylinker is inserted into the β-galactosidase gene (lac Z). Transformants containing non-recombinant plasmid grew as blue colonies on X-gal plates due to the enzymatic action of lac Z on X-gal to produce a blue indolyl derivative. Transformants containing recombinant plasmid in which the insert has inserted into the polycloning site, and therefore disrupted the lac Z gene, grow as white colonies. Sequencing of insert is facilitated by the presence of sequence homologous to the forward and reverse sequencing primers (Section 2.15).

**pUC 13/18** (Kay and McPherson, 1987) is a vector containing a symmetric polylinker (constructed by fusion of Xmn I-Hind III of pUC13 to Hind III-Xmn I of pUC18) with unique Hind III site in the centre which enables the insertion of DNA into a site to become flanked by pairs of sites. This vector retains the lac Z gene, origin of replication and AmpT gene.
pMC1neo PolyA (Thomas and Capecchi, 1987) (Stratagene) contains the bacterial transposan Tn5 neo gene (with synthetic sequence to substitute the weak native translation initiation signal) attached to the HSV-tk gene promoter, including duplication of a synthetic 65 bp fragment from the Polyoma virus enhancer and a polyadenylation signal. This neo vector is found to be highly efficient at transforming ES cells to G418 resistance. The plasmid retains an origin of replication and Amp gene for amplification in bacterial cells.

SDM Blueneo (Dorin et al., 1989) contains a promoterless neo gene attached to the HSV-tk polyadenylation signal in bluescribe. The stop codon immediately 5' to the initiating codon has been removed to allow translational readthrough from sequence placed in-frame with the neo sequence.

pSV2neo contains the neo gene attached to the SV40 early region promoter and polyadenylation signal along with the small t intron of SV40.

pUC18HSV-tk contains the Herpes simplex virus type 1 thymidine kinase gene (Wagner et al., 1981) 2 kb Pvu II fragment cloned into the Hinc II site of the polylinker in pUC18 (obtained from Dr. R. Hill). The Bam HI site of the polylinker is positioned 5' to the gene.

pBLCAT2 (Luckow and Schutz, 1987) contains the bacterial chloramphenical acetyltransferase gene (CAT) attached to the HSV-tk promoter and SV40 t intron and polyadenylation signal in pUC18. This vector was used for the construction of CAT constructs to test for the regulatory effects of cloned DNA elements on a heterologous eukaryotic promoter.

pBLCAT6 is different from pBLCAT2 in that it does not contain a promoter. This vector was used for the construction of CAT constructs to test for promoter function of cloned DNA.

Hox-2.1 subclones covering the region from the EcoR1 site in Hox-2.2 to the Bam H1 site downstream of Hox-2.1 were available cloned into Bluescribe. The subclones used throughout the project are shown in Figure 3.1.
2.3.4 Preparation of competent JM83/GM161.

This method was obtained from Sambrook et al., (1989), pages 1.76-1.81. Transformation efficiencies of 5 X 10^7 to 1 X 10^8 transformed colonies per µg supercoiled plasmid DNA may be obtained by this method.

A single colony was used to inoculate a 10 ml overnight culture of L-Broth, shaking 37°C. 0.4 ml of this culture was used to inoculate 20 ml L-Broth. This culture was grown to OD_{530} = 0.3. 10 ml was used to inoculate 200 ml psi broth (L-Broth containing 10 mM NaCl and 25 mM KCl). This culture was grown to OD_{530} = 0.48.

The culture was incubated on ice for 15 mins and centrifuged at 5K for 5 minutes at 4°C. The pellet was resuspended in 60 ml filter-sterilized TFB1 (100 mM RbCl₂; 50 mM MnCl₂; 10 mM CaCl₂; 35 mM NaOAc and 15% v/v glycerol) (pH 5.8) and incubated on ice for 15 minutes. The cells were pelleted as before and resuspended in 8 ml filter-sterilized TFB2 (10 mM MOPS (pH 6.8), 10 mM RbCl₂, 75 mM CaCl₂ and 15% v/v glycerol). After incubation on ice for 20 mins 200 µl aliquots were snap frozen in liquid nitrogen and stored at -70°C.

2.4 Manipulation of plasmid DNA.

2.4.1 Transformation of competent JM83/GM161.

An aliquot of frozen competent cells was thawed on ice. Plasmid DNA to be transformed was added and the cells left on ice for 20 mins. The cells were then heat shocked at 42°C for 90 secs, incubated on ice for 1 min and 0.5 ml L-Broth were added. After 30 mins incubation at 37°C 100 µl aliquots were spread onto L-agar plates containing ampicillin and if required X-gal.

2.4.2 Small scale plasmid DNA preparation (alkali lysis).

Adapted from Birnboim and Doly (1979).

5-10 ml Terrific Broth, containing ampicillin, was inoculated with a single colony and incubated, shaking at 37°C for 6 hrs or overnight. Cells from 1.5 ml
culture were pelleted in an eppendorf. (Remaining culture was stored at 4°C for preparation of glycerol stock if necessary). The supernatant was discarded and the pellet resuspended by vortexing in 100 µl solution 1 (50 mM glucose, 0.05 M EDTA (pH 8) and 25 mM Tris-HCl (pH 8)), containing freshly added lysozyme at 2 mg/ml). The tube was incubated at room temperature for 10 mins. 200 µl solution 2 (0.2 M NaOH and 1% SDS) was added and the tube rocked gently from side to side before incubation on ice for 10 mins. 150 µl 3 M NaOAc (pH 5.5) was added, the tube vortexed and incubated on ice for 10 mins. After centrifugation for 10 mins the supernatant was transferred to a fresh eppendorf and neutralised by the addition of 5 µl 5 M NaOH.

One phenol extraction, one phenol:chloroform (1:1) extraction and one chloroform extraction was performed in succession. The resulting aqueous layer was transferred to a fresh eppendorf and plasmid DNA precipitated by the addition of 1 ml ethanol. The DNA was pelleted by centrifugation at 4°C for 15 mins and washed in 70% ethanol before drying under vacuum. The DNA was resuspended in 50 µl TE and RNase (previously heat treated to destroy DNases) was added to a final concentration of 0.2 mg/ml. 5 µl was digested with the appropriate restriction enzyme and visualised on an ethidium bromide-stained agarose gel.

Upon identification of a desired clone, a glycerol stock was prepared by vortexing a 1:1 mixture of culture and sterile glycerol in a Nalgene cryotube. The glycerol stock was stored at -70°C.

2.4.3 Large scale plasmid DNA preparation.

A single colony or 10 µl glycerol stock was used to inoculate 10 ml Terrific Broth, containing ampicillin. After shaking at 37°C for at least 6 hrs this culture was used to inoculate 200 ml Terrific Broth, containing ampicillin and the flask shaken at 37°C overnight.

The cells were pelleted by centrifugation at 6000 rpm for 10 mins. The supernatant was discarded and the pellet resuspended in 20 ml solution 1 (Section
2.4.2), containing freshly added lysozyme at a concentration of 10 mg/ml. After incubation at room temperature for 5 mins, 40 ml solution 2 was added and mixed gently by rocking from side to side. After incubation on ice for 10 mins 30 ml solution 3 (3M KOAc, ph 4.8 with acetic acid) was added and the mixture vortexed. A further incubation on ice for 30 mins was followed by centrifugation at 12000 rpm for 30 mins. The supernatant was poured through muslin into a fresh centrifuge tube and 0.6 volumes isopropanol added. After incubation at room temperature for 15 mins the precipitate was pelleted by centrifugation at 12000 rpm for 30 mins. The pellet was washed in ether, dried under vacuum and resuspended in 3 ml TE. 3 g CsCl and 120 μl (10 mg/ml) ethidium bromide were added and the mixture transferred to a 3.5 ml ultracentrifuge tube. The tube was heat-sealed.

The plasmid DNA was banded by centrifugation on a Beckman tabletop ultracentrifuge at 100000 rpm for at least 3 hours. The plasmid band (visible under UV due to intercalation of ethidium bromide) was transferred using a syringe and needle to an eppendorf tube. The volume was measured and named x.

The ethidium bromide was removed by a series of extractions with water-saturated butan-2-ol, until the organic phase was colourless. The final volume of aqueous phase was measured and named y. x volumes H₂O and y volumes isopropanol were added to precipitate the plasmid DNA. After incubation at -20°C the DNA was pelleted by centrifugation, washed in 70% ethanol, dried under vacuum and resuspended in 0.5 ml TE. DNA concentration was determined by OD analysis as before (Section 2.2.2).

2.5 Manipulation of DNA by enzymes.

2.5.1 Restriction enzyme digestion of DNA.

Digestion of plasmid or genomic DNA by restriction enzymes was carried out under the conditions (buffers provided and temperature) recommended for complete digestion by the manufacturer (Boehringer Mannheim).

When DNA was to be double-digested with two enzymes an enzyme buffer
favourable to both enzymes was used. If conditions required were dissimilar the DNA was first digested with the enzyme requiring low salt and the concentration adjusted by the addition of NaCl for the second enzyme digestion. One unit of enzyme activity is defined as the amount of enzyme that will cut at all its specific sites in 1 μg of lambda DNA in 1 hr.

5-10 μg genomic DNA was digested in a 40 μl volume with 3 units enzyme per μg DNA for 4 hrs minimum.

Heat-sensitive enzyme reactions were stopped by heating to 65°C for 10 minutes while heat-stable enzyme activity was stopped by phenol extraction.

Digests to be loaded onto agarose gels were stopped by the addition of loading dye, containing EDTA (100 mM EDTA (pH 8), 20% w/v Ficoll and orange G).

2.5.2 Dephosphorylation of DNA.

Cloning vector linearised with a single enzyme for the ligation of an insert with compatible ends was dephosphorylated using bacterial alkaline phosphatase (BAP) to remove the 5' phosphates and thus prevent self-religation.

10 μg plasmid DNA was digested with the desired restriction enzyme in a 50 μl volume. A 2 μl aliquot was visualised on an ethidium bromide-stained agarose gel to ensure digestion was complete before the enzyme action was stopped. 5 or 10 μl 10X BAP buffer (0.5 M Tris (pH 8) and 0.5 M NaCl) was added depending on whether the digest had been heated or extracted on the previous step. 2 units BAP was added and the volume made up to 100 μl with distilled H₂O. The reaction was incubated at 65°C for 1 hr. The BAP was destroyed by the addition of Proteinase K at a final concentration of 100 μg/ml and incubation at 37°C for 30 mins. The DNA was extracted once with phenol, once with chloroform and precipitated with 1/10 volume 3 M NaOAc (ph5.5) and 2.5 volumes ethanol. The DNA was pelleted by centrifugation, washed in 70% ethanol, dried under vacuum and resuspended in 100 μl TE.
2.5.3 Conversion of 5' overhang to blunt end.

5' overhangs were filled in using the 5' to 3' polymerase activity of the Klenow fragment of *E. coli* DNA polymerase I.

After digestion with the required restriction enzyme, the enzyme activity was stopped either by heat or phenol extraction. 1 μl 200 μM dNTPs, 1 μl 10 X Klenow buffer (0.5 M Tris (pH7.6) and 0.1 M MgCl₂) and 1 unit Klenow enzyme were added to 1 μl digested DNA (1mg/ml) in a volume of 10 μl. The reaction was incubated at 37°C for 30 mins. The Klenow activity was stopped by heat.

2.5.4 Phosphorylation of double-stranded oligonucleotide linker.

T4 polynucleotide kinase (PNK) catalyses the transfer of the γ-phosphate from ATP onto the terminal 5'-OH group of DNA.

The following were added to 1 μl linker (100 ng/μg): 1 μl 10 X PNK buffer (50 mM Tris (pH 8), 10 mM MgCl₂ and 5 mM DTT), 0.5 μl 0.5 mM ATP and 5 units PNK in a 10 μl volume. The reaction was incubated for 37°C and the kinase destroyed by heat.

2.5.5 Ligation of DNA.

T4 DNA ligase catalyses the ligation of DNA ends by forming a phosphodiester bond between the 3'-OH group of one end and the 5' phosphate of another end (Weiss, 1968). This enzyme was therefore used to ligate DNA molecules under conditions which preferentially generate self-ligated products (inverse PCR) or the ligation of a DNA insert into plasmid.

The buffer conditions used were 66 mM Tris (pH 7.6), 5 mM MgCl₂, 5 mM DTT, and 1 mM ATP.

For inverse PCR ligations, 5 μg digested genomic DNA was ligated in a 1 ml volume with 10 units ligase (Boehringer Mannheim) and the reaction incubated at 16°C overnight.

For insert into plasmid ligations, 100 ng insert and 20 ng plasmid were
incubated together in a 20 µl volume containing 1 unit ligase at 16°C overnight for blunt-end ligation or room temperature for 2 hours for sticky-end ligation.

2.6 Amplification of DNA by the polymerase chain reaction (PCR).

PCR enables the selective amplification of a specific DNA sequence between two hybridising oligonucleotides (Saiki et al., 1985). All PCRs were carried out using Taq DNA polymerase (Promega), a thermostable DNA polymerase isolated from *Thermus aquaticus* (Saiki et al., 1988).

2.6.1 PCR on plasmid or genomic DNA.

Reactions were carried out in a 50 µl volume containing: 5 µl 10 X buffer (supplied by Promega), 50 pmoles of each oligonucleotide primer, 0.05 mM dNTPs, 1 µg genomic DNA template (or 10 ng plasmid DNA) and 1 unit Taq polymerase. The reaction volume was overlaid with liquid paraffin.

The reactions were performed on a Techne PHC-2 PCR machine. The reactions were heated to 94°C for 2 mins before 30 cycles of 94°C denaturing for 15 seconds, annealing at the calculated annealing temperature for 15 seconds and 72°C extending for 1 min/kb to be amplified. The annealing temperature was determined by the GC richness of the oligonucleotides by the following equation: 5°C below the melting temperature (*T_m*) where *T_m* at 1 M Na+ concentration = 4(G+C) + 2(A+T) (Wedell et al., 1990). 5 µl product was visualised on an ethidium bromide-stained agarose gel.

2.6.2 Inverse PCR.

This method allows the amplification of unknown sequence flanking known sequence (Ochman et al., 1988). Linearised DNA is ligated under conditions which promote the formation of circles and PCR primers orientated in opposite directions on the linear DNA can now amplify across the junction of the circle.

5 µg genomic DNA was digested with the chosen restriction enzyme in a 30 µl
volume. The enzyme activity was stopped by heating to 65°C. The volume was made up to 1 ml with ligation mix and 10 units T4 DNA ligase added. The reaction was incubated at 16°C overnight. The ligase was destroyed by heating to 65°C for 10 mins. The DNA was ethanol precipitated, pelleted and dried under vacuum. 1 µg DNA was added to 50 µl PCR mix and 30 cycles PCR amplification were performed.

10 µl product was visualised on an ethidium bromide-stained agarose gel. The gel was blotted and the blot hybridised with an end-labelled oligonucleotide, internal to the expected product. The band hybridising with the internal oligonucleotide was cut from a LMT agarose gel and 30 cycles of PCR performed on 1 µl of the gel slice, either between the original primers or between one primer internal to the product and the second primer.

2.7 Gel electrophoresis of nucleic acid.

2.7.1 Solutions.

**20 X TBE.**

216 g Tris (1 M), 18.6 g EDTA (1M) and 110 g boric acid (20mM) made up to 1 L with distilled H$_2$O and autoclaved to prevent precipitation.

**5 X MOPS buffer.**

0.2M MOPS (pH 7)

50 mM sodium acetate

5 mM EDTA (pH 8)

made up to 1 L with distilled H$_2$O and autoclaved.

**DNA loading buffer/stop mix.**

100 mM EDTA (pH 8), 20% ficoll and enough orange G for visualisation on gel.

**RNA loading buffer.**

50% glycerol, 1 mM EDTA (pH 8), 0.4% bromophenol blue, 0.4% xylene cyanol blue.
DEPC treated water.

Water used for RNA was treated by adding diethyl pyrocarbonate (DEPC) to a final concentration of 0.5 v/v and leaving overnight before autoclaving.

2.7.2 Agarose gel electrophoresis for resolving DNA molecules.

Agarose gels were used to resolve DNA fragments from 100 bp to 30 kb using concentrations of agarose from 2% to 0.8% in 1 X TBE buffer. 4% Nusieve gels were used to resolve 100 bp to 400 bp DNA fragments. Ethidium bromide was added to the agarose at a concentration of 0.5 µg/ml.

Samples were prepared by adding 1/10 volume loading buffer and loaded onto the gel in the tank. Gels were run in 1 X TBE at 3-12 V/cm. Hind III digested lambda DNA (BRL) or Hae III digested φX DNA (BRL) were used as molecular weight markers.

DNA was fractionated on LMT agarose for the isolation of fragments for ligation, further amplification by PCR, Genecleaning for direct PCR sequencing or for random prime labelling as hybridisation probes.

2.7.3 Purification of DNA from agarose.

DNA was purified from LMT agarose by Geneclean (BIO 101 Inc.). The procedure was as specified by the manufacturer.

DNA was recovered from type II agarose by centrifugation through blotting paper (Weichenhan, 1991). This method was suitable for inserts for ligation. A square piece of Whatman 3 MM paper (2 x 2 cm) was wrapped round the blunt end of a micropipette tip to form a beaker and was inserted into a punctured eppendorf tube. The agarose slice was placed inside the blotting paper. The tube was mounted into a second eppendorf and centrifuged for 1 min. The DNA, in TBE, was collected in the lower eppendorf. 1 µl of this was used for ligation reactions.

2.7.4 Formaldehyde gel electrophoresis of RNA.

RNA gels were run according to the method of Sambrook et al. (1989), 7.43-
7.45.

1.2% agarose gels were prepared by melting 1.34 g agarose in 70 ml H₂O, cooling to 60°C and adding 22.4 ml 5 X MOPS buffer and 20 ml deionised formaldehyde.

RNA samples were prepared by incubation of 4.5 µl (up to 20 µg) in 2 µl 5 X MOPS buffer, 3.5 µl formaldehyde and 10 µl deionised formamide, at 55°C for 15 mins. 2 µl loading buffer was added before loading onto the gel. Gels were run in 1 X MOPS buffer at 60 V.

1 µl of 1 mg/ml ethidium bromide was added to the RNA sample if visualisation was required.

2.7.5 Polyacrylamide gel electrophoresis.

Denaturing polyacrylamide/urea gels were used to resolve DNA sequencing reactions. Polyacrylamide gels were used to resolve gel retarded products.

The glass plates for the sequencing gel were cleaned thoroughly in detergent and water. The surface of the top/eared plate was wiped with distilled H₂O and then dimethyl dichlorosilane. The plate was left to dry in the fume hood. The bottom plate was wiped with ethanol and allowed to dry. Spacers were sandwiched between the plates and the edges and bottom of the plates sealed with tape.

40 ml instagel (6% acrylamide:N,N'-methylenebisacrylamide (19:1)), 7.5 M urea and 1 X TBE) was mixed with 0.33 ml 10% ammonium persulphate and 50 µl TEMED. This mixture was poured between the plates and a sharkstooth comb inserted before the gel was allowed to set. The gel was run at a limiting power of 26 W for 1 to 4 hrs.

Gel retard experiments were run on a Pharmacia LKB 2001 vertical 16 cm (1.5 mm thickness) electrophoresis system. Plates were washed in ethanol before assembling as manufacturers instruct.

40 ml 10%, 0.5 X TBE gels were made up from 19:1 stock acrylamide:N,N' methylene bisacrylamide with 400 µl 10% ammonium persulphate, TBE and H₂O.
The gel was run at a limiting voltage of 150-200 V in 0.5 X TBE.

After electrophoresis the tape and top plate were removed and the gel immersed in fix (12% methanol and 10% acetic acid) for 10 mins. The fix was drained from the plate and the gel transferred to a piece of 3MM paper by laying the paper on top and peeling off slowly. The gel was covered with clingfilm and dried under vacuum at 80°C on a gel drier. The gel, with clingfilm removed, was exposed to X-ray film at room temperature.

2.8 Transfer of nucleic acid to nylon membrane.

2.8.1 Southern transfer of DNA.

DNA was transferred from agarose gels to nylon membranes (Hybond-N, Amersham) by the capillary blotting method outlined by Southern (1975).

Gels, aligned with a ruler, were photographed on a UV transilluminator. The DNA was denatured by shaking the gel in denaturing solution (1.5 M NaCl and 0.5 M NaOH) for 45 mins and neutralised by shaking in neutralising solution (1.5 M NaCl, 0.5 M Tris and 1 mM EDTA) (pH 5.5) for 45 mins.

The gel was placed on a wick made from Whatman 3MM paper pre-soaked with 20 X SSC (3 M NaCl and 0.3 M sodium citrate) (pH 7) and dipped into a reservoir of 20 X SSC. Exposed regions of the wick were covered with clingfilm. A piece of Hybond-N was cut and placed flat on top of the gel. 2 pieces of 3MM paper were cut, presoaked in 20 X SSC and rolled flat on top of the Hybond-N. A 2 inch layer of paper towels, a glass plate and a weight were placed on top.

Genomic DNA blots were left for a minimum of 5 hrs.

The DNA was cross-linked to the membrane by exposure to UV (3 mins on UV transilluminator or 1200 μjoules from stratalinker (Stratagene)) and baking at 80°C for 30 mins.

Hybridised DNA may be removed from Hybond-N filters for re-probing. Filters were stripped by immersing in boiling 0.01 M Tris (pH 8) and 0.1% SDS and shaking until the solution had cooled to room temperature.
2.8.2 Dry-blotting of DNA.

This method was used as a quick way of transferring plasmid or PCR product from mini-gel to nylon. The mini-gel was denatured and neutralised as above and placed on a piece of clingfilm. A piece of Hybond-N was laid on top, followed by 3MM paper and paper towels as before. The blot was left for a minimum of 1 hour. DNA was crosslinked by exposure to UV (1200 μjoules from the stratalinker).

2.8.3 Transfer of plasmid DNA from bacterial colonies to nylon membrane.

This method is taken from Sambrook et al. (1989), pages 1.90-1.99. It was used to screen large numbers of bacterial colonies for recombinant plasmid.

Bacterial colonies were streaked in duplicate, using sterile toothpicks, onto a gridded master L-agar plate and onto a gridded circle of Hybond-N, which was laid flat on an L-agar plate. Both plates were incubated at 37°C for a minimum of 6 hours until bacterial colonies were observed to grow. The master plate was stored at 4°C until required.

The Hybond-N circle was transferred to a piece of Whatman 3MM paper presoaked in 10% SDS for 3 mins. Then it was transferred to a piece pre-soaked in denaturing solution, left for 5 mins and transferred to a piece pre-soaked in neutralising solution for 5 mins. The filter was rinsed in 3 X SSC. The DNA was cross-linked as before and hybridised to an end-labelled oligonucleotide specific for the insert in the recombinant plasmid. Upon identification of a recombinant clone a culture was grown from the streak on the master plate corresponding to the positive hybridising colony on the filter.

2.8.4 Northern blotting of RNA.

RNA formaldehyde gels were rinsed twice for 20 mins in H₂O. The RNA was then capillary blotted onto Hybond-N, gel turned upside down, on an apparatus similar to that described for Southern blotting. 10 X SSC was used for the transfer.
2.9 Radiolabelling of DNA.

2.9.1 Random primed oligonucleotide-labelling with Klenow.

DNA probes excised from LMT gels were radiolabelled with [α-32P] dCTP by random priming from hexadeoxyribonucleotides using the Klenow fragment of E. coli DNA polymerase I (Feinberg and Vogelstein, 1983 and 1984).

1 ml distilled H2O per gram of DNA agarose slice was added and boiled for 10 mins to denature. At this stage the DNA probe could be stored at -20°C and melted at 65°C before subsequent labellings. Total plasmid probes (25 ng in 11 μl H2O) were boiled for 10 mins.

11 μl (approximately 30 ng) DNA probe was labelled in a 20 μl volume using the Boehringer Mannheim kit. 1 μl each dATP, dGTP, dTTP, 2 μl reaction mix (containing hexanucleotides), 30 μCi [α-32P] dCTP and 2 units Klenow were added. The reaction was incubated at 37°C for a minimum of 30 mins.

Unincorporated dCTP was removed by either spinning through a sephadex G50 column (1600 rpm for 5 mins) in 3 X SSC or by passing through a Nick column (Pharmacia) following the instructions given by the manufacturer.

2.9.2 End labelling of DNA oligonucleotides.

Oligonucleotides were end labelled by the transfer of the α32P-labelled γ-phosphate from [γ-32P] ATP onto the terminal 5'-OH group.

50 ng oligonucleotide was labelled in a 20 μl volume containing 1 X PNK buffer (5 mM Tris (pH 8), 1 mM MgCl2 and 0.5 mM DTT), 50 μCi [γ-32P] ATP and 10 units PNK. The reaction was incubated at 37°C for 30 mins and stopped by heating to 65°C for 10 mins.

2.10 Nucleic acid hybridisation.

2.10.1 Hybridisation solutions.

Quick oligonucleotide hyb mix

5 X SSC, 0.05% w/v BSA, 0.05% w/v Ficoll, 0.1% w/v SDS, 0.05% w/v
polyvinyl pyrrolidone, and 0.1% w/v sodium pyrophosphate.

Hybridisation fluid

4 X SSC, 0.2% w/v SDS, 0.1% w/v sodium pyrophosphate, 50 mg/ml denatured salmon sperm DNA, 5 X Denhardts and 2 mM EDTA. This fluid was stored as 50 ml aliquots at -20°C.

100 X Denhardts

2% w/v polyvinyl pyrrolidone, 2% w/v BSA and 2% w/v Ficoll.

Church buffer

(Church and Gilbert, 1984)

7% SDS, 0.5 M NaPO₄ (pH 7.2)

2.10.2 Hybridisation methods

Hybridisation of end labelled oligonucleotide was carried out in quick oligonucleotide hyb mix at the Tₘ-5 of the oligonucleotide. Hybridisation of random primed DNA probes was carried out in the above hybridisation fluid (Southerns) or Church buffer (Northerns) at 68°C.

Filters were prehybed in hybridisation solutions in sealed plastic bags for a minimum of 30 mins at 68°C, submerged in a shaking water bath.

End labelled oligonucleotide was added directly to the bag, the bag resealed and the hybridisation incubated in a shaking water bath for a minimum of 4 hrs. Filters were washed in 4 X SSC and 0.1% SDS, 2 washes of 10 mins at room temperature and one wash for 2 mins at hybridisation temperature.

Random primed probes were boiled for 10 mins to denature before adding to the bag. One random primed reaction was usually added to two large filters in 30 ml hybridisation fluid. The hybridisation was incubated in a shaking water bath for a minimum of 16 hrs. Filters were washed at 68°C once for 20 mins in 3 X SSC, 0.1% SDS, once for 20 mins in 2 X SSC, 0.1 % SDS and once for 20 mins in 1 X SSC, 0.1% SDS.

53
2.11 Autoradiography.

Filters were dried on 3MM paper and wrapped in clingfilm.

$^{32}$P labelled filters were exposed to Kodak X-AR5 autoradiograph film in cassettes, which had tungsten intensifying screens, at -70°C. Filters were stored from 20 mins up to 2 weeks before developing the film using an automatic X1 X-GRAPH processing machine.

2.12 DNA sequencing.

2.12.1 Double-stranded DNA sequencing of plasmid DNA.

DNA sequencing was performed using the chain termination method of Sanger et al. (1977). The Sequenase version 2.0 kit (United States Biochemical) was used to perform reactions on double-stranded DNA template.

1-2 µg miniprep or maxiprep plasmid DNA was denatured in the presence of 2.5 ng oligonucleotide primer in 0.2 M NaOH. After 2 mins this was neutralised by the addition of 1.4 M ammonium acetate (pH 7.6) and the DNA precipitated with 4 volumes ethanol. After 15 mins at -20°C the DNA was pelleted, washed in 70% ethanol and dried under vacuum. The primer will have become annealed to the template.

The annealled template-primer was resuspended in 10 µl distilled H$_2$O. 2 µl 5 X sequenase buffer (200 mM Tris (pH7.5), 100mM MgCl$_2$ and 250 mM NaCl), 1 µl 0.1 M DTT, 2 µl 1 X labelling mix (0.75 µM dGTP, dCTP and dTTP), 5 µCi [$\alpha$-$^{35}$S] dATP and 2 µl Sequenase enzyme (diluted 1:8 in enzyme dilution buffer) were added. The reaction was incubated at room temperature for 5 mins, during which time primer is extended by the polymerase activity of Sequenase and $^{35}$S incorporated into the newly synthesized strand. 3.5 µl aliquots were added to 4 tubes labelled T,C,G and A, each of which contained 2.5 µl of the corresponding T,C,G or A termination mix (50mM NaCl, 80 µM of three dNTPs and 8 µM of the limiting ddNTP). The reactions were incubated at 37°C for 5 mins before the addition of 4 µl stop mix. The newly synthesized strand of labelled DNA terminates with a ddNTP,
which prevents further extension due to the lack of a 3'-OH. Each tube contains varying lengths of labelled DNA molecules terminated at each complementary base to the ddNTP in the sequence (within a few hundred bases of the primer).

The DNA was denatured by incubation at 80°C for 2 mins before running the 4 reactions side by side on a polyacrylamide/urea sequencing gel.

To obtain sequence information close to the primer 1 µl Mn buffer (0.15 M sodium isocitrate and 0.1 M MnCl₂) was added to the labelling reaction with the Sequenase. The addition of Mn²⁺ to normal (Mg²⁺) sequencing reactions reduces the length of DNA synthesized and therefore intensifies the bands corresponding to sequence close to the primer. Sequencing gels for the visualisation of such reactions were run until the bromophenol blue dye had run 3/4 the length of the gel.

2.12.2 Direct sequencing of PCR products.

2/5 PCR product (20 µl) was excised from an agarose gel and Genecleaned prior to sequencing using the Sequenase version 2.0 kit (Winship, 1989).

3-5 µl Genecleaned PCR product was boiled for 3 minutes with 140 ng primer (either PCR primer or one internal to amplified product), 2 µl sequenase buffer, 1 µl DMSO in a 10 µl volume. The tube was snap-freezed by dropping into liquid nitrogen.

The DDT, dilute labelling mix, [α-³⁵S] ATP and sequenase were added as in plasmid DNA sequencing and the labelling reaction incubated for a maximum of 90 seconds before the addition of 3.5 µl reaction to the four termination mixes. 10% DMSO had been previously added to the termination mixes. The termination reaction was incubated at 37°C for 5 mins as before.

2.13 CAT (chloroamphenical acetyltransferase) assays.

The ability of cis-acting elements to regulate gene expression may be assayed by attachment to a reporter gene such as CAT (Gorman et al., 1982). CAT modifies and inactivates chloramphenical by mono- and diacetylation. The CAT activity was
assayed by thin-layer chromatography (TLC) (Sambrook et al., 1989; 16.56-16.61).

1 X 10^7 ES cells were electroporated with 40 µg CAT construct under the conditions described in section 2.2.3. The cells were plated onto two 10 cm diameter dishes and incubated for 2 days. After two washes in PBS the cells were harvested by scraping into 1 ml PBS. The cell suspension was transferred to an eppendorf and the cells pelleted by centrifugation for 10 seconds. The cells were resuspended in ice-cold PBS and repelleted. The PBS was aspirated from the eppendorf using a vacuum line. The cells were resuspended in 100 µl 0.25 M Tris (pH 7.8) and vortexed. The cells were disrupted by three cycles of freezing in a dry ice/ethanol bath and thawing at 37°C. The cells were pelleted by centrifugation for 5 mins and the supernatant transferred to a fresh eppendorf. 50 µl of the extract was stored at -20°C.

The second 50 µl aliquot was incubated at 65°C for 10 mins to inactivate deacetylases and any debris pelleted by centrifugation. 96 µl CAT reaction mix (56 µl 1 M Tris (pH 7.8), 20 µl 0.25 M (57 µCi/mmol) 14C-labelled chloramphenical and 20 µl 3.5 mg/ml acetyl coenzyme A) was added to the supernatant and the reaction incubated at 37°C for 2 hrs. 1 ml ethyl acetate was added and the mixture vortexed for three periods of 10 seconds each. After centrifugation for 5 mins, 900 µl of the upper phase (containing the acetylated forms of chloramphenicol) was transferred to a fresh eppendorf. The ethyl acetate was evaporated by placing in a rotating SpeedVac for 1 hr.

The reaction products were redissolved in 25 µl of ethyl acetate and applied to the origin of a 25-mm silica gel TLC plate. The plate was placed in a TLC chamber containing 200 ml chloroform:methanol (95:5) and allowed to run until the solvent front had moved 75% of the distance to the top of the gel.

After air drying the plate was exposed to X-ray film at room temperature.

The chloramphenical acetylated spots were scraped from the plate and 5 ml Aquasol 2 scintillation fluid added before liquid scintillation counting.
2.14 Gel retardation.


Cells were harvested by trypsinisation and the trypsin neutralised by washing the cells twice in medium. The cells were washed three times in PBS and pelleted.

The pellet was resuspended in 2.5 ml lysis buffer (10 mM Tris (pH 7.5), 10 mM NaCl, 2 mM MgCl\textsubscript{2} and 0.05% NP40). The lysate was incubated for 5 mins on ice and forced up and down through a 21 gauge needle. Release of nuclei was checked for by inspection under the microscope. The lysate was transferred to a corex tube and the nuclei pelleted by centrifugation at 3000 rpm for 1 min.

Nuclear protein was extracted in the buffer used by Strauss and Varshavsky (1984). The nuclear pellet was resuspended in 0.5 ml buffer (0.35 M NaCl, 5 mM EDTA, 10 mM β-mercaptoethanol, 10 mM Tris-HCl (pH7.5) and 0.2 mM PMSF (proteinase inhibitor). After incubation for 30 mins on ice with occasional vortexing, the suspension was centrifuged for 15 mins and the supernatant transferred to a fresh eppendorf. 15% glycerol was added and the protein extract stored at -70°C.

The protein concentration was measure using a Biorad dye-binding assay. The acidic solution of Coomassie Brilliant Blue G-250 shifts absorbency from 465 to 595 nm in the presence of protein. Bovine gamma globulin was used as a protein standard. The protein extract was found to have a concentration of 1.2 mg/ml.

2.14.2 Gel retardation.

Oligonucleotides were labelled with \textsuperscript{32}P as described previously and annealed by cooling slowly from 65°C. Annealing was checked by running the oligonucleotides on a 12% polyacrylamide gel.

5 µl proteins and 1 µl 1/100 dilution of labelled oligonucleotides (20000 counts per minute/ml) were incubated on ice for 1 hr in gel shift buffer (Frain \textit{et al}., 1989; gifted by Peter Budd) (20 mM Tris HCl (pH 7.6), 12% glycerol, 25 mM KCl, 4 mM MgCl\textsubscript{2}, 4 mM spermidine, 1 Mm EDTA and 0.5 mM BSA) with 4 µg dldC (non-specific competitor) before loading on a 10% polyacrylamide gel. Tracking dye
(50% glycerol, 1 X TBE and 1% bromophenol blue) was loaded into the end tracks. The gel was run at a limiting voltage of 150 V until the bromophenol blue had run 3/4 the length of the gel, fixed and dried as described previously.

2.15 Oligonucleotides.

Most oligonucleotides were synthesized by Doreen Chambers on an Applied Biosystems 381A oligonucleotide synthesizer, with the exception of the forward and reverse sequencing primers (Biolabs) and oligonucleotide 598 (Oswel DNA Service).

Oligonucleotides were precipitated from ammonium stocks with 1/10 volume 3 M sodium acetate and 2.5 volumes ethanol at 20°C for 1 hr. The oligonucleotides were pelleted by centrifugation for 15 mins, washed 3 times in 80% ethanol and dried under vacuum. They were resuspended in 100 μl TE. The concentration was determined by OD analysis, where OD260 = 1 at 20 μg/ml single-stranded DNA.

The oligonucleotides which were used for PCR, probes, sequencing, in CAT constructs or gel retards are listed below:

- neo18 CCTCTCCACCCAAGCGGC  anti-sense neo
- neo19 ACGATCCTCCTCCTGTCTC  anti-sense neo
- 192 AAATCGTGAAGCACAGGG  39-56 Hox-2.1 cDNA sense
- 314 GCTACCAGACCCCTGGAGC  711-729 Hox-2.1 cDNA sense
- 315 GTCAGGTAAGCTTGAAGG  745-763 Hox-2.1 cDNA anti-sense
- 598 TCTTCAGAGAGATCTCCC  spans BglII site 3' to Hox-2.1
- B2  GGTGGCTGCTGTGGGCTA  101 bp 3' to oligonucleotide 598
- B85  GGACTGAGCCAGGCTGAG  5' insertion in 7B2: HTS sequence 40-58
- B218 GGTGAACCTCACAGCCTCCC  3' insertion in 7B2: HTS sequence 435-417
- B407 CCTAGGGTAGATATCAGA  3' insertion in 7B2: HTS 1177-1160
- B460 GAGTCTCGGATATCATAC  3' insertion in 7B2: HTS 481-496
- B461 GGATGAGAAATGTGTTCC  3' insertion in 7B2: HTS sequence 922-905
- -40 GTTTTCCCAGTCAGGAC  forward sequencing primer
rev  AACAGCTATGACCATG  reverse sequencing primer
671  GTAAAAACGACGGCCAGT  Bluescribe primer: -41 to -25
B539  GATCCCAACCGCTCGCAG  F2 locus: 193-209
B267  AGCAACAGTTTCCAGGCTTCCCC  7B2 direct repeat
B268  GGGGAAGCCTGGAAACTGTGGCT  complement of B267
B277  AGCAACTGACTTCAGGCTTCCCC  negative control for B267
B278  GGGGAAGCCTGAAGTCAGTTGCT  complement of B277
CHAPTER 3

TARGETING *HOX-2.1* IN ES CELLS
3. TARGETING HOX-2.1 IN ES CELLS

3.1 Introduction.

The targeting frequency of a vector may be expressed as the number of correctly targeted homologous recombination events as a fraction of the total number of G418-resistant (G418r) clones. Targeting frequency is reduced by random integration. Various parameters have been shown to affect targeting efficiency - in particular the construct design.

There are two basic types of targeting vector - the replacement and the insertion type. Their modes of action are described in Section 1.4 and illustrated in Figure 1.2. Early studies by Thomas and Cappecchi (1987) showed that these vectors had similar, but rather low targeting frequencies of 1/1000 G418r clones. Homologous recombination occurs rarely against a large background of non-homologous integration events.

Two strategies have been used to enrich for homologous recombinant events by reducing the amount of random insertion events surviving selection. By introducing a promoterless neo into the construct (Doetschman et al., 1988; Dorin et al., 1989 and Sedivy and Sharpe, 1989) only integration events that lie close to an active promoter will allow neo expression. This strategy relies on expression of the targeted gene in ES cells. In the alternative positive-negative selection (PNS) strategy (Mansour et al., 1989) the HSV-tk gene is introduced into the construct at the end of target gene sequence homology. Upon homologous recombination and replacement of target DNA, the HSV-tk gene is lost; while random integration of the whole construct results in integration of the HSV-tk gene. Cells in which the HSV-tk gene has integrated are sensitive to the base analogue, gancyclovir (ganc), and therefore, may be selected against. These enrichment strategies have been discussed in more detail in Section 1.4.

During this work, it was reported by Dr. Alan Bradley (personal communication) that insertion vectors had high targeting frequency, comparable to
those reported for PNS vectors. A recent paper by Hasty et al. (1991b) confirms that insertion vectors may be up to 9-fold more efficient than replacement events (with no enrichment). This may be because one cross-over is required instead of two, the proximity of two adjacent homologous free ends or because of uninterrupted homology.

Five different targeting vectors have been used in this project to target *Hox-2.1* in ES cells, based on the replacement, PNS and insertion vectors: a replacement vector with promoterless *neo* (construct 2); two PNS vectors with either the *Hox-2.1* promoter driving *neo* expression (*Hox2.1ptk*) or a promoterless *neo* (*Ω2BBtk*) and two insertion vectors (construct 6 and *HoxTneo*).

L cells were transfected as a test system for the replacement vectors until ES cells were being grown successfully in the laboratory. For the same reason one transfection of ES cells with construct 2 was made in Dr. Robin Lovell-Badge's laboratory. Clones were grown by Robin Lovell-Badge and Michaela Riddel, cells pelleted and posted to Edinburgh for DNA extraction and analysis.

Initially G418^R clones were analysed by PCR amplification between an oligonucleotide (192) upstream of the *Hox-2.1* sequence present in the replacement targeting construct and the sense *neo* primer (*neol8*) on crude cell lysates prepared from a few cells. Only cell lines in which homologous recombination between the incoming vector and the endogenous locus has occurred will have the sequence required for this amplification. However, in testing out the PCR primers on control plasmid (*Ω2*) contamination from PCR product became a problem. Positives were identified which upon further analysis of restriction fragment lengths were shown to be random integration events and therefore false positives on the PCR screen. There have been reports of PCR screens giving false negatives, where homologous recombinants have failed to PCR amplify but have been identified on restriction mapping. This may be caused by an inhibitor of PCR present in the crude cell lysate. PCR may also be subject to artifact from random integration events (see Section 1.4, Mansour, 1990 and Kim et al., 1991). It was decided that all subsequent clones
should be analysed by restriction enzyme mapping at the *Hox-2.1* locus to detect any alteration due to insertion of *neo*. It was thought that enrichment for homologous recombination events by the promoterless *neo* and PNS would reduce the numbers needed for screening. Although this analysis is more time-consuming (since firstly greater numbers of cells are needed and secondly the procedure is lengthier) the results are more reliable.

All the vectors were designed to introduce a null mutation into one copy of *Hox-2.1*.

### 3.2 Replacement vector with promoterless *neo*.

Construct 2 (Figure 3.2d) was cloned from a larger targeting construct, Ω2 (Figure 3.1c/3.2c), which had been previously constructed and tested for targeting efficiency by Dr. Ian Jackson. Ω2 has 400 bp *Hox-2.1* sequence (polylinker Eco R1 of cDNA clone (Figure 3.1b) (obtained from Dr. R Krumlauf) to Bam HI within coding sequence) cloned upstream of a promoterless *neo* gene and 10 kb *Hox-2.1* sequence (Hind III site within intron to 3' Bam HI) cloned downstream. The 1.2 kb *neo* sequence replaces the 600 bp Bam HI to Hind III exon to intron fragment of *Hox-2.1*. Thus a null mutation is introduced into *Hox-2.1* upon homologous recombination.

It was thought at the time of study that the synthetic Bam HI site present at the 5' end of the upstream *Hox-2.1* sequence and that deletion of the 1 kb Hind III fragment from within the 3' *Hox-2.1* sequence in Ω2 might decrease or hinder targeting by this vector. Therefore, the 5 kb Sst I fragment cut from Ω2 and identified as the *neo*-containing fragment by hybridisation to the neo18 oligonucleotide was cloned into the Sst I site of Bluescribe to generate the targeting vector, construct 2. Construct 2 has 100 bp of *Hox-2.1* sequence cloned upstream of *neo* and 4 kb of sequence cloned downstream. Although it has a short region of homology at the 5' end, the synthetic Bam HI site has been removed. Construct 2 was digested with Sst I to release insert, prior to electroporation.
**Figure 3.1**

Subclones available from the *Hox-2.1* locus aligned with *Hox-2.1* genomic restriction map (a) and cDNA (b).

c. $\Omega 2$
d. *Hox2.1-2.2*
e. pH15BR
f. pMI

Enzymes are abbreviated as: E=EcoRI, H=Hind III, B=Bam HI, S=Sst I and Bg=BgII II.
a.

b. 

synthetic BamHI

intron

1 kb

deletion

c.

d.

e.

f.

polylinker KpnI

HSV-tk

polylinker KpnI

g.

h.
£2 (and hence construct 2) contains a promoterless neo derived from the SDM Blueneo of Dr. Julia Dorin (see Section 2.3.2). The stop codon immediately 5' of the initiating codon of the neo has been mutated by site directed mutagenesis to allow readthrough from upstream coding sequence. £2 was constructed by the inframe fusion of neo 3' to 400 bp of Hox 2.1 sequence (ligation of Bam HI Hox-2.1 fragment to Bcl I cut Blueneo). The fused sequence is presented in Figure 3.3. No promoter activity has been identified from the 400 bp Hox-2.1 sequence. In order to obtain neo expression the constructs (£2 or construct 2) must integrate close to an active promoter and allow translation from the neo AUG or produce a functional neo fusion protein.

It has been shown that Hox-2.1 is expressed at a low level in ES cells (Dr. R. Krumlauf, personal communication) by nuclear run-off analysis. Homologous recombination of construct 2 should therefore allow the transcription of Hox-2.1-neo fusion transcript.

Translation may start from the neo AUG or from a 5' AUG if the open reading frame into the neo is maintained. In order to generate G418-resistant clones the latter event must produce a functional neo-fusion protein. The Hox-2.1 AUG is present in £2 but is absent from construct 2 (see Figure 3.3). Upon homologous recombination Hox-2.1-neo fusion protein should be generated.

In order to confirm that the Hox-2.1-neo fusion transcript could provide G418R a third construct (Hox2.1pneo) was made. Hox2.1pneo (Figure 3.2e) contains 1.8 kb of Hox-2.1 sequence (1.7 kb 5' untranslated sequence and 100 bp of coding sequence) fused inframe to the promoterless neo. This 1.8 kb fragment was thought likely to contain the Hox-2.1 promoter function. Hox2.1pneo was constructed by cloning the 1.8 kb Bam HI fragment (gel purified from subclone Hox2.1-2.2 (Figure 3.1d)) into Bcl I linearised and BAP-treated SDM Blueneo. The orientation of the insert in relation to neo was checked by double digestion with Bam HI and Hind III. Numbers of G418R foci generated upon transfection of cells with Bam HI linearised
GAA GTA CAG TGC ATC GCT ATA ATT CAT TAA TAC ATC ATA AAT CGT
GAA GCA CAG GGT TAT AAC GAC CAC GAT CCA CAA ATC AAG CCC
TCC AAA ATC ACC CAA ATG AGC TCG TAC TTT GTA AAC TCC TTC TCG
GGG CGT TAT CCA AAT GGC CCG GAC TAT CAG TTG CTA AAT TAT
GGC AGT GGC AGC TCT CTG AGC GGC TCT TAC AGG GAT CAA GAG
ACA GGA GGA TCG TTT CGC ATG neo

Figure 3.3

Hox-2.1-neo sequence fusion in Ω2. neo sequence is in bold print. The T of the TGA translation stop codon -15 bp from the neo initiation codon has been mutated to a G (underlined) to allow in-frame readthrough from the Hox-2.1 initiation codon.

The Hox-2.1 initiation codon (underlined) has been deleted from construct 2 since it is 2 bp 5' to the GAGCTC Sst I site used to subclone construct 2 from Ω2.
Hox2.1pneo were compared to the numbers generated upon transfection with the promoterless neo vector, construct 2 (tabulated as G418\(^f\) frequency in Table 3.4). The frequency of G418\(^f\) induced by Hox2.1pneo was found to be 80-fold higher than that induced by construct 2. It may be concluded from these data that Hox2.1pneo contains promoter function and that either the Hox-2.1-neo fusion protein is capable of neomycin phosphotransferase activity or that translation starts from the neo AUG.

Construct 2 derived G418\(^f\) clones were analysed for homologous recombination by restriction mapping. The predicted alteration at the Hox-2.1 locus due to the integration of neo and deletion of 600 bp of Hox-2.1 sequence upon replacement of endogenous sequence with construct 2 is shown in Figure 3.4b. DNA extracted from cell lines was restricted with EcoRI, BamHI or Hind III and the blots generated probed with the Hox-2.1 1 kb Sst I fragment (gel purified from Hox2.1-2.2 subclone) from upstream of the construct sequence. Hybridisation to only the endogenous/normal band indicates random integration of the construct. Hybridisation to both the normal band and to the predicted diagnostic band would indicate homologous recombination in that cell line. These two bands will be of equal intensity and half as intense as the normal signal from the Hox-2.1 locus. The predicted fragment band sizes are summarised in Table 3.1. The neo (insert gel purified from pMC1neo polyA) probe should hybridise to the same diagnostic band in homologous recombinants since the neo and Hox-2.1 probes are separated by DNA that is uncut by the chosen restriction enzymes. Upon hybridisation with a neo probe, random restriction fragment lengths are observed in cell lines derived from random insertion events.
Figure 3.4
Predicted alteration to the restriction fragment lengths at the Hox-2.1 locus upon homologous recombination with construct 2 (b), Ω2BBtk (c) or Hox2.1ptk (d) compared to the endogenous Hox-2.1 restriction map (a). Construct sequence is represented by the thick line. The stippled lines represent the probes used to detect the restriction fragments. The enzyme abbreviations are as in Figure 3.1.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Endogenous Hox-2.1</th>
<th>Homologous recombinant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eco RI</td>
<td>5.0</td>
<td>5.6</td>
</tr>
<tr>
<td>Hind III</td>
<td>2.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Bam HI</td>
<td>1.8</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 3.1

The restriction fragment lengths (kb) at the endogenous Hox-2.1 locus and the predicted diagnostic fragment lengths for correct targeting of construct 2, upon hybridisation with the 1 kb Sst I probe 5' to the construct sequence. The same homologous recombinant diagnostic bands will hybridise to the neo probe.

3.3 Replacement vector based on PNS.

Two PNS vectors were constructed.

Ω2BBtk (Figure 3.2f) has the same 400 bp Hox-2.1 sequence segment as Ω2 fused 5' to the promoterless neo and 3.5 kb Hox-2.1 sequence (Hind III in intron to 3' Bgl II) cloned downstream. The HSV-tk gene is positioned 5' to the upstream Hox-2.1 sequence. Ω2BBtk was constructed by cloning the 5.5 kb (neo-containing) Bam HI to Bgl II fragment of Ω2 into Bam HI and BAP-treated pUC18HSV-tk. Ω2BBtk was either linearised with Bam HI or Kpn I. Upon linearisation with Bam HI, the HSV-tk gene and its promoter are positioned at the 3' end of the DNA molecule. Upon linearisation with Kpn I, the HSV-tk gene is positioned centrally, immediately 5' to the promoterless neo. The position of the HSV-tk gene may be important since upon random integration ends of the construct may be lost. HSV-tk positioned with the transcription start site at the end of the molecule may become non-functional upon integration and cell lines resistant to ganc will be generated.

Hox2.1ptk (Figure 3.2g) has the 1.5 kb (5' Hind III to Bam HI in coding sequence) fragment of Hox-2.1 fused inframe with the promoterless neo and the 4
kb (Bam HI in coding sequence to 3' Bgl II) Hox-2.1 fragment cloned downstream. The HSV-tk gene is positioned 3' to the Hox-2.1 downstream sequence. The 1.5 kb of sequence upstream of the neo has been shown to have promoter function (Table 3.4). The cloning steps involved in the construction of Hox2.1ptk are shown in Figure 3.5. Hox2.1ptk was linearised prior to elecroporation by digestion with Kpn I.

The predicted alterations to the Eco RI restriction length fragments caused by homologous recombination of Ω2BBtk and Hox2.1ptk into the Hox-2.1 locus are shown in Figure 3.4c and 3.4d. Homologous recombination leads to replacement of endogenous Hox-2.1 by a null copy, disrupted by the neo gene (as Ω2 and construct 2). G418r gancr cell line DNA was digested with EcoRI and the blots generated hybridised to the 1.5 kb Eco RI to Hind III fragment of Hox-2.1 (purified from Hox2.1-2.2) from upstream of the PNS construct sequences. Hybridisation to the normal band only, indicates random insertion of the construct, while hybridisation to the normal band and the predicted diagnostic band size, would indicate homologous recombination. The fragment lengths are summarised in Table 3.2. Cell lines were checked for integration of the construct by hybridisation of the blots to the neo probe.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Endogenous Hox-2.1</th>
<th>Ω2BBtk homologous recombinant</th>
<th>Hox2.1ptk homologous recombinant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eco RI</td>
<td>5</td>
<td>5.6</td>
<td>6.2</td>
</tr>
</tbody>
</table>

Table 3.2

The restriction fragment length (kb) at the endogenous Hox-2.1 locus and the predicted diagnostic fragment lengths for correct targeting of Ω2BBtk or Hox2.1ptk upon hybridisation with the 1.5 kb Eco RI to Hind III probe 5' to the construct sequence. The same diagnostic homologous recombinant bands will hybridise to the neo probe.
Figure 3.5
Steps involved in the construction of the PNS construct, Hox2.1ptk. Vectors and subclones are described in Section 2.3.3 or shown in Figure 3.1. A line across the enzyme name indicates the site is not reconstituted after the cloning step.
3.4 Insertion vectors.

Two insertion vectors had been previously constructed by Dr. Ian Jackson. Construct 6 (Figure 3.2h) contains the 1.3 kb (Bam HI to Eco RI of exon 1-intron-exon 2) fragment of Hox-2.1 cloned into the Bam HI and Eco RI sites of pSV2neo (Southern and Berg, 1982). HoxTneo differs from construct 6 by absence of the 100 bp exon-intron Hind III fragment from the Hox-2.1 sequence and absence of the Hind III to Bgl II fragment from the pSV2neo sequence.

Construct 6 was linearised through the Xho I site in the Hox-2.1 sequence and HoxTneo was linearised through the Hox-2.1 Hind III site, prior to electroporation. Construct 6 therefore has 360 bp 5' and 900 bp 3' and HoxTneo has 450 bp 5' and 700 bp 3' Hox-2.1 homology.

The predicted alterations in the Bgl II restriction fragment lengths at the Hox-2.1 locus upon integration of construct 6 or HoxTneo are shown in Figures 3.6a and 3.6b. Integration leads to duplication of the construct Hox-2.1 sequence at the locus. However, since the 5' copy of the gene is disrupted by plasmid/neo sequence and since the 3' copy lacks promoter/5' sequence, it is unlikely that functional Hox-2.1 is produced by this locus.

Insertion vector derived G418r cell line DNA was digested with Bgl II and the blots generated hybridised with the Hox-2.1 3 kb Eco RI to Bgl II fragment (gel-purified from subclone pM1 (Figure 3.1c)) downstream of Hox-2.1 sequence within the construct (3' probe). Hybridisation to the normal band only, indicates random integration of the construct, while hybridisation to the normal band and the predicted novel band size, would indicate homologous recombination. The blots were stripped and rehybridised with the neo probe to check for integration of the construct. The predicted fragment lengths are summarised in Table 3.3.
Figure 3.6
Predicted restriction lengths upon homologous recombination into the Hox-2.1 locus by construct 6 (b) or HoxTneo (c). Construct Hox-2.1 sequence is represented by a thick line, plasmid sequence represented by a checked line and probe by a stippled line. The Hox-2.1 cDNA is illustrated as in Figure 3.1. The deleted regions of the gene upon integration of the construct are indicated by the asterisks. Enzymes are abbreviated as in Figure 3.1.
Table 3.3

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Endogenous <em>Hox-2.1</em></th>
<th>Construct 6 homologous recombinant</th>
<th>HoxTneo homologous recombinant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bgl II</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 3.3

The restriction fragment length (kb) for the endogenous *Hox-2.1* locus and the diagnostic fragment lengths for correct targeting of construct 6 or HoxTneo upon hybridisation with the 3 kb Eco RI to Bgl II probe from 3' to the construct sequence are shown in kb. The same diagnostic homologous recombinant band will hybridise to the neo probe. The same diagnostic band should hybridise to the 5' probe in HoxTneo homologous recombinants.

3.5 Frequencies of G418 resistance.

The required concentration of G418 for selection of ES cells was determined for each batch of G418 by selection of cells on varying doses for 14 days. Plates from such a selection and the resulting kill curve are shown in Figures 3.7 and 3.8. 400 µg/ml G418 was chosen as the required dose for this batch. Too high a dose may lead to death of cells containing *neo* but the expression of which is low.

The numbers of G418<sup>r</sup> (or G418<sup>r</sup> and ganc<sup>r</sup> in the case of the PNS vectors) foci produced after electroporation of L cells or ES cells with each of the targeting constructs, Hox2.1pneo or pMC1neo polyA were counted after 10-14 days selection. The frequency of G418<sup>r</sup> is stated in Table 3.4 as a fraction of the number of cells electroporated.
Figure 3.7
Colonies remaining on plates after 10 days selection on 150, 200, 250, 300, 350, 400, 500 μg/ml G418.
Figure 3.8
Kill curve illustrating the % G418' after 10 days selection on varying concentrations of G418.
### Table 3.4

<table>
<thead>
<tr>
<th>Construct</th>
<th>L cell G418&lt;sup&gt;r&lt;/sup&gt;</th>
<th>L cell G418&lt;sup&gt;r&lt;/sup&gt; + ganc&lt;sup&gt;r&lt;/sup&gt;</th>
<th>ES cell G418&lt;sup&gt;r&lt;/sup&gt;</th>
<th>ES cell G418&lt;sup&gt;r&lt;/sup&gt; + ganc&lt;sup&gt;r&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>construct 2 (-p)</td>
<td>3 x 10&lt;sup&gt;-6&lt;/sup&gt;</td>
<td>3 x 10&lt;sup&gt;-6&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hox2.1pneo (+p)</td>
<td></td>
<td></td>
<td>2.5 x 10&lt;sup&gt;4&lt;/sup&gt;*</td>
<td></td>
</tr>
<tr>
<td>Ω2BBtk (-p)</td>
<td>7 x 10&lt;sup&gt;-6&lt;/sup&gt;</td>
<td>6 x 10&lt;sup&gt;-6&lt;/sup&gt;</td>
<td>1.5 x 10&lt;sup&gt;-7&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Hox2.1ptk (+p)</td>
<td>7 x 10&lt;sup&gt;-6&lt;/sup&gt;</td>
<td>1.1 x 10&lt;sup&gt;-6&lt;/sup&gt;</td>
<td>1.5 x 10&lt;sup&gt;-7&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>construct 6 (+p)</td>
<td></td>
<td></td>
<td></td>
<td>1.1 x 10&lt;sup&gt;-6&lt;/sup&gt;</td>
</tr>
<tr>
<td>HoxTneo (+p)</td>
<td></td>
<td></td>
<td></td>
<td>1.3 x 10&lt;sup&gt;-6&lt;/sup&gt;</td>
</tr>
<tr>
<td>pMC1neo polyA</td>
<td></td>
<td></td>
<td></td>
<td>1 x 10&lt;sup&gt;-5&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

The frequency of G418<sup>r</sup> or G418<sup>r</sup> + gancyclovir<sup>r</sup> (ganc<sup>r</sup>) after 10-14 days selection is expressed in number of surviving foci per number of cells transfected with the given construct.

*Transfection carried out by Dr. Sinead Jones. p=promoter.

Since frequencies may vary between experiments due to differences in cells or batches of G418, it is difficult to make valid comparisons between construct types. Also since the electroporation conditions etc. vary between laboratories it is difficult to compare frequencies with those reported in the literature.

The reported G418<sup>r</sup> frequency for pMC1neo polyA is 1 x 10<sup>-3</sup> (expressed as a fraction of the number of cells surviving electroporation) (Thomas and Capecchi, 1987). This frequency has not been reproduced. Other reports (Tybulewicz et al., 1991) claim a frequency equal to pSV2neo (Southern and Berg, 1982). pSV2neo induces a G418<sup>r</sup> frequency of 5 x 10<sup>-5</sup> (Thomas and Capecchi, 1987).

The G418<sup>r</sup> frequency after transfection with Hox2.1pneo indicates the presence of a strong promoter driving neo expression. This is in contrast to the observed expression levels of Hox-2.1 in ES cells and may reflect the lack of negative control regions from the Hox-2.1 locus in this vector or an increased...
stability of the Hox2.1neo mRNA over the natural Hox-2.1 mRNA. The Hox2.1-neo mRNA is functional and provides G418 resistance.

The G418\(^r\) frequency for construct 2 is as expected for a promoterless neo construct, being one third lower than the frequency obtained from pMC1neo polyA (comparable to the frequencies reported by Dorin et al. (1989)) and 80-fold lower than the G418\(^r\) conferred by Hox2.1pneo. This indicates that selection for integration next to an active promoter was relatively successful.

The G418\(^r\) frequency obtained upon transfection with Ω2BBtk is similar to that of construct 2, again indicating successful selection of integration into active promoters. Double selection with G418 and ganc reduces the number of selected foci by 40-fold. This is a considerably lower than the 2000-fold enrichment reported by Mansour et al., 1988. However, there have been varying reports on enrichment levels obtained by ganc selection of PNS vector derived clones (see Section 1.4) and 40-fold enrichment compares favourably with these. There is no difference in the frequency of G418 and ganc resistant foci obtained from the promoterless neo Ω2BBtk construct or Hox2.1ptk (which contains the same upstream region as Hox2.1pneo). This may suggest the presence of a negative enhancer element in the intron or 3' untranslated region of Hox2.1 (regions present in Hox2.1ptk but not in Hox2.1pneo).

The G418\(^r\) frequencies for construct 6 and HoxTneo (pSV2neo encoded neo) are 10-fold lower than the frequencies reported for pSV2neo. As transfections were not carried out with pSV2neo alone it is not possible to comment on the apparent reduction.

3.6 Screening of colonies by restriction mapping.

A number of G418\(^r\) colonies were picked after transfection with the different constructs. DNA was prepared for the analysis of restriction map alteration at the Hox-2.1 locus. Digests and probes were used as described in Sections 3.2-3.5. Table 3.5 shows the number of clones analysed for homologous recombination and which
the same integration event (ie. secondaries). None of the cell lines screened contained the diagnostic band for homologous recombination. Examples of Southern blot hybridisation analyses are presented in Figures 3.9, 3.10 and 3.11. By analysis of the neo restriction fragment length (Figures 3.9b, 3.10b and 3.11b) it was observed that most of the cell lines were derived from random integration of one copy of the construct.

Table 3.5

<table>
<thead>
<tr>
<th>Construct</th>
<th>L cells</th>
<th>ES cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>construct 2</td>
<td>5</td>
<td>31</td>
</tr>
<tr>
<td>Ω2BBtk and Hox2.1ptk</td>
<td>27</td>
<td>6</td>
</tr>
<tr>
<td>construct 6</td>
<td></td>
<td>37</td>
</tr>
<tr>
<td>HoxTneo</td>
<td></td>
<td>179</td>
</tr>
</tbody>
</table>

Table 3.5
The number of cell lines screened for homologous recombination at the Hox-2.1 locus by restriction fragment length analysis.
Figure 3.9
Southern blot of EcoRI/Hind III/Bam HI digested ES cell (E) DNA and G418<sup>f</sup> cell line DNAs (1-7), hybridised to 5' probe (see text) (a) or neo probe (pMC1neo polyA gel purified insert) (b). Cell lines were derived from construct 2 transfection.
Figure 3.10

Southern blot of Eco RI digested G418^cell line DNAs (1-4), hybridised to 5' probe (see text) (a) or neo probe (b). Cell lines were derived from Hox2.1ptk PNS vector transfections.
Figure 3.11

Southern blot of Bgl II digested G418R cell line DNAs (1-8), hybridised to the 3' probe (see text) (a) or neo probe (b). Cell lines were derived from HoxTneo insertion vector transfections.
3.7 Identification of a "pick-up" clone.

One G418R cell line (H4) derived from transfection with the insertion vector, HoxTNeo, had a novel Bgl II restriction fragment (20 kb) upon hybridisation to the 3' Hox-2.1 probe (Figure 3.12a). The band also hybridises to the neo probe (Figure 3.12b). However, this band is not indicative of correct homologous recombination and insertion of the construct because it is larger than predicted and is half as intense as the Hox-2.1 locus band (1 copy of new locus to 2 copies of endogenous locus). No extra band hybridises to the 5' Hox-2.1 probe (1 kb Sst I fragment purified from Hox2.1-2.2 subclone)(Figure 3.12c). Both 3' and 5' probes are expected to hybridise the same diagnostic 11 kb Bgl II band in homologous recombinant.

It is likely that clone H4 is derived from a "pick-up" and random integration event. Pick-up events have been reported to occur at equal frequency to homologous recombination events (personal communication from Dr. Alan Clarke). Similar events have been identified and reported by Adair et al. (1989); te Riele et al. (1990) and Itzhaki and Porter (1991). It is proposed that the construct aligns with the region of homology, picks up several kb of sequence from 5' or from 3' (not usually both) by a process of gene conversion (strand invasion and replication followed by breakage and joining), and integrates elsewhere in the genome.

Upon hybridisation of the 3' Hox-2.1 probe to a Hind III digest of H4, a novel 4.5 kb band is identified (Figure 3.13a). Since this band does not hybridise to the neo probe (Figure 3.13b) it may be concluded that the Hind III site at the 3' end of the construct has been reconstituted during the pick-up. Since the normal Hox 2.1 fragment length is 9 kb, this would suggest that less than 4.5 kb of Hox-2.1 sequence has been transferred to the new locus by the targeting construct.

In situ hybridisation to metaphase spreads of H4 cells was carried out, by Dianne Hanratty, in an attempt to map the integration site outside the Hox-2.1 locus. Labelled insert from the subclone Hox2.1-2.2 correctly mapped Hox-2.1 to the distal
Figure 3.12
Southern blot of Bgl II digested ES cell (E), H4 (1) and insertion vector derived G418^R cell line (2) DNAs hybridised to 3' probe (see text) (a), neo probe (b) or 5' probe (see text) (c).
Figure 3.13
Southern blot of Hind III digested ES (E) and H4 DNAs hybridised to 3' probe (see text) (a) or neo probe (b).
tips of two chromosomes (chromosome 11). No other hybridisation site was identified. This may be due to rearrangement of targeting vector sequence upon integration or integration of the targeting vector near the Hox-2.1 locus.

It is possible that during the pick-up event, a mutation was introduced into the endogenous Hox-2.1 locus. However, the H4 cell line had the abnormal karyotype of 44 chromosomes. Therefore, it was decided that characterisation of the Hox-2.1 loci, to determine whether the pick-up process had created a mutation, was not worthwhile in the context of this project.

No further pick-up clones were identified. However, only one third of the insertion vector derived clones were screened with the 5' Hox-2.1 probe so there may have been other pick-up events that were not detected.

3.8 Conclusions.

The reported targeting frequencies for replacement vectors containing a promoterless neo vary from 3/11 (Dorin et al., 1989) to 1/100 G418r clones (Sedivy and Sharpe, 1989). These frequencies may be expressed as a fraction of the total integration events when the enrichment factor is taken into account, so that 3/11 becomes 1/11 (3-fold enrichment) and 1/100 becomes 1/10000 (100-fold enrichment). 36 G418r cell lines were analysed for homologous recombination of construct 2. None were homologous recombinants. If the enrichment factor is assumed to be 80 (since Hox2.1pneo generated 80-fold more G418r colonies), 36 G418r clones is equivalent to 2880 total integrants. There are two possible reasons for the low targeting efficiency of this vector. Firstly, the region of homology at the 5' end of the neo is small (100 bp). It has been demonstrated that targeting frequency of a vector is increased by increasing the amount of homology to endogenous sequence (Thomas and Capecchi, 1987; see Section 1.4). Secondly, Hox-2.1 is expressed at a low level in ES cells - thus neo expression upon homologous recombination into the Hox-2.1 locus may be too low to confer G418r. Homologous recombinants may be selected against. The expression from the control vector,
Hox2.1pneo, could be high due to loss of negative control regions influencing endogenous Hox-2.1 expression. However, Hox-2.1 expression is believed to be low due to mRNA instability rather than weak promoter function (Dr R. Krumlauf, personal communication). An AU rich region similar to those leading to rapid breakdown of c-fos (Wilson and Treisman, 1988) and rabbit β-globin (Shaw and Kamen, 1986) is present at the 3' end of the Hox-2.1 mRNA. The Hox-2.1-neo fusion transcript does not contain the 3' AU region and therefore, is not subject to rapid breakdown.

The reported targeting frequencies for PNS vectors, expressed as number of homologous recombinants as a fraction of G418r gancr clones, vary from 19/24 for hprt (Mansour et al., 1988), 1/38 for immunoglobulin μ gene (Kitamura et al., 1991) to 1/500 for T-cell antigen receptor (Mombaerts et al., 1991). Taking into account the reported ganc enrichment factors, the above targeting frequencies may be expressed as a frequency of G418r clones: 1/2500 for Hprt, 1/304 for immunoglobulin μ, 1/1000 for T cell antigen receptor and 1/7150 for Wnt-1. The variation in the reported targeting frequencies shows the extreme variation in locus targetability. 27 L cell and 6 ES cell G418r Gancr clones were derived from the Hox-2.1 construct transfections. This number is equivalent to 1320 G418r clones when the ganc enrichment factor is taken into account.

Initial reports on targeting frequency for insertion vectors were in the order of 1/1000 (Thomas and Cappecchi, 1987) but recently higher frequencies have been obtained - 1/31 for Hox-2.6 (Hasty et al., 1991) and 1/170 for the CFTR locus (Dr. Julia Dorin, personal communication). 216 Hox-2.1 insertion vector derived clones were analysed. The identification of a pick-up clone indicates that HoxTneo is capable of finding homology to the Hox-2.1 locus. HoxTneo, however, may be incapable of homologous recombination because 100 bp have been deleted at the site of linearisation, Hind III. Pick up clones are reported to occur at an equal frequency to correctly targeted events (Alan Clarke, personal communication). However properties of the Hox-2.1 locus (such as silencing of neo expression, to be discussed
below) may lead to the identification of pick-up events rather than homologous recombination events. Construct 6 may have a low targeting potential due to the small region of *Hox-2.1* homology 5' to the linearisation site.

The overall targeting frequency of *Hox-2.1* may be stated as less than 1/4416 integration events (when enrichment factors are removed). Many genes have been targeted at higher frequency than this. Equally many genes have been targeted at a lower frequency.

Finally, it should be noted that the targeting literature is biased towards genes which have been successfully targeted. Since no reports are published on the unsuccessful targeting experiments it is difficult to tell whether the zero targeting obtained for *Hox-2.1* is unusual.
CHAPTER 4

IDENTIFICATION OF A SITE OF FREQUENT INTEGRATION IN ES CELLS
4. IDENTIFICATION OF A SITE OF FREQUENT INTEGRATION IN ES CELLS.

4.1 Introduction.

Of the 31 G418\' cell lines derived from transfection of ES cells with the promoterless neo replacement vector, construct 2, two (5B2 and 7B2) were identified as having the same integration sites. This was initially surmised from the hybridisation pattern of the neo probe to the Eco RI, Hind III and Bam HI restriction digests of the cell line DNAs - the band sizes were apparently equal for both the cell lines in each of the digests (Figure 4.1). On close inspection there is observed to be a slight variance in size between the band sizes, with the 5B2 band being slightly smaller in each digest. 5B2 and 7B2 were clones picked from separate plates which in turn had been plated from separate electroporation vials and are thus unlikely to be derived from a single integration event. DNA was re-extracted from freshly grown cells to confirm no mix-up had occurred during pelleting of cells or preparation of DNA. These cell lines were derived from transfections carried out by Dr. Robin Lovell-Badge.

It is proposed in this chapter that this site is a site of frequent integration (SOFI) and shall henceforth be referred to as such.

4.2 Restriction site mapping at the integration sites of 5B2 and 7B2.

Various digests were performed on 5B2 and 7B2 DNA in order to map restriction sites flanking the construct integration site. Sites were mapped 5' and 3' to the Hind III site in the construct by performing double digests with Hind III and various enzymes. The Southern blots were probed with the neo probe (5' to the construct Hind III) or the 2.4 kb Eco RI to Bgl II fragment from subclone pMI (see Figure 3.1) (3' to the construct Hind III). The hybridisation results are shown in Figures 4.2a and 4.2b. The 3' probe hybridises to both construct sequence and
Figure 4.1
Southern blot of Eco RI and Hind III digested G418r cell line DNAs (1-6), hybridised to neo probe. Cell lines were derived from construct 2 transfection. 5 and 6 denote 5B2 and 7B2, respectively.
Figure 4.2
Southern blot of restriction enzyme digested 5B2 (5) and 7B2 (7) DNAs hybridised to *neo* probe (a) or the pMI 3' *Hox-2.1* probe (b). Enzymes are abbreviated as: B=Bam HI, K=Kpn I, S=Sst I, H=Hind III and E=Eco RI.
endogenous *Hox-2.1* sequence - the construct sequence band was identified as the band half the intensity of the *Hox-2.1* band and by comparison to wild type DNA digests. The hybridisation patterns confirm that 5B2 and 7B2 were indeed integrations into similar loci. The restriction fragment lengths are tabulated in Table 4.1 and the restriction map calculated from these data shown in Figure 4.3.

**Table 4.1**

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Size 5' probe</th>
<th>Size 3' probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hind III</td>
<td>3.8</td>
<td>5.5</td>
</tr>
<tr>
<td>Eco RI</td>
<td>2.9</td>
<td>4.0*</td>
</tr>
<tr>
<td>Eco RI and Hind III</td>
<td>2.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Bgl II</td>
<td>6.4</td>
<td>6.4</td>
</tr>
<tr>
<td>Bgl II and Hind III*</td>
<td>3.2</td>
<td>3.2*</td>
</tr>
<tr>
<td>Sst I</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Sst I and Hind III*</td>
<td>3.8</td>
<td>5.5</td>
</tr>
<tr>
<td>Kpn I</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Kpn I and Hind III*</td>
<td>3.8</td>
<td>5.5</td>
</tr>
<tr>
<td>Bam HI</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Bam HI and Hind III*</td>
<td>3.8</td>
<td>5.5</td>
</tr>
</tbody>
</table>

* The 3’ Eco RI site is mapped in relation to the Eco RI site in the construct.
1. The Bgl II site has been retained at the 3’ end of the construct.
2. The Sst I sites have been lost from both ends of the construct.
3. The 5’ and 3’ SstI, Kpn I and Bam HI sites cannot be mapped since they are further 5’ or 3’ than the Hind III sites.
4. Data not presented.
Figure 4.3
Restriction map of the 5B2/7B2 integration site. Bold line represents the integrated construct. The SstI sites at the ends of the construct have not been reconstituted and are represented by a crossed "S". Enzymes are as abbreviated in Figure 4.2 and Bg=Bgl II. Neo probe is represented by a stippled line and the 3' probe by a striped line.
The restriction map was compared to the published maps for *Hox* gene loci in case the construct had integrated into these genes due to homology. No map was found to be similar.

4.3 **Inverse PCR to obtain flanking sequence.**

Flanking sequence, upstream and downstream of the construct sequence was amplified by inverse PCR (Ochman *et al.*, 1988) of 7B2 DNA. The steps involved in amplification of DNA sequence by inverse PCR are shown in Figure 4.4. Genomic DNA is linearised with a suitable restriction enzyme, the resulting fragments are self-ligated to form circles and PCR amplification is carried out between primers from the known sequence across the unknown junction sequence.

Eco RI was chosen as the enzyme for linearisation since the construct contains an Eco RI site and since Eco RI sites had been mapped 600 bp upstream and 1 kb downstream of the integration site.

The relative positions of the four primers involved in the amplification are shown in Figure 4.4. Figure 4.5 shows sequence obtained from the 3' end of construct 2 (*Hox-2.1* locus genomic sequence). This sequence was required to synthesize the oligonucleotide 598 for downstream amplification and oligonucleotide B2 for hybridisation to the downstream product. 598 is derived from sequence 1-18 and B2 from sequence 130-148. Since 598 spans the Bgl II site and since this site is known to have been retained upon integration of the construct, it was considered likely that 598 was present in the integrated sequence.

Upstream and downstream inverse PCR amplifications were carried out separately on 7B2 DNA.

The specific bands were identified by hybridisation to the internal oligonucleotides (neo19 for upstream and B2 for downstream products). Neo19 identified a 500 bp product amplified by the upstream oligonucleotides (neo18-314). B2 hybridised strongly to a 200 bp product and weakly to a 1.4 kb product.
Figure 4.4
Steps involved in inverse PCR amplification of flanking sequence. Construct sequence is represented by a thick line. Arrows indicate a 5' to 3' direction of the PCR primers. The final products of Upstream (314 to neo18) and downstream (598 to 315) amplification are schematically presented.
Figure 4.5

Sequence obtained from the 3' end of construct 2, by reading from the rev sequencing primer, spans the Hox-2.1 sequence from Bgl II to Sst I sites (see Figure 3.1a). Sequences, from which oligonucleotides 598 and B2 are derived, are underlined. Restriction enzyme sites are underlined by double lines.
amplified by the downstream oligonucleotides (598-315). The former is surmised to be the endogenous Hox-2.1 sequence inverse PCR product which is amplified in preference to the latter longer sequence. The latter sequence is surmised to be the construct-flanking sequence inverse PCR product.

The construct-flanking sequence bands were gel-purified and re-amplified. The re-amplified products were cut from LMT agarose and purified by Geneclean.

The upstream product ends were filled in with Klenow enzyme, digested with Eco RI to remove the 314 oligonucleotide end (see Figure 4.4) and cloned into Sma I-Eco RI cut Bluescribe. (From analysis of the cloned sequence it was discovered that the PCR product had been cut successfully with Eco RI but that the overhang had been filled in (presumably by residual Klenow, despite heat treatment) and the blunt-ended molecule ligated into Sma I cut Bluescribe, with the insert orientated with the neo end closest to the reverse sequencing primer).

The downstream product was digested with Bgl II (site within the 598 oligonucleotide) and Eco RI to remove the 315 oligonucleotide-end (see Figure 4.4). This fragment was cloned into Bam HI and Eco RI cut Bluescribe.

Confirmation that the clones contained flanking sequence to the integrated construct, was made by hybridisation of the cloned DNA to Eco RI or Hind III restricted 7B2 DNA (Figures 4.6b and 4.7b) The neo probe and upstream inverse PCR product clone hybridise to the same 2.9 kb Eco RI and 3.8 kb Hind III bands. The 3' probe (see Section 4.2) and the downstream inverse PCR product clone hybridise to the same 4.0 kb Eco RI and 5.5 kb Hind III bands. The bands observed to hybridise to the downstream probe in ES cell DNA (Figure 4.7b) identify the endogenous uninterrupted "target" locus. These bands are not clearly distinguishable on hybridisation with the upstream probe (Figure 4.6b) because of multiple band hybridisation. However, the same ES cell Eco RI and Hind III fragments should hybridise to both the upstream and downstream probes.

The multiple bands hybridised by the upstream product suggest this sequence is repeated several times in the genome. This raises the possibility that 5B2 and 7B2
Figure 4.6
Southern blot of Eco RI (E) and Hind III (H) digested ES cell (ES) and 7B2 (7) DNAs hybridised to neo (a) or upstream inverse PCR product clone (b) probes. The arrows indicates the bands that are hybridised by both the probes.
Figure 4.7
Southern blot of Eco RI (E) and Hind III (H) digested ES cell (ES) and 7B2 (7) DNAs hybridised to 3' construct probe (a) or downstream inverse PCR product (b) probes. The downstream product probe is the gel purified Pst I-Eco RI 250 bp fragment. The arrows indicates the bands that are hybridised by both the probes.
are integration events into the same sequence at different loci. However, since the bands appear to be single-copy strength, and it is the same band that is altered upon integration in both 5B2 and 7B2, it is likely that the events are into the same locus.

4.4 Sequencing of upstream flanking sequence.

The insert of the neo18-314 inverse PCR clone from 7B2 was sequenced from both the forward, reverse and the neo19 primers. The neo-Hox-2.1 fusion was identified and the construct-flanking junction sequenced through. 8 or 9 bp of 5' construct sequence were deleted from the construct upon integration in 7B2 (the exact number cannot be determined since bases at the site may be derived from either genomic DNA or construct DNA).

An oligonucleotide (B85) was synthesized from the 7B2 upstream flanking sequence and the B85-neo18 fragment was PCR amplified from 5B2 genomic DNA. The product was directly sequenced using the neo19 primer. Mn buffer was included in the sequencing buffer in order to read sequence close to the primer. 77 bp of 5' construct sequence were deleted upon integration in 5B2.

Sequencing data presented in Figure 4.8 shows the difference between 7B2 and 5B2 at the 5' genomic-construct DNA junctions. The upstream flanking sequence obtained from 5B2 and 7B2 is presented in Figure 4.9a.

4.5 Sequencing of downstream flanking sequence.

The insert of the 598-315 downstream inverse PCR product clone was sequenced using forward, reverse and B2 primers. The insert was completely sequenced from internal primers (B406, B460 and B461) prepared from the forward and reverse sequence. The construct-downstream flanking sequence junction was sequenced. 14-16 bp of 3' construct sequence were deleted upon integration of the construct in 7B2.

An oligonucleotide (B218) was synthesized from 7B2 sequence and the 598-B218 fragment PCR amplified from 5B2 genomic DNA. The product was
Figure 4.8
Autoradiographs of DNA sequencing gels from reactions:

a. 7B2 upstream PCR product clone primed from rev sequencing primer.

Integrated construct Hox-2.1 sequence reads into the integration site upstream flanking sequence. 68 bp of Hox-2.1 sequence, present in 7B2, has been deleted in 5B2.
Figure 4.9
The upstream (a) and downstream (b) flanking sequence from 5B2 and 7B2 are shown in bold print. Construct sequence is shown in non-bold print. The underlined sequence is shared by 5B2 and 7B2. The construct sequences not underlined are retained in 7B2 but are deleted from 5B2. The target sequences CC (a) and T (b) are retained in 5B2 but deleted from 7B2.
sequenced directly using the B2 primer, with Mn buffer in the sequencing reaction. 58 bp 3' construct sequence were deleted upon integration in 5B2.

In Figure 4.10 the sequencing data presented shows the difference between 7B2 and 5B2 at the 3' genomic-construct DNA junction. The downstream flanking sequence obtained from 5B2 and 7B2 is presented in Figure 4.9b.

4.6 Sequencing of the SOFI.

In order to obtain the complete sequence at the integration site, PCR amplification was carried out between the upstream flanking sequence primer, B85 and the downstream flanking sequence primer, B218, on ES cell genomic DNA. The 400 bp product was sequenced directly from the B85 and B218 primers. Sequencing data showing the difference between unintegrated genomic DNA and the 5B2 integration site is presented in Figure 4.11.

The sequence obtained for the 7B2 upstream, 7B2 downstream and the ES cell B85-B218 product combine to give 1.25 kb of sequence from the SOFI locus. This sequence is presented in Figure 4.12.

3-6 bp of sequence were deleted from the locus upon integration of the construct in 7B2. No sequence was deleted from the locus upon integration in 5B2.

The differences in the amount of construct or locus sequence lost or retained in 5B2 and 7B2 confirm that these two cell lines are derived from separate integration events. The 130 bp deleted from 5B2 and not from 7B2 would explain the observation of the slightly smaller restriction fragment lengths for 5B2 (Figures 4.1 and 4.2).
Figure 4.10

Autoradiographs of DNA sequencing gels from reactions:

a. 7B2 downstream PCR product clone primed from B2 primer.
b. 598-B218 5B2 PCR product primed from B2 primer.

Integrated construct *Hox-2.1* sequence reads into the integration site downstream flanking sequence. 58 bp of *Hox-2.1* sequence, present in 7B2, has been deleted in 5B2.
Figure 4.11

Autoradiographs of DNA sequencing gels from reactions:
a. ES cell B85-B218 PCR product primed from B218 primer.
b. 5B2.598-B218 downstream PCR product primed from B218 primer.
Downstream sequence reads into upstream sequence in the ES cell PCR product sequence (a).
Downstream flanking sequence reads into Hox-2.1 construct sequence in 5B2 PCR product (b).
Figure 4.12
Sequence of the SOFI locus. The integration site of the construct in 5B2 is indicated by arrow 5. The integration region in 7B2 is highlighted in bold print. The possible integration sites in 7B2 are indicated arrows labelled 7. CCT is lost from the site upon integration in 7B2; the C and AG on either side of the CCT may either be derived from the integration site or from the incoming construct in 7B2. The Eco RI sites at ends of the upstream and downstream inverse PCR products are underlined. The Pst I site used to generated the unique sequence probe from the region is underlined. Sequence showing 72% homology to retroviral LTR sequence is underlined and the AGTTTC motif is underlined twice.
1 AATTCTATGA ATTTATTTAT TTGTGCCTTC ATTTTGAAAG GACTGAGCCA
   Eco RI
51 GGCTGAGTCA C TAGACAGGC TTCAAACTGA CAGGACGTTG GAGTAGGGCT
101 AAGTTTTCTGT TTCCCCAGAT TAAGCAAGTT TGGGAAAGTC AGTCTCTCAC
151 AGGGAAAAAA AAAAAAGAAA AAAAAAAACA GTCCCTGGGC AGCCAACTCAG
201 AGGCTGTCCA GACACTTACG CTGGAGGGA AGCAATGAGT TAGCAACAGT
251 TTCCAGGCTT CCCCAGCAAC AGTTTCCAGG TTCCCCCAAC AAGAGTTCTG
301 GATTGTCCCC AGAGGTCTGAG CCAATCCCAG AGAACCACAT ACCCTTAGT
351 ACCAGACCAA CCAAATCAAGA TAAGGGTTAC CCACTCCCCT CCTGGAATCC
401 CCTAATTCTGC TTAAAAAGGA GGCTGAGGAT TCACCTGGGC CTCTCCATCT
451 GAGTGAAATGG TAGACCCCTG AATACTATAA TAGCTCGAGA TACATATTCT
501 CACCCCTACA AATGGTTACT TCATTGCTTT TTTTTTTTT TAAATGGTTT
551 TTGAGACAGG GTTTCTCGGC TTACCCTGGC TATCTGACTN CTCTGAGAT
601 TAGGCTGGTCA CCAATTTTAA ACGATCTGCC TGCCCTGGCC TATGAGGTGC
651 TGAGATTAAA GACATGTGCT GACACTACCACC GACTTCCTTC GTTTAATTAT
701 TTGATCGTTT GAGGCACTTC AGAGCCTAGA AGGGAAGGTG TCACCTGGGC
751 GATCAAAGAA AGAGGTACTGC CAGAGTGTAA GACACGTGCT AGCACACTGG
801 GAGGAACAGT CCTCAATGAC TGAACCTGAA CCTGGGACAG ATGTACAGAC
851 GCACATCCAT TAGATTGATC CAGGTTTGTG TGGGTTTTTT TGGTGGGCGGA
901 TAGGGGAACA CATTTCTCAT CCTAAAGCCC TCTATCGACT CTGTATGATT
951 TTATGAAGAA A AACACCATTG CCCCCGNAAC CCAATTTGTT CTCTGAGACT
1001 GTCTGCGAGTG CCAATAAATG GAAGACATTG CTGTCTGAGA GAAAGTACC
   Pst I
1051 CATGGCTGTA AAGCCCTCTC AGGAAAAAAGG CATCATAGAT AATGGAAAGAC
1101 AAGCAGTGTG CTATTGATTC CATTTCGGG TCAAAGGTACT TCTGAAACAA
1151 ACTCTTCCATT CTGGATATCAT CCACTGGAAT GCCTAAATAC AGGCCCTTTCA
1201 TTCTGATGTT TACCCTAGAT ACGACTGATC CACTAGATT C
   Eco RI
4.7 Analysis of the SOFI Sequence.

Sequence analysis was carried out using the University of Wisconsin Genetics Computer Group (UWGCG) software package (Devereux et al., 1984).

4.7.1 Search for DNA sequence homologies.

No extensive homology was found between the SOFI sequence and the ends of the targeting vector using the UWGCG BestFit program. Figure 4.13 illustrates the proposed sequence alignments of construct to genomic DNA prior to integration in 5B2 and 7B2. 3bp at the 7B2 junction are homologous, none are homologous at the 5B2 junction. It has been proposed that short sequence homologies (1-5 bp) may direct "illegitimate" integration in mammalian cells (Roth and Wilson, 1986; Murnane and Young, 1989 and Macleod et al., 1991). Since there is no sequence specificity contributing to the integration event in 5B2 it is proposed that this region possesses properties that enhance recombination.

Topoisomerase I cleavage sites AAT, CAT, CTC, CTT, GAT, GTC and GTT have been shown to be over-represented around illegitimate integration points or regions (Konopka, 1988). Two of these sites are present within 10 bp of the integration regions in 5B2 and 7B2, as is shown in Figure 4.13.

The sequence was used to search the GenBank and EMBL sequence databases, using UWGCG Wordsearch and FASTA programs. No significant homology to the 1.25 kb SOFI sequence was found. Limited homology to sequences in the databases was found to smaller regions of the sequence.

Sequence 241-295 shows 72% similarity to retroviral long terminal repeat (LTR) sequence, such as the match shown in Figure 4.14. The sequence 243-310 contains three copies of a 23 bp repeat (the third copy is degenerate) and the sequence 242-297 also contains two copies of an overlapping 27 bp direct repeat. These repeats contain a 6 bp AGTTTC motif which is present in LTR sequences (Golemis, et al., 1990). The AGTTTC motif is present twice at an interval of 23 bp, which is also the case in some proviral LTRs. The function of this motif is unknown. No further homology to retroviral sequence was found.
Figure 4.13
The proposed alignment of construct DNA to endogenous SOFI locus DNA in the generation of 7B2 (a) and 5B2 (b) integration sites. The construct sequence is represented by a box and the direction of neo transcription is indicated by the arrow. Boxed sequence is lost by the incoming construct. The bases homologous between construct and endogenous sequences are marked by vertical lines. The C and AG flanking the construct sequence in 7B2 may be derived from the endogenous or construct sequence. Underlined trinucleotide sequences are potential Topoisomerase I cleavage sites.
Figure 4.14
An example of SOFI to retroviral LTR sequence homology found by the UWCGC FASTA database search. The upper sequence is SOFI sequence (see Figure 4.12) and the lower sequence is Murine Leukemia Virus RadLV variant LTR sequence (Rassart et al., 1986). Base homologies are indicated by vertical lines. 72% similarity is observed across 55 bp of sequence, with the introduction of one gap.
The ladder of bands observed upon hybridisation of the upstream flanking sequence to genomic DNA digests (Figure 4.6b) suggests this sequence is repeated 20-30 times in the mouse genome (strain 129). It may be concluded that the repeats are dispersed since the bands on a number of different digests appear to be of single-copy strength.

An enzyme cutting once in the LTR and once in the viral gene sequence of a retroviral provirus will give rise to a fragment common to all proviral copies upon hybridisation to the LTR probe. There is an Eco RI site in the SOFI 5' to the LTR-like homology and an Eco RI site may be expected within about 4 kb. Since no strongly hybridizing Eco RI band is observed upon hybridisation to the upstream probe (Figure 4.6b) it is unlikely that this sequence identifies a novel retroviral sequence.

The sequence 241-295 may retain some function similar to that of the retroviral sequence.

Sequence 701-511 shows 76% similarity to the B1 SINE (short interspersed element). The homology between the SOFI and the B1 SINE consensus is shown in Figure 4.15. The B1 SINE, first characterised by Krayev, et al. (1980) and reviewed by Hastie in Lyon and Searle (1988), is the most highly repeated element (130 000 to 180 000 copies) in the mouse genome. Hybridisation of the total downstream sequence to genomic DNA digests resulted in hybridisation of the probe to total DNA - this confirms the presence a highly repetitive sequence.

B1 SINES are believed to derive from short RNA molecules by a retroposan-like reverse transcription and integration mechanism. A characteristic of the SINE is an A-rich region in the region corresponding to the 3' terminus of the RNA molecule. There is an A-rich region in the SOFI B1 element.

SINES are believed to be functionless but may have been important in the restructuring of the genome during evolution.
Figure 4.15
The B1 consensus sequence to SOFI sequence homology. The upper sequence is SOFI sequence (see Figure 4.12) and the lower sequence is B1 consensus sequence (Hastie, 1989). Base homologies are indicated by vertical lines. 76% similarity is observed across 195 bp of sequence, with the introduction of 5 gaps.
4.7.2 Translation of the SOFI sequence.

Upon translation of the sequence in all six reading frames using the UWGCG program Frames (Figure 4.16), no extensive open reading frames (ORF) are identified.

It was noted that in both 5B2 and 7B2 the neo sequence has integrated into frame 2 of the SOFI sequence (4.16b). This frame has no AUG initiating codon. It is possible that the sequence is spliced.

However, on the basis of the UWGCG Testcode plot (Figure 4.17) the sequence is unlikely to be coding. Testcode scores sequence coding potential on the basis of A, C, T, G content or codon position. The Testcode plot of SOFI sequence lies below the midline indicating the sequence is probably non-coding.

The peptide sequence translated from the six frames was used to search the EMBL and Genbank protein sequence databases using UWGCG Wordsearch and Fasta programs. The sequence from frame 2 was used to search the NBRF, Swissport databases and the translation of the EMBL database sequences (search carried out by Dr. D. Kipling). No extensive homology was found.
FRAMES of: ins.seq  Ck: 9967, 1 to: 1,241 September 26, 1991 11:53
ins.seq Length: 1241 July 18, 1991 16:29 Check: 9967

---

Diagram with labeled sections A, B, and C, showing 5' to 3' orientations and temporal markers.
TESTCODE of: ins.seq ck: 9967, 1 to: 1241
Window: 200 bp  September 26, 1991 11:58
4.7.3 Attempt to map the SOFI.

The position on the genetic map of a cloned piece of DNA may be determined by the use of recombinant inbred (RI) strains (Taylor in Morse, 1978). The BXD RI strains result from a cross between two progenitor inbred strains, C57BL/6J and DBA/2J.

The method relies on the presence of a polymorphic site which varies between the parental strains and is identified by the cloned DNA. Segregation of the polymorphic site in the RI strains allows the locus to be mapped in respect to previously mapped loci.

The unique sequence probe (Pst I to Eco RI fragment from the downstream inverse PCR product clone) from the SOFI was hybridised to a panels of C57 and DBA DNAs digested with 12 restriction enzymes (Taq I, Mspl, Hind III, Bam HI, Bgl II, Pst I, Eco RI, Sph I, Xba I, Sac I, Pvu II and Stu I). No polymorphic site was identified.

4.8 Identification of the SOFI locus promoter function.

The upstream flanking sequence is expected to contain promoter elements, necessary for the expression of the neo gene and selection of 5B2 and 7B2.

No consensus TATA box sequences (Corden et al, 1980) are present in the 350 bp of cloned upstream sequence.

Promoter function was tested for directly by two methods: 1. by attachment of the cloned upstream sequence to a promoterless neo and transfection into ES cells to test for G418 resistance and 2. by attachment of the cloned upstream sequence to the chloramphenical transferase (CAT) gene and transfection into ES cells for the assay of CAT activity.

To identify protein binding sites that may be associated with promoter sequence, gel retardation experiments with oligonucleotides derived from the upstream sequence were carried out.
4.8.1 Promoter elements from the SOFI locus drive neo expression.

The upstream inverse PCR product clone was digested with Eco RI and Bam HI to release the insert. The insert was gel-purified, Genecleaned and digested with Sau 3A. Sau 3A cuts at the Hox-2.1-neo junction. This fragment was cloned into Eco RI and Bel I cut SDM Blueneo to recreate the SOFI-neo sequence fusion of 5B2/7B2 (construct named SOFIneo). SOFIneo was digested with Eco RI and Hind III to release total insert, prior to electroporation into ES cells.

The G418r frequency generated by SOFIneo is compared to that obtained from the promoterless neo vector SDM Blueneo and is tabulated in Table 4.2.

<table>
<thead>
<tr>
<th>Construct</th>
<th>G418r frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOFIneo</td>
<td>3.8 x 10^-6</td>
</tr>
<tr>
<td>SDM Blueneo</td>
<td>1.5 X 10^-6</td>
</tr>
</tbody>
</table>

Table 4.2
G418r frequency expressed as the number of resistant foci as a fraction of the number of ES cells electroporated.

The G418r frequency obtained from the SOFIneo construct is 2.5 fold higher than that of the the promoterless neo vector. It may be concluded that the cloned SOFI sequence has some, though perhaps weak promoter function.

4.8.2 CAT assays.

The Eco RI to Sau 3A fragment from the upstream inverse PCR clone was cloned 5' to the CAT gene in pBLCAT2 (contains the HSV-<i>tk</i> promoter) and pBLCAT6 (promoterless) to generate the constructs SOFICAT2 and SOFICAT6.

The double-stranded (ds) B267-B268 oligonucleotide was also cloned into the CAT vectors to give the constructs 267CAT2 and 267CAT6. B267 is derived from the sequence 242-274 of the SOFI, shows homology to retroviral LTR and contains
the motif AGTTTC.

The cloning steps involved in the construction of the CAT constructs are presented in Figure 4.18. 267CAT2 and 267CAT6 were sequenced to check for correct insertion of the oligonucleotide sequence.

A plasmid containing the β-galactosidase gene driven by the β–actin promoter was transfected into ES cells by electroporation to test the efficiency of uptake. However, on staining the cells *in situ* with X-gal to monitor β-galactosidase activity, no cells stained blue. Other people in the lab have reported failure of this plasmid to express in mammalian cells. Therefore, no suitable transfection control was available in these experiments and the results must be regarded as preliminary.

pBLCAT2, 267CAT2, SOFICAT2, pBLCAT6, 267CAT6 and SOFICAT6 were electroporated into ES cells. 48 hrs later the cells were harvested and CAT assays performed. No chloramphenicol acetylation was observed when the cells were electroporated at 3 μF and 800 V. CAT assay results from cells electroporated at 960 μF and 200 V are presented in Figure 4.19 and the scintillation counts obtained from the CAT2 construct transfection acetylated spots are tabulated in Table 4.3.
Bluescribe cut HindIII
Klenow treated
cut BamHI

ligate and clone

$\frac{S}{S} B$ $\frac{S}{H}$

CAT

500 bp

pBLCAT cut
BamHI and SmaI
insert gel purified

upstream inverse PCR clone
cut EcoRI and BamHI
insert gel purified
Klenow treated
cut Sau3A

B267 and B268
annealed

cut SmaI
BAP

cut SmaI and BamHI

ligate, cut SmaI
and clone

PCR between bluescribe
oligonucleotide 671
and oligonucleotide B268
detects B267CAT constructs.

S

SOFI | CAT
Figure 4.19
Autoradiograph of CAT assay TLC plate. Numbers denote plasmid transfections:
1 = no plasmid; 2 = pBLCAT2; 3 = 267CAT2; 4 = SOFICAT2; 5 = pBLCAT6; 6 = 267CAT6 and 7 = SOFICAT6. The spot that has migrated the least distance from the origin is that of nonacetylated chloramphenical. Acetylated chloramphenical forms (acetylated at one or other of the potential acetylation sites) migrate to the higher positions on the TLC plate.
Table 4.3

<table>
<thead>
<tr>
<th>Construct</th>
<th>Scintillation counts per minute</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBLCAT2</td>
<td>5081</td>
</tr>
<tr>
<td>267CAT2</td>
<td>2574</td>
</tr>
<tr>
<td>SOFICAT2</td>
<td>22721</td>
</tr>
</tbody>
</table>

Table 4.3

The scintillation counts per minute obtained from the CAT2 construct chloramphenical acetylation spots.

The level of chloramphenical acetylation obtained is 4-fold higher from cells transfected with SOFICAT2 and 2-fold reduction from cells transfected with 267CAT2 constructs than from cells transfected with the control pBLCAT2. None of the other constructs show an increase in chloramphenicol acetylation over the control vectors. It may be concluded from these data that the cloned SOFI sequence contains an enhancer element but does not possess complete promoter function and that the 267-268 oligonucleotide sequence possesses no enhancer or promoter function and may even have a silencing effect on expression.

4.8.3 Gel retardation of sequence from the SOFI locus.

It was suggested by Dr. Jonathan Stoye that the AGTTTC motif from the LTR sequence that shows homology to the SOFI sequence might have enhancer activity and therefore might bind a transcriptional control protein. Gel retardation experiments were therefore carried out on the ds oligonucleotide sequence 267-268 (see Figure 4.12) with nuclear protein extracted from ES cells. Control ds oligonucleotide sequence 277-278, in which the AGTTTC motif was replaced by random sequence, was included in the experiments as a negative control.

The results of a gel retardation experiment are shown in Figure 4.20. Some of
Figure 4.20
Autoradiograph of 32P-labelled ds oligonucleotides 267-268 or 277-278, alone (A and C, respectively) or retarded by ES cell nuclear protein extract (B and D).
the labelled 267-268 oligonucleotide has been shifted to a higher molecular weight position in the presence of protein. No shift in the control oligonucleotide is observed.

Gel retardation experiments carried out in the presence of cold ds oligonucleotide as competitor gave unconvincing results. These results may be regarded as preliminary.

4.9 Analysis of expression from the SOFI locus.

The Pst I-Eco RI unique sequence downstream probe was hybridised to RNA from various tissues (kidney, liver, ovary, brain, salivary gland, heart, lung, small intestine, testes and spleen) (Northern kindly provided by Jane McConnel) and ES cells. The probe failed to hybridise. The upstream inverse PCR probe was also hybridised to ES cell RNA. The probe hybridised to the total RNA track (data not shown), suggesting identification of multiple transcripts.

4.10 Search for other integration events into the same sequence.

It is predicted that if this sequence has been targeted in 2 events out of 31 and if this site is indeed a site of frequent integration, similar integration events may be detected upon analysis of more G418r cell lines. Also if the upstream sequence is repeated 20-30 times in the genome that similar sequence at different loci may be targeted in different cell lines.

PCR amplification between the primer B85 (upstream of 5B2 and 7B2 integrated sequence) and the neo primer neo18 was performed on DNA from the 34 cell lines derived from the integration of construct 2 (including secondaries, in pools of 6) and 5 cell lines derived from integration of a second Hox-2.1 promoterless neo construct (obtained from Dr. Donald Macleod). The PCR products were hybridised to the internal oligonucleotide, neo19. One cell line, 7B3 gave a positive band of the same size as 7B2. Sequencing of the product confirmed the integration site was identical to that of 7B2. Since these clones were picked from the same plate, 7B2
and 7B3 are likely to derive from the same integration event (i.e. sibling clones).

Sst I cut construct 2 was transfected into ES cells to generate more G418\textsuperscript{r} clones for SOFI PCR analysis. PCR amplification between primers B85 and neo18 was carried out on the DNA from 15 clones (analysed in pools of 3 or 4). No positive was identified.

The overall frequency of targeting of the SOFI locus by the \textit{Hox-2.1} promoterless \textit{neo} vector is therefore 2/46 G418\textsuperscript{r} clones. Repeat copies of the same sequence have not been targeted.

Stu I cut SOFIneo (described in Section 4.8.1) was transfected into ES cells. This construct is expected to behave like an promoterless \textit{neo} insertion vector and target the endogenous SOFI sequence by homologous recombination. However, on analysis of 11 G418\textsuperscript{r} clones by PCR amplification between B461 primer (oligonucleotide made from 3' SOFI sequence not in targeting vector) and 671 primer (Bluescribe oligonucleotide, see Section 2.15), no correctly targeted events were identified.

4.11 Search for integration into the \textit{F2} locus.

The mouse \textit{F2} locus defines a region where three recombinant M-MuLV retroviruses have inserted into the genome of F9 EC cells (Barklis \textit{et al.}, 1986; Peckham \textit{et al.}, 1989 and Peterson \textit{et al.}, 1991). Barklis \textit{et al.} suggest integration into this locus is selected for since specific cellular sequence at this locus promotes viral \textit{neo} activity. Peckham \textit{et al.} demonstrate that sequence flanking the proviral insert is normally expressed, readthrough from the endogenous promoter into viral sequence takes place and that the mRNA product is spliced from a cellular splice donor to a proviral splice acceptor. The three proviral insertions map to intronic sequences at varying distances (500-1500 bp) from the \textit{F2} promoter. The \textit{F2} locus has been cloned and sequenced, but the function is unknown (Peterson \textit{et al.}, 1991).

No homology between \textit{F2} and SOFI sequences was found by the UWCGC BestFit program.
An oligonucleotide (B539) was made from the published F2 locus sequence (193-209; Peterson et al., 1991) and PCR amplification between B539 and neo18 was carried out on the 46 construct 2 derived G418r cell line DNAs. No positives were identified. Therefore, the integration into the F2 locus has not been preferentially selected for.

4.12 Conclusions.

The SOFI locus is defined as a region where the promoterless neo Hox-2.1 construct has integrated twice, in two independent events, to generate the ES cell lines 5B2 and 7B2. The overall targeting frequency is 2/46 G418r clones.

The sequence flanking the construct in 5B2 and 7B2 has been sequenced. This has enabled the SOFI to be sequenced. The integration points are within 3-6 bp and varying amounts of incoming vector have been deleted in the separate events.

In some cases of illegitimate recombination small regions of homology have been found at the cross-over junction and these are thought to promote recombination (reviewed by Roth and Wilson in Kucherlapati and Smith, 1988). 3 bp of homology between the construct and SOFI sequences are observed at the proposed construct-SOFI junction in 7B2. No homology is observed at the proposed junction in 5B2. Therefore, homology to SOFI sequence is not assumed to promote integration of vector into this locus.

Topoisomerase I has been implicated in the formation of some illegitimate recombination events (reviewed by Champoux and Bullock in Kucherlapati; Smith, 1988; Konopka, 1988 and Macleod et al., 1991). Topoisomerase I introduces a transient break into one of the two DNA strands and can join DNA strands from different molecules. Topoisomerase I sites are present in the vicinity of 92% illegitimate cross-over points (Konopka, 1988). It is proposed that Topoisomerase I may cause a break in DNA at the region, ends may be filled in or trimmed (since the recognition site is often a few bp away from the integration point), before recombination occurs. There are three Topoisomerase I recognition sites within 10
bp of the integration region in 5B2 and 7B2 - CAT (position -6); GTC (+6) and GTC (+11). These are not the preferred positions for these sites at illegitimate integration regions (Konopka, 1988) and they may be present by chance.

Barklis et al (1986) have described the F2 locus - a site for integration and expression of proviral sequence. Three retroviral integrations out of 17 were into the F2 locus. Barklis et al. propose that the F2 locus provides strong promoter function and a splice donor site for the expression of viral neo. Integration events into this locus are thus selected for on G418.

In contrast to the F2 locus, the promoter function of cloned SOFI sequence is found to be weak. Comparing G418r frequency with that of a promoterless neo construct a 2.5-fold increase is observed. The sequence is also observed to have a moderate enhancer activity from the analyses of CAT construct activity. It is unlikely that this site acts as a strong promoter (although this cannot be ruled out since only 350 bp from the locus was used in this study), and it is therefore unlikely that this integration event is enriched for by G418 selection, as is proposed for the identification of the F2 integrants. Also enrichment for promoter activity would not be expected to select for integration into such a specific (6 bp) region. The described integration events into the F2 locus vary by several hundred base pairs.

It is proposed that other properties of the SOFI sequence promote integration. It is possible that aspects of the secondary structure enhance recombination. Short indirect repeats, longer direct repeats and stretches of AT sequence are present. These sequences may form single-stranded bubbles which could increase the likelihood of recombination. The 55 bp retroviral homology may also be significant. It is possible that the sequence retains some LTR properties or functions, such as the binding of protein or promotion of recombination. The function of the retrovirally conserved motif, AGTTTC, present in the SOFI sequence, is unknown. Preliminary gel retardation results showed that this sequence may bind weakly to an ES cell protein. However, no enhancer or promoter activity was displayed by the sequence in CAT constructs.
No motif associated with recombinational hotspots was identified in the SOFI sequence (reviewed by Fischer Lindahl, 1991 and Steinmetz et al., 1987). This will be discussed further in Section 5.4.

Searches of the Genbank/EMBL sequence databases revealed no significant homology over the entire SOFI sequence. Two smaller regions of homology were found - 55 bp homology to retroviral LTR sequence and 190 bp homology to the B1 repeat sequence. There are no long ORFs present and the sequence is unlikely to be coding. No transcript was detected on a Northern of RNA from various tissues and ES cells when the 250 bp unique sequence probe from 650 bp downstream of the integration site was used as a probe. It is possible that this sequence is intronic and not detectable on Northerns. The probe from upstream of the integration point detects multiple transcripts in ES cell RNA. These transcripts are probably derived from several loci (Figure 4.6b).

The function of the SOFI promoter and the properties that make the SOFI sequence a hotspot for recombination remain unclear.
CHAPTER 5

DISCUSSION
5. DISCUSSION

5.1 Targeting Hox-2.1 in ES cells.

The aim of this project was to produce a mouse mutant for Hox-2.1. The mutant would be useful for the functional analysis of Hox-2.1 - to determine its developmental importance or role and determine its interaction with developmental control and downstream "regulated" genes.

The overall procedure for producing a mouse mutant for Hox-2.1 is presented in Figure 1.4. The first step in the procedure is the mutation of the gene in ES cells.

ES cells were successfully cultured during the project. Culture conditions were devised by myself, Dr. Julia Dorin, Dr. Sinead Jones and Wendy Kimber, such that the ES cells appeared to be healthy and in an undifferentiated state. Julia Dorin has recently shown that ES cells cultured under these conditions are capable of forming highly chimaeric mice, with ES cell germ line contribution. ES cells of an early passage number (20-22) were used for transfections.

L cells were initially used as a test system until ES cells were being grown successfully. However, it is thought that genes target with different efficiencies in different cell types. This may be due to different availability of enzymes for recombination or difference in expression or accessibility of the gene.

Five targeting vectors have been used to target Hox-2.1 during this project.

PCR amplification between flanking endogenous and vector sequence was initially devised as the strategy for rapid screening of homologous recombinant clones against the background of random integration events. However, contamination from control plasmid amplified products became a problem during the first transfection analyses and other vectors contained too large a region of Hox-2.1 sequence for reliable amplification by PCR. All clones were therefore analysed by Southern blot hybridisation.

The first targeting attempts were made with a promoterless neo replacement vector. No homologous recombinants were obtained after isolation of 5 L cell and 31
ES cell G418^r clones. 36 clones are equivalent to 2880 total integration events if the 80-fold enrichment factor for integration into active promoter sites is taken into account. Promoterless neo targeting vectors have been successfully used to target expressed genes at high frequency e.g. SV40 at 3/11 G418^r clones (1/11 total integration events)(Dorin et al., 1989) and pim at 34/40 G418^r clones (te Riele et al., 1990). However, some lower frequencies have been reported e.g. polyoma middle T antigen at 1/100 G418^r clones (1/10000 total integration events)(Sedivy and Sharpe, 1989) or c-abl at 1/34 G418^r clones (Schwarzberg et al., 1989).

Analysis of 36 clones may be too low a number to conclude that this targeting vector is incapable of targeting Hox-2.1. However, it is possible to speculate that the size (100 bp) of the 5' Hox-2.1 homology is a limiting factor. Decreased regions of homology decrease targeting frequency (Thomas and Capecchi, 1987). Also it is possible that the selectable marker, neo, is not expressed at a high enough level to confer G418 resistance on homologous recombinants. Negative control regions at the Hox-2.1 locus may have an inhibitory affect on neo expression from the Hox-2.1 promoter. It was shown that the construct, Hox2.1pneo, containing 1.5 kb upstream Hox-2.1 sequence fused inframe with the neo, was capable of conferring G418 resistance on ES cells.

Two PNS vectors were constructed. One was designed to act like a promoterless neo vector with 400 bp of Hox-2.1 homology at the 5' end; the second contained 1.5 kb 5' Hox-2.1 homology and therefore Hox-2.1 promoter function. Surprisingly the latter generated no more G418^r colonies than the former, which should enrich for promoter integrations. This may suggest the presence of a negative control region in the 3' Hox-2.1 sequence (sequence not present in the Hox2.1pneo) inhibiting expression of neo. The HSV-tk gene was cloned onto the end of the vectors so that random integration events could be selected against by sensitivity to the base analogue, gancyclovir. Gancyclovir selection reduced the number of G418^r colonies by 40-fold. This compares favourably with the level of enrichment reported in the literature. 25 L cell and 6 ES cell G418^r ganc^r clones were analysed
(equivalent to 1320 G418R clones). None were homologous recombinants. Again it is not possible to say whether these vectors are incapable or capable of targeting Hox-2.1. There is considerable variability in the reported targeting frequencies obtained with PNS vectors. Some of the frequencies published are: 1/38 G418R gancR clones (1/304 G418R clones) for immunoglobulin μ gene (Kitamura et al., 1991) and with double tk for enrichment 1/500 (1/1000) for T-cell antigen receptor (Mombaerts et al., 1991) or 1/275 (1/7150) for Wnt-1 (McMahon and Bradley, 1990). It must be noted again that both vectors rely on the expression of the Hox-2.1 promoter at the Hox-2.1 locus for expression of neo.

Finally, two insertion vectors were used to target Hox-2.1 in ES cells. These vectors contain the same region of Hox-2.1 but 100 bp is deleted from the Hox-2.1 sequence at the linearisation site of HoxTneo. Construct 6 has a smaller region of 5' Hox-2.1 homology (360 bp) upon linearisation. No homologous recombinants were detected in 226 G418R clones analysed. It is not possible to say whether these vectors can function to target Hox-2.1. Targeting frequencies reported for insertion vectors are variable: 1/31 for Hox-2.6 (Hasty et al., 1991) to 1/1000 G418R clones for Hprt (Thomas and Capecchi, 1987). It is possible that the 360 bp of Hox-2.1 homology at the 5' end of construct 6 limits targeting potential of this vector and it is possible that HoxTneo is incapable of homologous recombination due to deletion of Hox-2.1 sequence at the linearisation site. However, HoxTneo did generate a pick-up clone. It is thought that the targeting vector found homology to the target sequence but instead of undergoing homologous recombination, picked up adjacent sequence from the locus by a process of gene conversion, before randomly integrating into the genome.

Hox-2.1 has not been successfully targeted. A total of 253 ES cell and 32 L cell selected clones were analysed. It is not possible to say whether the failure to target is due to low targeting potential of the targeting vectors used (discussed above) or because of inherent low targetability of the Hox-2.1 locus. There is obviously a wide range in the targeting frequencies obtained for targeting of
different loci/sequences, as reported in the literature (or not reported in the literature!). These differences are not always due to vector design since very similar vectors have been used to target different genes with different targeting frequencies, e.g. PNS vectors. It may be that sequence at the Hox-2.1 locus has low recombinogenic potential. The same 6 kb of sequence was used to construct the different targeting vectors. Also it may be that the Hox-2.1 locus has an inhibitory effect on neo expression, such that any homologous recombinants are selected against by G418. This negative influence may be affecting neo expression from the Hox-2.1 promoter in the case of the replacement/PNS vectors or from the SV40 early region promoter of SV2neo in the case of the insertion vectors.

285 clones have been screened in this study. This number is equivalent to 4416 integration events. This is not a large number when compared to some of the numbers screened for the successful targeting of other genes. Many more clones need to be analysed before any firm conclusions may be drawn.

5.2 Future prospects for targeting Hox-2.1.

In order to screen larger numbers of clones more quickly a more rapid screening procedure needs to be devised. Two recent techniques, one for freezing cells and one for DNA isolation from cells, allow the growth of more clones to the DNA preparation stage more quickly. Both techniques require fewer numbers of cells than the standard techniques. Chan and Evans (1991) describe a technique for freezing of ES cells in situ in 2 cm² wells of a 24 well plate (see Section 2.1.7). Laird et al. (1991) describe a procedure for the isolation of DNA from a confluent layer of ES cells in a 2cm² well. This method also has the advantage of having fewer steps, with no centrifugation, so that large numbers of clones may be processed at once. The DNA is extracted by layering lysis buffer onto the cells and precipitating the DNA with isopropanol. The DNA is claimed to be of sufficient yield and quality for digestion by restriction enzymes for Southern blot analysis.

PCR amplification allows the screening of large numbers of clones, more
rapidly than Southern blot analysis. However, PCR amplification is subject to contamination and generation of false positives (as previously discussed, found by myself and Kim et al., 1991). Kim et al. suggest the generation of a pseudo-recombinant cell line for use as a positive control along with PCR screening. This cell line has an integrated plasmid containing the diagnostic PCR product fragment for some of the PCR primers but not all (so that PCR contamination may be detected). Pseudo-recombinant cell line PCR amplification will enable the identification of the correctly sized product and provides an intensity comparison for amplification in pools of non-recombinant clones. It is wise to take extra care when setting up PCR reactions so as to avoid contamination. Kwok and Higuchi (1989) outline some precautions which ought to be taken.

It is possible that the use of isogenic DNA between ES cell and targeting vectors will increase the targeting frequency obtained in some targeting experiments (te Riele, personal communication). There may be small sequence differences between different strains. Thus construction of a Hox-2.1 targeting vector from 129 strain DNA (isogenic with E14 ES cell line) may increase the likelihood of targeting.

The strength of the neo selectable marker gene in the targeting vector may also have a significant effect on targeting frequency. Tybulewicz et al. (1991) report a 13-fold increase in G418r colonies with a neo attached to the mouse phosphoglycerate kinase-1 (PGK-1) (Adra et al., 1987) compared to pMC1neo polyA and SV2neo. A strong promoter may allow neo expression independent of chromosomal integration site.

Thomas and Capecchi (1987) found that insertion and replacement vectors were equally efficient at targeting the Hprt gene. Recent data, of Hasty et al. (1991b), suggest that insertion vectors are 9-fold more efficient than replacement vectors (Section 3.1). Replacement vectors may also integrate into the target locus upon recircularisation or concatemerisation. Negative selection against integration events (ganc selection against HSV-tk) may thus eliminate 95% of targeted events (Hasty et al., 1991b). It may therefore be advisable to continue with the insertion
vector in future Hox-2.1 targeting experiments. However, a potential problem with insertion vectors is reversion of the introduced mutation - duplicated sequence at the insertion site may undergo a single intrachromosomal recombination event such that the normal gene sequence is restored.

It is possible that the Hox-2.1 locus sequence is refractory to homologous recombination either because of low recombinogenic nature or due to silencing of neo expression. Therefore, it may be useful to construct a targeting vector similar to that used by Mombaerts et al. (1991). A large region (10 kb) of Hox-2.1 endogenous locus sequence may be deleted by targeting with a replacement vector containing DNA from upstream and downstream of the locus. In this way, different DNA sequence is required for recombination and any negative transcription control region may be deleted from the site.

5.3 Future prospects for the study of Hox-2.1 function.

Once the targeting of Hox-2.1 has been achieved in ES cells, the cells will be used to generate a mouse mutant for Hox-2.1. The study of the Hox-2.1 mutant phenotype is a project in itself!

If a Hox-2.1 mutation causes an early embryonic lethality it may be necessary to create a double knock-out ES cell line for the analysis of contribution to chimaeric mouse development (te Riele et al., 1990). Alternatively a less severe phenotype may be observed due to redundancy of function between related gene products. Hox-2.1 mutants may be crossed to Hox-1.3 mutants to create double homozygous mutants. Since Hox-1.3 is the paralogue of Hox-2.1 a more severe phenotype may be observed in these mice.

The generation of cell lines or mice overexpressing Hox-2.1 may also be useful in the study of Hox-2.1 function.

5.4 Identification of a site of frequent integration.

The SOFI of ES cells is a site which has been integrated into twice by a
promoterless \textit{neo} targeting vector. The insert flanking sequences, for one clone (7B2), were amplified by inverse PCR and sequenced. The flanking sequence from the second clone (5B2) and the uninterrupted target sequence from ES cells were sequenced directly from PCR products. In the different integration events, different amounts of targeting construct sequence have been lost - therefore, different regions of the targeting vector were involved in the cross-over event. However, the construct has integrated into the same 6 bp region in both the clones. There are 3 bp homology between vector and endogenous sequence at the proposed cross-over junction in 7B2, but no homology in 5B2. It is unlikely that the vector integrated into this specific sequence due to homology.

The promoter function of the cloned upstream region has been tested in ES cells by measuring G418\textsuperscript{r} conferred by an attached \textit{neo} gene and by CAT assay. Weak promoter function is observed. It is therefore unlikely that integration into this sequence has been selected for due to high expression of \textit{neo} from the endogenous promoter. This is believed to be the reason for the isolation and identification of viral integrations into the \textit{F2} locus of EC cells (Barklis \textit{et al.}, 1986).

There are several sequence motifs found to be associated with hotspots for homologous recombination (reviewed by Fischer Lindahl, 1991 and Steinmetz \textit{et al.}, 1987). The Chi sequence, GCTGGTGGG, enhances homologous recombination in \textit{E. coli} and bacteriophage \textit{\lambda}. Chi is a recognition site for the RecBC enzyme, which cuts 3' to the Chi site as it unwinds DNA. Short repeated sequences similar to Chi (CAGA and CAGG) have been found to be associated with the mouse major histocompatibility complex (MHC) hotspots for recombination. Middle-repetitive elements (MT) (Heinlein \textit{et al.}, 1986), LTR-like sequences (Uematsu \textit{et al.}, 1986 and Edelman \textit{et al.}, 1989), a 57 bp direct repeat (Uematsu \textit{et al.}, 1986) and binding motifs for transcription factors (Shenkar \textit{et al.}, 1991) are also associated with recombinational hotspots. Solitary LTR elements (Wirth \textit{et al.}, 1983) are hotspots for recombination (Edelman \textit{et al.}, 1989). Edelman \textit{et al.} tested LTR recombination \textit{in vitro} and found two cis-acting sequences required for recombination. It is assumed
that DNA at the recombination hotspot interacts with proteins which regulate the rate-limiting step of recombination. Proteins may stabilise single-stranded DNA regions required for recombination.

It is not known what properties of the SOFI sequence promote the integration into this site. No motif associated with recombinational hotspots was identified in the SOFI sequence. However, there is a 55 bp of sequence present upstream of the integration site, which is 72% homologous to proviral LTR sequence and contains two copies of an AGTTTC LTR motif. It is possible that this sequence retains some LTR function and is associated with promotion of recombination.

Aspects of secondary structure may also contribute to the recombination at the SOFI locus. Indirect repeats and AT rich regions may form single-stranded DNA regions.

The function of the SOFI sequence and promoter in ES cells is unknown. No transcript was detected on a Northern blot of RNA from various tissues when the 250 bp unique sequence from 650 bp downstream of the integration site was used as a probe. It is possible that the sequence is intronic and not detectable on Northerns. Computer analysis of the sequence shows it is unlikely to be coding.

5.4 Application of inverse PCR.

The isolation of sequence from promoter-trap and enhancer-trap integration sites is necessary for the molecular characterisation of the gene. This may be achieved by screening lambda libraries cloned from cell line genomic DNA. Inverse PCR amplification of flanking sequence offers a more rapid method which can be applied to large numbers of clones. The use of inverse PCR amplification in the cloning and sequencing of flanking sequences to the insert in the cell line 7B2, has shown the power of this method for such applications.

A second method for PCR amplification of flanking sequence was initially tried. Vectorette PCR is described by Riley et al. (1990). Sequence between vector and an upstream/downstream restriction site may be amplified by digestion of DNA
with the restriction enzyme, attachment of synthetic linkers and PCR amplification between a vector-specific primer and a linker-specific primer. The method has been used to isolate terminal sequences from YACs (Riley et al., 1990). However, when I applied it to the amplification of flanking sequences from 7B2, no specific sequence was amplified. The added complexity of mouse genomic DNA over yeast DNA is probably the reason for this lack of amplification.

5.4 Future prospects for the analysis of the SOFI.

There are several lines of research that may be pursued for the characterisation of the SOFI.

In order to obtain a larger region from the SOFI locus a lambda library of mouse genomic DNA should be screened with the unique downstream flanking inverse PCR product clone probe (Pst I to Eco RI fragment). Subcloning and sequencing of sequence further upstream from the integration point will help characterise the promoter. A larger upstream region may be tested for promoter function in ES cells as before.

The SOFI locus may be mapped to the mouse genetic map using RI strains (Section 4.7.3; Taylor in Morse, 1978). The search for polymorphic sites around the locus may be continued with different enzyme digests and the unique sequence probe available or by isolation of and screening with new probes.

Probes around the region may also be used to hybridise Northern blots of ES cell and tissue mRNAs in order to detect a transcript from the locus. If a probe recognises a transcript, then a cDNA clone may be pulled out from a cDNA library. More information on the function of the SOFI sequence may be obtained from sequencing of a cDNA clone.

Chimaeric mice may be generated from the ES cell lines 5B2, 7B2 and 7B3 (secondary of 7B2). This has been attempted once by Robin Lovell-Badge. However, no extensive chimaerism was observed in the chimaeras born and none were germ-line. If these cells cannot form germ-line chimaeras (due to genetic
defects accumulated during culturing), it may be possible to generate new cell lines by targeting the SOFI locus homologously or again with the promoterless neo vector. If chimaeric mice are produced, homozygous mice for the SOFI integration may be studied for any phenotypic effects.

If the locus is indeed a site of frequent integration, it is predicted that integration into this locus may be repeated. Integration events into the locus may be detected by PCR as in Section 4.9. It is possible that other targeting vectors integrate into this locus - since no homology to the Hox-2.1 vector is observed. Analysis of other targeting vector illegitimate integration events can easily be carried out by PCR amplification on pools of DNA. It may be that random integration is not random at all, but into a few specific regions.

If the properties that make this sequence a site of frequent integration (e.g. sequence content or structure) can be characterised, useful information may be obtained for the targeting of other regions. If a specific gene is to be targeted, sequences in or around the gene may be analysed for similar sequence or structure to the SOFI sequence. Any region that shows similarity may be used to construct the targeting vector for that gene. Sequence similar to the SOFI sequence may have a high targeting frequency.
REFERENCES


